The goal of this study is to determine whether principal component analysis (PCA) can be used to process latitude-time ionospheric TEC data on a monthly basis to identify earthquake associated TEC anomalies. PCA is applied to latitude-time (mean-of-a-month) ionospheric total electron content (TEC) records collected from the Japan GEONET network to detect TEC anomalies associated with 18 earthquakes in Japan (\(M \geq 6.0\)) from 2000 to 2005. According to the results, PCA was able to discriminate clear TEC anomalies in the months when all 18 earthquakes occurred. After reviewing months when no \(M \geq 6.0\) earthquakes occurred but geomagnetic storm activity was present, it is possible that the maximal principal eigenvalues PCA returned for these 18 earthquakes indicate earthquake associated TEC anomalies. Previously PCA has been used to discriminate earthquake-associated TEC anomalies recognized by other researchers, who found that statistical association between large earthquakes and TEC anomalies could be established in the 5 days before earthquake nucleation; however, since PCA uses the characteristics of principal eigenvalues to determine earthquake related TEC anomalies, it is possible to show that such anomalies existed earlier than this 5-day statistical window.

1. Introduction

Recent research into earthquake-associated ionospheric anomalies has been able to show that associations do exist between earthquakes and ionospheric disturbance [3–6]. The precise mechanism(s) is not clear and a number of reasonable theories have been presented as to what may cause earthquake-associated TEC anomalies. These include a number of geochemical processes in the dilatant phase of earthquake preparation such as radon ionization, causing long-living ion clusters in the lower atmosphere [7], the activation of mobile positive holes in the earth’s crust, resulting in strong electric fields and activation of highly mobile electronic charge carriers in rocks near the earth’s surface [8, 9], and the release of CO\(_2\) gas [10] which could potentially destroy neutral clusters, resulting in the near ground layer becoming rich in ions [3]. The resultant lower-atmosphere electric field follows geomagnetic field lines into the ionosphere leading to the noted anomalies. Other potential causes are physical such as acoustic gravity waves. One possible cause of these is CO\(_2\) gas release creating turbulence in the lower atmosphere and another would be fine vibrations in the earth’s crust associated with earthquake preparation. Whatever the specific cause of these anomalies, researchers have been able to show that in many cases the observed anomalies have been earthquake related [3, 11]. This has been achieved by first determining the existence of an anomaly based on a mean value for TEC and then, assuming a normal distribution, looking for significant deviations from the mean. The noted anomalies (both enhancements and reductions in TEC) are processed to determine their association with earthquakes after the elimination of other potential causes of which there are many. In two previous works [1, 12], I used principal component analysis (PCA) to identify and confirm earthquake-associated anomalies in the 1 to 5 days prior to nucleation of a number of
Table 1

<table>
<thead>
<tr>
<th>Date</th>
<th>Latitude-Longitude</th>
<th>Magnitude (depth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000 October 6 (04:30:19 UTC)</td>
<td>(35.45N 133.13'E)</td>
<td>6.7 (depth ~10 km)</td>
</tr>
<tr>
<td>2003 May 26 (09:24:32 UTC)</td>
<td>(38.90N 141.45'E)</td>
<td>7.0 (depth ~68.2 km)</td>
</tr>
<tr>
<td>2003 September 25 (19:50:06 UTC)</td>
<td>(41.775'N, 143.904'E)</td>
<td>8.3 (depth ~27 km)</td>
</tr>
<tr>
<td>2003 October 08 (09:06:55 UTC)</td>
<td>(42.658'N, 144.487'E)</td>
<td>6.7 (depth ~32 km)</td>
</tr>
<tr>
<td>2003 October 31 (01:06:28 UTC)</td>
<td>(37.829'N, 142.634'E)</td>
<td>7.0 (depth ~10 km)</td>
</tr>
<tr>
<td>2004 May 29 (20:56:12 UTC)</td>
<td>(34.256'N, 141.385'E)</td>
<td>6.5 (depth ~38 km)</td>
</tr>
<tr>
<td>2004 September 5 (10:07:07 UTC)</td>
<td>(33.062'N, 136.608'E)</td>
<td>7.2 (depth ~14 km)</td>
</tr>
<tr>
<td>2004 September 5 (14:57:18 UTC)</td>
<td>(33.184'N, 137.071'E)</td>
<td>7.4 (depth ~10 km)</td>
</tr>
<tr>
<td>2004 September 6 (23:29:35 UTC)</td>
<td>(33.184'E, 137.201'E)</td>
<td>6.7 (depth ~10 km)</td>
</tr>
<tr>
<td>2004 October 23 (08:56:00 UTC)</td>
<td>(37.226'N, 158.779'E)</td>
<td>6.6 (depth ~16 km)</td>
</tr>
<tr>
<td>2004 November 28 (18:32:13 UTC)</td>
<td>(42.995'N, 145.056'E)</td>
<td>7.0 (depth ~39 km)</td>
</tr>
<tr>
<td>2004 December 6 (14:15:11 UTC)</td>
<td>(42.907'N, 145.200'E)</td>
<td>6.8 (depth ~35 km)</td>
</tr>
<tr>
<td>2005 March 20 (01:53:41 UTC)</td>
<td>(KYUSHU 33.805'N, 130.078'E)</td>
<td>6.6 (depth ~10 km)</td>
</tr>
<tr>
<td>2005 July 22 (07:34:57 UTC)</td>
<td>(35.506'N, 139.933'E)</td>
<td>6.0 (depth ~65.6 km)</td>
</tr>
<tr>
<td>2005 August 16 (02:46:28 UTC)</td>
<td>(38.251'N, 142.059'E)</td>
<td>7.2 (depth ~36 km)</td>
</tr>
<tr>
<td>2005 October 19 (11:44:43 UTC)</td>
<td>(36.383'N, 140.833'E)</td>
<td>6.4 (depth ~41.5 km)</td>
</tr>
<tr>
<td>2005 November 14 (21:38:51 UTC)</td>
<td>(38.101'N, 144.925'E)</td>
<td>7.0 (depth ~11 km)</td>
</tr>
<tr>
<td>2005 December 2 (13:13:09 UTC)</td>
<td>(38.122'N, 142.118'E)</td>
<td>6.5 (depth ~29 km)</td>
</tr>
</tbody>
</table>

This table shows the earthquakes (Richter magnitude ≥6.0) from 2000 to 2005 in Japan (Source: U.S. Geological Survey).

M ≥ 5.0 earthquakes on Taiwan reported by Liu et al. [11], Lin [12], and also anomalies associated with China’s 12 May, 2008 M = 7.9 Wenchuan earthquake that were noted by both Liu et al. [13] and Zhao et al. [14] and Lin [1]. The PCA results of those studies were independent of geomagnetic storm activity and long-term variance in TEC [12]. The reason for studying earthquake-associated anomalies reported by the above researchers was to give their identification a firm mathematical footing and examine the usefulness of the PCA method. PCA is a widely used method for finding relevant meaning in confusing datasets [15]. It is nonparametric and as such does not utilize deviations from a previously described average TEC to determine anomalies. The characteristics of assigned principal eigenvalues are used to determine TEC anomalies. In this case, when the principal eigenvalues are greater than 0.5 (in a normalized set) [12] they are considered large and a potential earthquake-associated TEC anomaly is established. Cross analysis is then conducted for other possible causes of these large principal eigenvalues, for example: solar flare and geomagnetic storm activity. In this paper, PCA is applied to latitude-time vertical ionospheric TEC records using data from Japan’s GEONET network system [16] (Figure 1) to assign principal eigenvalues to TEC anomalies occurring at and around the times 18 earthquakes of M ≥ 6.0 occurred in Japan from 2000 to 2005 (Table 1). Conceptually the benefits of using a large GPS network are similar to statements made by Garrison et al. [17] and Dyrud et al. [18]. Specifically, in this study, the advantage of using such a network is that detection of an earthquake-related anomaly can occur at any receiver within the network, and therefore, anomaly detection need not be just from a station nearby the epicenter. This means anomaly detection will be more sensitive and not reliant on a single station. TEC data is processed in a similar manner to median hourly electron content-latitude profiles of Walker et al. [19]; however, in this case a mean value of TEC for a month is examined. The purpose of this study is to examine a large amount of data over a 5-year period in which many large earthquakes occurred (18) in a similar geographical location to see if PCA can discriminate...
Figure 2: (a) shows the latitude-time TEC record in Japan for October 2000 (JST: Japanese Standard Time) (Unit: TECu). (b) shows 12 principal eigenvalues in 2000. Dates constitute the horizontal axis and corresponding eigenvalues are on the vertical axis. Peaks and troughs in eigenvalues have been plotted and graphed on a month-to-month basis to allow for interpolation. The largest principal eigenvalue is represented in October in which an earthquake occurred on October 6 (Table 1).

Figure 3: (a) shows the latitude-time TEC records in Japan for 2003 (Unit: TECu). (b) shows 12 principal eigenvalues in 2003. Three larger principal eigenvalues are represented in May, September, and October in which four earthquakes occurred on May 26, September 25, October 8, and October 31 (Table 1).
Figure 4: (a) shows the latitude-time TEC records in Japan for 2004 (Unit: TECu). (b) shows 12 principal eigenvalues in 2004. Five larger principal eigenvalues are represented in May, September, October, November, and December in which seven earthquakes occurred on May 29, September 5, September 6, October 23, November 28, and December 6 (Table 1).
Figure 5: (a) shows the latitude-time TEC records in Japan for 2005 (Unit: TECu). (b) shows 12 principal eigenvalues in 2005. Six larger principal eigenvalues are represented in March, July, August, October, November, and December in which six earthquakes occurred on March 20, July 23, August 16, October 19, November 14, and December 2 (Table 1).
earthquake-associated TEC anomalies based on the characteristics of principal eigenvalues. In addition, Lin [12] and Lin [1] were studies that focused on finding anomalies that had already been recognized by Liu et al. [11, 13] and Zhao et al. [14]. Those researchers used statistical methods that required precursor earthquake association of TEC anomalies close to the mainshock dates (5 days). However, it is possible that an earthquake-related TEC anomaly could occur earlier. The possibility of earlier occurrence of an earthquake-related TEC anomaly is important to establish in this paper because the work examines mean-of-a-month data. Therefore, PCA is applied to daily 1D TEC data relating to the Kyushu earthquake of 20 March 2005 (M = 6.6). That experiment shows an earthquake-related TEC anomaly (based on the characteristics of principal eigenvalues) occurring 11 days before the mainshock. Mean-of-a-month data includes the effects of geomagnetic storm, solar flare activity, and other ionospheric irregularities [20, 21]. Previous application of PCA found that PCA was ineffective at distinguishing earthquake-associated anomalies from other possible causes when earthquakes were smaller than M = 5 [12]. This present study only looks at earthquakes of M ≥ 6.0. It is assumed that large principal eigenvalues of greater than 0.5 (in a normalized set) are associated with earthquakes. Cross examination is conducted for months when the Dst index [22] indicates geomagnetic activity to see if this could have influenced the results.

2. Theory of Principal Component Analysis (PCA)

In the past, PCA has been widely used to detect and recognize the fine characteristics of signals; its physical meaning is indubitable [23, 24]. It is a simple nonparametric method that is frequently used to extract relevant data from confusing datasets. Three basic assumptions underpin the technique. They are linearity, a high signal to noise ratio (SNR), and orthogonal principal components. Linearity allows for the problem to be framed as a change of basis, a high SNR means that principal components with larger variance represent points of interest and those with lower variance represent noise (this assumption is strong and can be incorrect), and orthogonality makes PCA solvable with linear algebra. If these initial assumptions are correct, PCA allows for the underlying structure to be seen. TEC records form a matrix A with m rows (number of measurement types) and n columns (number of samples) with “a” being the data points (in this case vertical TEC data):

\[ A = [a]_{m \times n}. \]  

For each n, u is a unit vector. \( AA^T u = \lambda u \), the eigenvalues \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \). The maximum eigenvalue is \( \lambda_1 \), which represents the principal characteristics of signals.

3. Data Source and Processing for Japan 2000 to 2005

3.1. Data. The data used in this study is sourced from latitude-time vertical ionospheric TEC records from Japan’s GEONET network system for 18 earthquakes of M ≥ 6.0 that occurred in Japan from 2000 to 2005 (Table 1). Data processing follows a similar approach to the median hourly electron profiles of Walker et al. [19]; however, here the profiles are monthly. Mean-of-a-month data are representative of all TEC signals in a given month, including the influence of earthquakes, geomagnetic storm, and solar flare activity. However, previous study has shown that PCA can distinguish
between geomagnetic storm and solar flare activity [12]. In addition, the collection of data over such a wide area as that represented by Japan's GEONET network means that earthquake-associated anomalies that might occur far from the epicenter such as those detected by Zhao et al. [14] for China's 2008 Wenchuan earthquake will not be ignored. The focal depths of all the earthquakes are given in Table 1. These focal depths could have bearing to have recognizable anomalies according to Xia et al. [25]. In that study, it was shown that the focal depths of the hypocenter were less than 228 km for 20 earthquakes (M = 5 to 8) in the Tibetan Plateau and its neighboring regions from 1999 to 2008, and the earthquake-related disturbances could be registered for the earthquakes with focal depth exceeding 100 km. The focal depth affects the occurrence of anomalies while different magnitudes have little effect on it. However, researching its reason is beyond the scope of this paper. In this study, the depths of the researched earthquakes are less than 100 km.

3.2. TEC Data Processing of Japan's GEONET Data from 2000 to 2005. Data records form a matrix of dimensions 24
(24 hrs) × 1200 (1200 receivers) (GEONET consists of ~1200 receiver); that is, from Formula 1, \( m = 24 \) and \( n = 1200 \). The transforms are then conducted and principal eigenvalues assigned for every month from 2000 to 2005. The set of principal eigenvalues generated by this study are normalized by dividing all principal eigenvalues by the maximal principal eigenvalue discovered. This allows for relative comparisons to be made. Principal eigenvalues of greater than 0.5 are considered large [12].

4. Results

4.1. Japan 2000 to 2005. In October 2000, a large earthquake of \( M = 6.7 \) was recorded for western Honshu. Figure 2(a) shows a plot of the vertical latitude-time ionospheric record for October 2000 using data from Japan’s GEONET system for 1200 receivers. This plot is similar to the latitude-time-based TEC plots of Liu et al. [26]. Figure 2(b) gives a plot of principal eigenvalues assigned for each month of the year.
4.2. March 20, 2005 Kyushu Earthquake. A large earthquake occurred on Kyushu, Japan in March 2005. The corresponding TEC record is shown in Figure 6(a). This data is processed using one-dimensional data analysis (day-to-day basis) as per Lin [12] (TEC data of receiver near the epicenter, the receiver position is at 33.70°N, 129.70°E). The Kyushu earthquake occurred on 20 March, 2005 (local time) at a Richter magnitude scale M = 6.6 (Table 1). Figure 6(b) shows that on a day-to-day basis a large spike in principal eigenvalues can be seen on March 9, 11 days before the Kyushu earthquake.

The largest principal eigenvalue is represented in October. The Honshu earthquake occurred on October 6 (Table 1). Figures 3, 4, and 5 give the results of data processing for 2003, 2004, and 2005, respectively. All principal eigenvalues in Figures 2(b) to 5(b) are divided by the maximal principal eigenvalue in the bottom figure of Figure 3. This was the largest principal eigenvalue returned by the survey and is used for normalization. Principal eigenvalues of >0.5 [12] are considered large and representative of earthquake-related TEC anomalies.
5. Discussion

5.1. Possible Detection of Earthquake-Associated TEC Anomalies. This study examines a relatively long time series of data from 2000 to 2005. Figures 2 to 5 show that PCA applied to latitude-time TEC records for Japan (2000 to 2005) assigned large principal eigenvalues (>0.5) to TEC anomalies for all the listed earthquakes in Japan (Table 1). In addition, Figure 6 shows that the application of PCA to one-dimension daily TEC data for the March 20, 2005 Kyushu earthquake gave a large principal eigenvalue representative of an earthquake related anomaly 11 days before that earthquake. In addition, two recent studies of the China 12 May, 2008 Wenchuan earthquake [27] using empirical studies with normalized TEC data found possible earthquake-related TEC anomalies on days 6 and 13 before that earthquake. Their results as well as the result given in this paper are important in that they help support the integrity of the mean-of-a-month analysis for the 18 earthquakes of Table 1.
5.2. TEC Anomalies from Causes other than Earthquakes. The possibility that geomagnetic activity could have interfered with the results of this study is now examined. Figures 7, 8, 9, and 10 show the Dst indices for the study period. The Dst index is a geomagnetic index which monitors world wide magnetic storm levels based on the average horizontal component of the geomagnetic field from mid-latitude and equatorial magnetograms. Magnetic storms are indicated by negative Dst values whereby the more negative the Dst the more intense the magnetic storm. Figure 2 shows principal eigenvalues assigned to monthly mean electron content in the ionosphere for every month in 2000 while Figure 7 shows the monthly Dst index. From Figure 7, April and July both show geomagnetic storm activity [28]; however, PCA did not assign large principal eigenvalues for those months. A large earthquake occurred on Western Honshu on October 6, 2000 (local time) at a Richter magnitude scale M = 6.7 (Table 1) and that month did produce a large principal eigenvalue. There was also some geomagnetic activity in October 2000. However, given that April and July also had geomagnetic activity but neither produced a large principal eigenvalue, it is possible that the large principal eigenvalue in October is representative of earthquake activity. Similarly, the Dst index shows that geomagnetic activity was prevalent in August and November 2003 yet PCA only assigned relatively small principal eigenvalues for these months. May, September, and October of 2003, however, all had earthquakes and all produced large principal eigenvalues. The situation for 2004 and 2005 is similar to that described for 2000 and 2003.

6. Conclusion

In this study, PCA is applied to latitude-time (mean-of-a-month) ionospheric total electron content records collected from the Japan GEONET system to detect TEC anomalies associated with 18 earthquakes in Japan (M ≥ 6.0) from 2000 to 2005. PCA was able to discriminate clear TEC anomalies in the months when all 18 earthquakes occurred, and such anomalies existed earlier than this 5-day statistical window. It is possible that these TEC anomalies are earthquake related after examining the Dst index for geomagnetic storm activity over this same period. Application of PCA to examine the Kyushu earthquake of 20 March, 2005 (M = 6.6) revealed a large principal eigenvalue representative of an earthquake-associated TEC anomaly 11 days before that earthquake on March 9. This result supports the integrity of the mean-of-a-month analysis.

Acknowledgments

The author is grateful to Dr Louis (L.Y.) Tsai of the Graduate Institute of Applied Geology and Dr H. W. Chen of the Institute of Geophysics, National Central University for their useful references and data support.

References
