

Statistical characterization of ionosphere anomalies and their relationship to space weather events

Chunming Wang^{1,*}, Irvin Gary Rosen¹, Bruce T. Tsurutani², Olga P. Verkhoglyadova², Xing Meng², and Anthony J. Mannucci²

¹ Department of Mathematics, University of Southern California, Los Angeles CA 90089, USA

*Corresponding author: cwang@usc.edu

² The Jet Propulsion Laboratory, California Institute of Technology, Pasadena 91109, California, USA

Received 24 June 2015 / Accepted 25 December 2015

ABSTRACT

The statistical characterization of the relationship between thermosphere-ionosphere anomalies and space weather events, also referred to as space weather anomalies, such as solar coronal mass ejections (CMEs) and geomagnetic storms, is a crucial component in the development of a forecast capability for thermosphere-ionosphere disturbances. This manuscript presents a systematic statistical approach for analyzing historical ionosphere and space weather observations to derive a quantitative characterization of the relationships between the thermosphere-ionosphere anomalies and space weather anomalies. Based on 2 years of historical data, our analysis reveals the complex nature of the relationship between space weather disturbances and ionospheric responses.

Key words. Thermosphere-ionosphere anomaly – Statistical analysis – Space weather forecast

1. Introduction

The primary objective of our efforts in medium-range thermosphere-ionosphere storm forecasting (Mannucci et al. 2015) is to develop the capability to predict thermosphere-ionosphere anomalies starting with observations of the solar corona. Despite the availability of advanced first-principle based models linking solar activities to the dynamics of the thermosphere and ionosphere, the quantity of observations of the space environment is, at the moment, inadequate for calibrating the chain of models to a degree where they could be used to deliver a reliable forecast. Additionally, it is well known that anomalies in the thermosphere and ionosphere can also be triggered by troposphere waves (Pedatella et al. 2008; Immel et al. 2009; Lu et al. 2015). It is therefore necessary to characterize statistically the connection or correlation between major space environment disturbances such as solar corona mass ejections (CMEs) or magnetosphere storms and ionosphere anomalies.

There is an abundance of research efforts focused on identifying and explaining the effects of space environment disturbances on the ionosphere (Fagundes et al. 2008; Pokhotelov et al. 2010; Lee et al. 2011; Burns et al. 2012; Tsurutani et al. 2012). The identification of anomalous ionospheric features is often made within the context of a known space weather event, as well as by comparing ionospheric observations immediately following the event to preceding “quiet” period data. By “space weather event”, we refer specifically to periods when geomagnetic indices identify disturbed periods, or when solar wind conditions indicate a disturbance variation. Interestingly, a similar approach has been used in identifying thermosphere and ionosphere anomalies preceding or following an earthquake (Liu et al. 2004; Pulinets et al. 2006; Zhou et al. 2009) even though the latter lacks a convincing physical explanation linking the two events. A difficulty

with this approach is in determining and quantifying what exactly constitutes nominal conditions for the thermosphere and ionosphere. A corollary to this, is the question: how often do anomalies that are similar in nature to those attributed to either a space weather event or an earthquake, occur in the historical data record? The answers to this question are crucial for producing reliable forecasts with well-determined uncertainties.

In describing our findings, we will rely on terminology from Bayesian theory and statistical hypothesis testing. We will be interested in assessing the ability of linking the occurrence of an anomalous global ionospheric map (GIM) of vertical total electron content to the occurrence of an observed space weather disturbance, such as a CME. Thus we will be interested in calculating the *sensitivity*, the conditional probability of the occurrence of an anomalous VTEC map given that a space weather event occurred, the *specificity*, the conditional probability that an anomalous map did not occur given that a space weather event did not occur, and, most significantly for our purposes here, the *posterior* conditional probability that a space weather event occurred given that an anomalous global VTEC distribution was observed. These probabilities are closely connected with what are referred to in the statistical literature as the probability of a *Type I* error, that is, the conditional probability that an anomalous global VTEC distribution was observed given that a space environment disturbance did not occur and the probability of a *Type II* error, that is, the conditional probability that an anomalous global VTEC distribution was not observed given that a space weather disturbance did in fact occur. The probability of the complement of a Type II error, or the conditional probability that an anomalous global VTEC distribution was observed given that there was a space weather disturbance, is referred to as the *power* of the forecast.

With the availability of observations of the space environment and the thermosphere and ionosphere, a purely statistical

characterization of the relationship between these two types of anomalies is possible. One key aspect of the analysis required for this characterization is the classification of ionosphere anomalies. Recently there has been a great deal of progress in the development of new efficient data mining techniques in the context of a large variety of applications (Steinwart et al. 2005; Chandola et al. 2009; Sodemann et al. 2012; Grnitz et al. 2013; Mata et al. 2014). These techniques can be broadly grouped into *supervised* and *unsupervised* classification approaches. The most common analyses of storm effects on the ionosphere fall into the general category of supervised classification because the “quiet” or disturbed conditions are pre-identified by the absence or presence of space environment anomalies. There have also been several attempts at the automated monitoring of ionosphere anomalies (Forbes et al. 2000; Luo et al. 2005; Mandrikova et al. 2012) that could be viewed as examples of unsupervised detection. In the training of a supervised classification approach, a data mining algorithm searches for ways of separating data points labeled as “quiet” from data points labeled as “disturbed”. A trivial supervised classification of ionospheric condition consists of labeling an ionospheric condition as disturbed only when it is identical to conditions on one of the days during which solar or geomagnetic anomalies were observed. Given the natural variability of the ionosphere where no 2 days have identical conditions, the trivial classification may even be 100% correct when applied to a training data set. That is, at least for the training data set, the method would not classify any quiet day as disturbed. However, since it is well known that a portion of the solar or geomagnetic anomalies do not lead to significant disturbances in the ionosphere, a supervised classification algorithm risks relying on minor differences that are well within the natural variabilities of the ionosphere for the resulting classification. Since our goal is to characterize the connection between *significant* disturbances in the ionosphere and solar or geomagnetic anomalies, a classification of ionosphere disturbed conditions independent of the knowledge of solar or geomagnetic conditions is essential. Therefore, only unsupervised classification approaches are used to identify disturbed ionosphere conditions.

As a part of our effort in forecasting thermosphere and ionosphere anomalies and their connection to external space weather anomalies in the solar wind and magnetosphere, we developed techniques for a statistical characterization of the relationship between these two types of anomalies. In particular, we have attempted to derive quantitative estimates of the sensitivities and specificities of anomalous VTEC distributions and the posterior probabilities of the occurrence of an externally driven space weather event. Our analyses use two years, 2011 and 2012, of historical data of both ionospheric VTEC and observations of the space weather environment to develop our estimates. In this manuscript, we report on our efforts at establishing a quantitative characterization of the correlation between interplanetary space weather and thermosphere and ionosphere anomalies. An outline of our manuscript is as follows. In Section 2 we describe the data upon which our analysis will be based and in Section 3 we provide a detailed description of the key steps in our analysis. In Section 4 we discuss the implications of our findings.

2. Description of data sets

For the analyses presented in this paper, the primary data set we use for identifying anomalous conditions in the ionosphere

is the global vertical total electron content (VTEC) obtained from global ionosphere maps, or GIMs, produced by the Jet Propulsion Laboratory (Mannucci et al. 1998). The GIM-VTEC map is derived from data received from a network of 200–1,000 globally distributed ground-based Global Positioning Satellite (GPS) receivers. Although it is known that VTEC is not the most sensitive parameter to solar or geomagnetic disturbances, significant disturbances in the ionosphere often lead to unusual VTEC maps as well. Our primary reason for this selection is the global availability and the consistent quality (Komjathy et al. 2005a) of GIMs. As for the training of any automated classification algorithm, it is essential that the training data set is sufficiently large to include all possible scenarios. Daily GIM-VTEC maps have been operationally produced since the mid-1990s. Our analyses are based on global VTEC maps with a 1 by 1° resolution in latitude and longitude sampled at 15-min time intervals. Examples of these maps are shown in Figure 1. It is useful to remind ourselves that a single data point in our daily analysis consists of a collection of 96 global VTEC maps. Figure 1 only displays 12 of these maps. Although there are significant advantages of using a consistently produced gridded product, there are also drawbacks. In particular, as a map of VTEC, GIM inherently lacks many important descriptive parameters essential for a full characterization of the ionosphere such as the F_2 -region peak density, $NmF2$, and height $HmF2$. Alternatively, output of 3-dimensional electron density produced by data assimilation models, such as GAIM (Schunk et al. 2004; Wang et al. 2004), provides a complete description of the electron density vertical profile. However, the most advanced GAIM analyses are not currently operationally produced and readily available. In addition, a common shortcoming of GIM analyses and, for that matter, GAIM as well, is that they are not actual direct measurements. As a result, GIMs include artifacts associated with the analysis methods. For example, in regions where data are sparse or not available, these analyses are either extrapolated from other regions or products of climatological models. This aspect of the GIMs does not present a major problem for their use in anomaly detection because extrapolated or seasonally averaged data are not likely to be “out of family”. GIM and GAIM have already been used extensively in studies of disturbances, so their deviation from average behavior as represented in the data is well established (Pi et al. 2009; Immel & Mannucci 2013). In fact, truly anomalous GIM maps would have to contain their anomalies in data-rich regions. On the other hand, while the direct use of raw ionosphere measurements such as $NmF2$ and $HmF2$ derived from ionosondes or slant TEC measurements would be more desirable, they would also be more difficult to use due to the lack of absolute calibration and consistent availability. Therefore, we feel that the use of GIM as a proxy indicator of the state of the ionosphere presents an appropriate trade-off for demonstrating our general approach (Komjathy et al. 2005b; Datta-Barua et al. 2014).

There are many data sources that can be used to represent the state of the space environment (Menvielle & Berthelier 1991; Richardson & Cane 2010). However, since the main goal of our analysis is to establish a statistical characterization of the link between significant ionosphere anomalies and anomalies in the space environment, we decided to use classifications of the space environment widely accepted by the space weather community based on the solar wind, since solar wind conditions are in principle forecastable and are the basis of medium-range forecast research (Mannucci et al. 2015). In particular, we use the list of CMEs and ICMEs since 1996

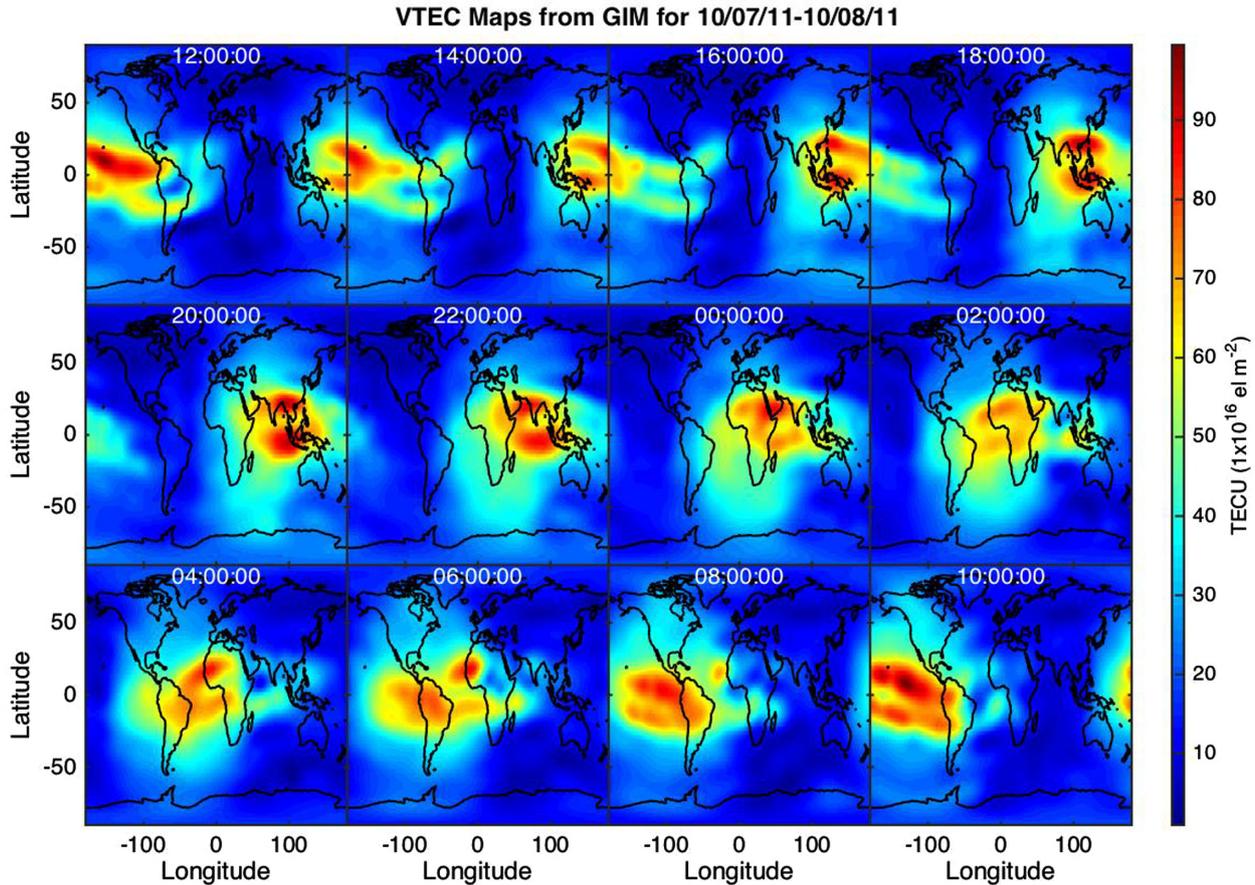


Fig. 1. Selected frames of daily GIM-VTEC for October 7–October 8, 2011

compiled by Richardson and Cane, which is available on the Internet at (<http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm>) to identify space weather anomalies associated with the solar wind. On the other hand, instead of directly using the Kp or Ap index for magnetosphere conditions, we rely on the quiet days, or *Q-days* and disturbance days, or *D-days* classification produced by the GFZ German Research Center for Geosciences as a global description of geomagnetic conditions. A description of the method used for selecting Q-days and D-days is provided by GFZ-Potsdam.¹ It should be noted that the determination of the occurrence of ICMEs (as defined in Richardson & Cane 2010) is not entirely unambiguous. Similarly, there is also significant nuance involved in the identification of Q-days and D-days (see Hibberd 1981).

In our analyses presented in this manuscript, we consider one calendar year of GIM-VTEC maps as a training data set. As a result, the classification of a normal or quiet day and an anomalous day is made in relationship to the rest of the days in the year. In particular, we have focused on 2011 and 2012, as two separate training sets, for our analyses. As a result, we also only use CME or ICME days and geomagnetic Q-days and D-days within these 2 years to investigate the correlations between the ionospheric and space environment anomalies. An interval of 1 year is considerably longer than the time intervals typically used in an episodic analysis of an isolated storm event. In those studies, on the other hand, the quiet

reference time interval usually spans less than 20 days. Restricting our historical data set to one calendar year inevitably limits the validity of our observations to only one specific phase within one specific solar cycle. However, as a proof-of-concept demonstration of our analysis methodology, we feel that our analysis time interval selection represents an appropriate trade-off between computational complexity and the statistical significance of our analyses.

3. Technical approach

Since an ionosphere anomaly is identified by comparing it to typical, or nominal, conditions, it is crucial to develop an unambiguous characterization of what can be considered to be the normal state of the ionosphere. For many previous space weather studies (Forbes et al. 2000), this requires only the characterization of the ionosphere during a time period absent of any geomagnetic storms or solar anomalies. This type of anomaly detection method is often referred to as *supervised detection* because two training data sets labeled “normal” and “abnormal” are considered available. However, the approach relies implicitly on two basic assumptions. First, we must assume that our observations of the geomagnetic storm and solar anomalies are complete, and second, we must believe that all ionosphere anomalies are consequences of either geomagnetic or solar anomalies. In fact, both of these assumptions are in dispute. Indeed, geomagnetic indices are indirect indicators of ionospheric conditions, and only very limited solar and interplanetary observations are available. At the same time, it is well known that troposphere waves can,

¹ Matzka, J., International Q-Days and D-Days, <http://www.gfz-potsdam.de/en/section/earths-magnetic-field/services/kp-index/explanation/qd-days/>.

in some cases, be the cause of ionosphere anomalies (Pedatella et al. 2008; Immel et al. 2009; Lu et al. 2015). Alternatively, an *unsupervised detection* approach relies primarily on the relative frequency of occurrence of normal conditions to distinguish them from abnormal conditions. The assumption that quiet conditions are a more common occurrence is implicit in many studies (see for examples, Tsurutani et al. 2004; Pokhotelov et al. 2010; Burns et al. 2012; Tsurutani et al. 2012). In developing an unsupervised classification scheme for ionosphere anomalies, we can attempt to establish possible correlations between ionosphere anomalies and externally driven space weather anomalies by examining the coincidence of occurrence of the two types of anomalies. In this section, we present the key steps in an unsupervised ionosphere anomaly classification approach and a statistical analysis for their correlation with space weather anomalies.

3.1. Cluster analysis and cluster radius

In unsupervised classification, data points are considered as belonging to the same class if the “distances” between them are small. Distance is determined by a metric that varies according to the degree of similarity between two VTEC distributions, and is zero for two identical distributions. For the purposes of classification of ionospheric conditions using VTEC maps, we consider a *data point* to be a collection of 96 global VTEC maps spanning a period of 24 h. As a result, for a training data set of one calendar year, there are 365 (or 366 in the case of 2012) such data points. Each data point can be considered to occupy a specific “location” in a multidimensional space of dimension $180 \times 360 \times 96$ (for $1 \times 1^\circ$ maps of 15-min cadence). The coordinates within this space are determined by the TEC values at each location and time.

There are many different ways of quantifying the difference between two sets of VTEC maps. We will discuss some of them in the next subsection. A selected metric, or distance function, gives us the distance $d_{i,j}$ between the i th day, or data point, and the j th data point in the data set. Therefore the 365×365 matrix D defined by $[D]_{i,j} = d_{i,j}$ is referred to as the distance matrix for a data set containing 365 data points. A graphic representation of a distance matrix is given in Figure 2. Since the metric used in Figure 2 is the average of the absolute VTEC differences, the distance is measured in TEC unit (TECU). A more precise definition of this metric is given in Section 3.2. Thus, a distance of 1 TECU indicates that the average difference of the two VTEC maps is 1 TECU for any longitude, latitude, and UT. One of the most noticeable features of the distance matrix is the relatively low values along and close to its principal diagonal (from the upper right corner to the lower left corner). The diagonal values are, of course, zero by definition, but the small values clustered around the diagonal are a consequence of the construction of the GIMs. Even though new data are used to refresh the VTEC maps, only a subset of ground receivers provide new data during any given 15-min sampling interval. Therefore the new refreshed map is the result of extrapolating data from the previous maps by taking into consideration the change of sun angle and geomagnetic field orientations. These approximations introduce what is in effect, a numerical diffusion phenomenon. As a result, the effect of any slant TEC measurement may persist over a day or two causing the VTEC maps to be statistically more similar to each other than are VTEC maps separated by several more days. It is also easily noticeable that

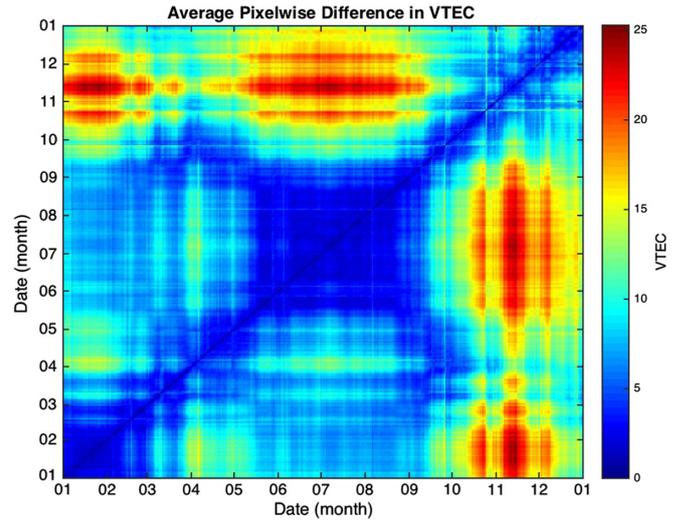


Fig. 2. Distance matrix for 2011 data set using the averaged absolute VTEC differences as a metric shown in a 2-dimensional color rendering in which red and blue represent, respectively, large and small distance between two data points. The most visible variations in the distance matrix, i.e., low values between data from June to September and high values between data from October to December, are most likely due to variations in geomagnetic activity levels.

substantially larger distances exist between maps for November and June than for January and June.

The goal of a cluster analysis based on a distance matrix is to identify data points that are surrounded by either a dense or sparse cluster of other data points. There are two main approaches to identify points in dense clusters. The first consists of counting the number of data points within a given radius r of a given data point. In the case of Figure 2, we can count the number of elements in a given column that have values less than r -TECU. The larger is the number of neighboring data points the more dense is the cluster of data around a given data point. We refer to the number of data points within r -TECU to a data point as the *popularity* of the data point. An example of popularity and location of closest neighbors is shown in Figure 3. Although not surprisingly most of the days within the 4.0-TECU radius are within a time interval of a few days to the given day, we note that not all days immediately before or after the given day are in the 4.0-TECU neighborhood. Conversely, there are also days several months before or after a given day that fall within a 4.0-TECU neighborhood. Alternatively, we could also calculate the largest distance, σ_i^n , of the n closest nearby data points to a given data point i . The smaller the value of σ_i^n the denser the population of data points around data point i . We refer to σ_i^n as the *radius* of the n -cluster around point i . The evaluation of σ_i^n is straightforward using the distance matrix D . In fact, we first sort the values in each column of the matrix D in increasing order. Then the $n + 1$ st value of the sorted distance matrix of column i is equal to σ_i^n . Figure 4 illustrates our approach for computing σ_i^{10} . A more direct depiction of the cluster radius is shown in Figure 5. We observe from Figures 4 and 5 that the cluster radii are substantially larger for time periods from March to May and from September to December than the time period from June to August. On the other hand, in each of these periods of time, there are days that stand out with a much larger cluster radius relative to other days. This may be due to enhanced geomagnetic activity during these months, that create anomalous

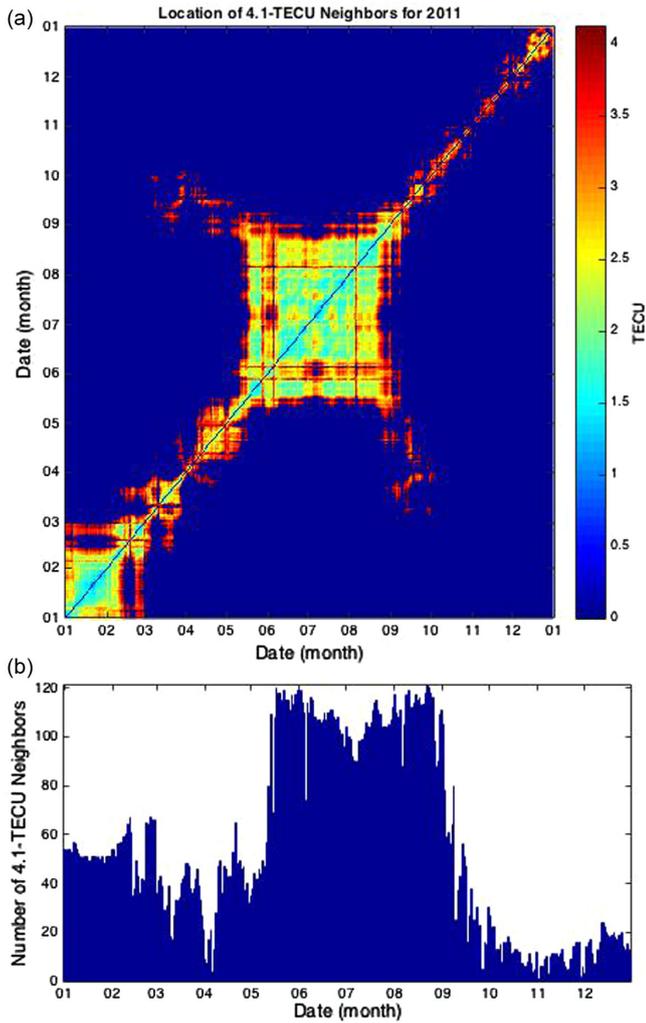


Fig. 3. (a) All entries of the distance matrix larger than 4.0-TECU are set to zero to help highlight the location of data points within 4.0-TECU radius to a data point. (b) The popularity of the dates in 2011.

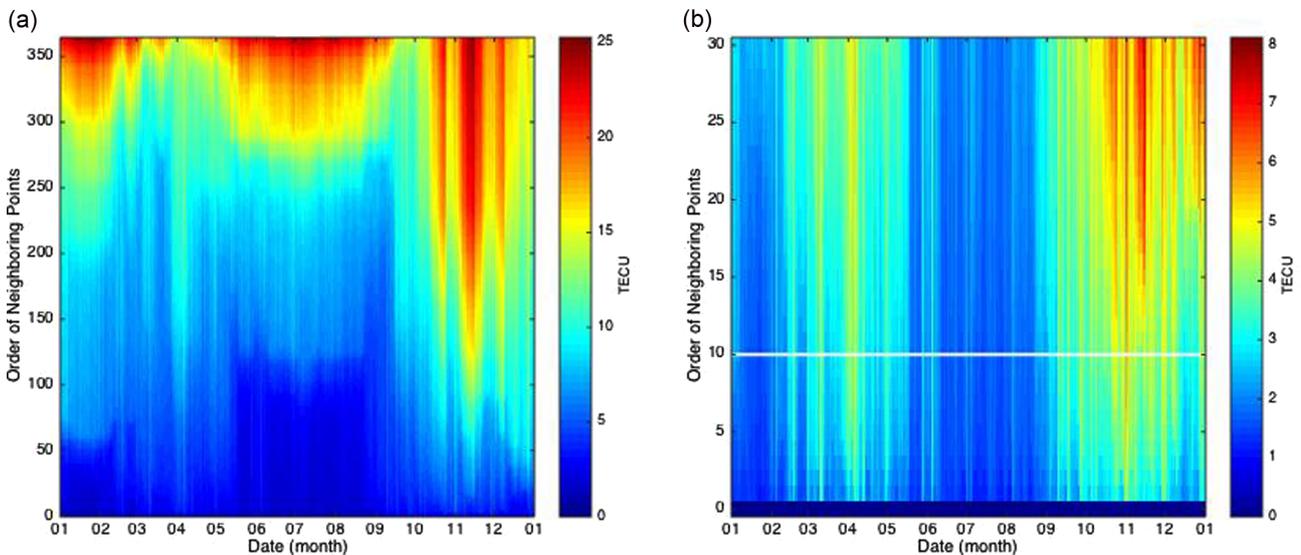


Fig. 4. In (a) the values in each column of the distance matrix D shown in Figure 2 are sorted in increasing order. Panel (b) shows only the first 31 rows of the sorted distance matrix. The color above the horizontal white line indicates the 10-cluster radius.

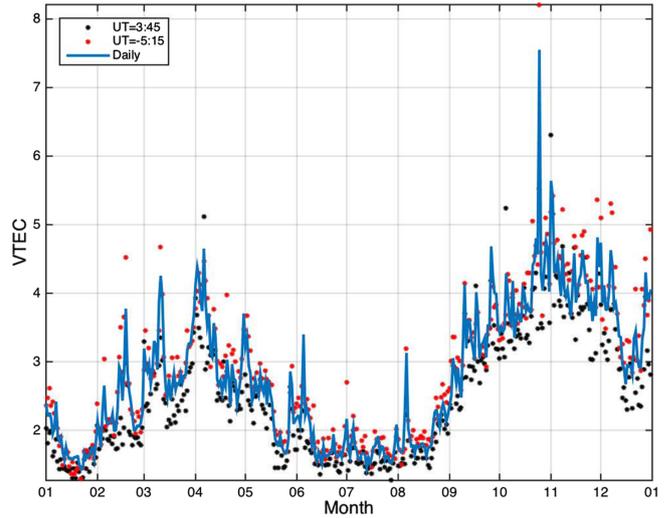


Fig. 5. Cluster radius σ_i^{10} for 2011 data set is computed using all VTEC maps in a day (solid line) or using maps for selected UTs only (5:15 red and 3:45 black dots).

ionospheres. Assuming that the metric we use is a valid way of comparing ionospheric conditions between any 2 days, the cluster radius shown in Figure 5 provides an indication of how unusual a given day's data is in relationship to other data points in the set. In Section 3.4, we shall present a method for identifying quiet and anomaly days using the cluster radii of the data points. However, in the next subsection, we first discuss the different metrics or distance measures we have considered in our analyses.

3.2. Metrics for comparing two VTEC maps

In the previous subsection the metric used in the examples presented was the average of the absolute value of pixelwise differences between 2 days of VTEC data. However, there are many ways to compare two sets of VTEC maps. In particular, we can define the following metrics.

dVTEC/dLat Maps from GIM for 10/07/11-10/08/11

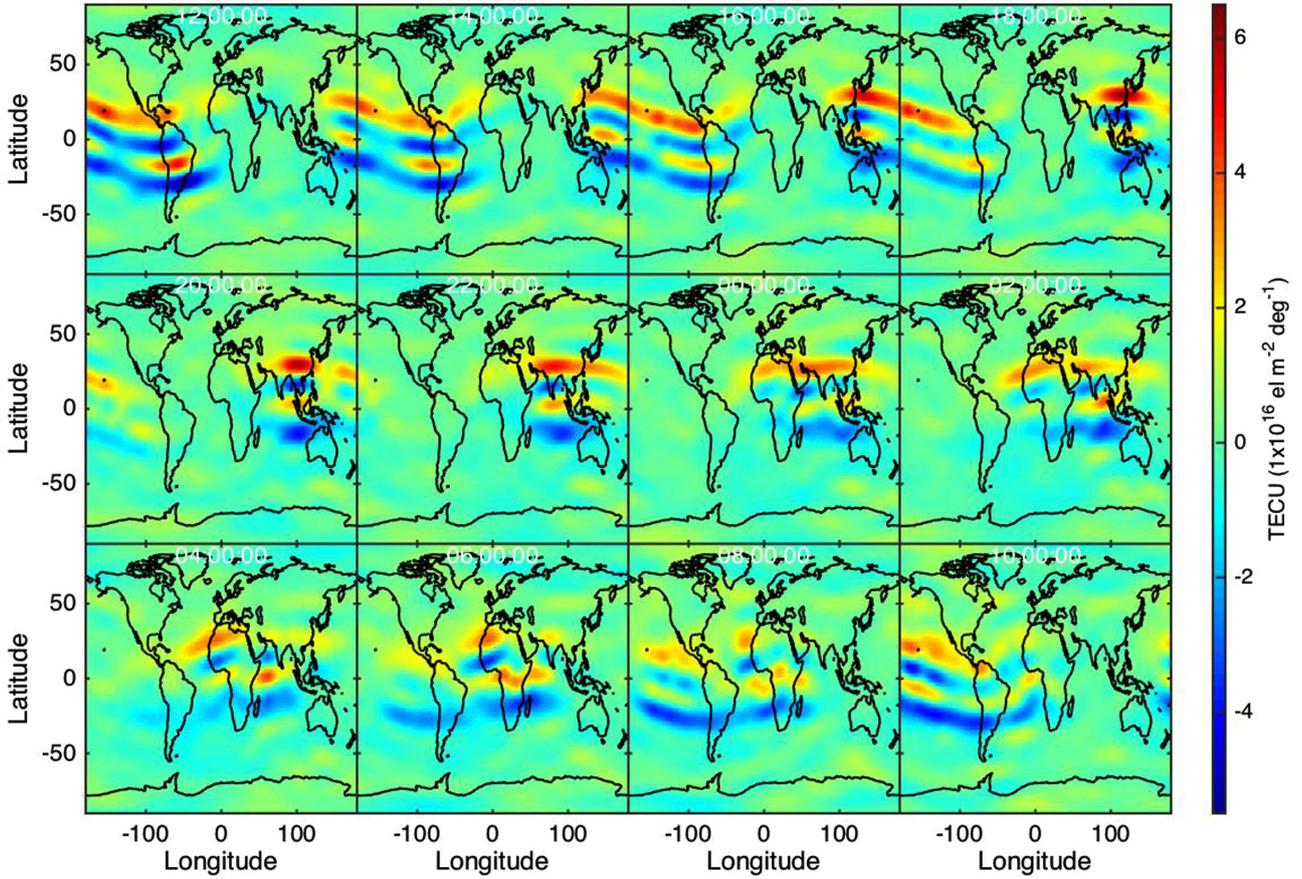


Fig. 6. Gradient in the latitude direction of VTEC maps for October 7–8, 2011 highlights the equatorial anomaly region.

$$d_{i,j}^{\text{abs}} = \frac{1}{N} \sum_{k=1}^N |v_k^i - v_k^j|, \quad (1)$$

$$d_{i,j}^{\text{rel}} = \frac{1}{N} \sum_{k=1}^N \left| \frac{v_k^i - v_k^j}{\max(|v_k^i|, |v_k^j|)} \right|, \quad (2)$$

$$d_{i,j}^{\text{relthr}} = \frac{1}{N} \sum_{k=1}^N \left| \frac{v_k^i - v_k^j}{\max(|v_k^i|, |v_k^j|, v_{\text{threshold}})} \right|, \quad (3)$$

where v_k^i represents the k -value of i -day of VTEC maps and N is the total number of VTEC values in these maps. The average of the absolute differences in VTEC, $d_{i,j}^{\text{abs}}$ in (1), is one of the most direct ways of measuring the differences between VTEC maps. However, the typical values of VTEC can vary significantly and systematically. For example, at local noon time near the equator, the VTEC value is much higher than the values in predawn high-latitude regions. As a result, $d_{i,j}^{\text{abs}}$ is always dominated by the differences in VTEC in the equatorial region. Conversely, the average of the relative difference, $d_{i,j}^{\text{rel}}$ in Eq. (2), normalizes the VTEC difference by the absolute value of VTEC. However, when VTEC is very low, as they are in predawn high-latitude regions, the difference in VTEC maps may be primarily due to VTEC measurement uncertainties. Indeed, when normalized by small VTEC values in these

regions, division by a relatively small value will tend to magnify measurement error. As a result, $d_{i,j}^{\text{rel}}$ could be artificially dominated by differences in regions with very low VTEC values. An intermediate approach is given in (3) where the normalization factor is at least as large as a pre-set threshold value $v_{\text{threshold}}$. In fact, $d_{i,j}^{\text{relthr}}$ is simply a multiple of $d_{i,j}^{\text{abs}}$ when $v_{\text{threshold}}$ is very large and $d_{i,j}^{\text{relthr}}$ is equal to $d_{i,j}^{\text{rel}}$ when $v_{\text{threshold}}$ is very small.

In addition to comparing the differences in the VTEC values, several researchers of ionosphere anomalies have indicated (see Luo et al. 2005) that differences in spatial and temporal gradient of VTEC could also be significant indicators of ionosphere anomalies. As a result, we can also define a metric that compares the gradients of VTEC maps by

$$d_{i,j}^{\text{H}_1} = \frac{1}{N} \sum_{k=1}^N \left[|D_{\text{lat}} v_k^i - D_{\text{lat}} v_k^j| + |D_{\text{lon}} v_k^i - D_{\text{lon}} v_k^j| + |D_{\text{UT}} v_k^i - D_{\text{UT}} v_k^j| \right], \quad (4)$$

where D_{lat} , D_{lon} , and D_{UT} are weighted difference operators in latitude, longitude, and UT, respectively. The selection of weights for these difference operators can definitely alter the results of an analysis. In our investigation, we select the weights to make the contributions of each of the three differences have comparable magnitude. In fact, each of the differences $D_{\text{lat}} v_k^i$, $D_{\text{lon}} v_k^i$, and $D_{\text{UT}} v_k^i$ is affected by different driving forces of the ionosphere. Figures 6–8 show examples

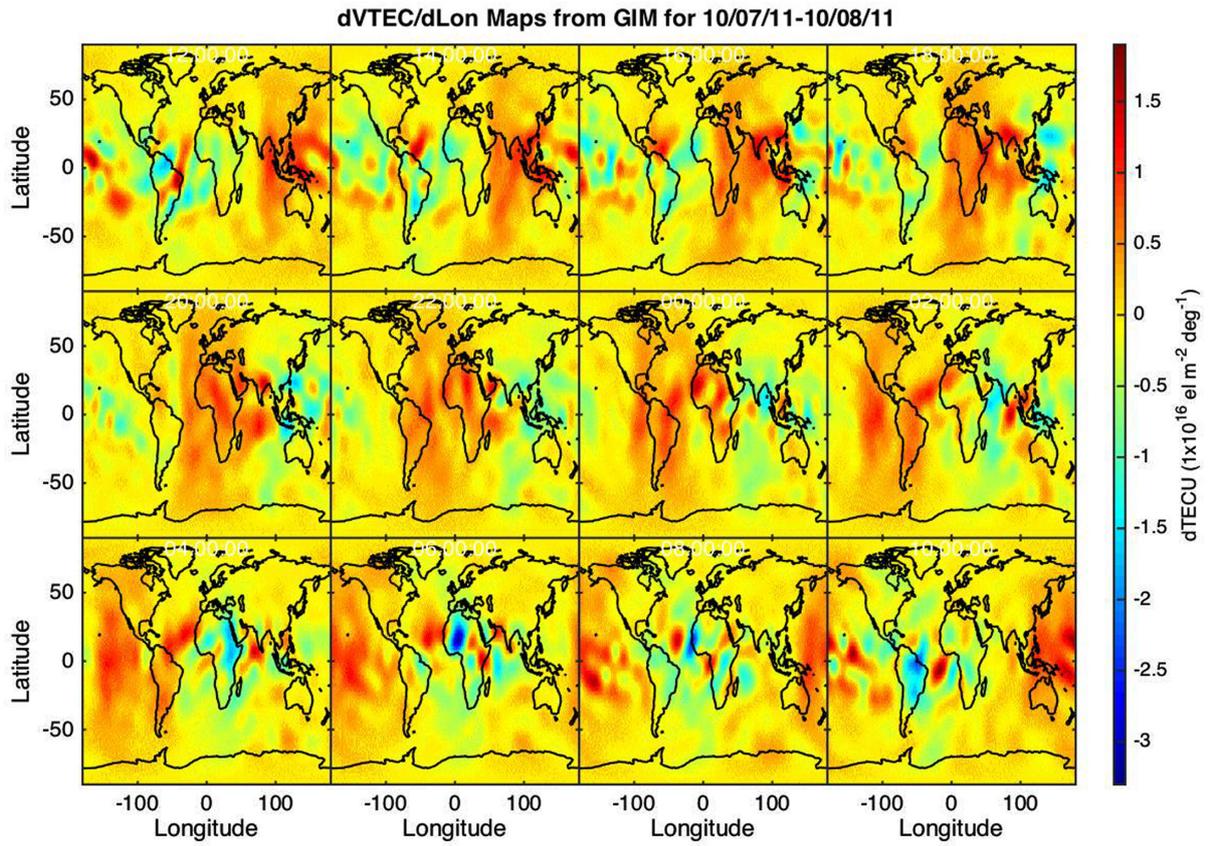


Fig. 7. Gradient in the longitude direction of VTEC maps for October 7–8, 2011 includes local time effects.

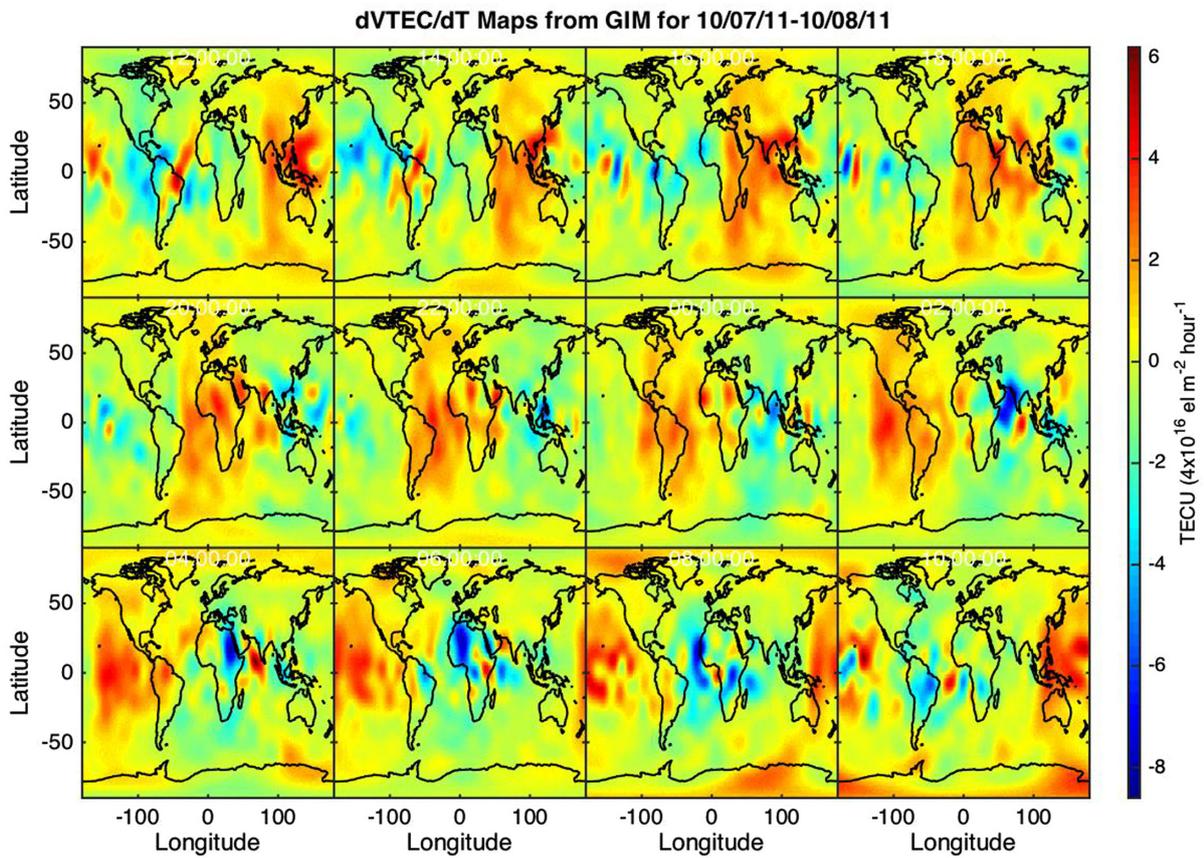


Fig. 8. Gradient in UT of VTEC maps for October 7–8, 2011 may also include variation in local time since ionosphere features are strongly dependent on solar zenith and azimuthal angles.

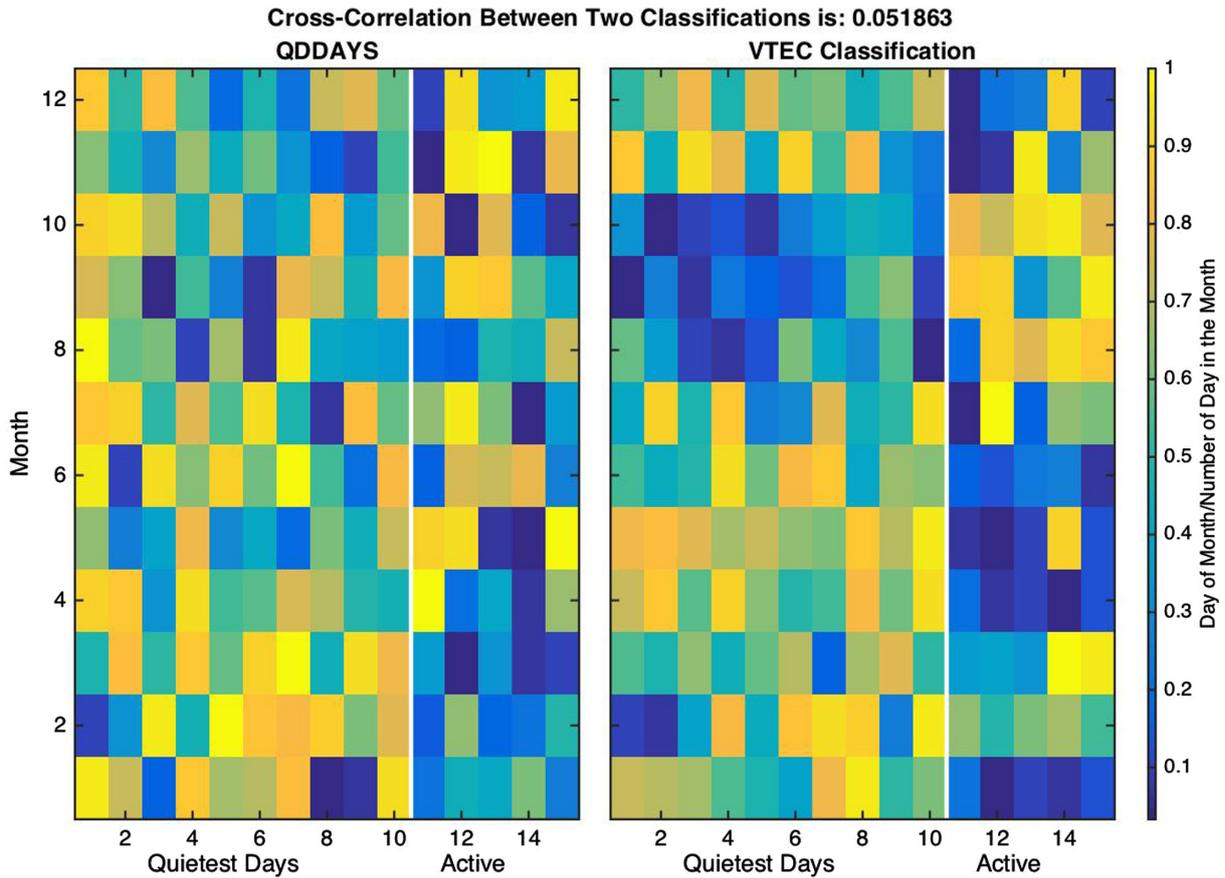


Fig. 9. Correlation between geomagnetic and VTEC Q-days and D-days is not very strong for the 2011 data set.

of these differences. For example, geo-electrical field strength strongly influences the ions' $E \times B$ velocity in the mid- and low-latitude regions. The resulting equatorial anomaly leads to the characteristic latitude-dependent features in the maps of $D_{lat} v_k^i$ shown in Figure 6.

In many investigations of the effects of solar and geomagnetic storms (see Tsurutani et al. 2004; Lee et al. 2011; Tsurutani et al. 2012) different latitude regions on the globe show different sensitivity. As a result, we can evaluate the metrics $d_{i,j}^{abs}$, $d_{i,j}^{rel}$, $d_{i,j}^{relthr}$, and $d_{i,j}^{H_1}$ by summing up the differences only over regions of interest. Specifically, we can examine the difference only over high-latitude regions; mid- and low-latitude regions; for selected local times, and for selected UTs. As a result, the combination of general metrics defined in Eqs. (1)–(4) and subsets of VTEC maps provide a large selection of metrics to compare VTEC maps for any 2 days. These metrics already include most commonly used approaches for identifying effects of solar and geomagnetic disturbances. In our subsequent statistical analysis, the sensitivity of these metrics to space weather disturbances is examined.

3.3. Correlation between geomagnetic Q-days and VTEC quiet conditions

The n -cluster radius for a given day derived in Section 3.1 is viewed as an indicator of how unusual are the VTEC maps for this day. It is therefore interesting to examine whether the

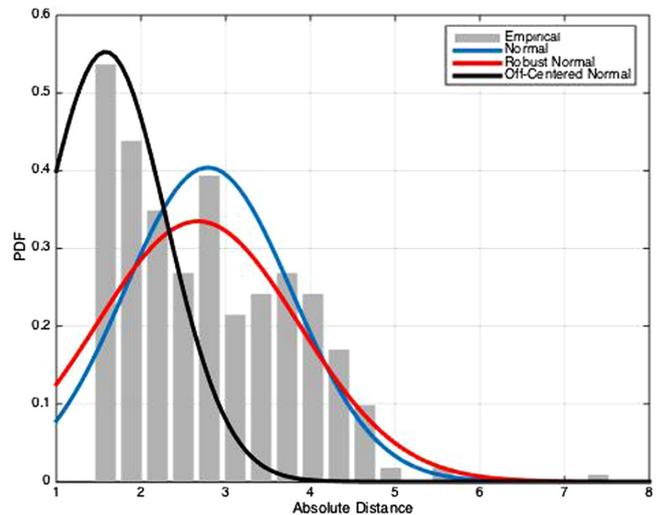


Fig. 10. Distribution of the 10-cluster radii associated with the average absolute difference in VTEC for 2011 deviates significantly from a normal distribution. The solid gray bars correspond to the empirical distribution of data. The three smooth curves represent three different attempts to approximate the empirical distribution by normal distributions with parameters derived by three different algorithms: standard (blue), robust algorithm (red) where data points far away from median are removed, and off-centered normal (black) where the median is replaced by the highest point of the empirical distribution in the rejection of outliers.

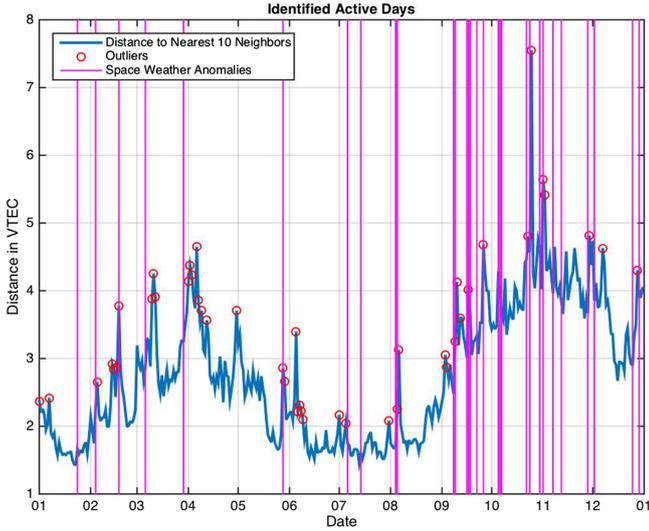


Fig. 11. The 10-cluster radius for the 2011 data set and identified outliers are shown in solid blue line and red circles. The vertical purple lines indicate days of ICME as compiled by Richardson & Cane (2010).

observed unusual nature of a particular map or set of maps is in anyway related to the geomagnetic activity level for this day. In particular, if for each month we select the 10 days with the smallest cluster radius, referred to as as *VTEC Q-days*, and 5 days with the largest cluster radius, referred to as as *VTEC*

D-days, we would like to determine whether or not there is a substantial overlap with the *Q-days* and the *D-days* compiled by GFZ-Potsdam. A graphic representation of our analysis is shown in Figure 9. The two panels in Figure 9 represent the geomagnetic *Q-days* and *D-days* in comparison to VTEC *Q-days* and *D-days*. Each row in these tables represents 10 *Q-days* and five *D-days* of a month. The colors of the small rectangles represent the position of a *Q-day* or *D-day* within a month. The dark blue color represents the first of a month and the bright yellow color corresponds to the last day of the month. The correlation of two image panels is an indication of the correlation between the two sets of *Q-days* and *D-days*. At least for the 10-cluster radius derived from average absolute VTEC differences, the correlation of the two panels does not appear to be very strong.

The level of correlation of the *n*-cluster radius determined *Q-days* and *D-days* using different comparison metrics can serve as a useful indicator of the nature of the sensitivity of ionosphere to geomagnetic disturbances. We shall present a comparison of the correlation coefficients for several distance metrics in the next section.

3.4. Identification of outliers

As explained in Section 3.1, our ionosphere anomaly classification approach relies on the cluster radius determined using the metrics defined in Section 3.2. In general, data points with exceptionally large cluster radii are considered as outliers. A common approach in outlier identification (see for example,

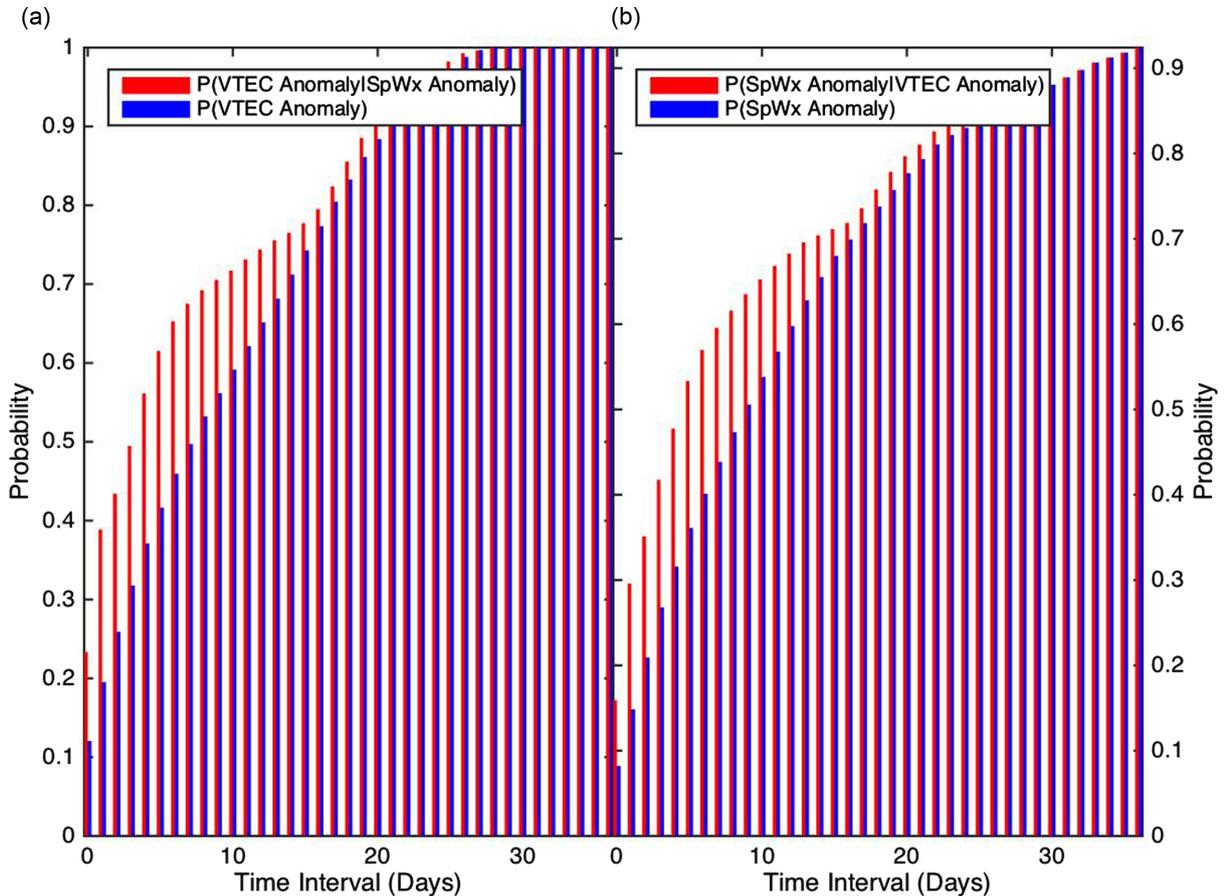


Fig. 12. Panel on the (a) shows a comparison between $P_{VTEC}(k)$ and $P_{VTEC}(k|ICME)$. Panel on the (b) shows a comparison between $P_{ICME}(k)$ and $P_{ICME}(k|VTEC)$ for the 2011 data set.

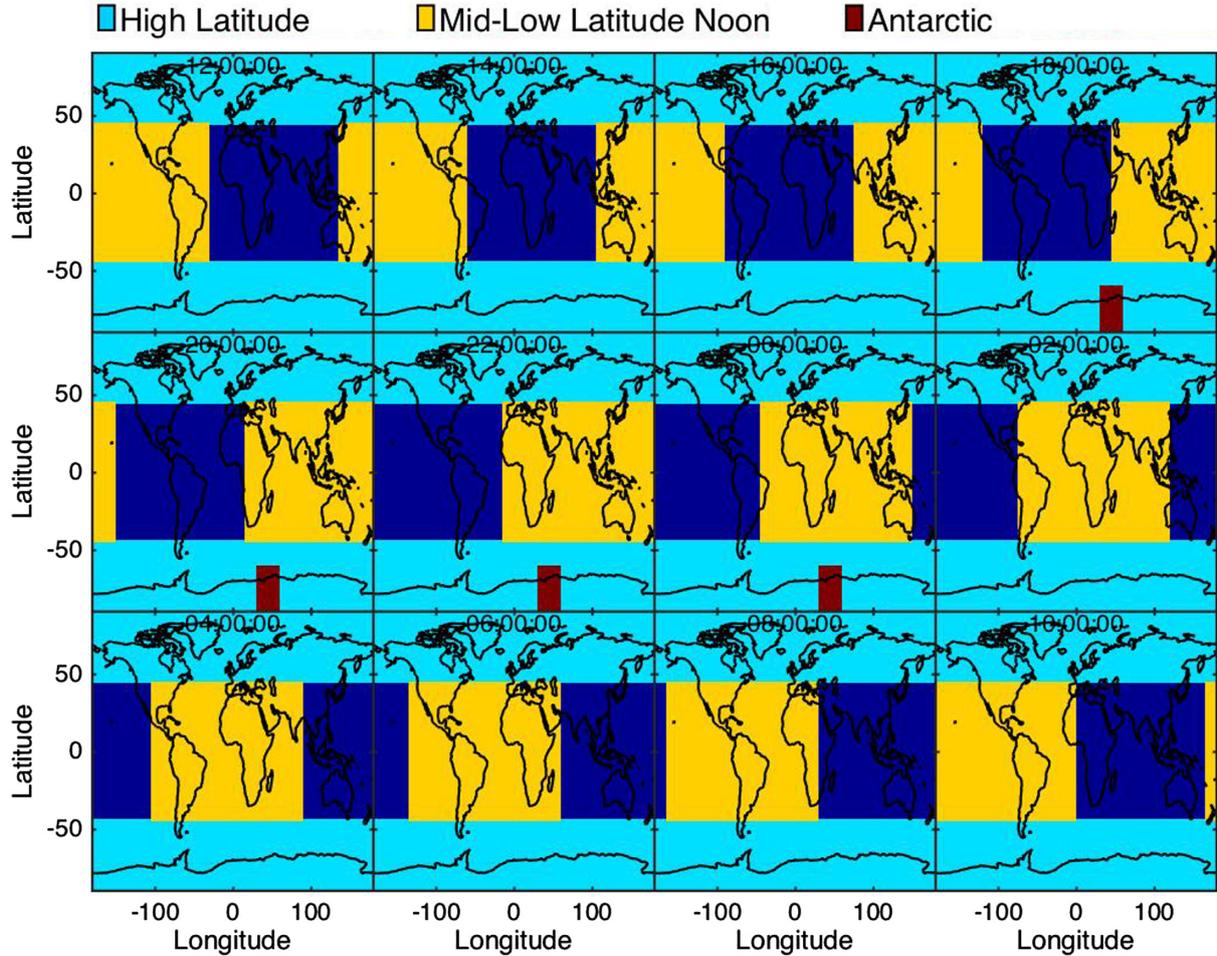


Fig. 13. The three masked regions are represented with different colors, light blue for high-latitude, yellow for mid-low-latitude noon, and brown for antarctic.

Sodemann et al. 2012) involves the comparison of a sampled empirical distribution with a normal approximation and labels the unexpectedly larger number of data points at the tails of the normal approximation as outliers. More precisely, for any level of significance α , a data point that is larger than $F^{-1}(1 - \alpha; \mu, \sigma)$, where F represents the cumulative normal probability distribution with expected value μ and standard deviation σ , are classified as an outlier. A close examination of the empirical distribution of the cluster radius in Figure 10 reveals significant deviation from a normal distribution. This is not entirely surprising since we can already observe in Figure 5 that during the months of September through November, the cluster radius of all points is substantially higher than during the period of June–August. As shown in Figure 10 three normal approximations of the empirical distribution of the 10-cluster radius are evaluated. The blue curve represents a normal approximation with expected value and standard deviation equal to the sample mean and standard deviation, respectively. This approximation leads to the classification of the large number of data points with small cluster radii as outliers. For the red curve, we use the median of the sample as the expected value for the normal approximation. If we let $r_q(p)$ represent the value of the cluster radius at quantile p , then the standard deviation for the normal approximation is given by

$$\sigma_r = \frac{\min(r_q(0.5 + p) - r_{\text{median}}, r_{\text{median}} - r_q(0.5 - p))}{F^{-1}(0.5 + p; 0, 1)}. \quad (5)$$

Since quantiles of a data set are in general less sensitive to outliers, σ_r is viewed as a more robust estimation of the standard deviation than the sample standard deviation. However, in the case of the cluster radii shown in Figure 10, the trimodal form of the empirical distribution makes σ_r even larger than the sample standard deviation. An alternative approach is to use the center of the bin in the histogram of the data set with the highest number of data points as an approximation to the expected value and then approximate the standard deviation by

$$\sigma_{\text{max}} = \frac{\Delta r_{\text{bin}}}{2F^{-1}(0.5 + n_{\text{bin}}/(2N); 0, 1)}, \quad (6)$$

where Δr_{bin} and n_{bin} are, respectively, the width and the number of points in the bin with the highest number of data points, and N is the size of the data set. This approximation is given by the black curve. It is evident that this approximation may lead to the classification of a large number of data points as outliers.

In order to avoid under- and overclassification of outliers, we use a hybrid global and local normal approximation of data. This approach consists of two levels of classifications. The first level relies on the normal approximation of the entire data set with mean μ equal to the sample mean, standard deviation σ equal to the sample standard deviation, and the labeling of all data points with cluster radius larger than $F^{-1}(1 - \alpha_g; \bar{r}, \sigma)$ as outliers. Then, for data in a moving time

Table 1. Identification numbers of metrics for VTEC map comparison.

Distance metric	Non-mask	High-latitude	Mid-low-latitude noon	Antarctic
$d_{i,j}^{\text{abs}}$	1	5	9	13
$d_{i,j}^{\text{rel}}$	2	6	10	14
$d_{i,j}^{\text{relthr}}$	3	7	11	15
$d_{i,j}^{\text{H}}$	4	8	12	16

window of width N_s days, a normal approximation for the empirical distribution of the subset of data is derived with μ and σ equal to the subsample mean and standard deviation, respectively. Then, all data points with cluster radius larger than $F^{-1}(1 - \alpha; \bar{r}, \sigma)$ are labeled as outliers. This permits outlier identification within a time neighborhood of several days. An example of this classification is shown in Figure 11.

Since the n -cluster radii depend on the metric used to measure the difference between the data points, a variety of classifications of outliers are possible.

3.5. Empirical conditional probabilities for VTEC anomalies and ICME conditions

The VTEC outliers identified using approaches presented in the previous section enable us to establish empirical probabilities of coincident occurrence of ionosphere VTEC anomalies and the ICMEs. In fact, let $n_{\text{ICME}}(k)$ be the number of days identified by Richardson & Cane (2010) that an ICME is observed within the previous k days. Although several parameters are provided in the list compiled by Richardson & Cane (2010), in our analysis we choose not to stratify our analysis results by these parameters at this moment. Let $n_{\text{VTEC}}(k)$ be the number of days for which an anomaly day in the previous

k days is identified by our method, and let $n_{\text{ICME+VTEC}}(k)$ be the number of days that both (1) an ICME is observed and (2) a day is identified as anomalous in the previous k days. The probability $P_{\text{ICME}}(k)$ that an ICME is observed within the previous k days is given by n_{ICME}/N where N is the number of days in a data set. For 2011, $N = 365$. Similarly if we are interested in forecasting a VTEC anomaly using the observation of an ICME, we first note that the sensitivity $P_{\text{VTEC}}(k|\text{ICME})$ of VTEC to ICME is the conditional probability of detecting a VTEC anomaly given that an ICME is observed in the previous k days. This quantity can be estimated as

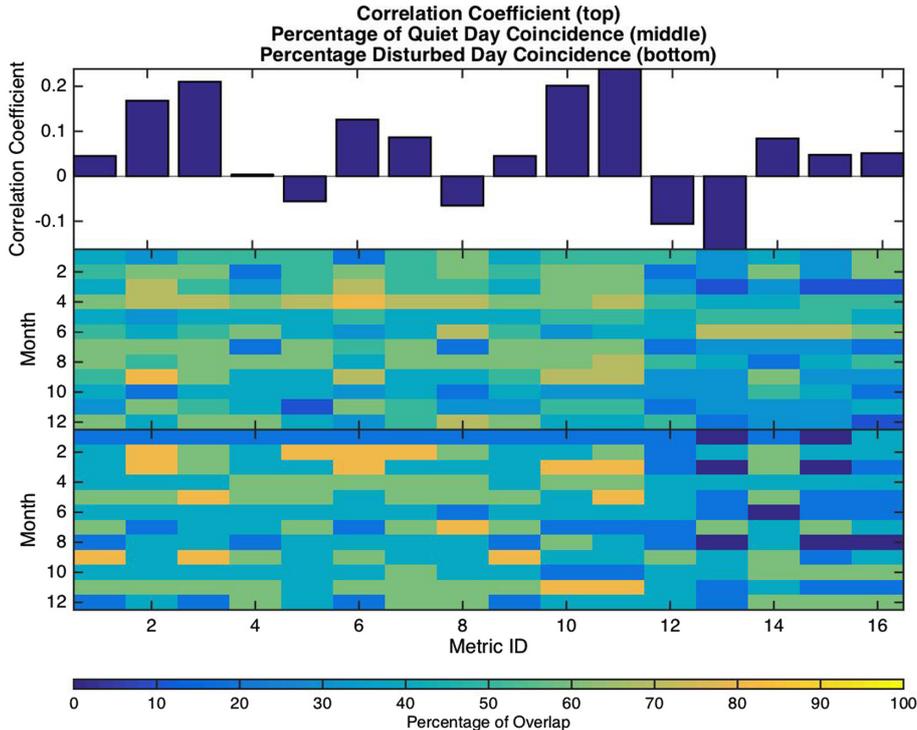
$$P_{\text{VTEC}}(k|\text{ICME}) = \frac{n_{\text{ICME+VTEC}}(k)}{n_{\text{ICME}}(k)}. \quad (7)$$

Conversely, the posterior probability of the VTEC anomalous response $P_{\text{ICME}}(k|\text{VTEC})$ to ICME is given by the conditional probability of observing an ICME given that an anomalous VTEC map is identified in the last k days. That is,

$$P_{\text{ICME}}(k|\text{VTEC}) = \frac{n_{\text{ICME+VTEC}}(k)}{n_{\text{VTEC}}(k)}. \quad (8)$$

The reason we choose to examine the coincident occurrence between the two types of anomalies over a time interval of several days, is so that we can attempt to account for the delayed response of ionosphere anomalies caused by an ICME. In this analysis, for simplicity, we do not require that an ICME must occur prior to the VTEC anomaly. A comparison between $P_{\text{VTEC}}(k)$ and $P_{\text{VTEC}}(k|\text{ICME})$ provides an indication of how predictive an ICME observation is for an occurrence of a VTEC anomaly. Figure 12 shows a comparison between the marginal probabilities $P_{\text{VTEC}}(k)$, $P_{\text{ICME}}(k)$ and the conditional probabilities $P_{\text{VTEC}}(k|\text{ICME})$, $P_{\text{ICME}}(k|\text{VTEC})$, respectively.

At a very basic level, the comparison in Figure 12 illustrates two obvious facts. First, there is not always an


Fig. 14. Summary of Q/D-day classification for 2011.

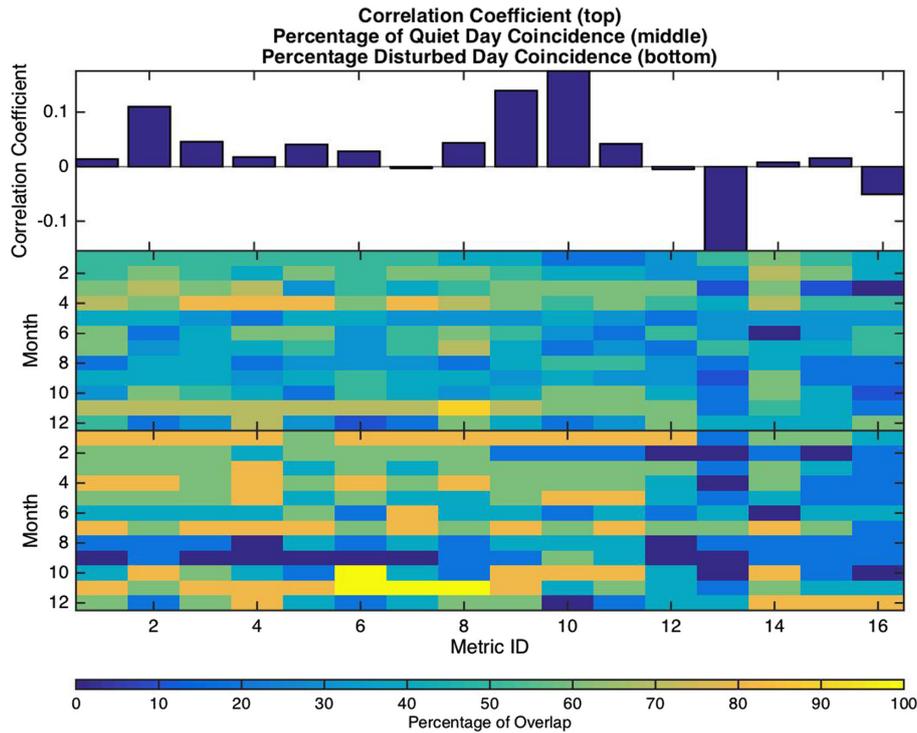


Fig. 15. Summary of Q/D-day classification for 2012.

occurrence of an ionosphere anomaly following an observation of an ICME. Although our definition of an ionosphere anomaly is dependent on many subjectively selected parameters such as the choice of a metric, the cluster size, and outlier detection threshold, the selection of these parameters is motivated by previous investigations (see for example Tsurutani et al. 2004; Burns et al. 2012; Tsurutani et al. 2012) in which similar metrics are used to explain the effects of an ICME. Conversely, it can also be observed that not all occurrences of anomalous VTEC maps are preceded by an ICME. A more comprehensive discussion of our findings will be presented in the next section.

4. Analysis of ionosphere anomalies and their connection with space weather events

As we have discussed in the preceding section, there are a multitude of ways that ionosphere VTEC maps can be compared and anomalies can be identified. In this section, we present our analyses for data sets from 2011 and 2012. Since in this study we have restricted ourselves to anomaly detection based on a single metric, one key issue we want to address is which metric provides the greatest sensitivity to either ICME or geomagnetic storms.

In addition to the general metrics defined in Eqs. (1)–(4), we also restrict the summations in these definitions to three different subsets, or *masks*, to create additional metrics. The primary purpose of restricting our comparison between two sets of VTEC maps to masked region is to potentially increase the sensitivity of the differences to space weather events. Indeed, many research efforts on thermosphere-ionosphere disturbances due to space weather events have focused on specific regions as well (Tsurutani et al. 2004; Lee et al. 2011; Tsurutani et al. 2012). These three masks are labeled as high-latitude, mid-latitude noon, and antarctic. Figure 13 provides a depiction of these regions. The two masked regions,

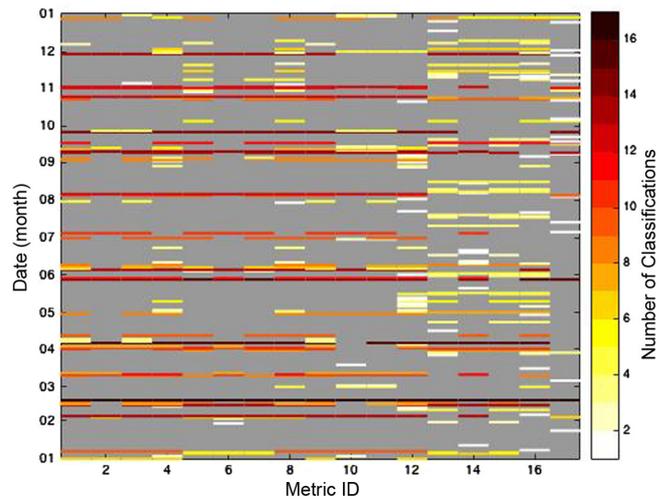


Fig. 16. Summary of VTEC anomalies identified by all 16 comparison metrics for 2011.

high-latitude and antarctic, are defined for fixed universal times. The mid-low-latitude region is defined for local time between 10:00 and 22:00 local time. As a result, a total of 16 different metrics are used for the classification of ionosphere anomalies. These metrics constitute a representative sample of the most often used criteria for identifying unusual features in a VTEC map. For future reference we assign a number to each of these metrics in Table 1.

For each of these 16 metrics, we evaluate the cross-correlation coefficients of the tables of Q/D-days determined by Kp index and that determined by the 10-cluster radius as presented in Section 3.3. Additionally, we tabulated the number of Q-days and D-days for each month that are identified by both classification methods as Q-days and D-days. The results are shown in Figures 14 and 15. The results in Figures 14 and

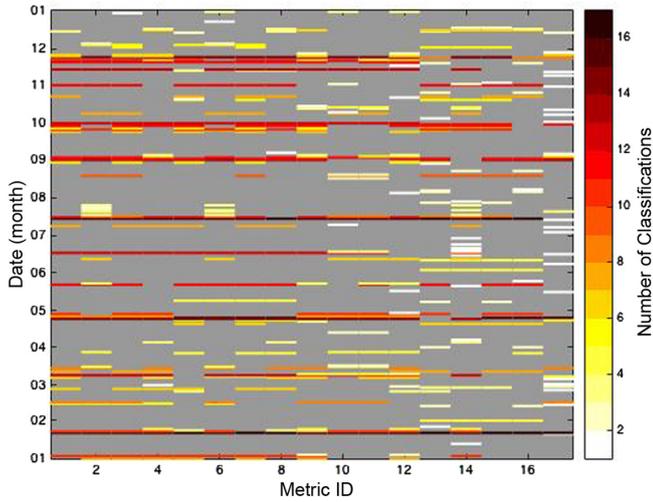


Fig. 17. Summary of VTEC anomalies identified by all 16 comparison metrics for 2012.

15 indicate that for most of the metrics we considered, although the correlations between the Q/D-tables are positive, they are, in general, quite weak. This observation is also confirmed by the examination of the percentage of quiet days that are classified as such by both methods. In general, the percentages of common classification for D-days are somewhat higher than those for Q-days. Also, metrics 1 and 2 seem to have higher sensitivity than do metrics 12–16.

For the classification of VTEC anomalous days, we use the identical parameters for outlier classifications with the

significance level of classification selected at 2.5% and a local outlier detection interval of 36 days. In **Figures 16 and 17** each day of the year is represented by a row of pixels in the image and the first 16 columns represent the classifications by the 16 metrics. When a day is classified by the k th metric as disturbed, the pixel on the k th column of the corresponding row is given a non-gray color. In fact, the color of the pixel corresponds to the total number of metrics by which the specific day is classified as disturbed. For example, if a white pixel is seen in a row which indicates that only one of 16 metrics classifies the day as disturbed, this must be the only non-gray pixel on this row. Conversely, if a black pixel is seen in a row which indicates all 16 metrics classify the day as disturbed, all 16 first pixels on the row must also be black. In column 17, the ICME days are shown. These figures show that although there are significant overlaps in the days identified by different metrics as anomalous, there are also significant differences among the classifications as well.

In **Figures 18 and 19** we show the marginal probability of a VTEC anomaly over an interval of width of n days, as well as, the conditional probabilities given the occurrence of an ICME during the same period. The results in **Figures 18 and 19** confirm our observation in **Section 3.5** that the majority of the comparison metrics show a certain degree of sensitivity to the occurrence of an ICME. In fact, the probability of detecting a VTEC anomaly based on several comparison metrics increases significantly from the marginal probability when an ICME is observed. While some of these increases are observed for both the 2011 and 2012 data sets, the behavior can be different for other metrics. For example, for metric 5 (or 3), a substantial increase in probability is observed for the 2011 data set, while the 2012 data appears to be quite insensitive to the occurrence of an ICME.

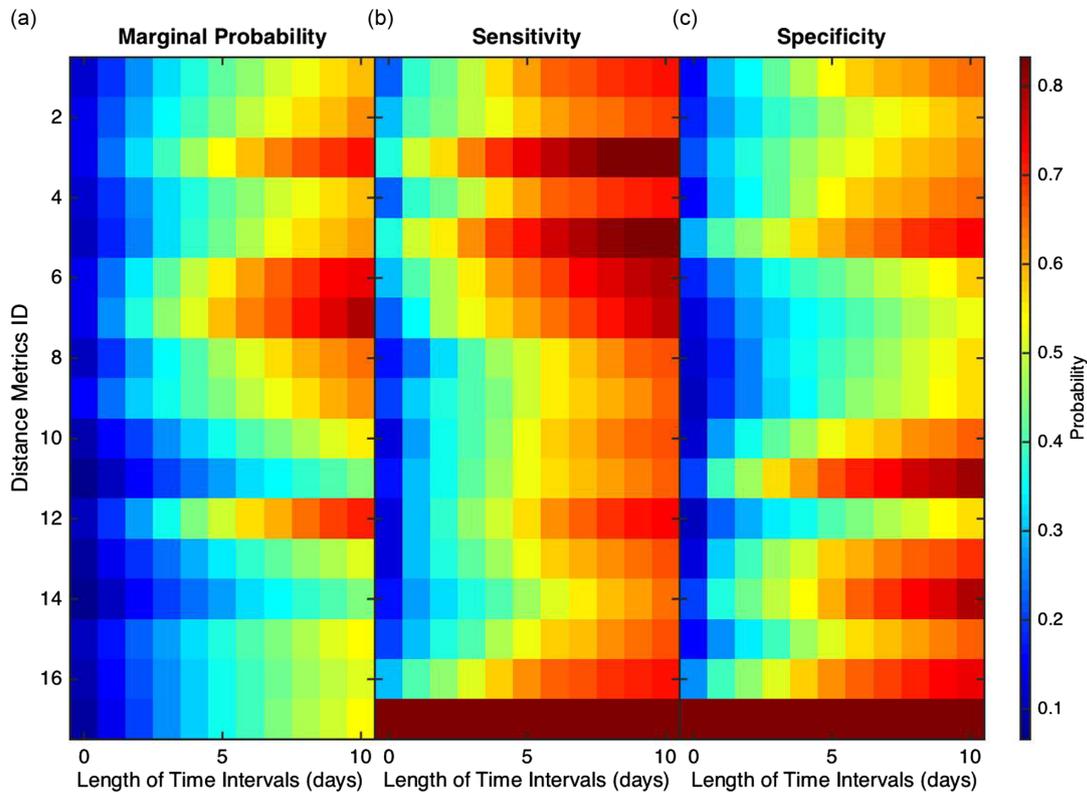


Fig. 18. Marginal probability of observing a VTEC anomaly during a time interval of n days (a), conditional probability of a VTEC anomaly given an ICME is observed during the same period, i.e., sensitivity (b), and conditional probability of observing an ICME given that a VTEC anomaly is observed during the time interval, i.e. posterior probability (c) for the 2011 data set. Note that the last row represents the probability of observing an ICME.

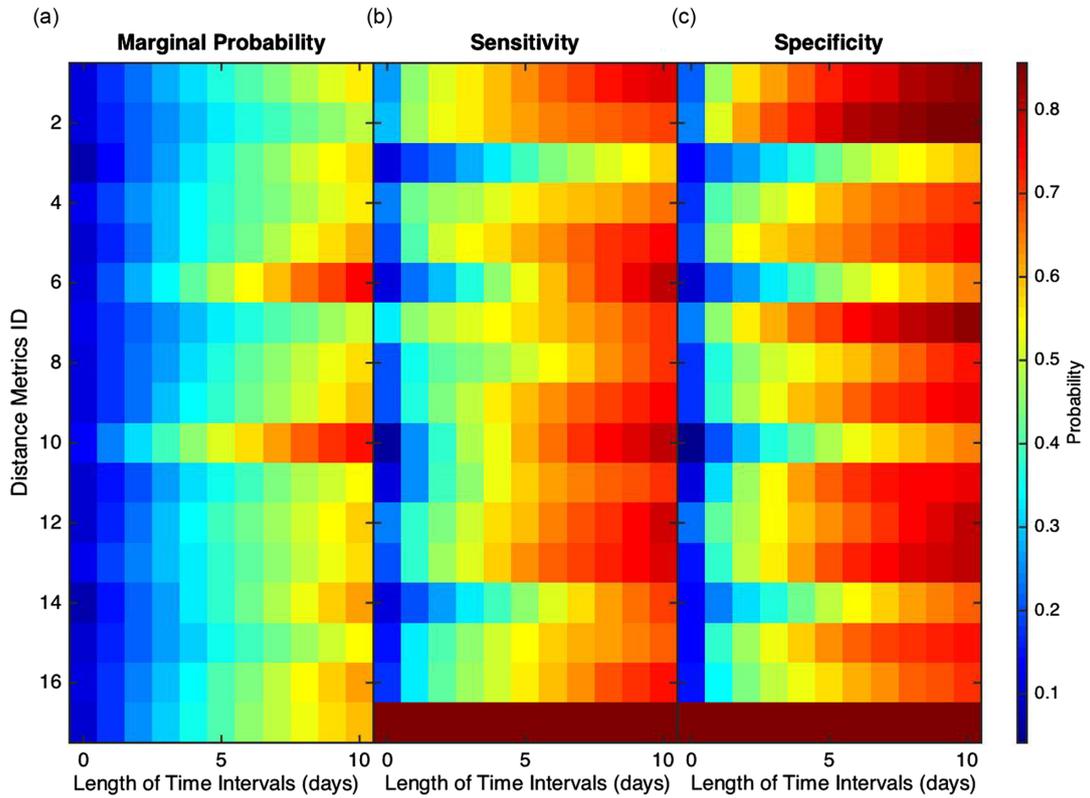


Fig. 19. Marginal probability of observing a VTEC anomaly during a time interval of n days (a), conditional probability of a VTEC anomaly given an ICME is observed during the same period, i.e., sensitivity (b), and conditional probability of observing an ICME given that a VTEC anomaly is observed during the time interval, i.e. posterior probability (c) for 2012 data set. Note that the last row represents the probability of observing an ICME.

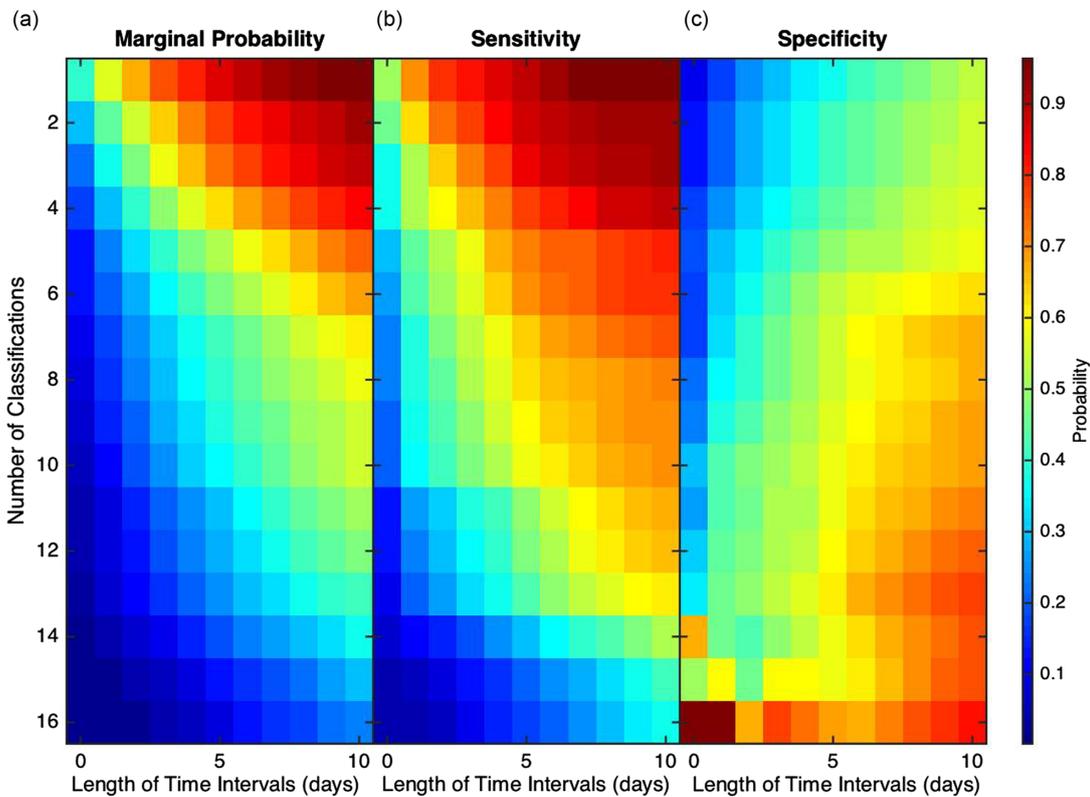


Fig. 20. Marginal probability of observing a VTEC anomaly during a time interval of n days (a), conditional probability of a VTEC anomaly given that an ICME is observed during the same period, i.e., sensitivity (b), and conditional probability of observing an ICME given that a VTEC anomaly is observed during the time interval, i.e. posterior probability (c) for the 2011 data set. Note that the row numbers represent the number of comparison metrics required for classification of a VTEC anomaly.

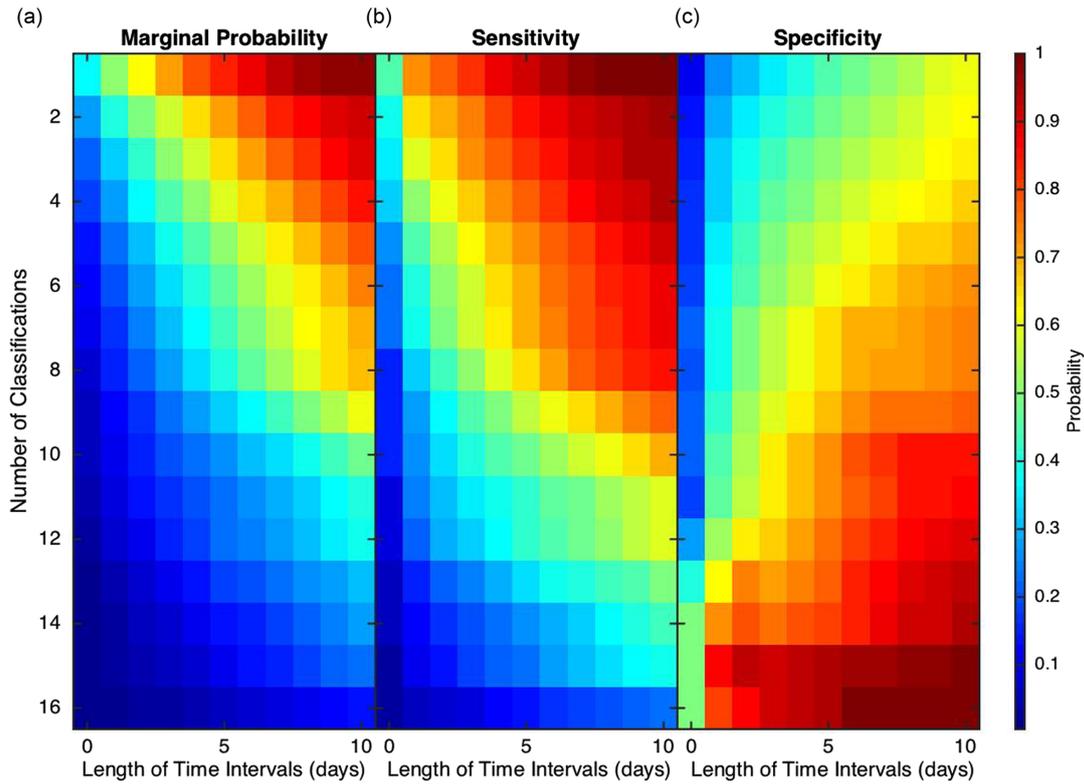


Fig. 21. Marginal probability of observing a VTEC anomaly during a time interval of n days (a), conditional probability of a VTEC anomaly given an ICME is observed during the same period, i.e., sensitivity (b), and conditional probability of observing an ICME given that a VTEC anomaly is observed during the time interval, i.e. posterior probability (c) for the 2012 data set. Note that the row numbers represent the number of comparison metrics required for classification of a VTEC anomaly.

In addition, to compute the marginal probability of observing a VTEC anomaly using a single comparison metric, we also computed the probability of observing a VTEC anomaly by at least k different comparison metrics for $k = 1, \dots, 16$. Naturally, as k increases, the marginal probability decreases because unless k comparison metrics all classified a data point as anomalous, the point will not be classified as anomalous. Figures 20 and 21 show the resulting marginal probabilities, as well as, the sensitivity and posterior probability of the anomalies to space weather events. The results in Figures 20 and 21 suggest that while the probability of identifying a VTEC anomaly by 4–10 metrics shows substantial sensitivity to the occurrence of an ICME, the classification by less than three metrics is overprescribed. These results also suggest that a more refined approach for using multiple comparison metrics has the potential to enhance both the sensitivity and posterior probability of VTEC anomaly detection. It is important to put the results presented in this section into perspective. First, the list of ICME events used in this analysis may include events that are not geo-efficient. Secondly, the differences in 2011 and 2012 results can simply be attributed to the difference in the phases of the solar cycle for these years.

5. Conclusion

The analysis results presented in the previous section constitute a demonstration of a technical approach for the statistical characterization of connections between space and ionosphere anomalies. Although these results already established quantitative estimates of sensitivity and specificity, these estimations

should be considered as preliminary due to the limited time span of data used in this study. However, the potential value of a quantitative characterization of the connections between space weather events and thermosphere-ionosphere anomalies is quite evident. For example, as indicated by the results in Figure 18, without knowledge of an ICME, the probability of observing a VTEC anomaly is below 10%. With the detection of an ICME, the probability of observing a VTEC anomaly can be, in some cases, increased to 40–50%. In fact, a much richer pool of solar and magnetosphere observations can be mined to provide a more detailed characterization of the relationship between space weather events and ionosphere anomalies. In our view, the ultimate goal of forecasting ionospheric weather that results from solar and magnetospheric disturbances is advanced by a quantitative understanding of the causal relationship between these types of anomalies.

Acknowledgements. Sponsorship of the Living With a Star Targeted Research and Technology NASA/NSF Partnership for Collaborative Space Weather Modeling grant is gratefully acknowledged. Portions of the research for this paper were performed at the Jet Propulsion Laboratory, California Institute of Technology under contract with NASA. The editor thanks two anonymous referees for their assistance in evaluating this paper.

References

- Burns, A.G., S.C. Solomon, L. Qian, W. Wang, B.A. Emery, M. Wiltberger, and D.R. Weimer. The effects of corotating interaction region/high speed stream storms on the thermosphere and ionosphere during the last solar minimum. *J. Atmos. Sol. Terr. Phys.*, **83**, 79–87, 2012.

- Chandola, V., A. Banerjee, and V. Kumar. Anomaly detection: a survey. *ACM Comput. Serv.*, **41** (3), 15, 2009.
- Datta-Barua, S., T. Walter, G.S. Bust, and W. Wanner. Effects of solar cycle 24 activity on WAAS navigation. *Space Weather*, **12** (1), 46–63, 2014, DOI: [10.1002/2013sw000982](https://doi.org/10.1002/2013sw000982).
- Fagundes, P.R., M.T.A.H. Muella, J.A. Bittencourt, Y. Sahai, W.L.C. Lima, et al. Nighttime ionosphere-thermosphere coupling observed during an intense geomagnetic storm. *Adv. Space Res.*, **41**, 539–547, 2008.
- Forbes, J.M., S.E. Palo, and X. Zhang. Variability of the ionosphere. *J. Atmos. Sol. Terr. Phys.*, **62**, 685–693, 2000.
- Grnitz, N., M. Kloft, K. Rieck, and U. Brefeld. Toward supervised anomaly detection. *JAIR*, **46**, 235–262, 2013.
- Hibberd, F.H. Day-to-day variability of the S_q geomagnetic field variation. *Aust. J. Phys.*, **34** (1), 81–90, 1981.
- Immel, T.J., and A.J. Mannucci. Ionospheric redistribution during geomagnetic storms. *J. Geophys. Res. [Space Phys.]*, **118**, 7928–7939, 2013, DOI: [10.1002/2013JA018919](https://doi.org/10.1002/2013JA018919).
- Immel, T.J., S.B. Mende, M.E. Hagan, P.M. Kintner, and S.L. England. Evidence of tropospheric effects on the ionosphere. *EOS*, **90** (9), 69–70, 2009.
- Komjathy, A., L. Sparks, B.D. Wilson, and A.J. Mannucci. Automated daily processing of more than 1000 ground-based GPS receivers for studying intense ionospheric storms. *Radio Science*, **40**, RS6006, 2005a, DOI: [10.1029/2005RS003279](https://doi.org/10.1029/2005RS003279).
- Komjathy, A., L. Sparks, A.J. Mannucci, and A. Coster. The ionospheric impact of the October 2003 storm event on wide area augmentation system. *GPS Solutions*, **9** (1), 41–50, 2005b, DOI: [10.1007/s10291-004-0126-2](https://doi.org/10.1007/s10291-004-0126-2).
- Lee, W.K., H. Kil, Y.S. Kwak, Q. Wu, S. Cho, and J.U. Park. The winter anomaly in the middle latitude F region during the solar minimum period observed by the constellation observing system for meteorology, ionosphere, and climate. *J. Geophys. Res.*, **116**, A02302, 2011, DOI: [10.1029/2010JA015815](https://doi.org/10.1029/2010JA015815).
- Liu, J.Y., Y.J. Chuo, S.J. Shan, Y.B. Tsai, Y.I. Chen, S.A. Pulinets, and S.B. Yu. Pre-earthquake ionospheric anomalies registered by continuous GPS TEC measurements. *Ann. Geophys.*, **22**, 1585–1593, 2004.
- Lu, G., M.E. Hagan, K. Häusler, E. Doornbos, S. Bruinsma, B.J. Anderson, and H. Korth. Global ionospheric and thermospheric response to the 5 April 2010 geomagnetic storm: an integrated data-model investigation. *J. Geophys. Res. [Space Phys.]*, **119**, 10358–10375, 2015, DOI: [10.1002/2014JA020555](https://doi.org/10.1002/2014JA020555).
- Luo, M., S. Pullen, S. Datta-Barua, G. Zhang, T. Walter, and P. Enge. LAAS study of slow-moving ionosphere anomalies and their potential impacts, ION GNSS 18th International Technical Meeting of the Satellite Division, Long Beach, CA, 2337–2349, 2005.
- Mandrikova, O.V., Y.A. Polozov, V.V. Bogdanov, and E.A. Zhizhiki. Method of detection abnormal features in ionosphere critical frequency data on the basis of wavelet transformation and neural networks combination. *JSEA*, **5**, 181–187, 2012, DOI: [10.4236/jsea.2012.512b035](https://doi.org/10.4236/jsea.2012.512b035).
- Mannucci, A.J., B.D. Wilson, D.N. Yuan, C.H. Ho, U.J. Lindqwister, and T.F. Runge. A global mapping technique for GPS-derived ionospheric total electron content measurements. *Radio Science*, **33** (3), 565–582, 1998.
- Mannucci, A.J., O.P. Verkhoglyadova, B.T. Tsurutani, X. Meng, X. Pi, et al. Medium-range thermosphere-ionosphere storm forecasts. *Space Weather*, **13**, 125–129, 2015, DOI: [10.1002/2014SW001125](https://doi.org/10.1002/2014SW001125).
- Mata, F., P. Zuraniewski, M. Mandjes, and M. Mellia. Anomaly detection in diurnal data. *Computer Networks*, **60**, 187–200, 2014.
- Menvielle, M., and A. Berthelier. The K-derived planetary indices: description and availability. *Rev. Geophys.*, **29** (3), 415–432, 1991.
- Pedatella, N.M., J.M. Forbes, J. Lei, J.P. Thayer, and K.M. Larson. Changes in the longitudinal structure of the low-latitude ionosphere during the July 2004 sequence of geomagnetic storms. *J. Geophys. Res.*, **113**, A11315, 2008, DOI: [10.1029/2008JA013539](https://doi.org/10.1029/2008JA013539).
- Pi, X., A.J. Mannucci, B.A. Iijima, B.D. Wilson, A. Komjathy, T.F. Runge, and V. Akopian. Assimilative modeling of ionospheric disturbances with FORMOSAT-3/COSMIC and ground-based GPS measurements. *Terr. Atmos. Ocean. Sci.*, **20**, 273–285, 2009, DOI: [10.3319/TAO.2008.01.04.01\(F3C\)](https://doi.org/10.3319/TAO.2008.01.04.01(F3C)).
- Pokhotelov, D., P.T. Jayachandran, C.N. Mitchel, and M.H. Denton. High-latitude ionospheric response to co-rotating interaction region- and coronal mass ejection-driven geomagnetic storms revealed by GPS tomography and ionosondes. *Proc. R. Soc. A*, **466**, 3391–3408, 2010, DOI: [10.1098/rspa.2010.0080](https://doi.org/10.1098/rspa.2010.0080).
- Pulinets, S.A., D. Ouzounov, L. Ciraolo, R. Singh, G. Cervone, et al. Thermal, atmospheric and ionospheric anomalies around the time of the Colima M7.8 earthquake of 21 January 2003. *Ann. Geophys.*, **24**, 835–849, 2006.
- Richardson, I.G., and H.V. Cane. Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): catalog and summary of properties. *Sol. Phys.*, **264**, 189–237, 2010, DOI: [10.1007/s11207-010-9568-6](https://doi.org/10.1007/s11207-010-9568-6).
- Chunk, R.W., L. Scherliess, J.J. Sojka, D.C. Thompson, D.N. Anderson, et al. Global Assimilation of Ionospheric Measurements (GAIM). *Radio Science*, **39**, RS1S02, 2004, DOI: [10.1029/2002RS002794](https://doi.org/10.1029/2002RS002794).
- Sodemann, A.A., M.P. Ross, and B.J. Borghetti. A review of anomaly detection in automated surveillance. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, **42** (6), 1257–1272, 2012.
- Steinwart, I., D. Hush, and C. Scovel. A classification framework for anomaly detection. *JMLR*, **6**, 211–232, 2005.
- Tsurutani, B., A.J. Mannucci, B. Iijima, M.A. Abdu, J.H.A. Sobral, et al. Global dayside ionospheric uplift and enhancement associated with interplanetary electric fields. *J. Geophys. Res.*, **109**, A08302, 2004, DOI: [10.1029/2003JA010342](https://doi.org/10.1029/2003JA010342).
- Tsurutani, B.T., O.P. Verkhoglyadova, A.J. Mannucci, G.S. Lakhina, and J.D. Huba. Extreme changes in the dayside ionosphere during a Carrington-type magnetic storm. *J. Space Weather Space Clim.*, **47** (2), A05, 2012, DOI: [10.1051/swsc/2012004](https://doi.org/10.1051/swsc/2012004).
- Wang, C., G.A. Hajj, X. Pi, I.G. Rosen, and B.D. Wilson. Development of the global assimilative ionospheric model. *Radio Science*, **39**, RS1S06, 2004, DOI: [10.1029/2002RS002854](https://doi.org/10.1029/2002RS002854).
- Zhou, Y., Y. Wu, X. Qiao, and X. Zhang. Ionospheric anomalies detected by ground-based GPS before the Mw7.9 Wenchuan earthquake of May 12, 2008, China. *J. Atmos. Sol. Terr. Phys.*, **71**, 959–966, 2009.

Cite this article as: Wang C, Rosen IG, Tsurutani BT, Verkhoglyadova OP, Meng X, et al. Statistical characterization of ionosphere anomalies and their relationship to space weather events. *J. Space Weather Space Clim.*, **6**, A5, 2016, DOI: [10.1051/swsc/2015046](https://doi.org/10.1051/swsc/2015046).