Analyzing significance of pre-seismic anomalies in Total Electron Content in Napa Valley, California

By Caleb M. Vatral, Pierre-Richard J. Cornely

Abstract

Earthquakes are natural phenomena that shake the Earth, often causing significant damage and loss of human life. One method proposed in previous studies for prediction of the time of arrival of large (>4.5 Mw) seismic activity is monitoring atmospheric Total Electron Content (TEC). In this study, we examine TEC data from 2014 during certain days near and on the date of the 6.0 Mw Napa Valley earthquake in California, USA. We show by adaptively truncated Hotelling's T^2 test that TEC in the Napa region for a known non-seismic baseline is statistically different than the TEC during several weeks surrounding the earthquake. This suggests a potential correlation between atmospheric TEC and major seismic activity. The overall goal of this research is to qualitatively and quantitatively assess potential correlation between a class of earthquake precursor signals and the arrival of major earthquakes. It may be possible to use such correlations to predict the location and time of large-scale earthquakes before they occur.

1. Introduction

Many recent studies have shown a connection between ionospheric anomalies and seismic activity.[1, 2, 3] This research has naturally led to the question of whether ionospheric activity can be used to forecast seismic activity. Though there are several competing theories concerning the exact physical mechanism responsible for ionospheric and other pre-seismic anomalies, the predominant theory, and the one assumed correct for the purposes of this study, was first presented by Freund (2002).[4] The theory posits that prior to an earthquake, due to pressure from seismic forces in the earth's crust, positive charge carriers are released from crustal rocks. These charge carriers begin traveling toward the earth's surface and radiating outward under the influence of strong electric fields. Eventually the charge carriers reach the ionosphere, beginning a series of combination and recombination processes that ultimately result in ionospheric disturbances.[5]

One parameter commonly used to characterize the ionosphere is the TEC. Previous studies have posited that measurement of TEC could be used to describe the ionospheric anomalies surrounding seismic activity.[1, 3] The present study focusses on the Mw 6.0 Napa Valley earthquake which occurred on August 24, 2014. We hypothesize that there are measurable and statistically significant variations in TEC on the date of the earthquake when compared to known non-seismic days in the months prior.

2. Methods

2.1 Measurement of Total Electron Content

The TEC is defined as the number of electrons along a path within a thin cylindrical tube between a receiver on the surface of the earth denoted *rx*, and a Global Positioning System

(GPS) satellite in orbit denoted *st*. TEC can be computed using a line integral from the receiver to the satellite, as described in (1).[1]

$$TEC(rx, st, t) = \int_{rx}^{st} d(r, \theta, \phi, t)dp \tag{1}$$

Where:

$rx =$ range or radial position (meters)

st $=$ latitude (degrees)

$t =$ time (seconds)

The TEC computed in (1) is the slant TEC (STEC), which is preferred for this study. It provides a realistic estimation of the variations in the number of electrons per unit of surface along the straight-line path.[3] This slant TEC includes many errors from the original measurements typically classified as Differential Code Bias (DCB) and propagation errors: slight changes in satellite orbit, delays due to interaction with species in the ionosphere and troposphere, timing errors due to inaccurate timing clocks, timing errors due to temperature variation, etc.

In order to be useful, STEC must be corrected for the measurement biases to find Unbiased TEC (UTEC). UTEC can be estimated using GPS pseudo-range measurement along with an estimate of the DCB and propagation errors. The differential pseudo-range can be modeled as in (2).[3]

$$P_{td} = 40.3 \times \left(\frac{1}{f_1} - \frac{1}{f_2}\right) \times TEC \times DCB^i + DCB_j + P\_Tropo + P\_Iono \tag{2}$$

Where:

$P_{td} =$ The differential pseudo-range measurement

$f_1, f_2 =$ The GPS measurement frequencies

$TEC =$ The total electron content

$DCB^i =$ The ith satellite bias

$DCB_j =$ The jth receiver bias

$P\_Tropo =$ The tropospheric delay

$P\_Iono =$ The ionospheric delay

The UTEC can then be calculated from (2) as

$$UTEC = -\frac{f_1^2 f_2^2}{40.3(f_2^2 - f_1^2)}\left(P_{td} - DCB^i - DCB_j - P\_Tropo - P\_Iono\right) \tag{3}$$

The GPS/GNSS TEC data was obtained from receiver independent exchange format (RINEX) data files from University NAVSTAR Consortium (UNAVCO), a non-profit organization formed under auspices of Cooperative Institute for Research in Environmental Sciences (CIRES) at the University of Colorado, Boulder. Differential code biases used in (3) were computed using the techniques developed in Jin et al. (2012).[6]

2.2 Adaptively Truncated Hotelling's T^2 Test

Of interest in this study is the question of whether TEC data are statistically the same from day to day when compared to a baseline TEC. More specifically, we wish to compare the TEC for each of several days leading up to seismic activity to that of a known non-seismic baseline. A widely accepted test for this sort of multivariate hypothesis testing is a two sample Hotelling's T^2 test, which tests the null hypothesis $H_0: \mu_1 = \mu_2$ where μ_i is an *n*-dimensional vector.[7] The T^2 test can be computed as

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{y_1} - \overline{y_2})'\widehat{\Sigma^{-1}}(\overline{y_1} - \overline{y_2}) \tag{4}$$

where n_1 and n_2 are the number of observations in each of the two samples, $\overline{y_1}$ and $\overline{y_2}$ are the vectorized sample means, and $\widehat{\Sigma^{-1}}$ is the pooled sample covariance matrix.[7]

For each active GPS receiver, the UTEC computation as described in section 2.1, is performed for each of the 32 active satellites, sampling once every thirty seconds (twice a minute). This data can be compiled into a 2880×32 matrix, where each column represents a time series of one satellite's measurements over a given 24-hour period as observed from one ground receiver station. These matrices can then be averaged row-wise to create a single 2880×1 time series vector representing the average TEC at that receiver. Multiple receiver time series are then compiled together into a single 2880 by *n* matrix, where *n* is the number of receivers for which we have data. The process of creating this data matrix is depicted in figured 1. This data matrix can be modeled as a random column variable of dimension 2880 with *n* observations per dimension, leading to the formulation of the problem as a multivariate hypothesis test. Since UTEC is functional data, the methods proposed in Lee, et al. (2015) for adaptive truncation of the Hotelling's T^2 test are employed. The test statistic in (4) then becomes:

$$T_k^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{y_1} - \overline{y_2})'\widehat{\Sigma^{-1}}^{*}(\overline{y_1} - \overline{y_2}) \tag{5}$$

where $\widehat{\Sigma^{-1}}^{*}$ is now the truncated Moore-Penrose pseudo-inverse of the pooled sample covariance matrix.[7] For the truncation of the pseudo-inverse, the methodology as presented in Lee, et al. (2015) of keeping only the p_0 greatest eigenvalues and their associated eigenvectors was employed, where p_0 is given by:

$$p_0 = max\{\, j: \widehat{\lambda_j}/\widehat{\lambda_1} > c \,\} \tag{6}$$

for some threshold, *c*, which was chosen as $c = 10^{-15}$ for this study. We compute the t-statistic in (5) several times, maintaining *k* eigenvalues and eigenvectors each time over $k = 1, \ldots, p_0$. The maximum t-statistic over the iterations is taken as the statistic value.

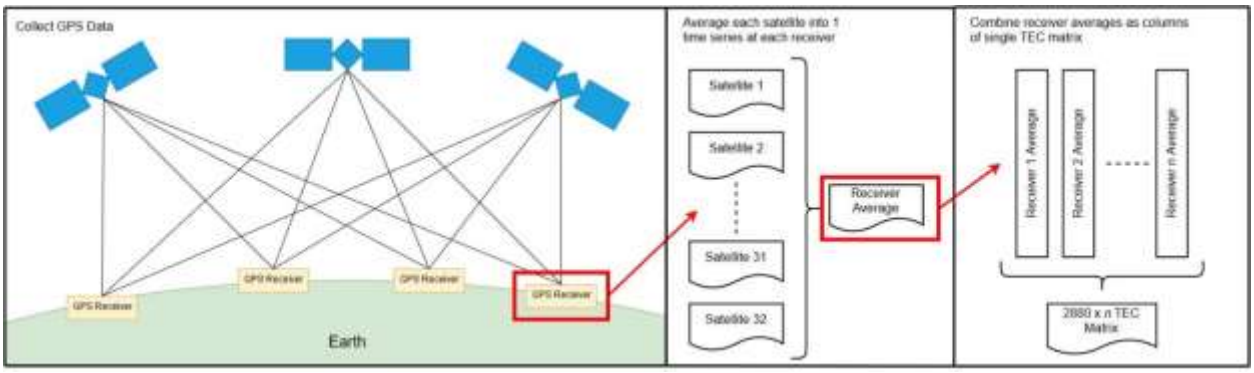

Figure 1. Cartoon displaying the GPS receiver geometry and workflow for creating the daily 2880 x *n* TEC matrix for the region, as described in section 2.2.

2.3 Data Selection

Since we desire to statistically compare several days leading up to seismic activity to a known non-seismic day, criterion for a non-seismic baseline day must be defined. For our study, we focus on the Napa Valley region in California, USA. Choosing a non-seismic day is a non-trivial task, as this region is heavily affected by small scale seismic activities. The best that can be done is to choose days which have "approximately normal background seismic activity". For this study, we consider two baseline signals. The first baseline was chosen as July 1, 2014. This date was considered approximately normal background because it meets the criteria of both no seismic activity above 3.0Mw within a 30-mile radius for a period between five days prior to and five days after, and no seismic activity above 5.0Mw within a 100-mile radius for a period between one month prior to and one month. The second baseline was chosen as the pointwise average of the daily TEC for each day in January, February, and March of 2014. Plots of these baselines are depicted in figure 2.

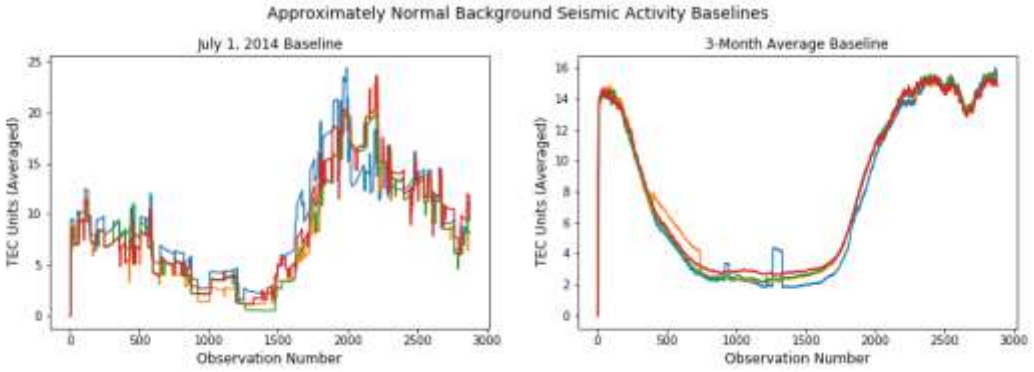

Figure 2. Plots of the two baselines utilized for this study. On the left is the July 1, 2014 baseline and on the right is the baseline constructed of the pointwise average of TEC in Jan, Feb, and March of 2014.

Since research to this point is inconclusive as to how far pre-seismic perturbations in the ionosphere extend geographically, GPS receivers were limited for this study to a 20-mile radius from the epicenter of the earthquake. For the August 24, 2014 Napa Valley earthquake, this included 4 GPS receiver stations for which TEC was calculated. Figure 3 shows a map of the

region in California with markers at the four selected stations and the epicenter of the earthquake.

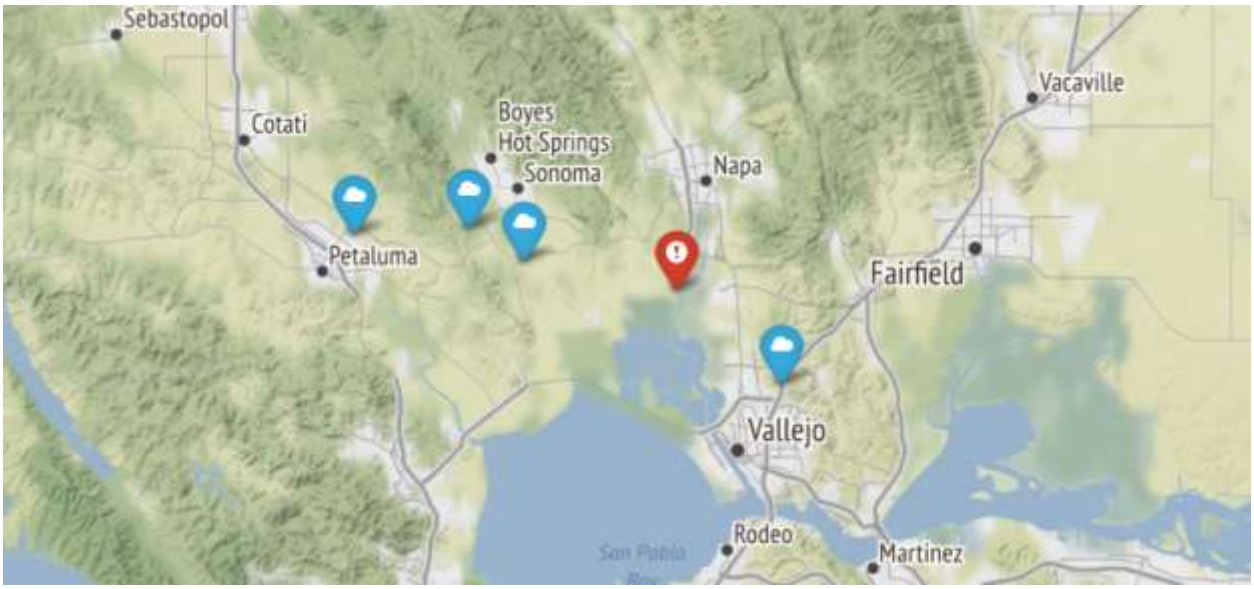

Figure 3. A map of the Napa Region in California, USA with markers displaying the locations of the four selected GPS receiver stations (blue) and the epicenter of the August 24 earthquake (red).

3. Results

Utilizing the adaptively truncated Hotelling's T^2 test, average TEC matrices were tested against each of the baselines. Each day approaching the earthquake was compared to the baselines, beginning with January 1st, 2014 and ending with December 30th, 2014.

For the July 1, 2014 baseline, a graph of the p-values from a selected time range of the tests, July 1 to September 30, is displayed in figure 4. Under seismically quiet times, the p-values exhibit a high standard deviation with no discernible pattern day to day. Most days under these conditions shows failure to reject the null hypothesis, indicating that they are statistically similar to the non-seismic baseline. During these same periods, there are certain days where the p-values drop to a statistically significant level; however, these drops are not sustained and occur only for a few days before returning to insignificant p-levels. Approaching the date of the earthquake, a large sustained drop in p-value to significant levels can be observed on July 14 and, other than two outlier spikes on July 31 and August 3, persists until September 9. This behavior in the p-values indicates that TEC becomes consistently anomalous for a significant period of time before and after the earthquake.

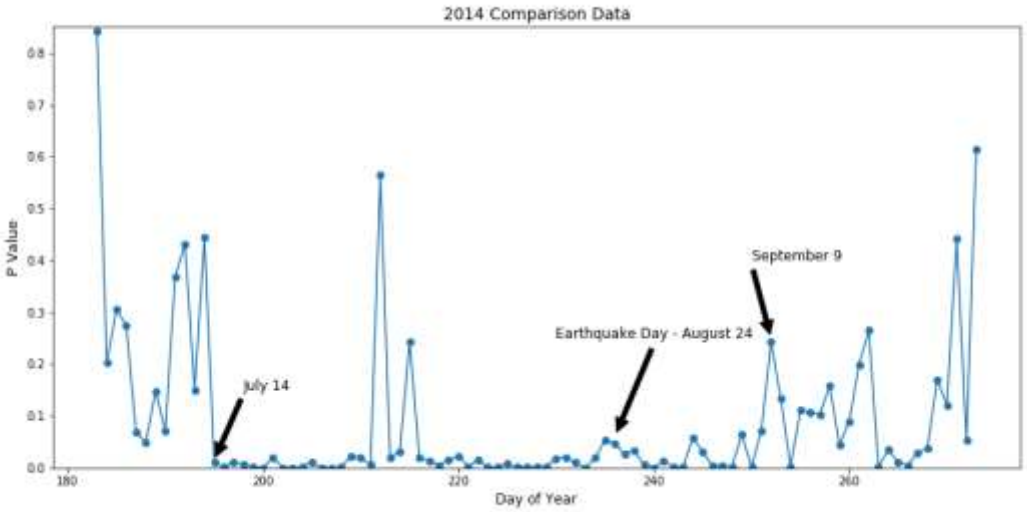

Figure 4. P-values plotted against day of year from the Hotelling's T^2 comparison of the July 1, 2014 baseline to days leading up to and following the August 24 earthquake. A large drop in p-value can be observed on July 14 and persisting until September 9, indicating statistically significant anomalies in TEC prior to the seismic activity.

To ensure that the kind of persistent drop in p-value observed in figure 4 is not common behavior, the same test was performed for the two years prior to the earthquake, 2012 and 2013, and one year after the earthquake, 2015, as shown in figure 5. Each of these three years displays behaviors similar to that of the quiet seismic times in 2014, with no instances of sustained low p-values over a significant period of time. This suggests that this behavior of a drop in p-values and its subsequent stabilization over a sustained period of time is unique to seismically affected times.

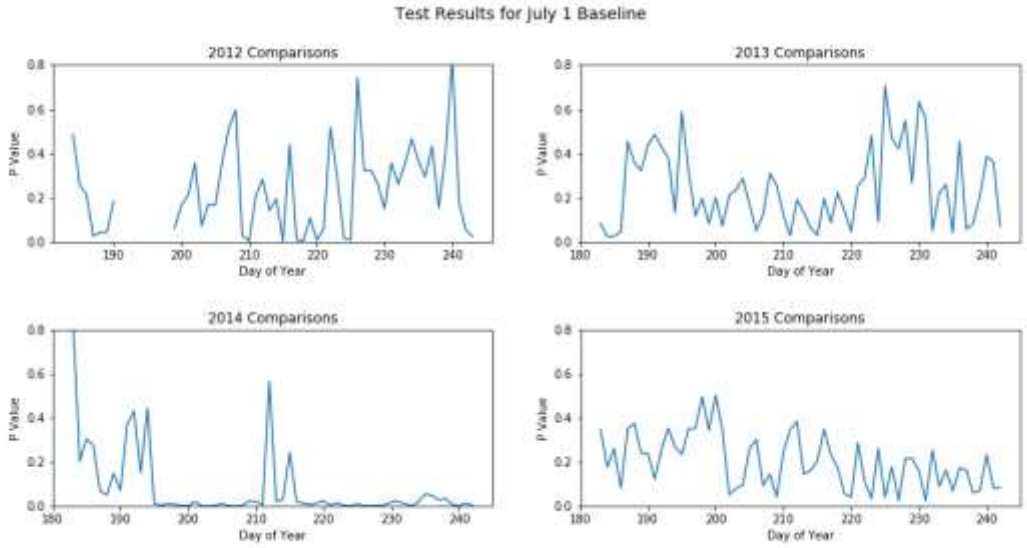

Figure 5. P-values for Hotelling's T^2 test plotted against day of year for two years prior to the earthquake and one year after the earthquake. The large drop and stabilization exhibited during the year of the earthquake, 2014, does not present in any of the other years, indicating that it is behavior unique to seismically affected times.

For the three-month average baseline, a graph of the t-statistic values from the complete year is displayed in figure 6. Under seismically quiet times, the t-statistics exhibit consistently low behaviors, fluctuating approximately around a nominal value of 10. During these same periods, there are certain days which spike to higher values; however, these spikes never exceed a value of 100 and are not sustained for more than one to two days. Beginning at approximately June 1, a slow rise in t-statistic value is exhibited, growing past the date of the earthquake on August 24 and peaking four days after the earthquake on August 28. The t-statistic values then begin to descend, returning to their normal non-seismic behavior by approximately September 26. This indicates that TEC has increasingly significant anomalies approaching the time of the earthquake on August 24.

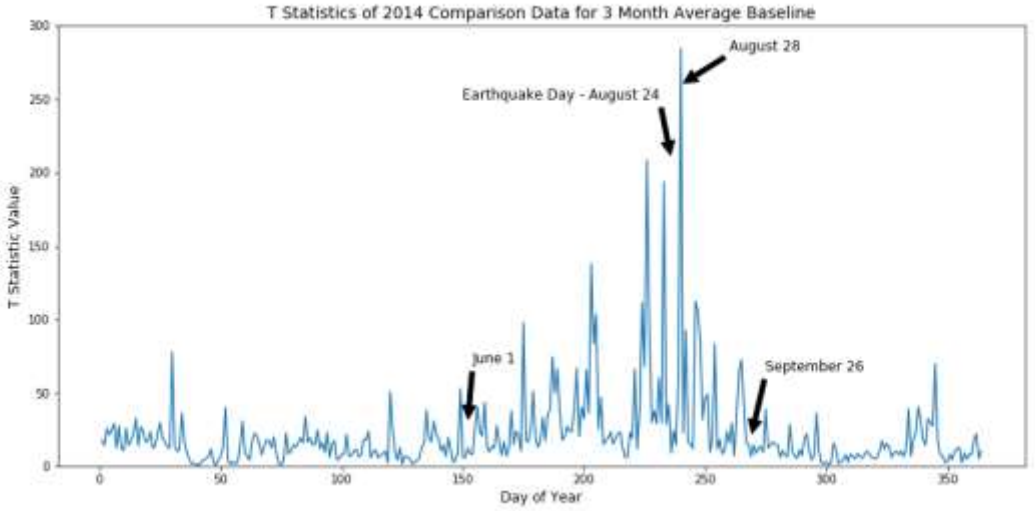

Figure 6. T-statistic values plotted against day of year from the Hotelling's T^2 comparison of the three-month average baseline to days leading up to and following the August 24 earthquake. A slow rise in t-statistic can be observed beginning approximately June 1, peaking four days after the earthquake on August 28, and returning to normal behavior on approximately September 26. This indicates increasingly significant anomalies in TEC approaching the earthquake.

To ensure that this kind of rise in t-statistic is not common behavior, the same test was once again performed for the two surrounding years, 2013 and 2015, as shown in figure 7. Both of the surrounding years exhibit behavior similar to that of 2014 during its quiet seismic times, with no instances of the rising behavior in t-values. This suggests once again that this behavior of the t-values is unique to seismically affected times.

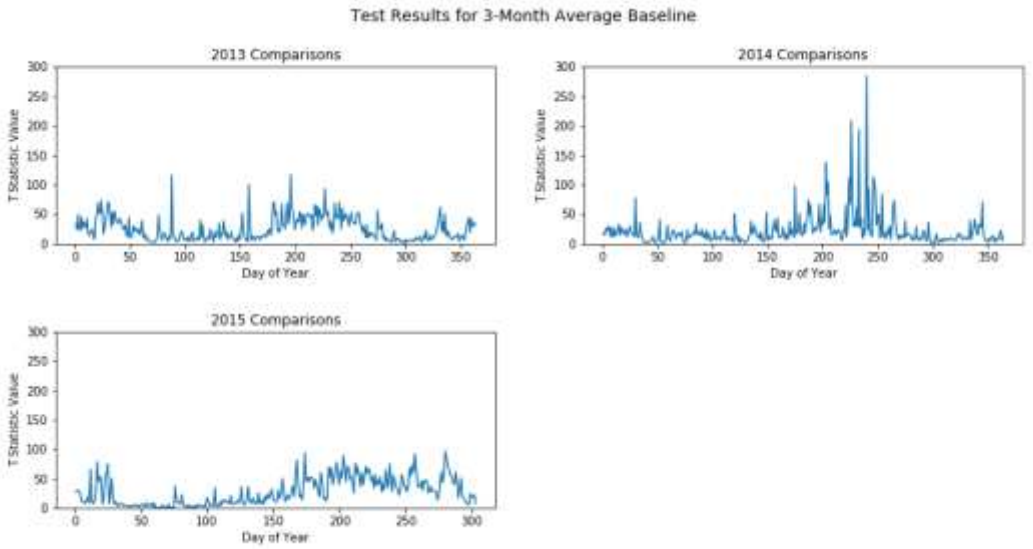

Figure 7. T-statistic values for Hotelling's T^2 test plotted against day of year for two years, 2013 and 2015, surrounding the year of the earthquake. The slow rise in t-statistic exhibited during the year of the earthquake, 2014, does not present in any of the other years, indicating that it is behavior unique to seismically affected times.

4. Conclusion

Previous studies have proposed that measurable anomalies in ionospheric TEC are caused by large scale seismic activity[1, 2, 3]. In this investigation, we utilized adaptively truncated Hotelling's T^2 test to compare the average UTEC for each day to two known baseline signals, July 1 2014 and a point-wise average of each day in the first three months of 2014. Each of these baselines were chosen as approximates to a normal background seismic activity in the region under study. Each day, beginning with January 1 and ending with December 30, was compared to each of the baselines using adaptively truncated Hotelling's T^2 test and the resultant t-statistic values and p-values were analyzed to determine if any statistically significant deviations or anomalies exist.

For the July 1, 2014 baseline, it was shown that under quiet seismic times, p-values exhibit a high standard deviation with no discernible pattern day to day. Approaching the earthquake, there was a sudden drop in p-value on July 14, 2014. This drop was sustained at the low value past the date of the earthquake on August 24 until it rose and resumed its normal behavior on September 9. This same test was performed on data from 2012, 2013, and 2015, and it was shown that this sustained drop in p-value was not exhibited in these years. This indicates that there are statistically significant anomalies in TEC that manifest in the weeks before large seismic activity, and that these anomalies are unique to seismic times.

For the three-month average baseline, it was shown that under quiet seismic times, t-statistic values exhibit a fairly consistent behavior, fluctuating around a low nominal value, which in the case of 2014 was approximately 10. Approaching the earthquake, at approximately June 1, a slow rise in t-statistic value can be seen, growing past the date of the earthquake on August 24, peaking four days after the earthquake on August 28, and returning back to its normal behavior by approximately September 26. This same test was performed on data from 2013 and

2015 with no instance of this rising behavior in either year. This indicates that increasingly significant anomalies in TEC appear during the weeks approaching large seismic activity, and that these anomalies are unique to seismic times.

In this study, we have shown that anomalies in TEC appear in the weeks approaching large seismic activity when compared to both baselines utilized for this study. In addition, these anomalies were shown to be statistically significant ($p = 0.05$). With further research, it may be possible to use such anomalies as part of an earthquake forecasting model which would warn of large seismic events ahead of time.

This study had limitations in not accounting for other environmental factors that could potentially influence UTEC. Further work should attempt to account for such factors before performing comparisons. Assumptions on what criteria define approximately normal background seismic activity were also a limiting factor in this study and should be examined in further depth in future studies. Further work should also include studies of the "fusion" of the TEC with other earthquake precursors. By analyzing and incorporating multiple precursors together in an earthquake forecasting model, the chances of a false prediction due to anomalies caused by other mechanisms besides seismic activity can be decreased.

5. References

[1]Hammerstrom, J and Cornely, P-R. (2016) "Total Electron Content (TEC) Variations and Correlation with Seismic Activity over Japan." *Journal of Young Investigators*. 32. 4. Pg 36-40.

[2]O'Brien, M and Cornely, P-R. (2015) "Analyzing Anomalies in the Ionosphere Above Haiti Surrounding the 2010 Earthquake." *Journal of Young Investigators.* 29. 5. Pg 13-16.

[3]Cornely, P-R and Hughes, J. (2018) "Unbiased Total Electron Content (UTEC), their fluctuations and Correlation with Seismic Activity over Japan." *Acta Geophysica.* 66. 1. Pg 51-70.

[4]Freund, F. (2002). "Charge generation and propagation in igneous rocks." *Journal of Geodynamics*. 33. 4-5. Pg 545-570.

[5]Freund, F, Takeuchi, A and Lau, B. (2006) "Electric currents streaming out of stressed igneous rock – A step towards understanding pre-earthquake low frequency EM emissions." *Physics and Chemistry of Earth.* 31. 4-9. Pg 389-396.

[6]Jin, R, Jin, S and Feng, G. (2012) "M_DCB: Matlab code for estimating GNSS satellite and receiver differential code biases." *GPS Solutions.* 16. Pg 541-548.

[7]Lee, J, Cox, D and Follen, M. (2015) "A two sample test for functional data." *Communications for Statistical Applications and Methods.* 22. 2. Pg 212-135.