New lightning-derived vertical total electron content data provides unique global ionospheric measurements.

Erin H. Lay[1], Jeffery D. Tippmann[1], Kyle C. Wiens[1], Sarah E. McDonald[2], Anthony J. Mannucci[3], Xiaoqing Pi[3], Anthea Coster[4], R. Marc Kippen[1], and Rob Redmon[5]

[1] Los Alamos National Laboratory, Los Alamos, NM 87545.

[2] Space Science Division, Naval Research Laboratory, Washington, DC.

[3] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109.

[4] MIT Haystack Observatory, Westford, MA, 01886.

[5] National Centers for Environmental Information, NOAA, Boulder, CO, USA.

Corresponding author: Erin Lay (elay@lanl.gov)

Key Points:

- New global gridded VTEC dataset derived from RF measurements from lightning emissions during entire year of 2018 has been released.
- VTEC product adds new global coverage in areas with few GNSS receivers (ocean and continental Africa).
- The VTEC values, generated from VHF RF data, show bias in Madrigral dataset and differences to JPL-GIM model derived VTEC.

Abstract

A newly-released, novel ionospheric dataset of global gridded vertical total electron content (VTEC) is introduced in this paper. This VTEC dataset, provided by Los Alamos National Laboratory (LANL), is derived from very-high frequency (VHF; defined as 30-300 MHz) broadband radio-frequency (RF) measurements of lightning made by U.S. Department of Defense sensing systems on board Global Positioning System (GPS) satellites. This paper presents the new dataset (LANL VTEC), discusses the errors inherent in VHF TEC estimation due to ionospheric dispersion, and compares the LANL VTEC to two community standard VTEC gridded products: Jet Propulsion Laboratory's Global Ionospheric Model (JPL GIM) and the CEDAR community's Open Madrigal VTEC gridded measurements of L-band GNSS (global navigation satellite systems) TEC. We find that the LANL VTEC data has an offset of 3 TECU from CEDAR Madrigal GNSS VTEC, and a full-width-half-maximum (FWHM) of 6 TECU. In comparison, the offset between LANL VTEC and the JPL GIM model is -3 TECU, but with a FWHM of 5 TECU. We also compare to Jason-3 VTEC measurements over the ocean, finding an offset of less than 0.5 TECU and a FWHM of < 5 TECU. Because this technique uses a completely different methodology to determine TEC, the sources of errors are distinct from the typical ground-based GNSS L-band (GHz) TEC measurements. Also, because it is derived from RF lightning signals, this dataset provides measurements in regions that are not well covered by ground-based GPS measurements, such as over oceans and over central Africa.

Plain Language Abstract

The ionosphere is a region of the atmosphere that is very important in communications between ground and satellite. For that reason, decades of scientific effort has been put towards developing models of the ionosphere so that we can more accurately predict what the state of the ionosphere is at any given location and time. A general product produced by many of these models is the vertical total electron content (VTEC), which is the vertically-integrated electron density at a particular location at a particular time. The majority of these models use measurements of TEC from ground-based receivers or instruments, meaning that abundant measurements that go into the models often lack data from over the oceans, or in technologically-limited regions of the world (e.g. Africa). Here we present a new VTEC dataset that is derived from lightning strokes detected with U.S. Department of Defense sensing systems on GPS satellites. Because the data set uses naturally-occuring lightning for its source, it does not have the same limitations as ground-based TEC measurements, and can provide an additional source of validation data for ionospheric models. We introduce the data set and compare it with community-accepted VTEC models and measurement.

1 Introduction

The ionosphere is a layer of plasma in the Earth's atmosphere (from 60 to more than 2000 km altitude) (Kelley 2009). The nature of this plasma impacts natural and man-made electromagnetic signals that interact with it. This interaction occurs in many forms, such as refraction, absorption, dispersion and scintillation. Being able to nowcast and forecast the ionospheric state and variability on a global scale is important for understanding associated communication impacts. Variability in the ionosphere is driven by several external sources, roughly ordered as follows from the most globally to regionally significant: solar forcing solar EUV flux (photoionization), solar flares, and geomagnetic storms (e.g. Shunk and Sojka, 1996; Mannucci et al., 2005; Yizengaw et al., 2006; Coster and Skone, 2008; Wang et al., 2010); and lower atmosphere forcing gravity waves, thunderstorms, earthquakes, and explosions (e.g. Lay 2018; Azeem et al., 2015; Lay et al., 2015; Galvan et al., 2011).

Empirical and physics-driven models have been developed to accurately analyze and forecast the state of the ionosphere (e.g. Mannucci et al., 1998; Mandrake et al., 2005, Scherliess et al., 2006; Nava et al., 2008, Scherliess et al., 2009; Bilitza et al., 2011). These models are built on decades of ionospheric observations. However, the ionosphere is extremely difficult to accurately measure continuously on a global scale, and therefore it is difficult to detect and predict variations due to individual events from the climatological average. Ground-based Global Navigation Satellite System (GNSS) total electron content (TEC) measurements are a widely-used global data source for use in ionospheric modeling based on their geographic ubiquity and nearly continuous operation. However, with current ground based capabilities, these measurements are sparse or unavailable over oceans and some regions of the world such as central Africa.

GNSS receivers calculate the integrated electron density along a line of sight from satellite to receiver (slant TEC) by measuring the dispersive contributions to two separate GHz-frequency satellite signals. TEC is then determined by combining the less precise but absolute pseudorange measurement with the precise but ambiguous phase differential measurement. Inherent timing uncertainties in the GHz frequency bands limits the accuracy of the absolute TEC derived from this method to 2-5 TECU, where 1 TECU = 1 x 10^{16} electrons/m^2 (Burrell et. al, 2009). Slant TEC measurements are often converted to vertical TEC (VTEC) values by using a geometrical mapping function based on the satellite elevation angle and an assumed ionospheric height (Burrell et al., 2009; Mannucci et al., 1998), typically between 350 and 400 km.

We present a novel dataset that provides an independent comparison of the much-utilized GNSS TEC measurements and, in addition provides TEC measurements in low-coverage areas (oceans, economically disadvantaged areas). The new data is derived from unique measurements of lightning events, each of which produces a broadband radio signal that gets dispersed through the ionosphere before it is detected on satellite receivers (Jacobson et al., 1999; Roussel-Dupre et al., 2001). Each measured lightning provides a snapshot of the ionospheric conditions at that instant, and many lightning measurements over time and around the globe provide unique measurements for ionospheric science. With 93% of the LANL VTEC measurements covered by existing GPS ground stations, these matching measurements can be used for improving calibration of the GPS TEC measurements in space weather models. The remaining 7% of the LANL VTEC measurements that add new coverage, while small in quantity due to the non-continous nature of the source events, can still be used to fill the gaps in models to improve the global scale TEC outputs in these regions during those time periods.

In this paper, we introduce the new Los Alamos National Laboratory vertical total electron content (LANL VTEC) product. We then compare the LANL VTEC data with GPS-measured VTEC from the CEDAR Madrigal GNSS VTEC product (Rideout and Coster, 2006; Vierinen et al., 2016), JPL GIM VTEC (Mannucci et al., 1998; Iijima et al., 1999), and Jason-3 VTEC measurements derived from ocean altimetry satellites.

2 Datasets

2.1 LANL VTEC Data

Los Alamos National Laboratory (LANL) vertical TEC measurements are derived from U.S. Department of Defense broadband radio-frequency (RF) sensing systems on Global Positioning System (GPS) satellites that measure transient events in the very-high frequency (VHF, defined as 30-300 MHz) range. Detection of an event by multiple satellites allows determination of the event location, based on minimization of the time of arrival at each satellite. As RF signals from transient broadband lightning events travel through the ionosphere, they are dispersed, so lower frequencies arrive later in time than higher frequencies (Lay et al., 2011; Moses and Jacobson, 2004; Jacobson et al., 1999). This dispersion can be used to determine the slant total electron content (STEC, or the integrated electron density along the line of sight between each lightning location and the GPS satellite in medium Earth orbit (MEO)). To determine the STEC from a recorded lightning signal, we fit the time delay versus frequency spectrogram to the first order approximation of the Appleton-Hartree Equation (Lay et al., 2011; Eqn. 6, and Eqn. 7a):

$$n^2 = 1 - \frac{X(1-X)}{(1-X)-\frac{1}{2Y_T^2}+\mathrm{s}\sqrt{\frac{1}{4Y_T^4}+(1-X)^2Y_L^2}}; \qquad \text{(Eqn. 1)}$$

where $X = f_p^2 / f^2$; $Y = f_{ce} / f$; $Y_L = Y \cos\theta$; $Y_T = Y \sin\theta$; f_p is the plasma frequency of the ionosphere, f is the wave radio frequency, f_{ce} is the electron cyclotron frequency and θ is the angle between the propagation and the magnetic field vectors. The variable $s = \pm 1$, represents the fast and slow modes in the plasma. When X and Y are small compared with unity, this relation can be estimated by a Taylor expansion, taken to first order, $n \approx 1 - \frac{1}{2}\frac{f_p^2}{f^2} + \cdots$. This approximation gives a first order time delay of

$$\tau(f) = \frac{R}{c} + \frac{C_1}{f^2} + \cdots \qquad \text{(Eqn. 2)}$$

where R is the source-to-sensor distance, c is the speed of light, and $C_1 = e^2 / 8\pi^2 c\varepsilon_0 m_e \int_o^L N_e(l)dl$. The integral of electron density, N_e, in this term is taken along the line-of-sight path between the source and sensor, and thus represents the line-of-sight STEC (Lay et al., 2011). LANL VTEC is derived from individual STEC measurements determined by using the first-order approximation to the A-H Equation (Eqn. 2), and implementing a broadband fitting routine similar to that described in Lay et al., 2011. Each lightning stroke can produce several measurements of STEC along the various lines-of-sight to all satellites detecting it.

A mapping factor of $VTEC = STEC\sqrt{1 - \left[\cos(el)/\left(1 + \frac{H}{R_e}\right)\right]^2}$ (Burrell et al., 2009; Jakowski et al., 2011), then is used to project the STEC to a vertical TEC (VTEC), where *el* is the elevation angle of the satellite with respect to the lightning location, H is chosen to be 350 km, the assumed height of the ionospheric pierce point (IPP) (though altitudes between 350 to 450 have been used by different groups depending on specific model approach), and R_e is the radius of the Earth. VTEC measurements from individual lightning events are combined in 5° latitude × 5° longitude geographical bins and 1-hour time bins. Latitude bins are limited to

144 between -60° and 60°, since very few lightning strokes occur outside those limits. The median
145 VTEC in a given bin is provided in the LANL VTEC product. This binning creates a sparse
146 global VTEC product from lightning events. An example time period is shown in Figure 1
147 (bottom panel). The currently released data includes all of 2018, globally. There are 143,541 grid
148 cells with LANL VTEC measurements over this time period. The comparisons in this paper of
149 LANL VTEC to Madrigal VTEC and JPL GIM VTEC are on this entire data set.

150

151 **2.2 Estimated Error on LANL VTEC**

152 While the same first order approximation is made in order to determine GNSS TEC and
153 VHF TEC, the error introduced by the approximation is larger at lower frequencies due to
154 decreased fidelity of the $X << 1$ and $Y << 1$ assumptions. Also, the signal refracts much more
155 significantly at lower frequencies, meaning that the line-of-sight approximation is less valid as
156 well (Lay et al., 2011; Roussel-Dupre et al., 1999). Previous studies show that STEC error can be
157 larger than 10 TECU for frequencies lower than 45 MHz, but the effect is mitigated when using a
158 broadband signal to fit time-delay versus frequency for many frequencies at once. Because the
159 assumptions on X and Y are valid in the GHz regime (GNSS TEC), the majority of the error on
160 GNSS slant TEC estimation comes from uncertainty in resolving the absolute time delay in the
161 GPS signal due to hardware limitations (Sardón and Zarraoa, 1997), rather than uncertainty due
162 to dispersion and ray bending effects.

163 In this work, we estimate the ionospheric-induced error on LANL VTEC based on a
164 realistic sampling of lightning events around the world. We use a similar ray-tracing technique to
165 that of Lay et al., 2011, but with realistic ionospheric electron density profiles from the

166 NeQuick2 model (Nava et al., 2008) for the specific day and location of the simulation. The
167 locations and times of the simulated lightning signals are generated from a random sample of
168 lightning locations detected by the World Wide Lightning Location Network (WWLLN)
169 (Dowden et al., 2002; Lay et al., 2004; Abarca et al., 2010) during two days in 2014: 14 January
170 and 14 July. The year 2014 was chosen due to high solar activity at that time, and, thus, high
171 electron density levels to bound the error on the high end of the TEC distribution. Most years
172 will have lower overall VTEC values, leading to lower VTEC errors.

173 To estimate a realistic error distribution, we created an electron density altitude profile
174 along lines of sight for the chosen random sample of WWLLN lightning events (location and
175 time) to all GPS satellites within view. Along these lines of sight, we used the NeQuick2 model
176 (Nava et al., 2008) with 1-km vertical spacing to provide the profile shape, including the height
177 of the peak electron density (hmF2). The NeQuick2 model was chosen for its fast execution
178 time, ease of use via command line interface, and ease of modeling the electron profile along the
179 line of sight, as opposed to vertically. We then used the Utah State University Global
180 Assimilation of Ionospheric Measurements (USU GAIM) model (Schunk et al., 2004; Decker
181 and McNamara, 2007) to scale the overall profile as follows. An estimate of the STEC along the
182 simulated line-of-sight was made by projecting vertical TEC USU GAIM onto each line of sight
183 given above. We then scale the profile to give the calculated USU GAIM STEC along the line of
184 sight. This mapping, as opposed to integrating USU GAIM along the line of sight, was also
185 chosen for computational speed. Our goal was to model a realistic profile in terms of shape, peak
186 electron density altitude (hmF2), and peak electron density (nmF2) at the time and location.

187 These STECs and electron density profiles form the “truth” ionosphere that we use to
188 compare with the estimated ray-traced TEC along each line of sight. The collection of lightning-

to-sensor paths used in the simulation represent a realistic configuration of possible on-orbit measurements of lightning, with as realistic ionospheric conditions as possible. These paths are only used for this error simulation and representative of solar activity, and associated ionospheric variation, in the year 2014 only.

We use a ray tracing algorithm through this "truth" ionosphere that uses the Bouguer's Formulation to produce a frequency-dependent time-delay using the full A-H index-of-refraction with no approximations to the dispersion relation (Born and Wolf, 1999; Lay et al., 2018). A broadband synthetic signal is produced that accounts for dispersion and ray-bending. Both modes of the Appleton-Hartree Equation are propagated and combined to produce the simulated signal arriving at a satellite through a known "truth" ionosphere. This dispersed signal is created with the highest possible fidelity to approximate a real signal arriving at a satellite. This synthetic signal is then processed through a first-order (Eqn. 2) matched filter to produce an estimate of the STEC variability due to natural ionospheric and geometric variability.

From this simulation, the known line-of-sight STEC determined from the "truth" ionosphere is compared with the estimated ray-traced STEC. We then convert these STEC values to VTEC based on the geometry of source-to-sensor and the mapping function above. Figure 2 shows a probability density plot (counts per TECU x TECU bin / total counts) of the "truth" VTEC from the known ionosphere versus the estimated ray-traced VTEC determined as described above. This comparison gives an indication of error introduced in the LANL VTEC measurements due to ionospheric variability and geometry. Other sources of error in the uncertainty budget are not considered here. For VTEC less than about 30 TECU, where the majority of the points lie, the error introduced to LANL VTEC by ionospheric and geometric approximations is on the order of 1-2 TECU. As VTEC increases above 30 TECU, the LANL

VTEC gridded data begins to overestimate the true VTEC, with errors as large as 10 TECU above VTEC values of 50 TECU. These higher VTEC, and hence higher STEC, values are often correlated with lines of sight with lower elevation angles. At lower angles, the estimated STEC has inherently larger errors due to the fact that the second order approximation does not account for ray bending. The lower frequencies of VHF band bend significantly more than L-band frequencies, leading to larger errors in estimated TEC. These findings are consistent with our previous work (Lay et al., 2011). In addition, the geomagnetic mapping factor from STEC to VTEC loses fidelity at lower elevation angles.

2.3 Madrigal VTEC

Vertical total electron content (TEC) data, calculated from the GNSS satellite constellations, are provided through the Madrigal distributed data system (http://www.openmadrigal.org). Currently only the GPS and GLONASS constellations are being processed, although it is planned to include additional constellations. The algorithms used to compute the total electron content are described in Rideout and Coster (2006) and Vierinen et al. (2015). All values are then mapped to an ionospheric pierce point defined to be 350 km, and a pierce point latitude and longitude is derived. Data from approximately 6000 GNSS dual-frequency receivers world-wide are used. The data in this paper is derived from the gridded TEC product in Madrigal, where the TEC is stored in $1° \times 1°$ bins at a 5-minute cadence. Each TEC value represents the median TEC of all values within the bin. An estimate of the error is also provided.

Because the Madrigal VTEC data (Figure 1, middle panel) is measurement-based, not all grid cells contain data, similar to the LANL VTEC data. Empty grid cells give an indication of where ground-based GPS coverage is limited globally. This gridded TEC product is available online starting in the year 2000 through the present. For comparison to LANL VTEC, we take the median of all Madrigal VTEC values contained in a given 5° × 5° x 1 hour bin corresponding to the LANL VTEC grid. This averaging can combine up to 300 Madrigal VTEC measurements into one measurement for comparison with LANL VTEC, with the median value being 43 measurements. Nevertheless, the standard deviation of all Madrigal measurements within one LANL VTEC cell has a median value of less than 1 TECU. This indicates that this averaging technique produces a reliable estimate for comparison.

2.4 JPL GIM VTEC used in LANL VTEC comparison

A technique for producing global ionospheric maps (GIM) of TEC has been developed in the 1990s (Mannucci et al., 1998; Iijima et al., 1999). It makes use of GNSS data collected from hundreds of globally distributed ground-based stations. To produce GIM, the GNSS dual-frequency pseudorange and carrier phase data are processed first to fix phase breaks and adjust the level of the precise but ambiguous phase data to the noisier but absolute range data. This phase-smoothed data, which contains much less noise than the pseudorange data alone, is then used to compute relative line-of-sight or slant TEC. The slant TEC data are modeled as vertical TEC multiplied by a geometric scaling function that depends on elevation angle, plus receiver and satellite instrumental biases. The vertical TEC is fitted to a set of 330 localized basis functions on a global grid to form a vertical TEC “surface”, which has a continuous second derivative. The satellite and receiver biases are additional parameters of the fit. The basis

functions are defined in a spherical sun-fixed longitude and dipole-based geomagnetic latitude reference frame that accounts for two major sources of ionospheric variability: photoionization and geomagnetic control of ionospheric dynamics. The sun-fixed frame also allows to update the different grid points with data from the same geographic longitude at different universal time, which helps to augment spatial coverage. A Kalman filter is used with the data to solve for a time series of the coefficients of the basis functions and for the satellite as well as receiver inter-frequency instrumental biases. The basis functions with the time-dependent coefficients can then be used to estimate vertical TEC globally at any location and time. The fitted biases can be removed from slant TEC measurements, and bias-removed slant TEC at any elevation angle can be reconstructed by applying the slant-to-vertical mapping function.

GIM is routinely produced by the Ionospheric and Atmospheric Remote Sensing (IARS) group at the Jet Propulsion Laboratory (JPL) to support NASA's Deep Space Network and space missions, solid Earth and ocean altimeter missions, and space weather research. One of the GIM products is distributed in IONEX format. The files consist of globally gridded TEC maps with 5° × 5° (geographic latitude and longitude) spatial resolutions and 2-hour cadence. The IONEX TEC map data is delivered to the NASA CDDIS data archive center (refer to the above weblink) on a daily basis. A higher-resolution version of 2° × 2° and 1-hour cadence is also available, which is used in this study. The IONEX maps, which are included in the Jason series altimetry data products, contain an offset of +2 TECU added to GIM in post-processing. This offset was based on early comparisons to TOPEX/Poseidon TEC in the late 1990s, which suggested that GIM maps were biased low. After years of GIM comparisons, it is not clear that GIM are biased low, but the offset for IONEX has been maintained to ensure a consistent record for altimetry, which is concerned with long-term sea level changes.

279 For this study, we use the JPL GIM or JPL VTEC term throughout this paper to represent
280 the IONEX VTEC data of the JPL GIM at 2° × 2° resolutions and 1-hour cadence used in this
281 analysis, though other GIM products are also generated at JPL. For comparison to LANL VTEC,
282 we first interpolate onto a 1° × 1° x 1-hour grid, and then take the median of JPL VTEC values
283 contained in a given 5° × 5° x 1-hour bin corresponding to the LANL VTEC grid (Figure 1, top
284 panel). Because JPL VTEC data are already given on a time cadence of 1 hour, this averaging
285 method only combines about 7 JPL VTEC measurements into each LANL VTEC grid cell. The
286 standard deviation of JPL VTECs within one LANL VTEC grid cell is less than 0.5 TECU.
287
288 **2.5 Jason-3 VTEC**
289 Jason-3, launched in January 2016, is the most recent mission in a series of satellites
290 (TOPEX/Poseidon, Jason-1, and Jason-2) that include dual-frequency altimeters, operating at
291 13.575 GHz (Ku-band) and 5.3 GHz (C-band), to measure the height of the ocean surface to high
292 accuracy. Corrections must be applied to these measurements due to the dispersive nature of the
293 atmosphere that results in path delay of the radar signal. The ionospheric correction, or delay, is
294 directly proportional to the electron content along the ray path and inversely proportional to the
295 frequency (f) squared of the signal. The difference in delay between the altimeters' dual-
296 frequency measurements can be used to calculate the total electron content in the nadir direction
297 (VTEC) from the spacecraft at 1354 km altitude to the surface over the oceans (Imel, 1994). TEC
298 is calculated using the following formula:
299 Ionospheric TEC (electrons/m^2) = -dR * f^2/40.3
300 where dR is the Ku-band ionospheric range correction in meters provided in the Jason-3
301 geophysical data records (GDRs). The sampling rate of the Jason-3 instruments is 1Hz; however

as recommended by Imel (1994) and the Jason-3 Handbook (Dumont et al., 2017), the ionospheric range correction should be smoothed over 100 km or more to reduce instrument noise. To calculate the Jason-3 TEC used in this study, we have averaged the measurements over 18 seconds, which gives us TEC with a resolution of ~2°, or ~200 km. The Jason-3 satellite is in an orbit with a 66º inclination and a 10-day repeating reference orbit, advancing approximately 2º per day. To cover all local times takes about 90 days. While Jason-3 does not provide a dense set of measurements, it does provide a direct measure of VTEC up to an altitude of 1354 km. Altimeter data has been used extensively to validate TEC models and other measurement techniques (e.g., Mandrake et al., 2005; Yasyukevich et al., 2010). In comparing between GNSS TEC and Jason-3 TEC, it is important to account for the fact that GNSS TEC will include integrated electron density up to GNSS altitude in mid-Earth orbit (MEO: 20,200 km for GPS satellites) above the Jason-3 altitude (or plasmaspheric TEC), and thus, would be expected to be slightly higher than Jason-3 TEC by 1-2 TECU.

6 Data Analysis

For comparison of Madrigal VTEC and LANL, we first determine which grid cells had measurements for both data sets. The LANL VTEC contained 143,541 total cells. For the entire year of 2018 there are 134,084 joint cells between LANL VTEC and Madrigal VTEC, or 93.4% of all LANL VTEC cells are also covered by Madrigral VTEC cells. We refer to these cells as "matched" cells. Figure 3 shows a probability density of the differences (LANL VTEC – Madrigal VTEC; solid blue curve). The distribution has an offset of 2.5 TECU with a full-width, half-maximum of 6 TECU. Figure 4a shows a probability density plot of Madrigal VTEC (x-axis) versus LANL VTEC (y-axis) for all matched grid cells. Again, the slight offset of about 2.5

TECU for LANL VTEC is evident compared to Madrigal VTEC, but the distribution clusters well around a line of slope = 1 for the majority of TEC values. The LANL VTEC higher estimate for VTEC values is also evident in this figure, as well as the LANL VTEC tendency to produce larger overestimates at higher VTEC, as shown in Figure 2.

We then do the same comparison of joint grid cells between JPL GIM VTEC and LANL VTEC. Because JPL GIM is a global model it can be compared with LANL VTEC both where Madrigal VTEC data is existing (143,541 cells) and where LANL VTEC adds new coverage compared with GNSS measurements (9,457 cells). Figure 3 shows the probability density function of LANL VTEC – JPL GIM VTEC is plotted for all LANL grid cells (black solid line). While it is not plotted here, we have looked at the probability density for the new coverage separately from the cells matched to Madrigal, and the plots are nearly indistinguishable from the probability density of all LANL cells shown in Figure 3. This indicates that biases and widths are very similar regardless of whether the JPL GIM cell was well-covered by measurements or not. The offset between the LANL VTEC and matched JPL GIM data sets is -3 TECU with a FWHM of 4 TECU. The negative sign on the offset means that JPL GIM generally has a higher TEC value when compared with LANL VTEC. Given that a constant of 2 TECU is added to the IONEX data of JPL GIM due to the bias concern when compared with the altimeter TEC data, the difference between LANL VTEC and JPL GIM would be -1 TECU if that offset were removed. Figure 4b shows the density plot of all JPL GIM VTEC versus LANL VTEC cells.

Given that JPL GIM fills in all cells globally, the comparison of the LANL VTEC data is a comparison to a global fit rather than specific measurements at certain locations as with the Madrigal VTEC. Furthermore, while JPL GIM model is driven by global GNSS TEC measurements, it uses a different set of about 200 IGS receivers than Madrigal GNSS, which

uses all available receivers. The similarity between the LANL VTEC comparisons to Madrigal GNSS and JPL GIM model gives credibility to the LANL VTEC data set.

While JPL GIM and Madrigal GNSS VTEC are dependent on ground-based GPS measurements, and, thus, restricted to land locations, the Jason-3 VTEC measurement is only made over ocean, so can provide a comparison to LANL VTEC coverage over the oceans. However, since both LANL VTEC and Jason-3 VTEC are sparse data sets, the number of matched cells is limited (1594 matching out of 266,630 Jason-3 measurements in 2018). Figure 2 (dashed magenta line) shows the probability density function of the difference between LANL VTEC and Jason-3 VTEC, with an offset of less than 1 TECU, and a FWHM of about 4 TECU. Figure 4c shows the 2-D probability density comparison, indicating that these two data sets agree well for all VTEC values in common.

Because the matched number of events between Jason-3 and LANL VTEC are low, we also plot the probability density of the comparison between Jason-3 and JPL GIM (Figure 3, dash-dotted red line). The comparison between Jason-3 and JPL GIM is nearly identical to that between LANL and JPL GIM, giving further evidence that the LANL VTEC data agrees extremely well with the Jason-3 measurements.

Figure 5 shows the number of counts for which each geographical grid cell contained LANL VTEC data but not Madrigal VTEC data. The red dots show locations of the GNSS receiver sites used by Madrigal. As expected, LANL VTEC is able to add data in locations with limited or no GNSS receivers. For the particular time period shown in Figure 1, LANL VTEC is able to fill in a critical part of the map (central Africa) that was not measured with existing ground-based GNSS receivers. The IGS receiver placement is also tied to accessible land

370 locations with maintainable internet connects, so we assume that similar regions should be
371 lacking IGS data as were lacking Madrigal data.

372

373 **6 Summary and Discussion**

374 This paper has introduced the LANL VTEC lightning data set, and compared it to the
375 community data set of Madrigal VTEC from GNSS receivers, the community ionospheric model,
376 JPL GIM, and oceanic VTEC measurements from Jason-3. Because the sources of error are
377 independent for these various data sets, the comparisons presented here add confidence to all
378 datasets. Specifically, these results show the validity of the novel LANL VTEC dataset.

379 We find the LANL data set is offset 3 TECU higher than Madrigal VTEC with a
380 distribution FWHM of 6 TECU. This finding is consistent with the estimated ionospheric error
381 of 1-2 TECU from the LANL ray-tracing comparison, and the estimated 2-4 TECU error on
382 GNSS L-band absolute TEC. Because the comparisons are made between measurements at the
383 same location and time, this agreement adds confidence to LANL VTEC values globally and
384 over a wide range of TEC values and ionospheric conditions. This comparison indicates that the
385 Madrigal VTEC may be leveled (absolute bias) slightly lower than it should be. A further
386 indication of a lower bias than reality is the finding that Madrigal GNSS VTEC, which includes
387 plasmaspheric TEC contributions, is found to be lower than Jason-3 VTEC, which does not
388 include plasmaspheric TEC. More comparisons must be done before fully understanding how
389 each dataset might need to be adjusted.

390 We find the LANL data set and JPL GIM dataset have an offset of -3 TECU, indicating
391 that JPL GIM VTEC are typically higher than LANL VTEC by 3 TECU. This includes a bias of

+2 TECU on the JPL GIM VTEC based on past comparisons. Thus, the JPL GIM /LANL VTEC offset may actually be less than 3 TECU. The overall FWHM spread between LANL VTEC and JPL GIM is 5 TECU. This indicates that JPL GIM may be leveled slightly higher than they should be, but very comparable with LANL VTEC.

While LANL VTEC and Jason-3 VTEC have very limited overlap in their coverage, the agreement between the two is very good, with an offset of less than 1 TECU. Most grid cells with overlapping coverage have less than 20 TECU, so this comparison does not address higher VTEC values. However, the comparison with better statistics between Jason-3 VTEC and JPL GIM gives similar results to the comparison between LANL VTEC and JPL GIM. This indicates that it is likely the good agreement between LANL VTEC and Jason-3 VTEC would hold for a wider range of grid cells.

Finally, this paper shows that the LANL VTEC provides additional ionospheric measurements in regions currently lacking data for global ionospheric models, such as over the oceans and central Africa. Following papers will delve more deeply into case studies, particularly looking into geographic differences between LANL VTEC and other data sets.

Acknowledgments and Data

This work was supported by the DARPA Defense Sciences Office and the Defense Nuclear Nonproliferation Research and Development Office of the National Nuclear Security Administration. We gratefully acknowledge the U.S. Dept. of Defense for operating and providing data from the sensing systems that this work is based on. Research conducted at the Jet

Propulsion Laboratory, California Institute of Technology, is under a contract with the National Aeronautics and Space Administration.

The LANL VTEC data are publicly hosted by the National Oceanic and Atmospheric Administration in NetCDF format and can be found at https://www.ncei.noaa.gov/archive/accession/0241206. MIT Haystack Madrigal VTEC data can be found at http://www.openmadrigal.org. JPL IONEX data can be found at https://cddis.nasa.gov/Data_and_Derived_Products/GNSS/atmospheric_products.html#iono. Jason-3 data are available at https://www.ncei.noaa.gov/data/oceans/jason3/gdr/gdr/.

GPS TEC data products and access through the Madrigal distributed data system are provided to the community (http://cedar.openmadrigal.org) by the Massachusetts Institute of Technology (MIT) under support from US National Science Foundation grant AGS-1952737. Data for TEC processing is provided from the following organizations: UNAVCO, Scripps Orbit and Permanent Array Center, Institut Geographique National, France, International GNSS Service, The Crustal Dynamics Data Information System (CDDIS), National Geodetic Survey, Instituto Brasileiro de Geografia e Estatística, RAMSAC CORS of Instituto Geográfico Nacional de la República Argentina, Arecibo Observatory, Low-Latitude Ionospheric Sensor Network (LISN), Topcon Positioning Systems, Inc., Canadian High Arctic Ionospheric Network, Centro di Ricerche Sismologiche, Système d'Observation du Niveau des Eaux Littorales (SONEL), RENAG : REseau NAtional GPS permanent, GeoNet - the official source of geological hazard information for New Zealand, GNSS Reference Networks, Finnish Meteorological Institute, and SWEPOS - Sweden.

435 **References**

436 Abarca, S. F., Corbosiero, K. L., & Galarneau, T. J. Jr. (2010). An evaluation of the Worldwide
437 Lightning Location Network (WWLLN) using the National Lightning Detection Network
438 (NLDN) as ground truth. *J.of Geophys. Res., 15*, D18206, doi:10.1029/2009JD013411.

439 Azeem, I., Yue, J., Hoffmann, L., Miller, S. D., Straka, W. C. III, & Crowley, G. (2015).
440 Multisensor profiling of a concentric gravity wave event propagating from the troposphere
441 to the ionosphere. Geophysical Research Letters, 42, 7874–7880,
442 doi:10.1002/2015GL065903.

443 Bilitza, D., L.-A. McKinnell, B. Reinisch, and T. Fuller-Rowell (2011), The international
444 reference ionosphere today and in the future, J. Geod., 85:909-920, doi:10.1007/s00190-
445 010-0427-x.

446 Born, M. and E. Wolf (1999), *Principles of Optics.* Cambridge University Press, 7th ed.

447 Burrell, A. G., N. A. Bonito, and C. S. Carrano (2009), Total electron content processing from
448 GPS observations to facilitate ionospheric modeling, *GPS Solutions, 13*(2),
449 doi:10.1007/s10291-008-0102-3.

450 Coster, A. and S. Skone (2009), Monitoring storm-enhanced density using IGS reference station
451 data, *J. Geod., 83*:345–351, doi:10.1007/s00190-008-0272-3.

452 Decker, D. T., and L. F. McNamara (2007), Validation of ionospheric weather predicted by
453 Global Assimilation of Ionospheric Measurements (GAIM) models, Radio Sci., 42, RS4017,
454 doi:10.1029/2007RS003632.

455 Deng, A., & Stauffer, D. R. (2006), On improving 4-km mesoscale model simulations. *Journal*
456 *of Applied Meteorology and Climatology*, *45*(3), 361–381, doi:10.1175/JAM2341.1.

457 Dowden, R. L., Brundell, J. B., & Rodger, C. J. (May 2002). VLF lightning location by time of
458 group arrival (TOGA) at multiple sites. *J.of Atmos. and Solar-Terr. Phys., 64*(7), 817–830,
459 doi:10.1016/S1364-6826(02)00085-8.

460 Dumont, J., Rosmorduc, V., Carrere, L., Picot, N., Bronner, E., Couhert, A., Guillot, A., Desai,
461 S., Benekamp, H., 2017b. Jason-3 Products Handbook, CNES: SALP-MU-M-OP-16118-
462 CN, Issue 1 rev 4, January 16th 2017.

463 Iijima, B.A., I.L. Harris, C.M. Ho, U.J. Lindqwister, A.J. Mannucci, X. Pi, M.J. Reyes, L.C.
464 Sparks, B.D. Wilson (1999), Automated daily process for global ionospheric total electron
465 content maps and satellite ocean altimeter ionospheric calibration based on Global
466 Positioning System data, *J. Atmos. and Solar-Terr. Phys. 61*, pp.1205-1218,
467 doi:10.1016/S1364-6826(99)00067-X.

468 Imel, D.A., 1994. Evaluation of the TOPEX/POSEIDON dual-frequency ionosphere correction.
469 *J. Geophys. Res. 99* (C12), doi:10.1029/94JC01869.

470 Jakowski, N., C. Mayer, M. M. Hoque, and V. Wilken (2011), Total electron content models and
471 their use in ionosphere monitoring, *Radio Sci., 46*, RS0D18, doi:10.1029/2010RS004620.

472 Kelley, M. C. (2009). *The Earth's ionosphere*. London: Academic.

473 Lay, E. H. (2018). Ionospheric irregularities and acoustic/gravity wave activity above low-
474 latitude thunderstorms. *Geophys. Res. Lett., 45*, doi:10.1002/2017GL076058.

475 Lay, E. H., P.A. Parker, M.E. Light, and P. Colestock (2018), Estimate of errors induced by first
476 order Appleton-Hartree approximation, Los Alamos National Laboratory, LA-UR-18-
477 20440.

478 Lay, E. H., X.-M. Shao, A. K. Kendrick, and C. S. Carrano (2015), Ionospheric acoustic and
479 gravity waves associated with midlatitude thunderstorms, *J. Geophys. Res. Space Physics,*
480 *120*, doi:10.1002/2015JA021334.

481 Lay, E. H., S. Close, P. Colestock, and G. Bust (2011), Development and error analysis of
482 nonlinear ionospheric removal algorithm for ionospheric electron density determination
483 using broadband RF data, *J. Geophys. Res., 116*, A02316, doi:10.1029/2010JA015862.

484 Lay, E. H., Holzworth, R. H., Thomas, J. N., Dowden, R. L., Rodger, C. J., & Pinto Jr., O.
485 (2004). WWLL global lightning detection system: Regional validation study in Brazil.
486 *Geophys. Res. Lett., 31*, L03102, doi:10.1029/2003GL018882.

487 Mandrake, L., B. Wilson, C. Wang, G. Hajj, A. Mannucci, and X. Pi (2005), A performance
488 evaluation of the operational Jet Propulsion Laboratory/University of Southern California
489 Global Assimilation Ionospheric Model (JPL/USC GAIM), *J. Geophys. Res., 110*, A12306,
490 doi:10.1029/2005JA011170.

491 Mannucci, A. J., B. D. Wilson, D. N. Yuan, C. H. Ho, U. J. Lindqwister, and T. F. Runge (1998),
492 A global mapping technique for GPS-derived ionospheric total electron content
493 measurements, *Radio Sci., 33*(3), pp.565-582, doi:10.1029/97RS02707.

494 Mannucci, A. J., B. T. Tsurutani, B. A. Iijima, A. Komjathy, A. Saito, W. D. Gonzalez, F. L.
495 Guarnieri, J. U. Kozyra, and R. Skoug (2005), Dayside global ionospheric response to the

major interplanetary events of October 29–30, 2003 “Halloween Storms”, *Geophys. Res. Lett., 32*, L12S02, doi:10.1029/2004GL021467.

Moses, R. W., and A. R. Jacobson (2004), Ionospheric profiling through radio-frequency signals recorded by the FORTE satellite, with comparison to the International Reference Ionosphere, *Adv. Space Res., 34*, 2096–2103, doi:10.1016/j.asr.2004.02.018.

Nava, B., Coïsson, P., & Radicella, S. M. (2008). A new version of the NeQuick ionosphere electron density model. *J. of Atmos. and Sol.-Terr. Phys., 70*(15), 1856–1862, doi:10.1016/j.jastp.2008.01.015.

Rideout, W. and A. Coster, (2006), Automated GPS processing for global total electron content data, *GPS Solutions*, doi:10.1007/s10291-006-0029-5.

Roussel-Dupre, R.A., Jacobson, A.R., Triplett, L.A (2001). Analysis of FORTE data to extract ionospheric parameters, *Radio Sci. 36* (6), 1615–1630, doi:10.1029/2000RS002587.

Sardón, E., and N. Zarraoa (1997), Estimation of total electron content using GPS data: How stable are the differential satellite and receiver instrumental biases?, *Radio Sci., 32*, 1899-1910, doi:10.1029/97RS01457.

Scherliess, L., D.C. Thompson and R.W. Schunk (2009), Ionospheric dynamics and drivers obtained from a physics-based data assimilation model, *Radio Sci., 44*, RS0A32, doi: 10.1029/2008RS004068.

Scherliess, L., R.W. Schunk, J.J. Sojka, D.C. Thompson and L. Zhu (2006), Utah State University Global Assimilation of Ionospheric Measurements Gauss-Markov Kalman filter

model of the ionosphere: Model description and validation, *J. Geophys. Res., 111*, A11315, doi: 10.1029/2006JA011712.

Schunk, R. W., et al. (2004), Global Assimilation of Ionospheric Measurements (GAIM), *Radio Sci., 39*, RS1S02, doi:10.1029/2002RS002794.

Schunk, R. W., and J. J. Sojka (1996), Ionosphere-thermosphere space weather issues*, J. Atmos. Sol. Terr. Phys.*, *58,* 1527–1574, doi: 10.1016/0021-9169(96)00029-3.

Vierinen, J., Coster, A. J., Rideout, W. C., Erickson, P. J., and Norberg, J.: Statistical framework for estimating GNSS bias, *Atmos. Meas. Tech., 9*, 1303-1312, 10.5194/amt-9-1303-2016, 2016.

Wang, W., J. Lei, A. G. Burns, S. C. Solomon, M. Wiltberger, J. Xu, Y. Zhang, L. Paxton, and A. Coster (2010), Ionospheric response to the initial phase of geomagnetic storms: Common features, *J. Geophys. Res., 115*, A07321, doi:10.1029/2009JA014461.

Yasyukevich, Yu.V., E.L. Afraimovich, K.S. Palamartchouk, P.V. Tatarinov (2010), Cross testing of ionosphere models IRI-2001 and IRI-2007, data from satellite altimeters (Topex/Poseidon and Jason-1) and global ionosphere maps, *Advances in Space Research, 46*(8), doi:10.1016/j.asr.2010.06.010.

Yizengaw, E., M. B. Moldwin, A. Komjathy, and A. J. Mannucci (2006), Unusual topside ionospheric density response to the November 2003 superstorm, *J. Geophys. Res., 111*, A02308, doi:10.1029/2005JA011433.

536 **Figure Captions**

537

538 **Figure 1.** 1-hour snapshot at 18:30:00UT 09 June 2018 of gridded global VTEC for JPL GIM
539 (top) Madrigal GNSS (middle) and LANL VTEC (bottom).

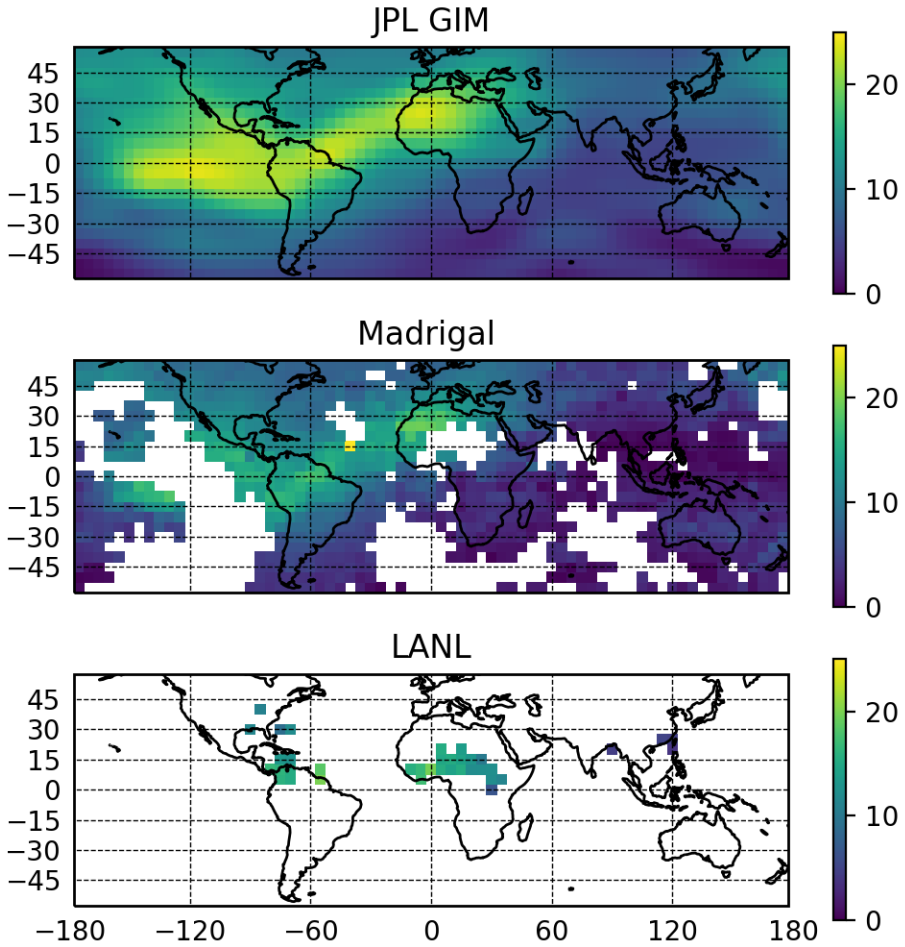
540 **Figure 2**. A 2-D probability density plot of TEC difference introduced by ionospheric variability
541 and geometry between the known VTEC (x-axis) and the estimated LANL VTEC (y-axis).

542 **Figure 3**. A probability density plot generated from the difference between LANL VTEC and:
543 Madrigral VTEC for matched cells (blue thin solid), JPL GIM (black thick solid), Jason-3
544 (magenta dashed). Red dash-dotted line shows Jason-3 – JPL GIM VTEC.

545 **Figure 4**. 2-D probability density plots of (a) Madrigal VTEC, (b) JPL-GIM VTEC, and (c)
546 Jason-3 VTEC on x-axis and LANL VTEC on y-axis. Colorbar represents density in counts per
547 sq. TECU unit area / total counts.

548 **Figure 5**. Counts of grid cells with LANL VTEC without comparable Madrigal VTEC gridded
549 data. The most notabled areas of increased coverage are in continental Africa, southwest of
550 Mexico, and over the oceans. The first two areas correlate well to significant lightning activity.
551

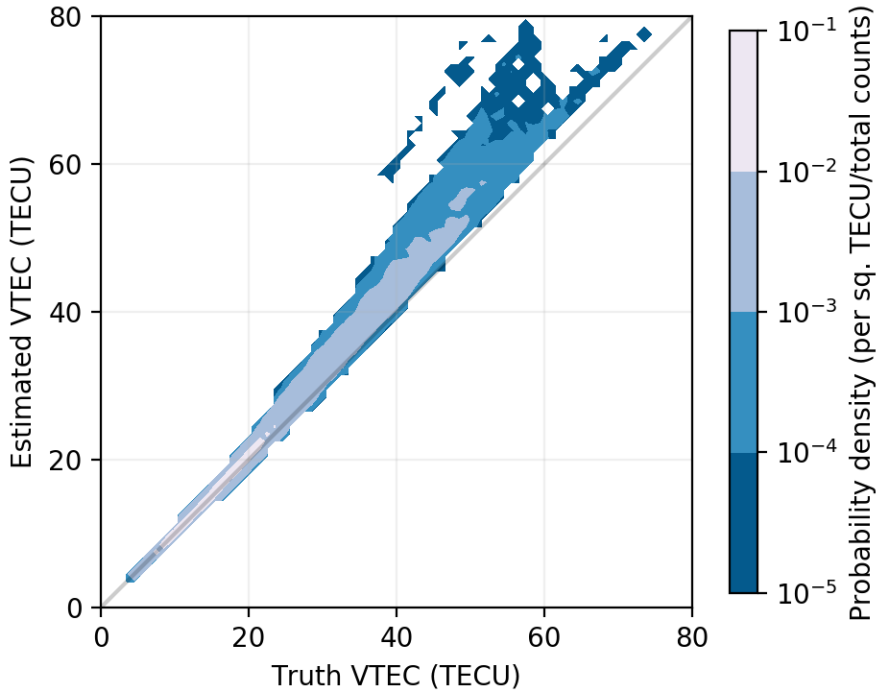
552 **Figures**



553

554 **Figure 1.** 1-hour snapshot at 18:30:00UT 09 June 2018 of gridded global VTEC for IONEX

555 TEC data of JPL GIM (top) Madrigal GNSS (middle) and LANL VTEC (bottom).

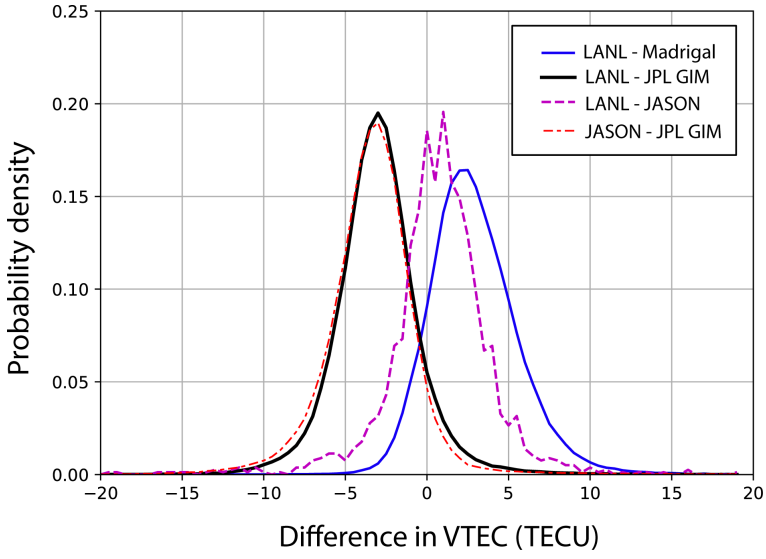


556

557 **Figure 2**. A 2-D probability density plot of TEC difference introduced by ionospheric variability

558 and geometry between the known VTEC (x-axis) and the estimated LANL VTEC (y-axis).

559

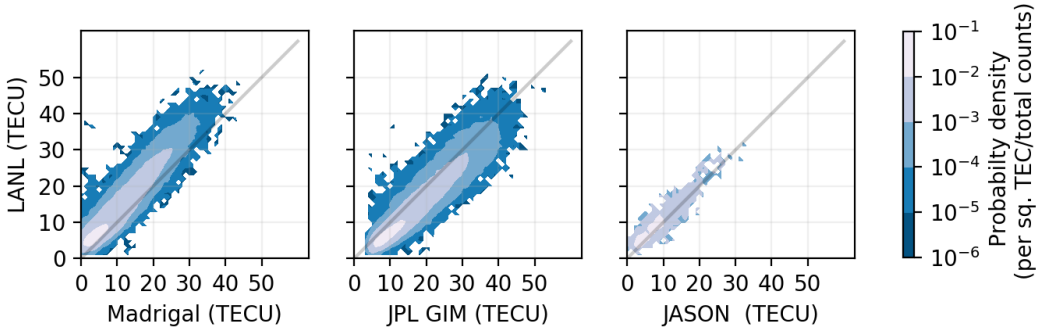


560

561 **Figure 3**. A probability density function generated from the difference between LANL VTEC
562 and: Madrigral VTEC for matched cells (blue thin solid), JPL GIM (black thick solid), Jason-3
563 (magenta dashed). Red dash-dotted line shows Jason-3 – JPL GIM IONEX VTEC.

564

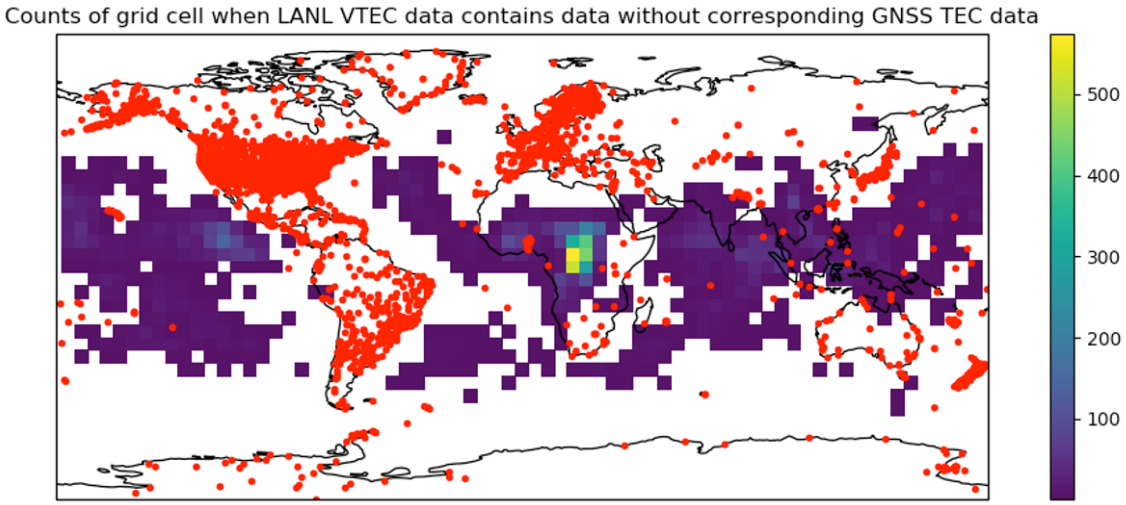
565



566

567 **Figure 4**. 2-D probability density plots of (a) Madrigal VTEC, (b) JPL-GIM IONEX VTEC, and
568 (c) Jason-3 VTEC on x-axis and LANL VTEC on y-axis. Colorbar represents density in counts
569 per sq. TECU unit area / total counts.

570



571

572 **Figure 5**. Counts of grid cells with LANL VTEC without comparable Madrigal VTEC gridded
573 data. The most notabled areas of increased coverage are in continental Africa, southwest of
574 Mexico, and over the oceans. The first two areas correlate well to significant lightning activity.