On Constructing a Realistic Truth Model Using Ionosonde Data for Observation System Simulation Experiments

Joseph Hughes[1*], Victoriya Forsythe [2†], Ryan Blay [1], Irfan Azeem [1], Geoff Crowley [1], Walter "Junk" Wilson [3‡], Eugene Dao [4§], Jonah Colman [4], and Richard Parris [4]

[1]Orion Space Solutions
[2]Naval Research Lab
[3]Booz-Allen-Hamilton
[4]Air Force Research Lab

Key Points:

- The amplitude and spatio-temporal extent of ionospheric variabilities are quantified across altitude, local time, season, and solar cycle.
- A method to add these variations into a smooth model is demonstrated and replicates ionospheric data well.
- This noisy truth model enables more accurate observation system simulation experiments.

*282 Century Place, Suite 1000, Louisville, CO, 80027
†4555 Overlook Ave SW, Washington, DC 20375
‡745 Space Center Dr., Colorado Springs, CO, 80915
§3550 Aberdeen Avenue SE. Bldg 427. Kirtland AFB, NM 87117-5776

Corresponding author: Joseph Hughes, joe.hughes@orionspace.com

Abstract

The ionosphere contains many small-scale electron density variations that are under represented in smooth physics-based or climatological models. This can negatively impact the results of Observation System Simulation Experiments, which use a truth model to simulate data. This paper addresses this problem by using ionosonde data to study ionospheric variability and build a new truth model with empirically-driven variations. The variations are studied for their amplitude, horizontal and vertical size, and temporal extent. Results are presented for different local times, seasons, and at two different points in the solar cycle.

We find that these departures from a smooth background are often as large as 25% and are most prevalent near 250 km in altitude. They have horizontal spatial extents that vary from a few hundred to a few thousand kilometers, and typically have the largest horizontal extent at high altitudes. Their vertical extents follow the same pattern of being larger at high altitudes, but they only vary from 10s of km up to 200 km in vertical size. Temporally, these variations can last for a few hours. The procedure for using these spatial and temporal distributions to add empirically-driven variance to a smooth truth model is outlined. This process is used to make a truth model with representative variations, which is compared to ionosonde data as well as GPS Total Electron Content (TEC) data that was not used to inform the model. The new model resembles the data much better than the smooth models traditionally used.

1 Introduction

The ionosphere is a region of Earth's upper atmosphere that extends from near 100 km to 1,000 km in altitude. It is lightly ionized and is influenced by the Sun, Earth's magnetic field, atmospheric tides, and other processes in the troposphere which propagate up to high altitudes.

There are two central reasons to specify the state of the ionosphere: in order to test and extend our knowledge through application of the scientific method, or to construct an operational system to support users. In either case, decisions must be made about how many sensors, of what type, and at what locations to deploy in order to gather observations. A minimum set of observations are usually required to meet a quality metric related to the operational system. For example, there are a minimum number of GPS satellites required to provide a user geolocation support. A cost effective way to evaluate the related sensor requirements is to perform an Observing System Simulation Esperiment (OSSE). An OSSE is a simulated experiment that attempts to quantitatively assess different sensor architectures and how well they meet the metrics. Such an assessment can also be coupled to a cost model in order to provide a cost benefit analysis. OSSEs are common for operational systems requiring tropospheric specification e.g. (Zeng et al., 2020) but are relatively new for the ionosphere. There are three fundamental steps to performing an OSSE. First, a truth model must be constructed that is capable of being sampled to generate synthetic measurements of the proposed sensor architectures. In the troposphere, the truth model is often referred to as a "nature run". Second, the simulated sensor data are provided to the operational model (or a reasonable facsimile) that is used to provide information about the metrics under consideration. Third, the different architectures are compared according to the quantitative requirements of the users.

Many of the limited number of OSSEs performed for ionospheric applications have relied on physics-based models as truth models. Such models cannot currently faithfully represent all of the relevant time and space scales typically observed by an ionospheric sensor. This is not surprising given the computational limitations but also because the models are designed to represent some phenomena while ignoring others. This limitation can lead to overly optimistic assessments of the utility of a given sensor architec-

ture. For example, an OSSE using a smooth truth model will show improvements between two sensor locations even though no measurements are taken there because the truth model is well-approximated by the interpolant between these measurements. Since the real-life ionosphere is more variable and is not well-modeled by an interpolant, this is an optimistic assertion of the improvement in the specification. It is therefore crucial to have a truth model with realistic variances to ensure accurate OSSE results. In other words, it is impossible to know if a sensor system is capable of resolving small-scale features if there are no small-scale features in the truth model.

In order to illustrate some of these issues in a more quantitative fashion, consider the critical frequency of the ionospheric F2 layer above Boulder as observed by an ionosonde and as represented by the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIE-GCM). TIE-GCM (Richmond et al., 1992) is considered a state of the art physics based model. The ionosonde measurements are taken with a DGS-256 ionosonde using Artist for autoscaling. The left panel of Fig. 1 shows a time series for four days in June of 2018. The two sharp spikes in the model-based foF2 are due to the edge of the Aurora incorrectly extending down to Boulder which causes the maximum frequency to increase. It is immediately apparent that TIE-GCM results are much smoother than the ionosonde observations and do not faithfully represent the sensor output.

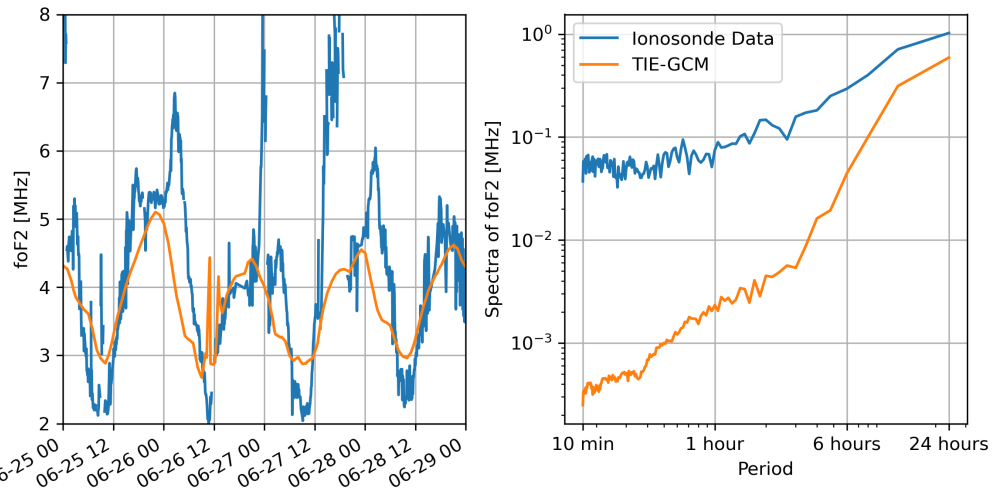


Figure 1. Observed and modeled foF2 timeseries (left) and spectra (right) above Boulder, CO.

The right panel of Figure 1 shows the spectra of the foF2 data for the months of June, July, and August 2018. Welch's periodogram is used to robustly estimate the amplitude of oscillations at short periods. This spectral analysis method splits up a long time series into small segments, takes the Fast Fourier Transform (FFT) of all these pieces, and averages the spectra together. The spectra for both the model and observed foF2 data are shown with orange and blue lines respectively.

The amplitude of the diurnal oscillation is slightly weaker in the model than the data, which is also evident in the time series where the model foF2 has about a 1/2 MHz amplitude (1 MHz peak-to-peak) daily oscillation in contrast to the ionosonde data which shows closer to 1-2 MHz amplitude. More importantly, there is more high-frequency variability in the data than the model. On 1-hour time scales, the data has 0.1 MHz variations, while the model only has ∼0.002 MHz variations. This means that the model is 50 times smoother than the data. Over a 10-minute period, this factor is closer to 200.

Clearly, the model is temporally much smoother than the real ionosphere measured by the ionosonde. In order to perform accurate ionospheric OSSEs, the truth model must faithfully represent the spatial and temporal variances associated with the sensor measurements. In addition to enabling realistic truth models, studying the spatial and temporal variability scales can inform the correlation lengths and times used by many assimilative models, and provide scientific insight to the underlying physics.

The ionospheric temporal, vertical, and horizontal correlation have been studied in many prior studies, primarily driven by the need for accurate estimation of the correlation matrix for assimilation. The horizontal correlation length is studied using Total Electron Content (TEC) data in (Gail et al., 1993; Klobuchar & Johanso, 1977; Shim et al., 2008; Yue et al., 2007; Forsythe, Azeem, & Crowley, 2020). The horizontal correlation length was studied with foF2 data rather than TEC (McNamara, 2009) using small clusters of ionosondes in Europe and South Africa. The vertical correlation distance has only been the subject of two studies known to the authors (Yue et al., 2007; Forsythe, Azeem, Crowley, & Themens, 2021). Both of these studies used Incoherent Scatter Radar (ISR) data. The ISR profiles are fit with a Chapman profile in (Yue et al., 2007), so there is no information on small variations. The temporal correlation time is only studied in one prior published study (Forsythe, Azeem, Crowley, Makarevich, & Wang, 2020) to our knowledge. Lastly, the 2017 conference abstract (Holmes et al., 2017) presents the first simultaneous and consistent study of horizontal and temporal variations for the purpose of making a better truth model for use in an OSSE. Four ionosondes in North America are used to study the spatial and temporal correlation of the critical frequency foF2. This paper builds off of this prior work to create a simultaneous and consistent study of temporal, vertical, and horizontal correlation of the electron density as well as the amplitudes of the variations. We include more than twenty ionosondes over two years to analyse the spatial and temporal density variations. The parameters are found as functions of altitude, local time, season, and solar cycle point. We use mid-latitude data from 2014 and 2018 to capture the ionosphere at solar minimum (2018) and solar maximum (2014). We also present a method for using these correlation distances to create a truth model with realistic variances. This is inspired in part by (Holmes et al., 2017) but contains much that is novel since foF2 was the only parameter varied in (Holmes et al., 2017) while this work varies electron density at all altitudes.

This paper is organized as follows: Section 2 describes the ionospheric data and its processing. Section 3 shows the amplitude and vertical, horizontal, and temporal extent of the variance not resolved by the model. Section 4 describes a method to use these parameters to add this variance to a smooth model to make a more realistic truth model. Section 5 validates this new noisy truth model against real measurements. Finally, conclusions and recommendations for future work are made in Section 6.

2 Data Processing

The National Oceanic and Atmospheric Administration (NOAA) maintains a large quantity of ionosonde data on their FTP site (`ftp://ftp.ngdc.noaa.gov/ionosonde/data`). This data is compiled from a number of different sensor, antenna, and software configurations and is not necessarily of uniform quality. Only cursory checks of the data were performed for this study. The majority of these ionsondes are DPS-4D models which use Artist 5002 for scaling. Data from 2014 and 2018 is used and is divided based on season: summer (June, July, and August), winter (December, January, and February), and equinox (March, April, May, September, October, and November). Only data from ionosondes between 20 and 55^{o} in magnetic latitude are included. Between 21 and 28 ionosondes meet these requirements depending on the year and season. Only data taken when the planetary K index (K_P) is less than 4 are included.

Many of these ionosondes have different measurement cadences, and all of them exhibit data gaps. Even though ionosondes only measure the bottom side of the ionosphere, the auto-scaled profiles that are available usually contain model-driven data for the topside. Some auto-scaling software packages also use model-derived data for the E layer. In this study, only data below the altitude of the maximum plasma frequency (hmF2) and within the frequency measurement range of the ionosonde are analyzed. Since the altitudes and frequencies that are directly measured change with time even for the same ionosonde, careful data management is required to avoid accidentally analyzing model driven data. Once the extrapolated data is removed, the remaining data is interpolated onto a uniform 10 km grid. This is done for each measurement from each ionosonde.

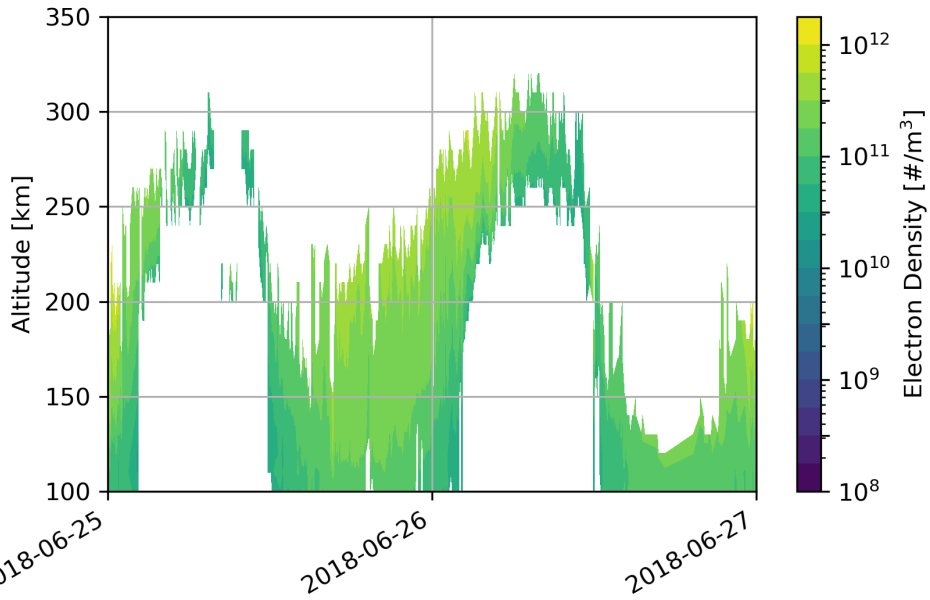


Figure 2. Observed electron densities in Boulder. Only data within the ionosonde measurement range and below hmF2 are shown.

A sample of the processed ionosonde data from Boulder, CO is shown in Fig. 2. In this figure, time is shown on the x axis, altitude on the y axis, and the electron density is shown with color. Only two days of data from June 25th to June 27th, 2018 are plotted here to show detail. Since the nominal sampling cadence for this ionosonde is 5 minutes, this type of plot shows hundreds of profiles at once. Both the temporal gaps and the variable measurement altitudes are visible. The highest altitudes directly measured by the ionosonde occur during local night. There are no direct measurements at low altitudes during this time when the E region disappears and the plasma frequency below ∼225 km is too low for the ionosonde to measure. The obvious outlier on the first night where the recorded densities are all between 200 and 220 km is likely the result of an auto scaling failure. This example is typical in that such failures often occur in clusters and and among data drop outs.

The TIE-GCM model is interpolated to the place and time of each ionosonde measurement and is shown for Boulder in Fig. 3a. The Aurora which causes the two foF2 spikes in Fig. 1 can be seen here as well. In principle, any smooth climatological or physics-based model can be used for this work. The most important model requirements are that

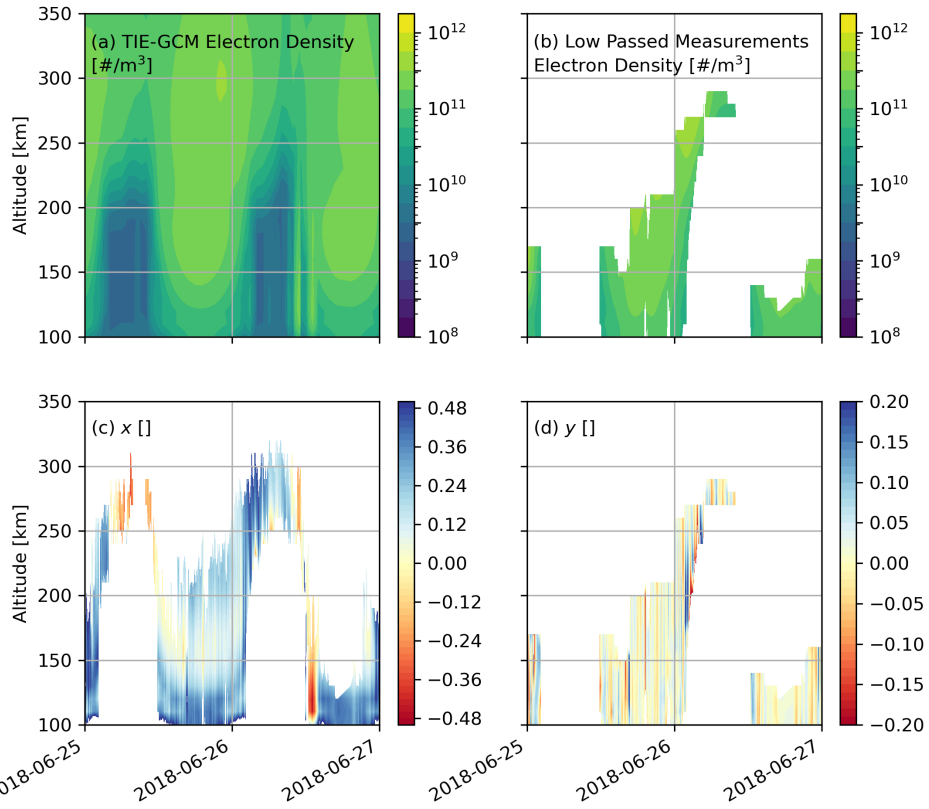


Figure 3. Processing technique used to make the parameters x and y. (a) The TIE-GCM model for the same place and time as the ionosonde data shown. (b) A low-passed version of the ionosonde measurements. (c) The parameter x. (d) The parameter y.

it reproduce the smooth features faithfully and with minimal bias. This will allow us to add unresolved features as normalized perturbations in order to construct the truth model. The TIE-GCM model is used as the basis for this study because some of the authors have used it previously for OSSEs and it may capture some features not present in a median climatology. Here we define the perturbation in terms of a variable x, the difference of the log densities of the ionosonde and corresponding TIE-GCM electron densities which can also be expressed as a ratio in eqn. (1).

$$x = \log_{10}(N_{eI}) - \log_{10}(N_{eM}) = \log_{10}\left(\frac{N_{eI}}{N_{eM}}\right) \tag{1}$$

In this equation, N_{eI} is the electron density from the ionosonde, and N_{eM} is the electron density from the model. The variable x is 0 when the ionosonde and model densities are equal, ∼0.3 when the ionosonde density is double that of the model, and about -0.3 when the model is double the ionosonde density. A lognormal basis is chosen because electron densities are lognormally distributed in the ionosphere (Garner et al., 2005), which

means that x can be well described by a Gaussian distribution. This basis also implicitly assumes that the ionosonde data is as likely to be half the model density as it is to be twice the model density. This variable is computed for each ionosonde for all available times and altitudes and is the heart for all but one of the unresolved variance characteristics. Figure 3c shows x between 100 and 350 km in altitude. A bifurcating color scale is used to better show deviations from zero. Often, x is near zero, as evidenced by the yellow shaded regions. However, sometimes it is as high as 0.5, which means that the ionosonde electron density is $10^{0.5} \sim 3.16$ times larger than the model density. Other times, x is as low as -0.5 when the ionosonde densities are ~ 3.16 times lower than the model densities.

The parameter x captures the unresolved variance well, but occasionally has large, low-frequency differences with the ionosonde model. For example, there is a lot more blue (indicating higher measurement than model densities) than red in Fig. 3c. If x was used to compute the amplitudes of the variations, very high results would be produced because of these long-period differences which are not the subject of this study. For cases where such low frequency differences would impact the results, such as a consistent diurnal bias, we compute and use the variable y where the ionosonde data is compared to a low-passed version of itself rather than the model results. This approach ensures that only the high-frequency differences are used to make the amplitudes. The low pass filter is a sixth-order Butterworth filter with a cutoff frequency corresponding to a 3-hour period and is applied to each altitude separately. Data with less than 1.75 hours of continuous measurements are not used. This also has the effect of removing periods where the autoscaling software failed as discussed for Fig. 2. The low passed ionosonde data are shown in Fig. 3b where there are frequent gaps in coverage due to the continuous time requirement.

$$y = \log_{10}\left(\frac{N_{eI}}{\text{Low Pass}(N_{eI})}\right) \tag{2}$$

Once the low-passed version of the measurements is created, the variable y is used to quantify the amplitude of high-frequency variations. The parameter y is defined similar to x, but with the low passed data in the denominator. The variable y is also well-modeled with a Gaussian distribution.

We acknowledge that the parameters used to create y are arbitrary, and that it is in principle possible to perform the following analyses using only y instead of x. However, the resulting variances may not be ideal to add back onto a TIE-GCM model to make a more realistic truth model. Such further investigations are exciting avenues for further study.

3 Characteristics of the Unresolved Variance

This section presents the vertical, horizontal, and temporal characteristics as well as the amplitude of the unresolved variance. Results are presented for four local time bins corresponding to the morning sector (3 to 9 LT), noon sector (9 to 15 LT), afternoon sector (15 to 21 LT), and the nighttime sector (21 to 3 LT). Data from the solar maximum (2014) and solar minimum (2018) are used and divided into three seasons: Summer (June, July, August), Equinox (September, October, November, March, April, May), and Winter (December, January, February). The characteristics are studied as a function of altitude from 100 to 400 km inside each of these local time bins for each season and year combination where the data is available. Referring to Fig. 3c for example, low-altitude data is expected to be sparse in the nighttime sector.

It should be noted that the ionosonde data contains both instrument noise and processing artifacts in addition to the real ionospheric measurement. It is difficult to separate these contributions and know whether a variation has instrumental of ionospheric

origin. In each subsection, we discuss our approach to attributing variations in x to either the instrument or to the ionosphere.

3.1 Vertical Correlation

The vertical correlation length describes how well x at one altitude is correlated to x at the same time, latitude, and longitude, but a different altitude. At the vertical resolutions used in this study (10 km), we expect both the model and ionosonde data to be relatively smooth and therefore expect that x will be a relatively smooth function of altitude. In Fig. 3c, x does not change rapidly with altitude. To study this quantitatively, the Pearson correlation coefficient r is calculated for x at each possible pair of altitudes from the covariance and the standard deviations as follows:

$$r = \frac{\mathrm{Cov}(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}} \tag{3}$$

Here, the subscripts 1 and 2 refer to the two different altitudes, and σ is the standard deviation. Figure 4 shows the square of this value for the Boulder ionosonde for summer 2018. The four local time bins are shown as separate panels. Only cells with at least 10% of the maximum possible data are considered. This reduces the impact of outliers when the total data volume is low.

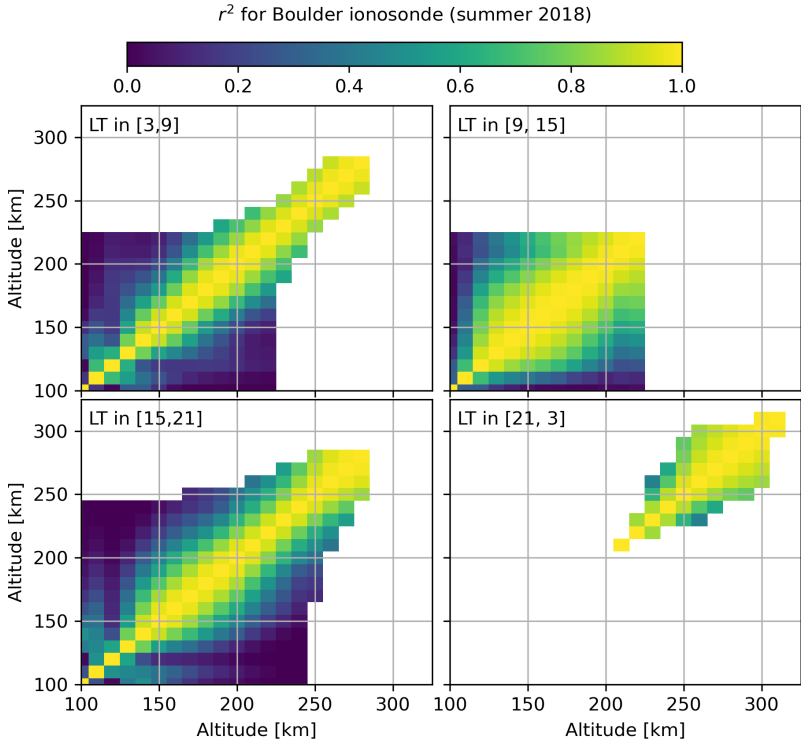


Figure 4. Correlation coefficients between different altitudes for the Boulder ionosonde for summer 2018. The four panels show different local time bins indicated in the upper left corner of the panels.

The correlation coefficient is high when the two altitudes are close, and exactly 1 when the two altitudes are the same. Altitudes far from each other are poorly correlated

as evidenced by the deep blue regions in the upper left and lower right corners of the plots. altitude pairs with insufficient simultaneous data are shown as blank cells. Referencing Fig. 3, we expect the nighttime LT sector to only have high altitude data. The correlation drops more quickly with altitude difference at some altitudes and local times than others. This can be seen by the wider yellow central band in the noon sector than the morning sector for example.

$$r^2 = e^{-|\Delta a|/d_v} \tag{4}$$

For each ionosonde and every height, the correlation coefficient is plotted against the altitude difference and fit with an exponential decay function described by eqn. (4) where d_v is the vertical correlation length and Δa is the difference in altitude. Since this function is guaranteed to be 1 when $\Delta a = 0$, we implicitly assume that all the unresolved variance is attributed to the ionosphere and none to the instrument when calculating the vertical correlation length.

Altitude and LT sectors with fewer than 1/10th of the possible data are not considered. Since the absolute value of the altitude difference is used, correlation to altitudes above is not separated from correlations to altitudes below. The median value of the vertical correlation length is shown as a solid line in Fig. 5 for all seasons and both years corresponding to solar maximum (2014) and solar minimum (2018). The range between the 25% and 75% quantiles (the middle half) is shown as a shaded region of the same color to give a sense of the variation from one ionosonde to another. Some of this variation is undoubtedly the result of autoscaling errors as evidenced by the systematically large nighttime values. A maximum of 200 km is enforced to keep the optimizer from choosing a very high correlation length in the nighttime sector where there is poor altitude coverage to compare against. The correlation length is generally small at lower altitudes, which indicates the prevalence of small-scale variations. This means that the differences between the ionosonde data and the model data will be more consistent over larger altitude ranges at high altitudes than at low altitudes.

Prior work from (Forsythe, Azeem, Blay, et al., 2021) found similar structure in the vertical correlation length in that the correlation length increases with altitude and is typically between 50 and 200 km, however they do not use a 200 km limit so their high altitude correlation lengths are much larger. Our use of exclusively bottomside ionosonde data also contributes to this difference. (Yue et al., 2007) also used ISR data to study the vertical correlation length and and found a nearly linear increase in the vertical correlation length from near 60 km to 200 km.

3.2 Horizontal Correlation

Pairs of ionosondes are used to compute the horizontal correlation length. The values of x at one ionosonde are correlated with the values of x at the same time and altitude but at a different ionosonde. The correlation coefficient from this pair of ionosondes is recorded along with the horizontal distance between the ionosondes. The horizontal distance used here is the great circle distance, e.g (Vallado, 2013) Chapter 11.3. This process is similar to that used by (McNamara, 2009) to study the horizontal correlation length of foF2 deviations.

This process is repeated for all pairs of ionosondes, and for every year, season, local time sector, and altitude. Since many ionosondes do not have the same sampling cadence and do not always have data for the same time and altitude, considerable data wrangling is required. Time differences smaller than 1 minute 26 seconds (1/1000th of a day) are ignored and results are compared directly. If there are fewer than 10 points in common for a given pair of ionosondes in a given altitude-LT bin, the pair is not considered.

An example of the output is shown in Fig. 6. Different colors represent different LT bins and the size of the symbol indicates the number of measurements that were used

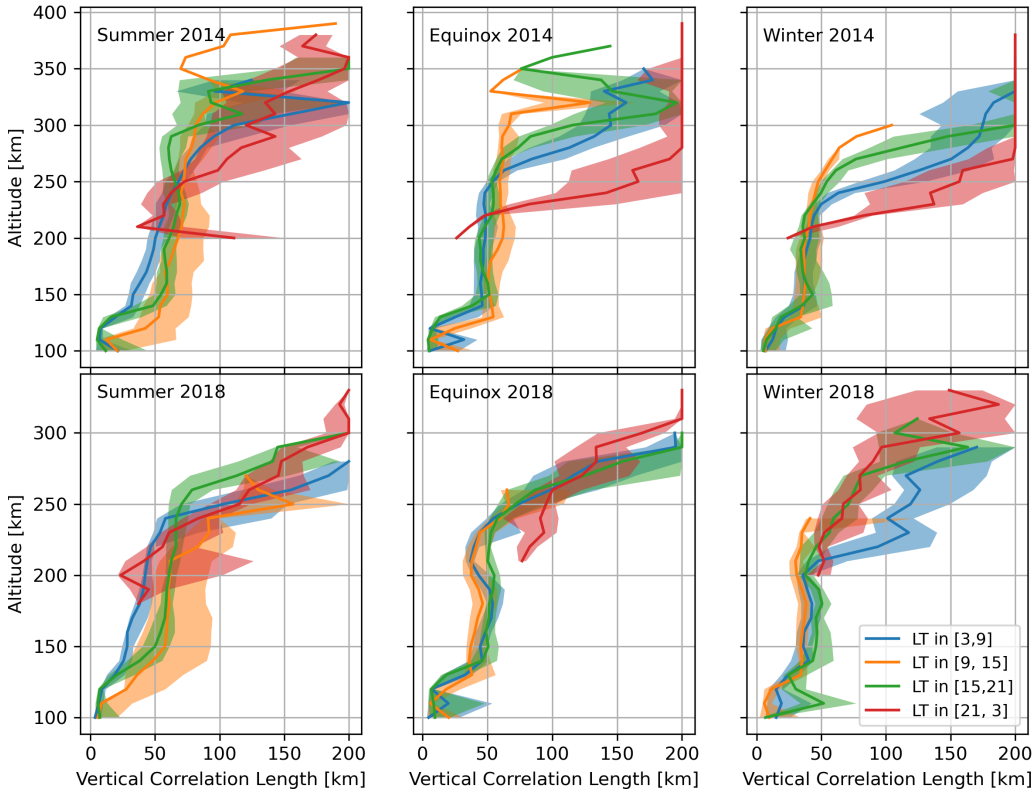


Figure 5. Vertical correlation length for all ionosondes in two years and three seasons. Solid lines show the median value while shading shows the region between the 25th and 75th percentiles or middle half of the data.

to make the correlation coefficient. Large dots indicate values that are less susceptible to outlier-induced errors. Since any two ionosondes must be no further than six hours apart in local time to be in the same local time sector, there is an upper limit of one-quarter of the earth's circumference on the maximum horizontal distance, which explains why there are no points further away than about 8,000 km.

The data from each LT bin and altitude are fitted with an exponential decay function similar to eq. (4) but with the ratio of horizontal separation to horizontal correlation length in the exponent. The number of points used to make the correlation coefficient are used as a weighting in the curve fit. These correlation lengths are used to make the solid curves in Fig. 6 for each LT bin. Although the exponential decay fits capture the trend of the data, much variability still exists. Since the fit is guaranteed to be 1 for no horizontal separation, we also implicitly assume that all the unresolved variance is attributed to the ionosphere and none to the instrument when calculating the horizontal correlation length. Since the closest pair of ionosondes are about 110 km apart, we cannot comment on how strong or weak an assumption this is.

The fit horizontal correlation lengths are shown in Fig. 7 with color indicating the LT bin. A shaded region shows the $\pm 2\sigma$ range for the correlation length that is provided by the curve fit. Note that the x scale varies for each panel. This scale cuts off some data, but shows detail well and allows for easy comparison across seasons.

For many seasons, altitudes, and LT bins, the correlation lengths are larger during solar maximum (2014) than solar minimum (2018). This indicates that deviations

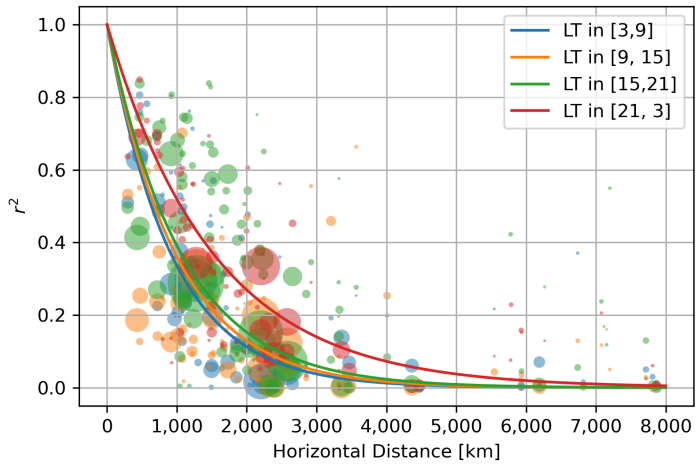


Figure 6. Scatter plot of correlation coefficient as a function of horizontal separation for all possible pairs of ionosondes. Color of dots indicates the LT sector and size indicates the amount of data used to make correlation coefficient. An altitude of 250 km is used for summer 2014.

of the ionosonde data from the smooth TIE-GCM model are larger in horizontal size during solar maximum. The correlation length also increase with altitude generally, which indicates that small scale variations are more common at low altitudes. This trend is most apparent in summer and least apparent in winter. The largest horizontal scales are typically found at night, especially at high altitudes. The horizontal correlation length peaks near 300 km during the afternoon in equinox seasons, but not in other local time bins or seasons.

All of the prior comparable work for horizontal correlation using GNSS-derived TEC data ((Gail et al., 1993; Klobuchar & Johanso, 1977; Shim et al., 2008; Yue et al., 2007; Forsythe, Azeem, & Crowley, 2020)) is difficult to compare since the TEC will average out small vertical structures that will impact our approach using ionosonde electron density. In addition, none of these models use TIE-GCM as a baseline to compute deviations. Many of them also separate the meridonal and zonal distances, which we do not separate. Nevertheless we note that our approach yields values between 1,000 and 3,000 km in most cases near the F2 peak which contributes the most to the TEC. This range is similar to that found in the majority of the TEC-based methods taking into account differences across season and year. The work in (Holmes et al., 2017) used foF2 values, and found a correlation length near 1,000 km for summer 2015 while (McNamara, 2009) found values between 700 and 1500 km depending on whether the zonal or meridional distance is considered and the solar activity. There is no known work to compare our results at non-foF2 altitudes.

3.3 Temporal Correlation

In addition to vertical and horizontal correlation, the parameter x also has temporal structure. If x is high at one time, it is likely to still be high a few minutes later since large electron density gradients are not expected to occur in time. As a case study, the calculated noise values in the ionosonde data shown in Fig. 3c do not abruptly change in time. To study the spectral characteristics, the ionosonde data are once again binned in altitude and LT. Since each LT bin is only 6 hours long, there is never more than 6 hours of continuous data. This limitation, in addition to the frequent gaps in data, makes

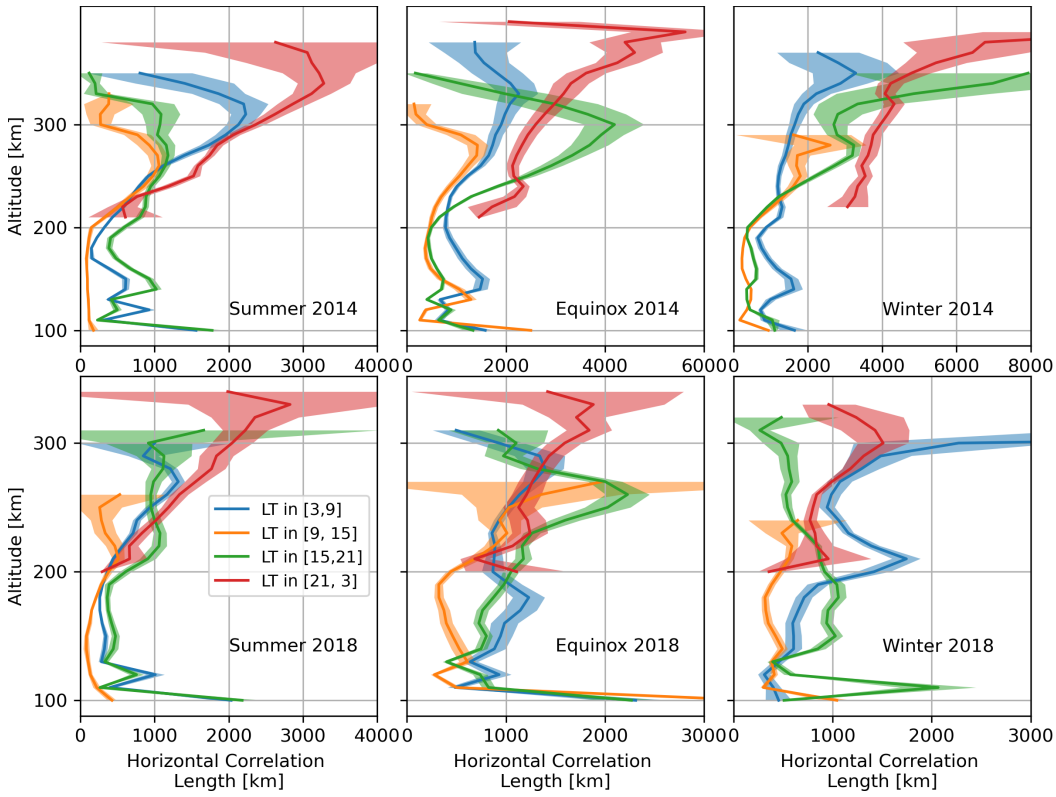


Figure 7. Horizontal correlation length over two years and three seasons. Solid lines show the fit value while shading shows the $\pm 2\sigma$ region.

a traditional Fourier transform infeasible. The repeated pattern of 6 hours of data followed by 18 hours of no data introduces spurious signals to a Lomb-Scargle Peridogram (LSP) at harmonics of 18 and 6 hours. Using multiple LSPs for every 6 hour block and averaging (a modification of Welch's Peridogram) also does not provide robust results. The most stable results come from a standard Fourier transform of data interpolated onto a regular sampling grid. For every ionosonde, local time, and altitude, x is split into segments whenever there is a gap larger than 30 minutes. Each of these segments are interpolated onto a regular grid with the median cadence and Fourier transformed. After this is done for every segment, an exponential decay function is fit to all the spectra. This is shown just for the Boulder ionosonde at 150 km altitude in summer of 2018 in Fig. 8.

The left panel shows a time series of x for a short time period at 150 km altitude. Both short and long-period trends can be seen. There are no nighttime measurements at this altitude which is why only three local time sectors are shown. The right panel shows the spectra of the entire season of data (not just the few days shown in the left panel). Since there are multiple short segments for each LT bin, there are many separate spectra which are all shown as scatter points. Since these segments do not all have the same number of points, the periods from the Fourier transform are not identical. These points are fit with a modified exponential decay function:

$$Ae^{-f/f_d} + A_0 \tag{5}$$

where A is the amplitude, f is the frequency, f_d is the decay frequency, and A_0 is the offset. This equation differs from eqn. (4) in that there are additional parameters to describe the amplitude and offset. The amplitude is needed since x is by no means constrained to be 1 at a frequency of 0, in contrast to the correlation coefficient based anal-

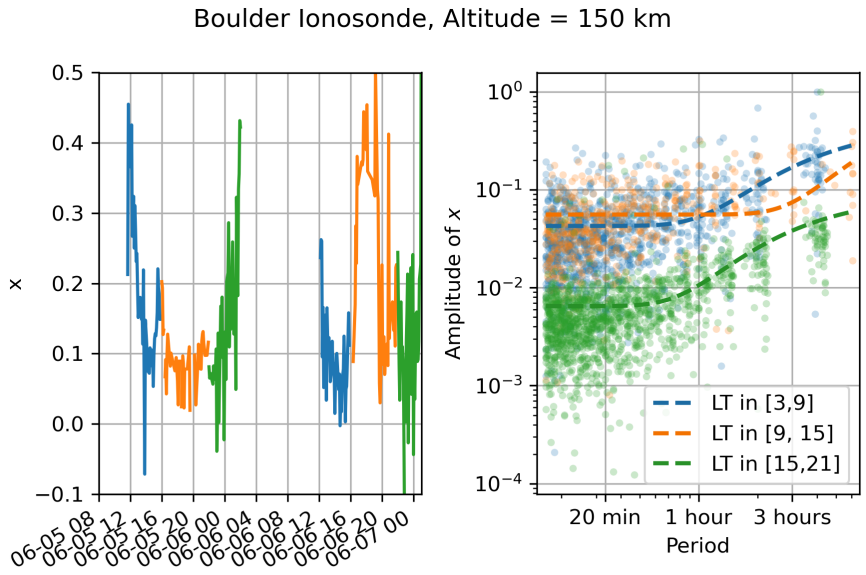


Figure 8. Time series and spectra of Boulder ionosonde data in the summer of 2018. left panel shows x for one and a half days while right panel shows a modified Fourier Transform of x for the whole season at different local times.

ysis where the correlation is constrained to be 1.0 for no altitude difference. The fits for all LT sectors are shown as dashed lines in the right panel of Fig. 8. The offset A_0 is needed to account for the variability that is not temporally correlated. While much of this white noise is undoubtedly instrumental in nature, some may be from the ionosphere itself. This approach of allowing for non-correlated noise differs from the vertical and horizontal analysis. High temporal sampling rates show a transition to non-correlated noise that is not possible with the coarse spatial sampling for the horizontal correlation lengths.

Since the amplitude is described separately (see Sec. 3.4), A is not saved. Since we assume the non-temporally-correlated noise (described by A_0) comes from instrument errors, it can be neglected. However, an OSSE will not produce realistic ionosonde measurements using this truth model if a sophisticated ionosonde error model is not also used. In order to produce realistic OSSEs without a sophisticated ionosonde error model, we save A_0/A to be able to introduce this noise in the truth model.

The decay period is shown for all years, seasons, local times, and altitudes in Fig. 9. As with Figs. 7 and 5, solid lines show the median value for all ionosondes, and shading shows the middle half of the distribution. Columns have the same year while rows have the same season. A low decay period means that there is still significant power at high frequencies (short periods). A high decay period indicates that there is not significant power at short periods. The decay periods for most local times, seasons, and years are highest near 125 km, indicating that the ionosphere is especially temporally smooth there. This trend is particularly evident for the morning sector, which is also the calmest sector for many seasons and years, particularly below ∼250 km. The most short-period variability typically occurs near 225 km, although this is more dramatic in winter than in other seasons.

The one comparable study of ionospheric correlation time (Forsythe, Azeem, Crowley, Makarevich, & Wang, 2020) used TEC data rather than ionosonde-provided electron density. They also defined the correlation time by when the correlation coefficient

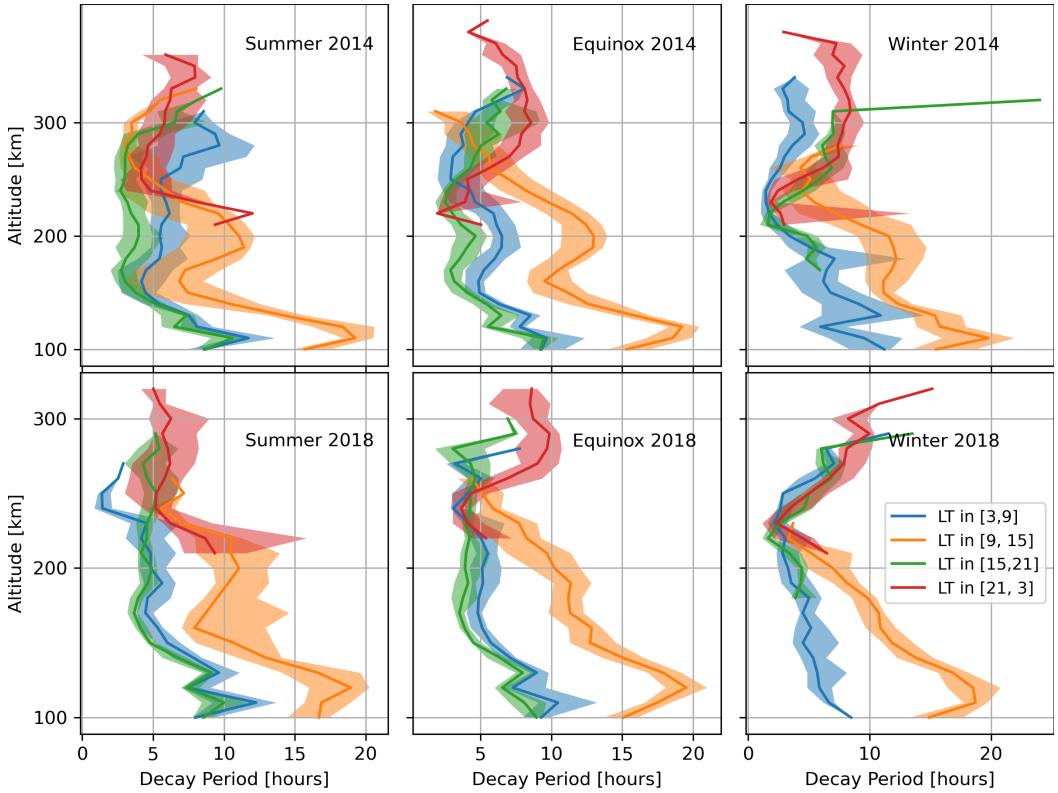


Figure 9. Decay periods for all ionosondes in two years and three seasons. Solid lines show the median value while shading shows the region between the 25th and 75th percentiles or middle half of the data.

dropped below 0.8, whereas we use the time at which is drops below $e^{-1} \approx 0.368$. Therefore, to translate the results from (Forsythe, Azeem, Crowley, Makarevich, & Wang, 2020) to ours, they must be multiplied by $-1/\ln(0.8) \approx 4.48$. For mid-latitudes they find times from 2 to 2.5 hours, which translate to between 9 and 11 hours in our system. These times are higher than our results in the F region, which will contribute the majority of electron density to the TEC.

Figure 10 shows the ratio of A_0, which describes the amplitude of the temporally uncorrelated noise, to A which describes the amplitude of the temporally correlated noise. The format is the same as Figs. 9, 7 and 5. Throughout different seasons and years, this ratio is typically near 0.1 with a peak near 150 km in altitude. Solar maximum (2014) has slightly higher values than solar minimum (2018). There are only small differences in local time. While it would be easy to interpret this plot to say that 10% of the noise is temporally uncorrelated, the situation is more subtle. The total power for the correlated (P_C) and uncorrelated (P_U) noise is given by:

$$P_C = \int_0^{f_M} e^{-f/f_d}\, df = f_d(1 - e^{f_M/f_d}) \approx f_d \tag{6}$$

$$P_U = \int_0^{f_M} A_0/A\, df = f_M \frac{A_0}{A} \tag{7}$$

where f_M is the measurement frequency. For measurement cadences on the timescale of minutes, e^{f_m/f_d} is very small and can be neglected with little impact. At very high measurement cadences, f_M dominates and the majority of the unresolved variance is uncorrelated. For example, if A_0/A is 0.1 and the decay period is 5 hours ($f_d = 0.2$ cy-

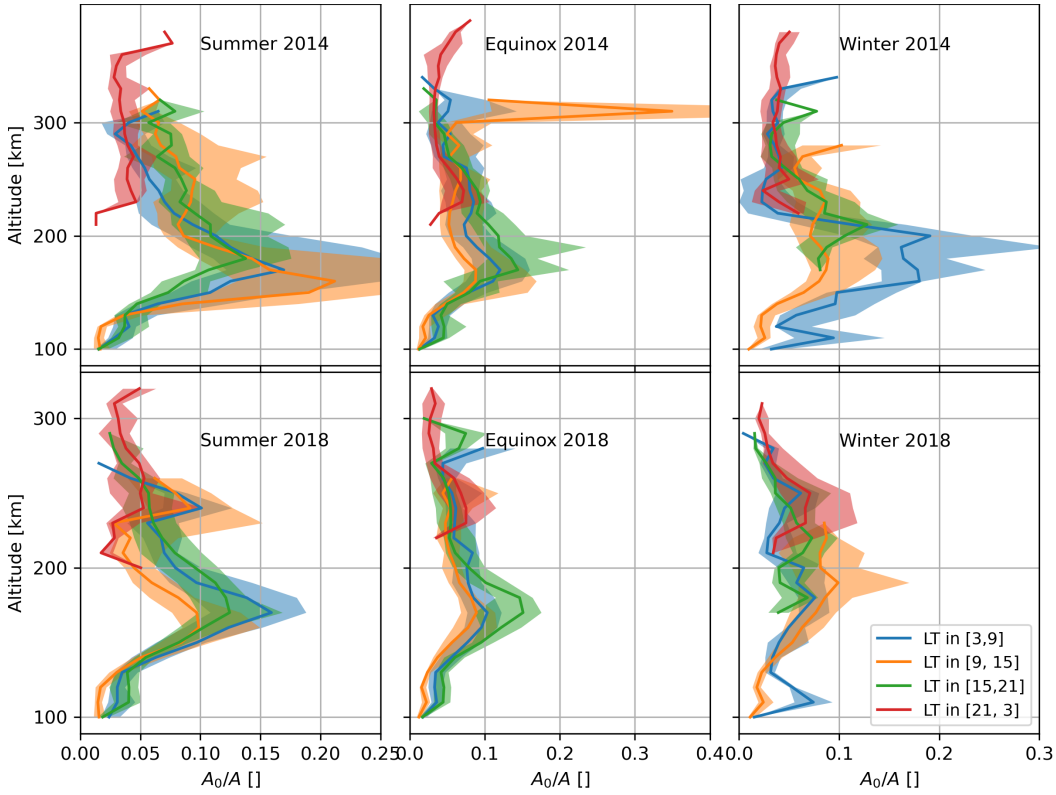


Figure 10. Ratios of uncorrelated to correlated noise for all ionosondes in two years and three seasons. Solid lines show the median value while shading shows the region between the 25th and 75th percentiles or middle half of the data.

cles per hour), then the variance will be equal parts correlated and uncorrelated when $0.1\, f_M = 0.2$. Thus, the majority of the variance will be temporally uncorrelated as long as the measurement cadence is lower than 30 minutes (2 cycles per hour). If the measurement cadence is 15 minutes, then 2/3 of the unresolved variance is temporally uncorrelated and 1/3 is correlated and so forth. We are not aware of any prior work to compare to for this analysis.

3.4 Amplitudes

To estimate the amplitude of the variance of the ionosonde data not resolved by the model, we use the variable y (see eqn. 2) which better captures the high-frequency (shorter than 3-hour period) noise and is less sensitive to the low-frequency differences between the ionosonde and model data. Since the ionosonde data is non-uniform and contains gaps, computing a low-pass of the ionosonde data is a non-trivial task. For each ionosonde, data at each altitude bin with intervals having no gaps longer than 6 hours are identified. The ionosonde electron density in each of these periods is interpolated onto a uniform time grid. This uniformly grided data is then low passed with the filter described in Sec. 2 and then interpolated back onto the original measurement times. The original electron density is divided by the low passed electron density, and the log of this ratio is taken to compute y.

The relatively large gap threshold (six hours) makes these periods long enough that edge effects from the filter are minimal, and since y is only computed at the actual measurement times, linearly interpolating through these gaps does not corrupt the data. Af-

ter y is formed, it is binned in local time and altitude, and its variance is found. The square root of this value, the standard deviation, is shown in Fig. 11. As in Figs. 5, 7, and 9, columns show seasons, rows show years, and colors show local time sectors. Shading shows the middle half of the data while solid lines show the median value.

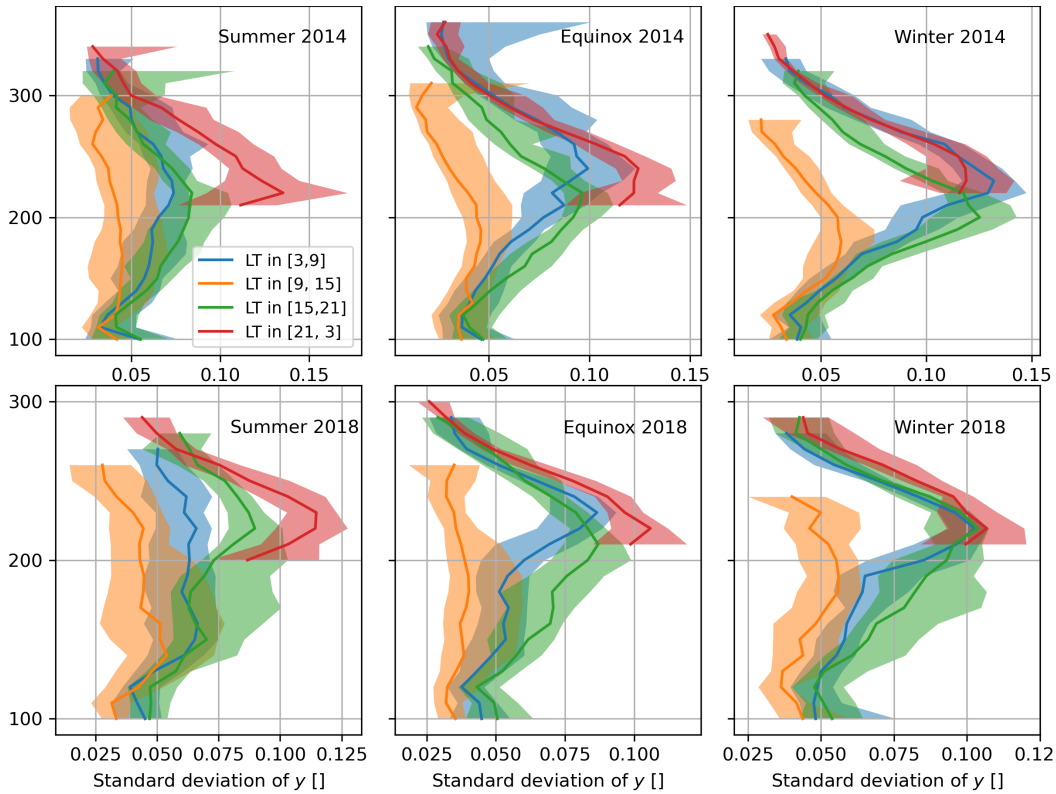


Figure 11. Noise amplitudes for all ionosondes in two years and three seasons. Solid lines show the median value while shading shows the region between the 25th and 75th percentiles or middle half of the data.

The amplitude of variations peaks near 250 km for many local times in all seasons and years. The notable exception is the noon sector which tends to have a less dramatic peak, and that peak tends to be lower (e.g. Winter and Equinox 2014). The nighttime sector tends to have has the highest amplitudes, followed by the morning and evening, and finally followed by the noon sector. The amplitudes max out at $\sim$0.1 but tend to be a little higher during solar maximum in 2014. The amplitude of 0.1 for y means that one should expect a $10^{0.1}-1 \approx 25\%$ increase above or decrease below the smooth background relatively often. Rarer 3σ events will indicate a doubling or halving of the electron density relative to the smooth background. We are unaware of any prior work to compare to for the amplitudes of the variations. This work does not consider resolved variance that is simultaneously temporally and spatially correlated which we would expect from a Traveling Ionospheric Disturbance (TID). This could be an exciting avenue for further study.

This section has studied the vertical (Sec. 3.1), horizontal (Sec. 3.2), and temporal (Sec. 3.3) characteristics of ionospheric variabilities as well as their amplitudes (Sec. 3.4). The variations have larger vertical and horizontal scales at high altitudes, but highest amplitudes near 250 km. The persist for longer times near 150 km, but are fairly short-lived elsewhere. The next section will use these four parameters - temporal, vertical, and

horizontal scales and their amplitudes - to create a variable truth model from a smooth physics-based one.

4 Truth Model Creation

A realistic truth model is created by beginning with a smooth truth model and multiplying by a factor of 10^x. There are four main steps in creating this factor so that it has the correct amplitude as well as spatial and temporal characteristics. First, uncorrelated Gaussian white noise is created with the correct amplitude. Secondly, a Gaussian kernel is created and used to enforce the spatial correlation. Thirdly, this smoothed noise is Fourier transformed and a frequency-dependent gain is applied before inverting it back into the time domain. Finally, this factor is applied to the smooth truth model.

The first step is to compute an array of Gaussian white noise (N_W) of shape $[n_t, n_a, n_{\text{lat}}, n_{\text{lon}}, n_{\text{LT}}]$ for an array with n_t times, n_a altitudes, n_{lat} latitudes, and n_{lon} longitudes. This noise is unscaled (all having a standard deviation of 1, and uncorrelated in time or space).

4.1 Spatial Smoothing

The second step in computing the truth model is to enforce vertical and horizontal correlation onto x. This is done by convolving the uncorrelated noise with a 3D Gaussian kernel. To compute the smoothed noise at a given time, altitude, latitude, longitude, and local time bin, the white noise at the same time and local time bin, but across all vertical and horizontal space are averaged, with points closer to the point in question contributing more to the average. This process is similar to a rolling "boxcar" mean in three dimensions, except that points closer to the center of the boxcar have a stronger contribution to the mean. In an equation, this is given by:

$$N_S[t_i, a_i, la_i, lo_i] = \sum K \times N_W[ti, :, :, :] \tag{8}$$

where N_S is the spatially smoothed noise, t_i, a_i, la_i, lo_i are the indices corresponding to a given time (t_i), altitude (a_i), latitude (la_i), and longitude (lo_i), and K is the spatial smoothing kernel. The kernel has shape $[n_a, n_{\text{lat}}, n_{\text{lon}}]$ and the sum is performed over all dimensions specified with an colon (:) to provide a scalar for N_S. The kernel is given by:

$$K = e^{-\left((\Delta x/l_H)^2 + (\Delta v/l_V)^2\right)/2} \tag{9}$$

where Δx is the horizontal distance between the chosen point (at a_i, la_i, lo_i) and every other point in the grid, Δv is the vertical distance between the chosen point and every other point in the grid, and l_V and l_H are the vertical and horizontal correlation lengths found in sections 3.2 and 3.1.

If the horizontal distance between a point and its neighbors was constant across the globe and the vertical and horizontal correlation lengths the same for all local times and altitudes, this convolution would be quite simple, and could be performed as a multiplication in the frequency domain. However, a grid in latitude and longitude is used, so points near the poles are much closer (when measured in kilometers) to their neighbors than points near the equator. Additionally, the correlation lengths change with altitude and local time, which is a function of both universal time (described by t_i) and longitude. To account for these real-world problems, the smoothing is done in a four-dimensional nested for loop. The innermost two loops are executed inside a function which is accelerated with the numba python library to make use of just-in-time (JIT) compilation and parallel processing.

After the first two loops (over latitude and longitude), the horizontal great circle distance to every other point is computed and has shape $[n_{\text{Lat}}, n_{\text{Lon}}]$. In the third loop (over altitude), the vertical distance from the point to every other point is computed as

$\Delta a = |\text{alt}[ai] - \text{alt}|$ which just has shape $[n_a]$. Once time is specified in the fourth and innermost loop, the local time is computed and the vertical and horizontal correlation lengths are computed via a standard interpolation in altitude, and a more complex interpolation over local time – to avoid sharp discontinuities near local time boundaries, the correlation lengths transition over the period of 1 hour. For example, if the horizontal correlation length is 100 km in the morning sector and 200 km in the noon sector, the correlation length would be 100 km at 8:30, 150 km at 9:00, and 200 km at 9:30. Once the correlation lengths are calculated, the kernel is formed by multiplying the vertical and horizontal kernels with broadcasting:

$$K_H = e^{-\left((\Delta x / l_H)^2\right)/2} \tag{10}$$

$$K_V = e^{-\left((\Delta v / l_V)^2\right)/2} \tag{11}$$

$$K = K_H[\text{None}, :, :] \times K_V[:, \text{None}, \text{None}] \tag{12}$$

In many situations, the model extent is much larger than the vertical and horizontal correlation lengths, so K is mostly zeros. To accelerate computation, K is only computed for three correlation lengths which ensures that the largest values of K being neglected are less than $e^{-9/2} \sim 0.011$. However, the corner points which are three correlation lengths in latitude, longitude, and altitude are measured out to a much further resolution. The kernel is normalized by its sum, and the spatially smoothed noise is computed for that time, altitude, latitude, and longitude.

After the noise is spatially smoothed, its amplitude is adjusted inside a three-dimensional for loop over altitude, time, and longitude. The local times are computed and the amplitudes from section 3.4 are interpolated for the given altitude and local time using the previously described method to avoid harsh gradients at local time boundaries. Even with the dramatic speedup provided by the numba library, this spatial smoothing step is by far the most computationally intensive.

4.2 Temporal Smoothing

Even after the spatial smoothing, the noise is still "white" in that it has equal power at all frequencies. To enforce the correct spectral structure, the smoothed noise is Fourier transformed, multiplied by a frequency-dependent gain, and then inverse transformed back into the time domain. Since this gain is a function of local time (and therefore universal time), a larger array of size $[n_t, n_a, n_{\text{Lat}}, n_{\text{Lon}}, n_{LT}]$ is created and used to store the results where n_{LT} is the number of local time bins (4 here). For each altitude and local time bin, the correct decay period from Sec. 3.3 is selected and the gain is given by either:

$$g(f) = e^{-f/f_d} \qquad \text{or} \qquad g(f) = e^{-f/f_d} + A_0/A \tag{13}$$

where f is the frequency, f_d is the decay frequency, and A_0/A is the offset needed to account for uncorrelated variance that could be provided either by a combination of the truth model and a detailed sensor model. To keep x at a mean of zero, the gain at zero frequency is set to zero before multiplying and inverting. The gain is also normalized by its mean to not drastically change the amplitude of the unresolved variance. Even with this modification, x still needs to be re-normalized after inverting back into the time domain. Each altitude and local time bin is multiplied by a factor to re-normalize the standard deviation to the desired amplitudes found in section 3.4. After this, the temporally smoothed noise is constrained to lie between the minimum and maximum of the spatially smoothed noise to avoid any large outliers.

4.3 Inversion

There is not full coverage of noise parameters across local time and altitude. For example, there are insufficient low altitude measurements in the night sector to allow us to derive the required parameters of variation. To prevent sudden gradients when transitioning from altitude/local times without data to those with data, x is varied from 0 to its predicted value over 25 km of altitude for each local time sector. After this, longitude and time are used to determine the local time and therefore which copy of the noise to use in the final x array. The final multiplicative factor is then given by 10^x, and is of shape $[n_t, n_a, n_{\text{lat}}, n_{\text{lon}}]$.

The smooth model used as the background is TIE-GCM which is run on uniform pressure levels and then interpolated onto a uniform height grid between 80 and 400 km altitudes. To extrapolate up to Global Positioning System (GPS) altitudes (20,000 km), an exponential decay function is fit to the last few points and extended up to 20,000 km. This exponential decay model under-predicts plasmaspheric electron densities, which contribute a few TECU to the vertical TEC (Total Electron Content). To assess this deficiency, the exponential decay model can be transitioned to a more accurate model of the plasmasphere described in (Fridman et al., 2006) based on hydrostatic equilibrium from 400 to 700 km altitude.

4.4 Data Size Considerations

The size of the resulting truth model can quickly become prohibitively large if sufficiently high resolution is used to avoid aliasing. Since the minimum horizontal correlation length is 82 km (summer 2018, 190 km altitude, morning sector), a truth model should have a latitude and longitude resolution of (6378 km/82 km)/2 $\times$ (180^o/π) $\approx$ 0.35^o to meet the Nyquist criteria at the equator. The smallest vertical correlation length is 3.7 km (summer, 2018, 100 km altitude, morning sector) so a vertical resolution of half this should be used to ensure that all variability is well modeled. To estimate the temporal sampling frequency, we require that we model 99% of the power so that $e^{-f_s/f_{dm}} = 0.01$ where f_N is the sampling frequency and f_{dm} is the largest decay frequency (19.2 cycles per day, summer, 2018, alt = 240 km, morning sector). This requirement translates to a sampling frequency of $f_s = -19.2$ cycles per day $\times \ln(0.01) = 88.1$ cycles per day = 16.3 minute period. Picking round numbers near these theoretical limits gives the size of a global, 24 hour truth model as:

$$N = \frac{300 \text{ km}}{2 \text{ km}} \times \frac{24 \text{ hours}}{15 \text{ min}} \times \frac{360^o}{0.25^o} \times \frac{180^o}{0.25^o} \approx 15 \text{ Billion Numbers} \approx 60 \text{ GB} \qquad (14)$$

A value of 4 Bytes per float has been assumed for this calculation. A range of 300 km has been used since the truth parameters span from 100 to 400 km in altitude. Remembering that the Fourier transformed data requires an array four times as large (to account for local time bins) gives a value near 240 GB. Clearly, this is above the memory (and possibly storage) capabilities in even high performance modern computers. This is why the presented code favors explicit for loops over vectorized operations, so less data needs to be held in storage at once. There are two methods to reduce the computational burden. First, a small region can be simulated rather than the entire globe. Second, these sampling requirements can be relaxed. The minimum horizontal and vertical correlation lengths from Figs. 7 and 5 have been used here, but apply to only a few of the altitudes and local times.

5 Truth Model Validation

This section visualizes the truth model and compares it to both the smooth truth model and measured data. The results presented use a vertical resolution of 2.5 km, a temporal resolution of 15 minutes, and an angular resolution of 2 degrees inside a small region over North America. Since the truth model is four-dimensional, it is difficult to visualize all at once. First, we look at just one time and two altitudes. In Fig. 12, the electron density is shown as a function of latitude and longitude at 18 UT for both the new noisy truth model (top row) and the original smooth model (bottom row). The left column shows the electron density at 150 km while the right column shows the electron density at 220 km. The color scale is the same for all four plots and is shown in the colorbar at the top.

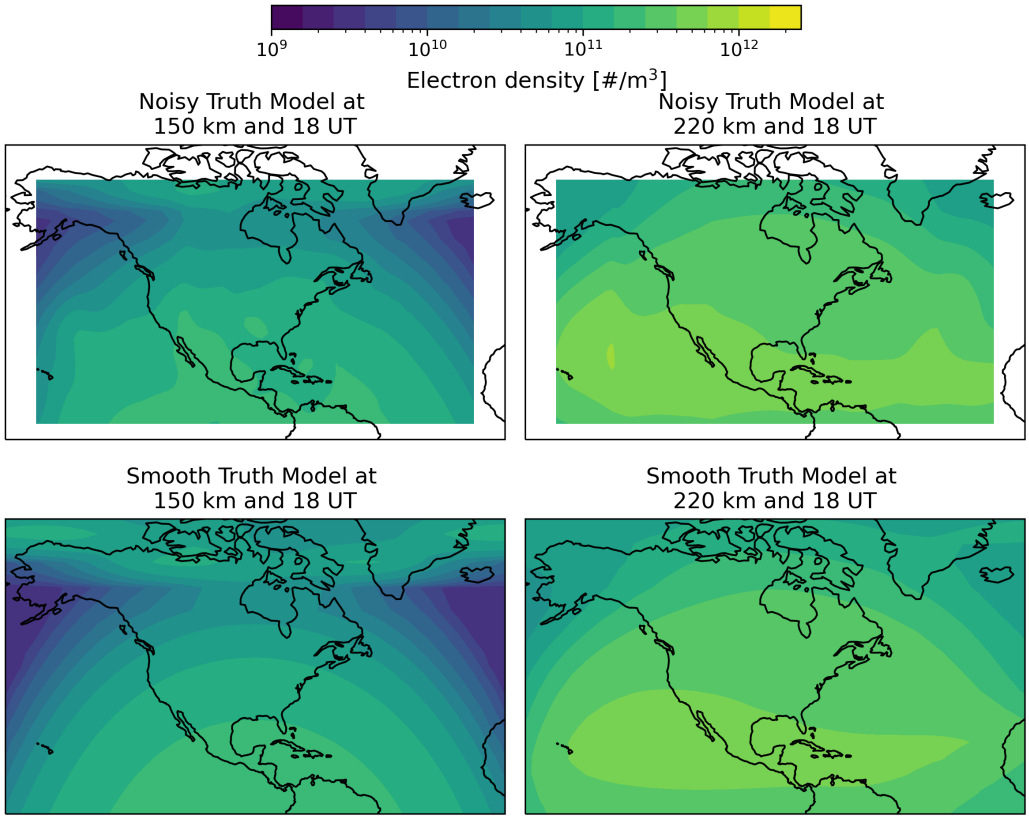


Figure 12. Electron density at 150 and 220 km for both the new and original truth model.

At 150 km, the smooth model is explained almost completely by the solar zenith angle and the aurora. In contrast, the noisy truth model has much more complex structure with many small regions of increased or decreased electron density. The size of these regions is governed by the horizontal correlation length for this altitude and local time.

At 220 km, the equatorial ionization anomaly is clearly visible in the smooth model, but harder to see in the noisy model. The regions of increased and decreased electron density are both more dramatic (larger changes relative to the background), as well as physically larger (larger extent in latitude and longitude) compared to 150 km. This expected, since the horizontal correlation length (see Fig. 7) and amplitudes (see Fig. 11) are larger at 220 km.

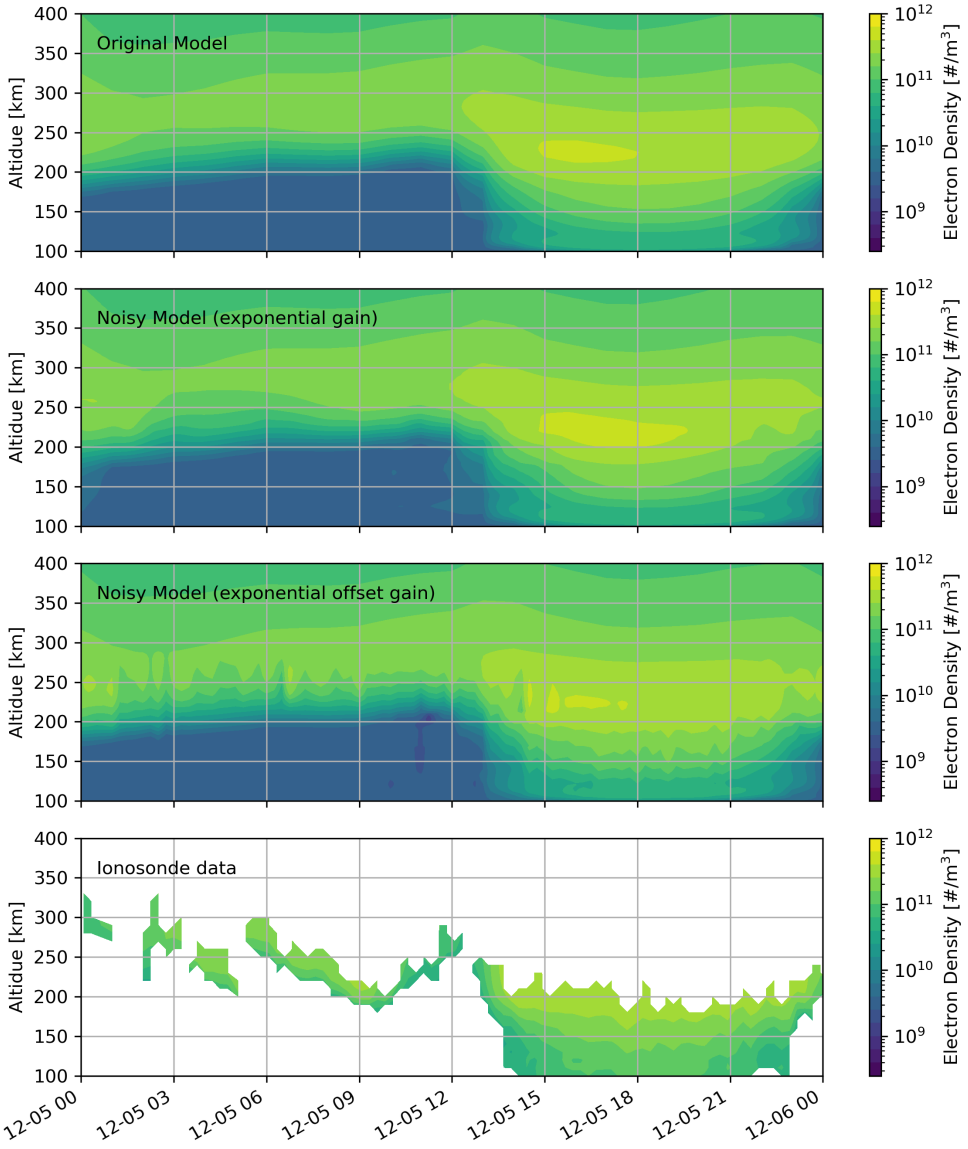


Figure 13. Electron density as a function of time and altitude for different models.

Next we visualize the truth model at just one latitude and longitude, but at all altitudes and times in Fig. 13. As with Figs. 2 and 3, time is on the x axis, altitude is on the y axis, and electron density is shown as color. There are four panels, each showing the ionosphere above the Austin, TX ionosonde. The top panel shows the original smooth truth model and the second panel shows a noisy truth model with a pure exponential temporal gain ($g = e^{-f/f_d}$). The third panel shows a noisy truth model with a temporal gain that includes a white noise offset ($g = e^{-f/f_d} + A_0/A$) to better capture the temporally uncorrelated noise in the absence of a sophisticated ionosonde error model. The last panel shows actual measurements from the Austin ionosonde. Although the ionosonde makes a measurements every 5 minutes, it is decimated to a cadence of 15 minutes to match the cadence of the truth models. As with earlier processing, only the parts of each ionosonde profile that are in the measurement range and below the F2 peak are used.

There are many small variations in the ionosonde data not present in smooth truth model. This should not be surprising given the spectral analysis shown in Fig. 1. The second panel shows a truth model which used a gain with no white noise offset. While this model introduces some temporal variations, they are all quite slow. For example, note the different shape of the F region between 15 and 18 UT, and a slightly faster oscillation between 21 and 24 UT and near 250 km in altitude. Since no ionosonde error model is used to make this figure, we do not expect this second panel to match the data (fourth panel) well. However, the third panel shows a truth model made with the white noise offset which matches the ionosonde data much better in the absence of a sophisticated ionosonde error model with which to view the truth model.

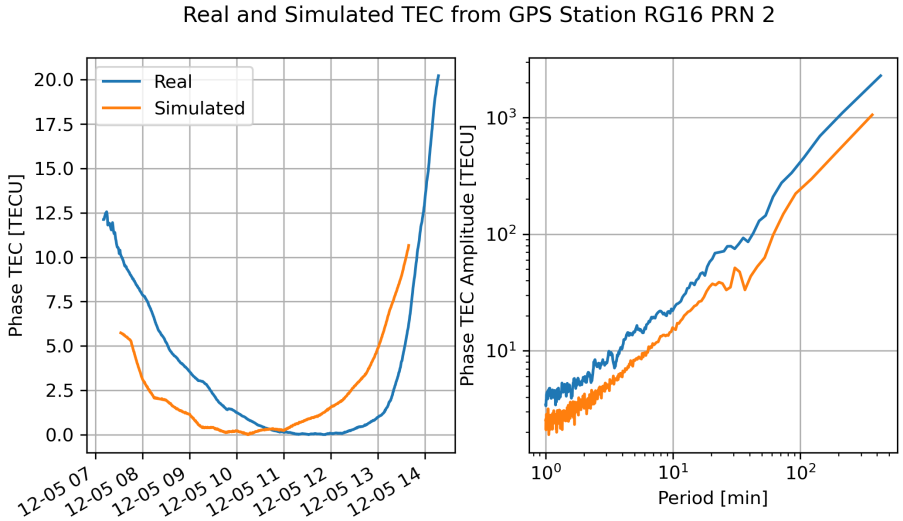


Figure 14. Time series (left) and spectrum of TEC from a real and simulated GPS ground station.

As a final comparison, consider Fig. 14 which shows a time series and spectrum of TEC from a real GPS ground station and simulated TEC using the new noisy truth model. This comparison is especially valuable since TEC data was not used in the formation of the truth model and is therefore an independent measurement. The chosen GPS station (RG16) is near Kremling, Colorado, and the data was for December 5th, 2018 to match the truth model. The data have a 30 second cadence in both cases, although the simulated data uses interpolation to reach this cadence. In both plots, TEC derived from the phase advance rather than code delay is used. Since this measurement type is ambiguous, we level both measurements so that the minimum TEC is zero. There are differences in the overall shape of the TEC profile, which come from the background model differing from reality and do not concern us. Rather, we are concerned with the size and frequency of the small variations. These can be quantified using the spectra in the right subplot which shows the amplitude of oscillations as a function of the period. While there are higher amplitudes in the real data than the simulated data at nearly all frequencies, the slope of both curves are incredibly close. Contrasting this to the spectra and time series shown in Fig. 1 shows significant improvement.

6 Conclusions

A truth model with realistic variances is needed to properly assess the ability of a sensor system and an assimilation system to specify the ionosphere. If a smooth truth model is used, one could expect better specifications than would be obtained in the real

world. While this use case motivates the work here, the resulting study of ionospheric variability is scientifically interesting in its own right. Section 2 describes the data processing techniques used to study the spatial and temporal structure of the ionosphere. This involves significant "data wrangling" since the data are not uniformly spaced and have frequent gaps. Section 3 uses ionosonde data to find the amplitude of small ionospheric density variations, and their temporal and spatial size. Section 4 discusses the algorithms used to compute the truth model. This is difficult primarily because of the large size of the arrays needed to adequately capture the small spatial and temporal scales. Finally, Section 5 compared the truth model to real data and found it much closer than the original smooth model.

This work represents the first published method to introduce empirically driven variations to a smooth model for more realistic OSSEs. These variations have been studied as a function of altitude, local time, season, and year (as a proxy for solar cycle). There are two primary avenues for future work on this topic. First, the variations could be studied during periods of active space weather and at different high latitudes. The work here only considers geomagnetically quiet ($K_P < 4$) conditions and mid magnetic latitude (between 20 and 55°). We expect the parameters to change substantially at equatorial and polar latitudes as well as during storms.

Secondly, additional data sources could resolve many of the limitations in this model. Since ionosondes do not measure above the F2 peak or in the valley region, it is impossible to study those regions or compute representative noise statistics. Using Incoherent Scatter Radar (ISR) data or GPS TEC data either from ground stations or Radio Occultation (RO) data could help illuminate some of these regions. Additionally, there are not many very close pairs of ionosondes, so it is difficult to study variations with small (< 100 km) horizontal sizes.

7 Acknowledgments

This work was funded partially under the MASE program (contract number FA807520F0021). The authors gratefully acknowledge NOAA for providing the ionosonde data `ftp://ftp.ngdc.noaa.gov/inosonde/data`. The GPS data was retrieved from `https://doi.org/10.7283/T5WQ021D`. The background is provided by TIE-GCM e.g. (Qian et al., 2014). We thank Dr. Wanli Wu for his assistance running these models and creating the truth models.

References

Forsythe, V. V., Azeem, I., Blay, R., Crowley, G., Gasperini, F., Hughes, J., ... Wu, W. (2021). Evaluation of the new background covariance model for the ionospheric data assimilation. *Radio Science*, *56*(8), e2021RS007286. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021RS007286` (e2021RS007286 2021RS007286) doi: https://doi.org/10.1029/2021RS007286

Forsythe, V. V., Azeem, I., & Crowley, G. (2020). Ionospheric horizontal correlation distances: Estimation, analysis, and implications for ionospheric data assimilation. *Radio Science*, *55*(12), e2020RS007159. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020RS007159` (e2020RS007159 10.1029/2020RS007159) doi: https://doi.org/10.1029/2020RS007159

Forsythe, V. V., Azeem, I., Crowley, G., Makarevich, R. A., & Wang, C. (2020). The global analysis of the ionospheric correlation time and its implications for ionospheric data assimilation. *Radio Science*, *55*(12), e2020RS007181. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020RS007181` (e2020RS007181 10.1029/2020RS007181) doi:

https://doi.org/10.1029/2020RS007181

Forsythe, V. V., Azeem, I., Crowley, G., & Themens, D. R. (2021). Ionospheric vertical correlation distances: Estimation from isr data, analysis, and implications for ionospheric data assimilation. *Radio Science*, *56*(2), e2020RS007177. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020RS007177` (e2020RS007177 10.1029/2020RS007177) doi: https://doi.org/10.1029/2020RS007177

Fridman, S. V., Nickisch, L. J., Aiello, M., & Hausman, M. (2006). Real-time reconstruction of the three-dimensional ionosphere using data from a network of gps receivers. *Radio Science*, *41*(5). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005RS003341` doi: https://doi.org/10.1029/2005RS003341

Gail, W. B., Prag, A. B., Coco, D. S., & Coker, C. (1993). A statistical characterization of local mid-latitude total electron content. *Journal of Geophysical Research*, *98*(A9), 15717–15728. doi: 10.1029/92JA01597

Garner, T. W., Taylor, B. T., Gaussiran II, T. L., Coley, W. R., & Hairston, M. R. (2005). On the distribution of ionospheric electron density observations. *Space Weather*, *3*(10). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005SW000169` doi: https://doi.org/10.1029/2005SW000169

Holmes, J., Egert, A., Dao, E., Colman, J., & Parris, T. (2017). Adding spatially correlated noise to a median ionosphere. In *Agu fall meeting.*

Klobuchar, J., & Johanso, J. M. (1977). Correlation distance of mean daytime electron content. *AFGL-TR-77-0185, Air Force Geophys. Lab., Hanscom Air Force Base, Mass.*.

McNamara, L. F. (2009). Spatial correlations of fof2 deviations and their implications for global ionospheric models: 2. digisondes in the united states, europe, and south africa. *Radio Science*, *44*(02), 1-5. doi: 10.1029/2008RS003956

Qian, L., Burns, A. G., Emery, B. A., Foster, B., Lu, G., Maute, A., ... Wang, W. (2014). The ncar tie-gcm. In *Modeling the ionosphere–thermosphere system* (p. 73-83). American Geophysical Union (AGU). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/9781118704417.ch7` doi: https://doi.org/10.1002/9781118704417.ch7

Richmond, A. D., Ridley, E. C., & Roble, R. G. (1992). A thermosphere/ionosphere general circulation model with coupled electrodynamics. *Geophysical Research Letters*, *6*, 601-604.

Shim, J. S., Scherliess, L., Schunk, R. W., & Thompson, D. C. (2008). Spatial correlations of day-to-day ionospheric total electron content variability obtained from ground-based GPS. *Journal of Geophysical Research*, *113*(A09309). doi: 10.1029/2007JA012635

Vallado, D. A. (2013). *Fundamentals of astrodynamics and applications* (4th ed.). Microcosm Press.

Yue, X., Wan, W., Liu, L., & Mao, T. (2007). Statistical analysis on spatial correlation of ionospheric day-to-day variability by using gps and incoherent scatter radar observations. *Annales Geophysicae*, *25*(8), 1815–1825. Retrieved from `https://angeo.copernicus.org/articles/25/1815/2007/` doi: 10.5194/angeo-25-1815-2007

Zeng, X., Atlas, R., Birk, R. J., Carr, F. H., Carrier, M. J., Cucurull, L., ... Zhang, F. (2020). Use of observing system simulation experiments in the united states. *Bulletin of the American Meteorological Society*, *101*(8), E1427 - E1438. Retrieved from `https://journals.ametsoc.org/view/journals/bams/101/8/bamsD190155.xml` doi: 10.1175/BAMS-D-19-0155.1

Figure 1.

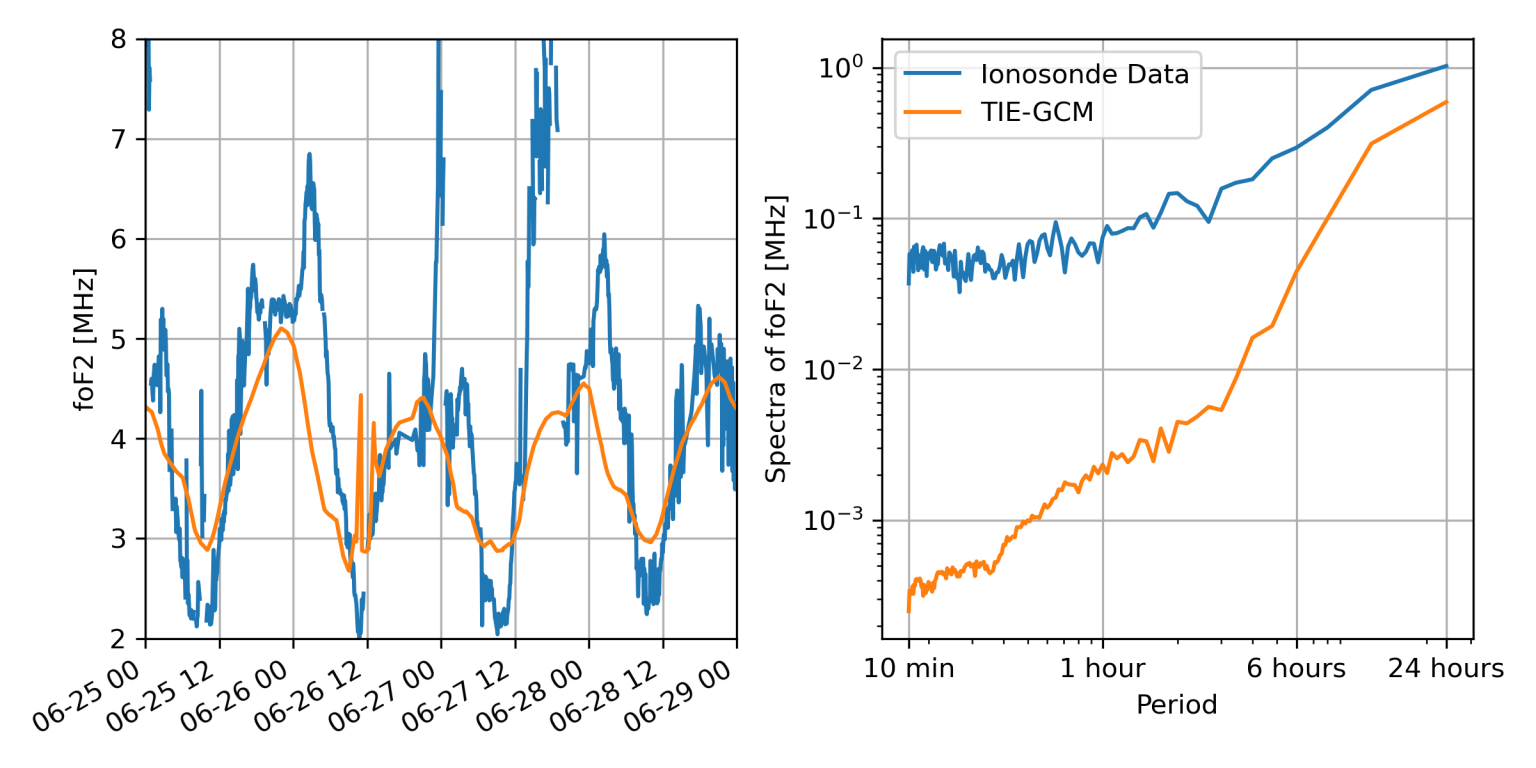

foF2 [MHz]
8
7
6
5
4
3
2
06-25 00
06-25 12
06-26 00
06-26 12
06-27 00
06-27 12
06-28 00
06-28 12
06-29 00
Spectra of foF2 [MHz]
Ionosonde Data
TIE-GCM
10^{0}
10^{-1}
10^{-2}
10^{-3}
10 min
1 hour
6 hours
24 hours
Period

Figure 2.

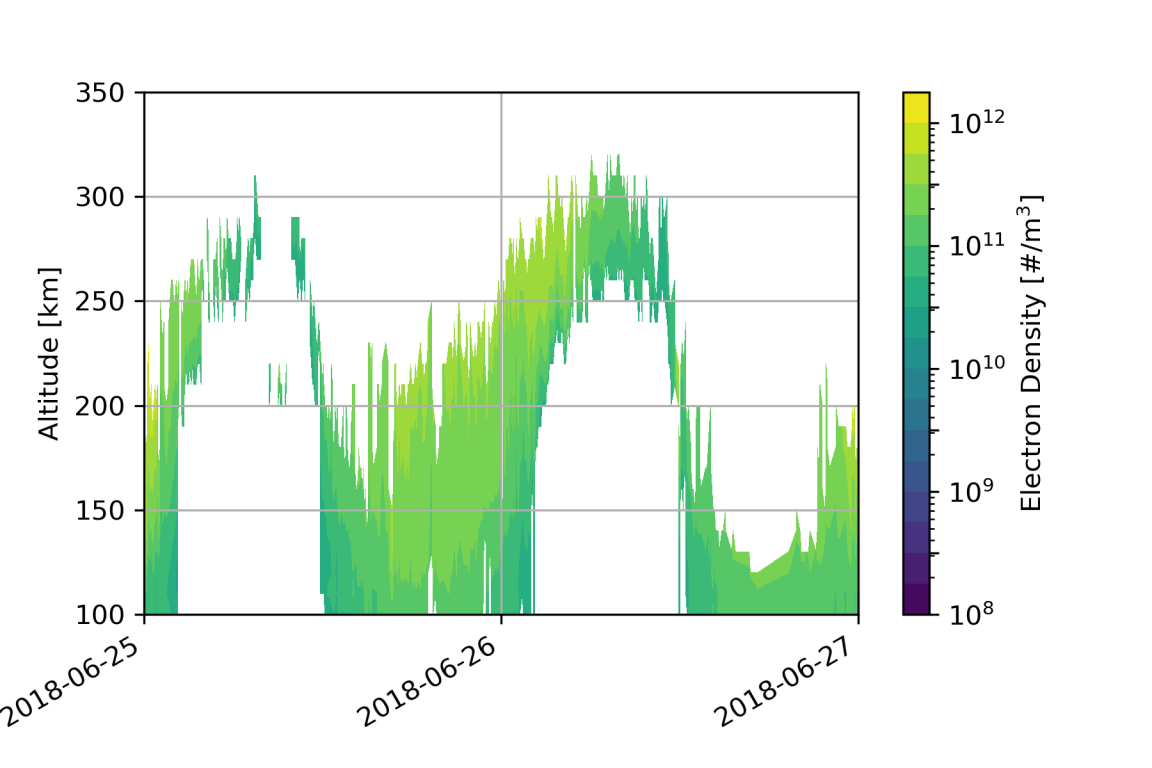

350
300
250
200
150
100
Altitude [km]
2018-06-25
2018-06-26
2018-06-27
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
Electron Density [#/m^3]

Figure 3.

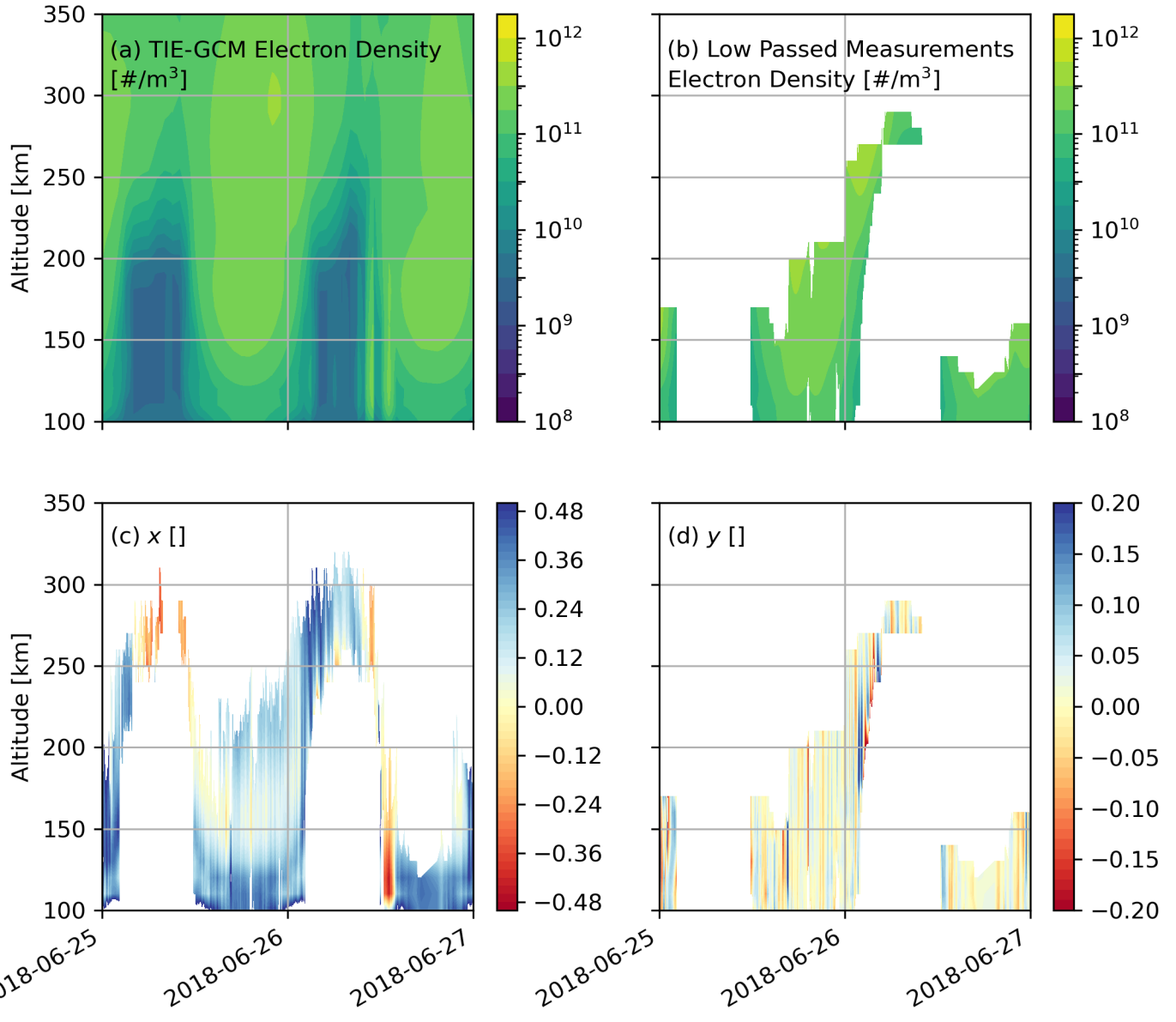

(a) TIE-GCM Electron Density [#/m^3]
(b) Low Passed Measurements Electron Density [#/m^3]
(c) x []
(d) y []
Altitude [km]
350
300
250
200
150
100
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
0.48
0.36
0.24
0.12
0.00
−0.12
−0.24
−0.36
−0.48
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
2018-06-25
2018-06-26
2018-06-27

Figure 4.

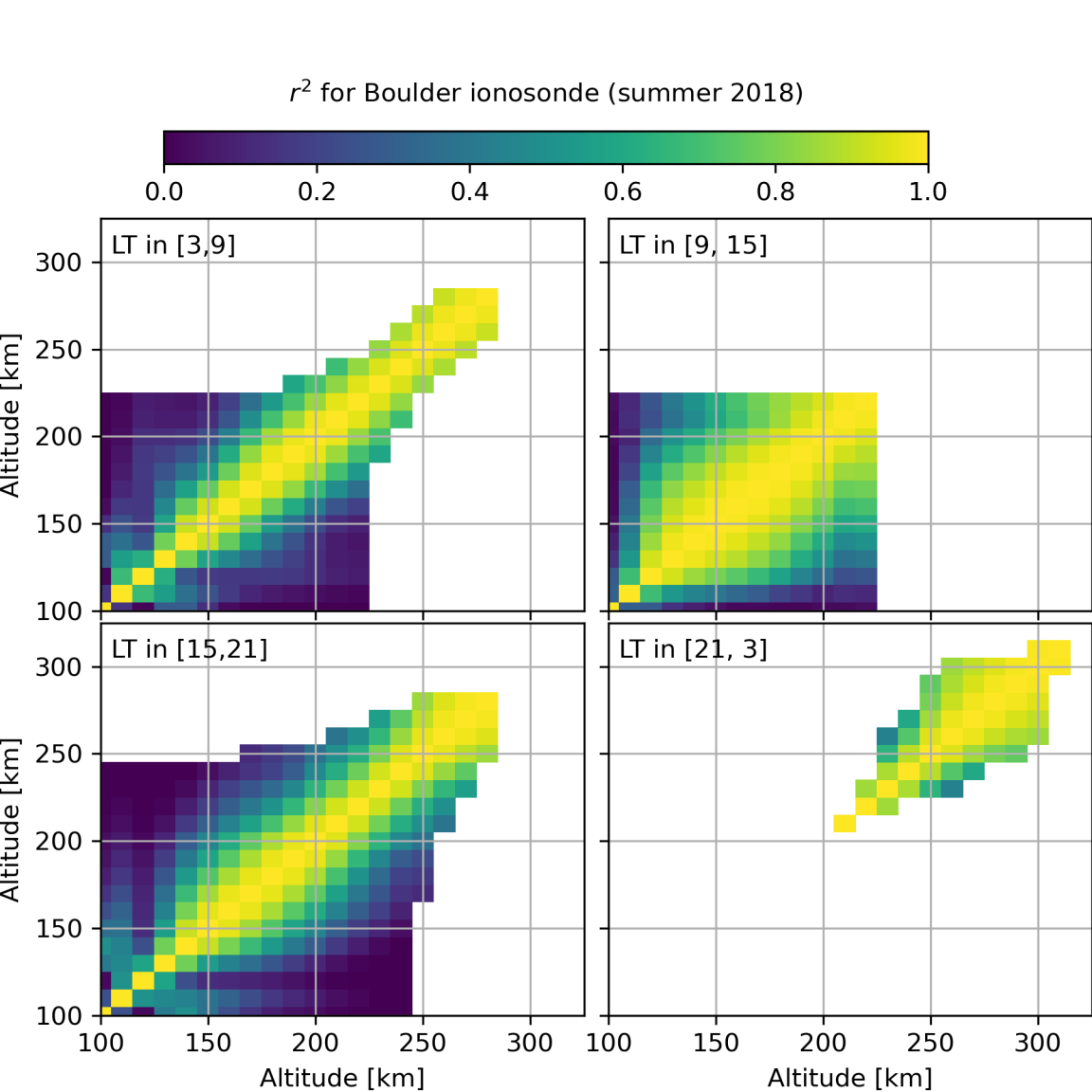

r^2 for Boulder ionosonde (summer 2018)
0.0
0.2
0.4
0.6
0.8
1.0
LT in [3,9]
LT in [9, 15]
LT in [15,21]
LT in [21, 3]
Altitude [km]
Altitude [km]
100
150
200
250
300
100
150
200
250
300
Altitude [km]
Altitude [km]

Figure 5.

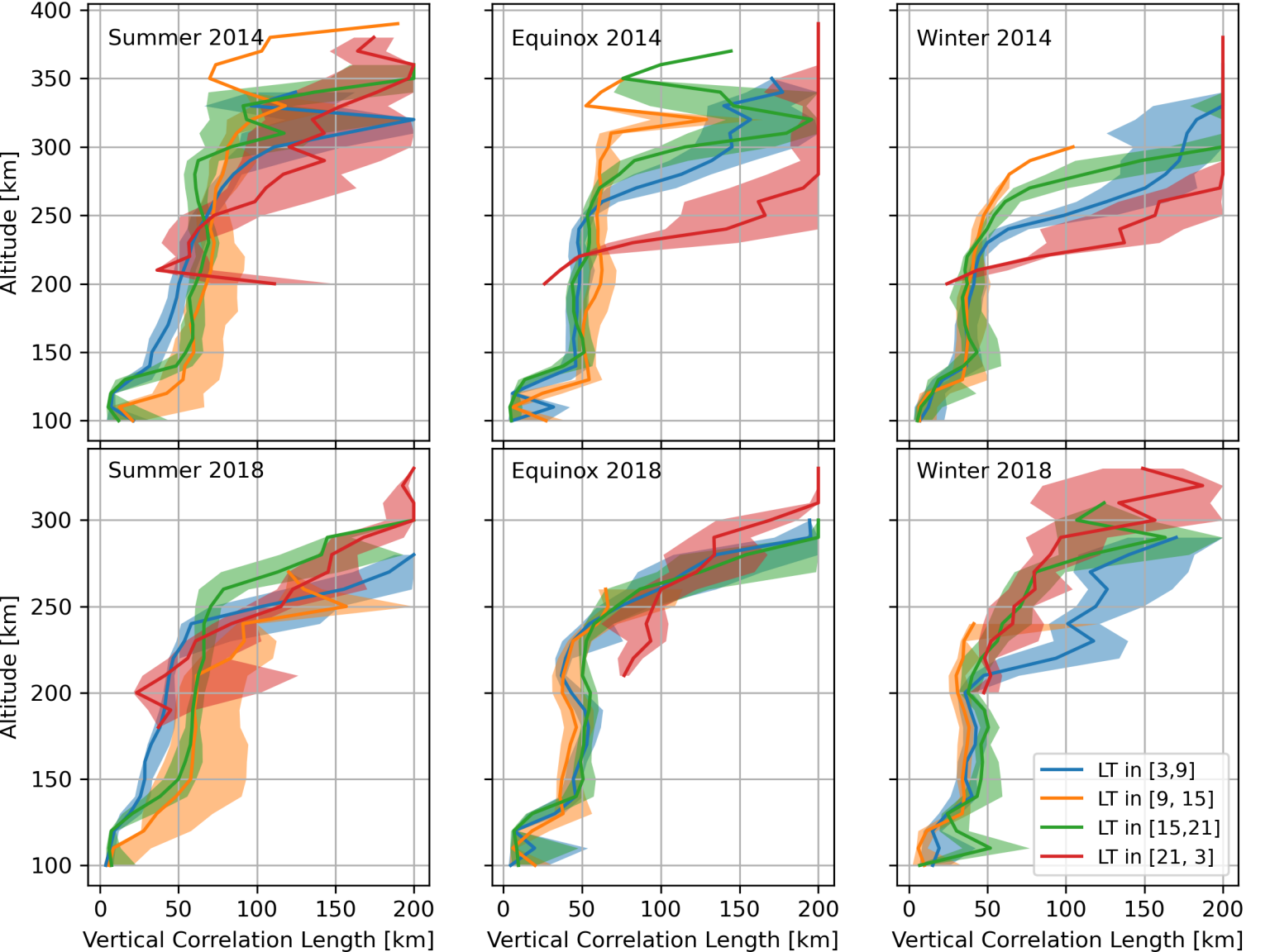

Summer 2014
Equinox 2014
Winter 2014
Summer 2018
Equinox 2018
Winter 2018
Altitude [km]
Vertical Correlation Length [km]
LT in [3,9]
LT in [9, 15]
LT in [15,21]
LT in [21, 3]

Figure 6.

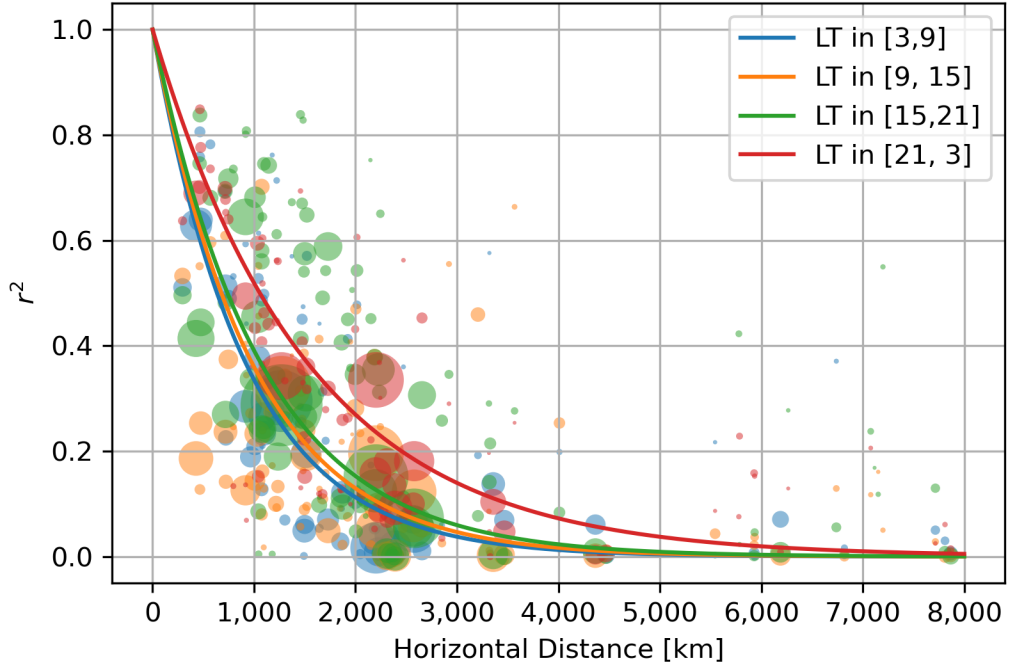

LT in [3,9]
LT in [9, 15]
LT in [15,21]
LT in [21, 3]
1.0
0.8
0.6
0.4
0.2
0.0
r^2
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
Horizontal Distance [km]

Figure 7.

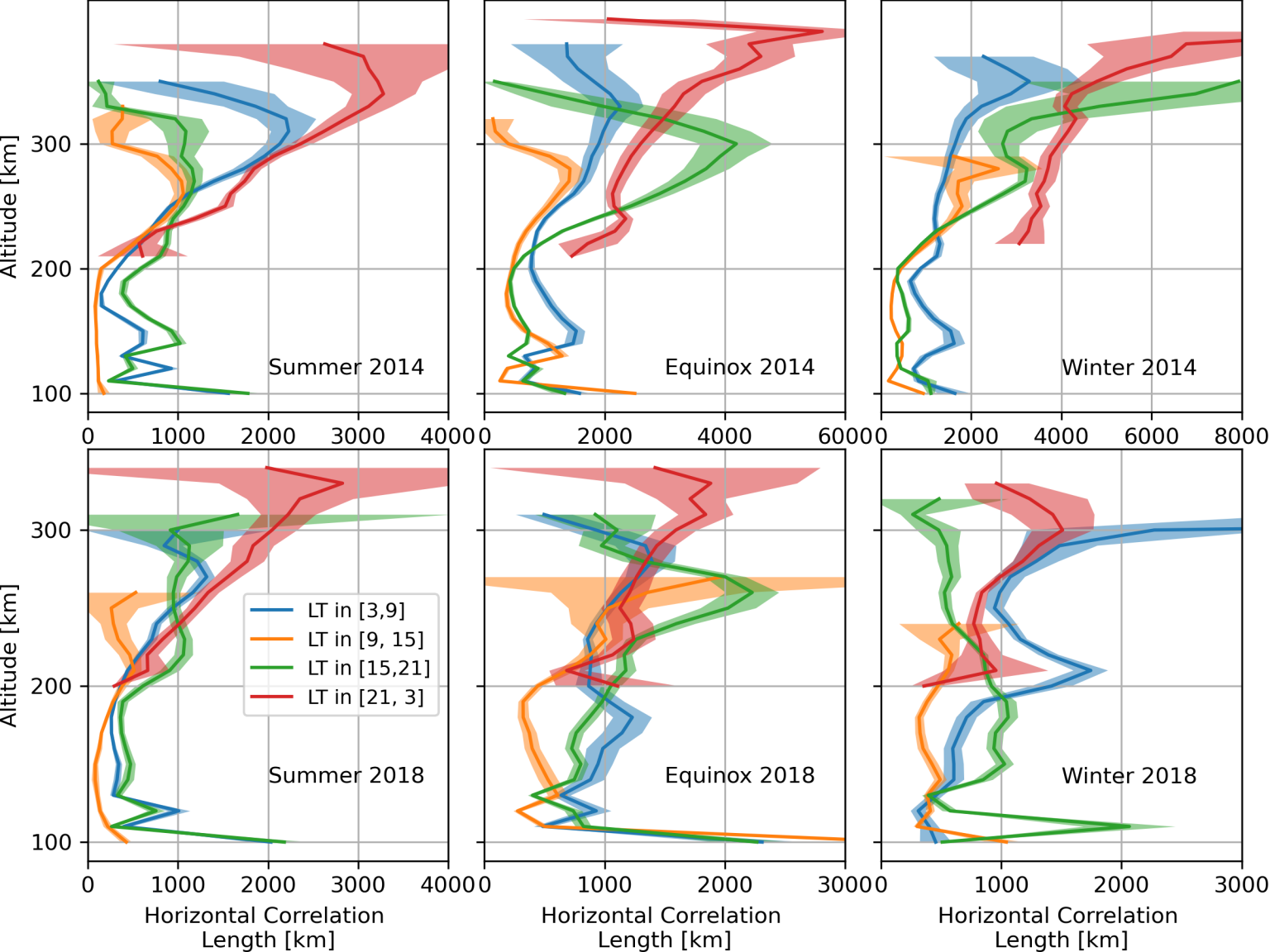

Summer 2014
Equinox 2014
Winter 2014
Summer 2018
Equinox 2018
Winter 2018
LT in [3,9]
LT in [9, 15]
LT in [15,21]
LT in [21, 3]
Altitude [km]
Horizontal Correlation Length [km]

Figure 8.

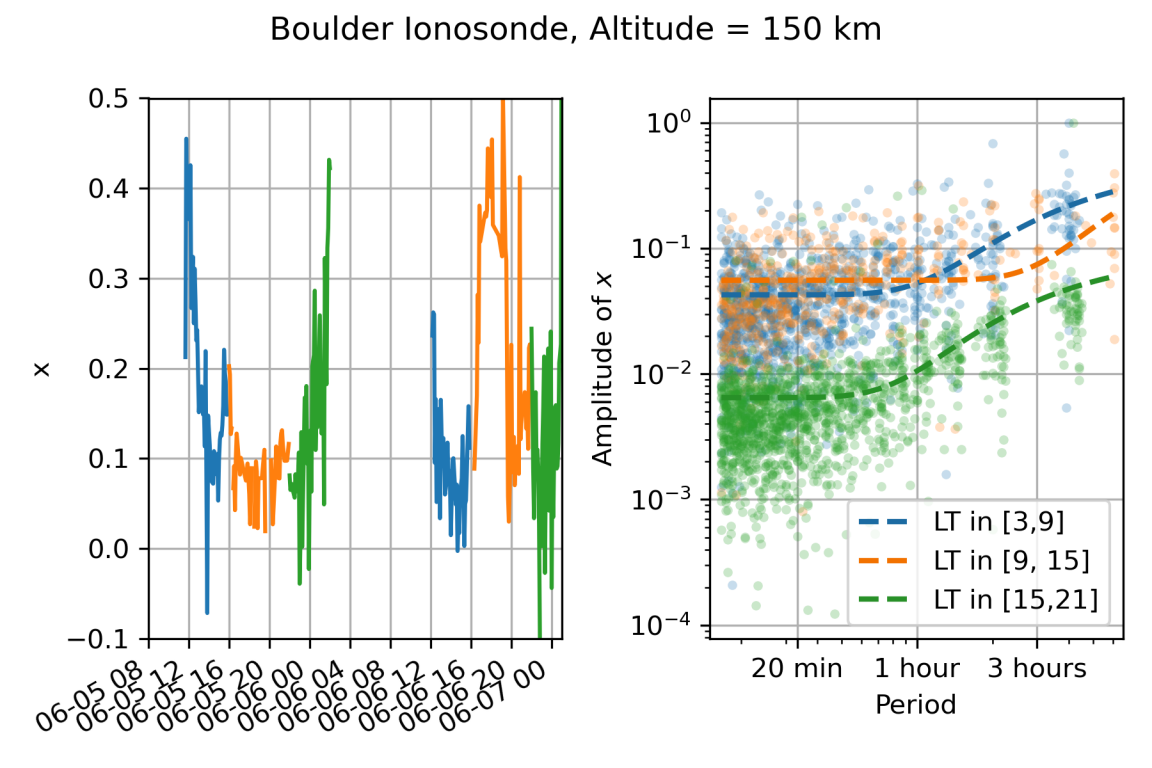

Boulder Ionosonde, Altitude = 150 km
x
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
06-05 08
06-05 12
06-05 16
06-05 20
06-06 00
06-06 04
06-06 08
06-06 12
06-06 16
06-06 20
06-07 00
Amplitude of x
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
LT in [3,9]
LT in [9, 15]
LT in [15,21]
20 min
1 hour
3 hours
Period

Figure 9.

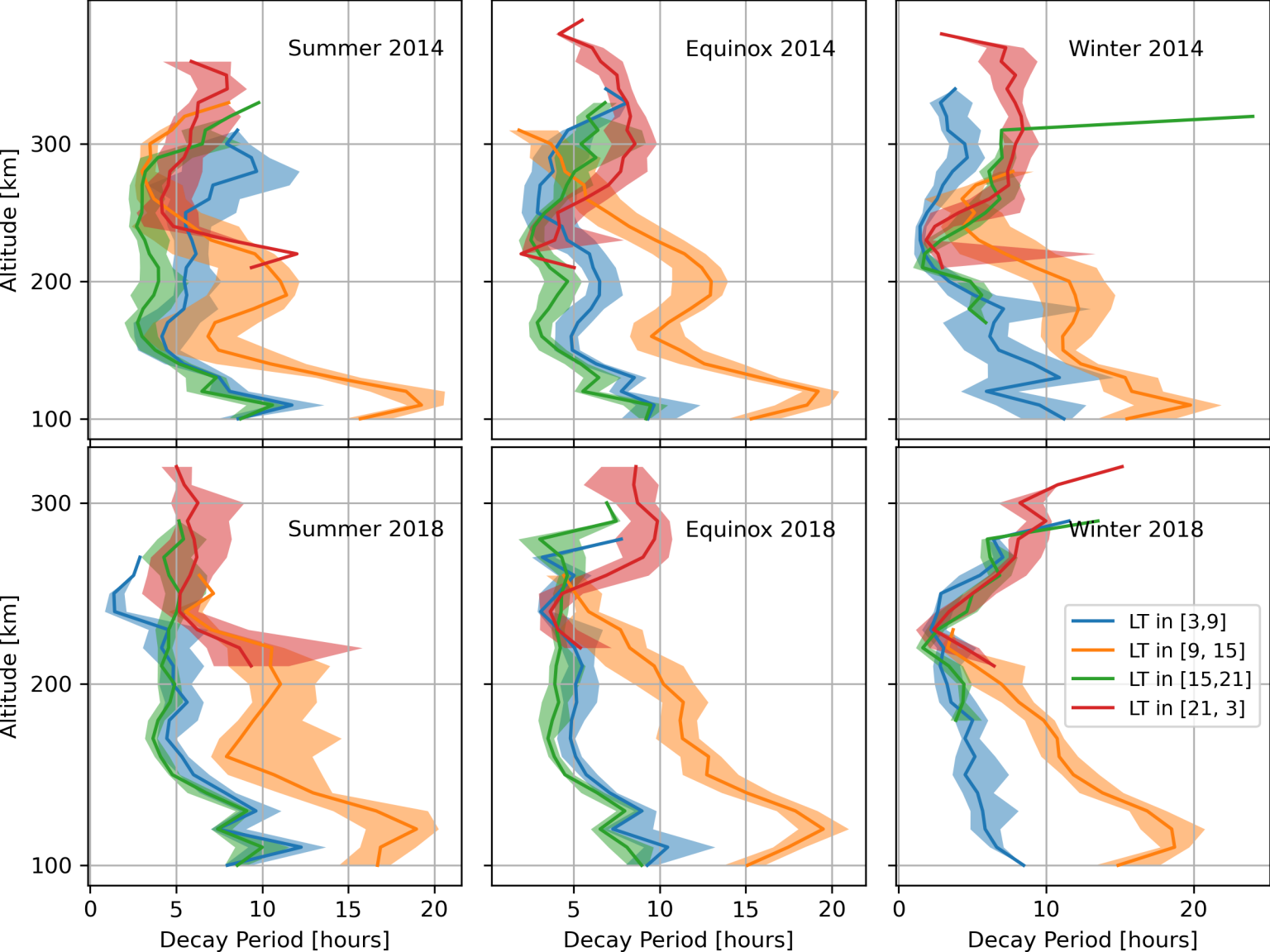

Summer 2014
Equinox 2014
Winter 2014
Summer 2018
Equinox 2018
Winter 2018
Altitude [km]
100
200
300
0
5
10
15
20
Decay Period [hours]
LT in [3,9]
LT in [9, 15]
LT in [15,21]
LT in [21, 3]

Figure 10.

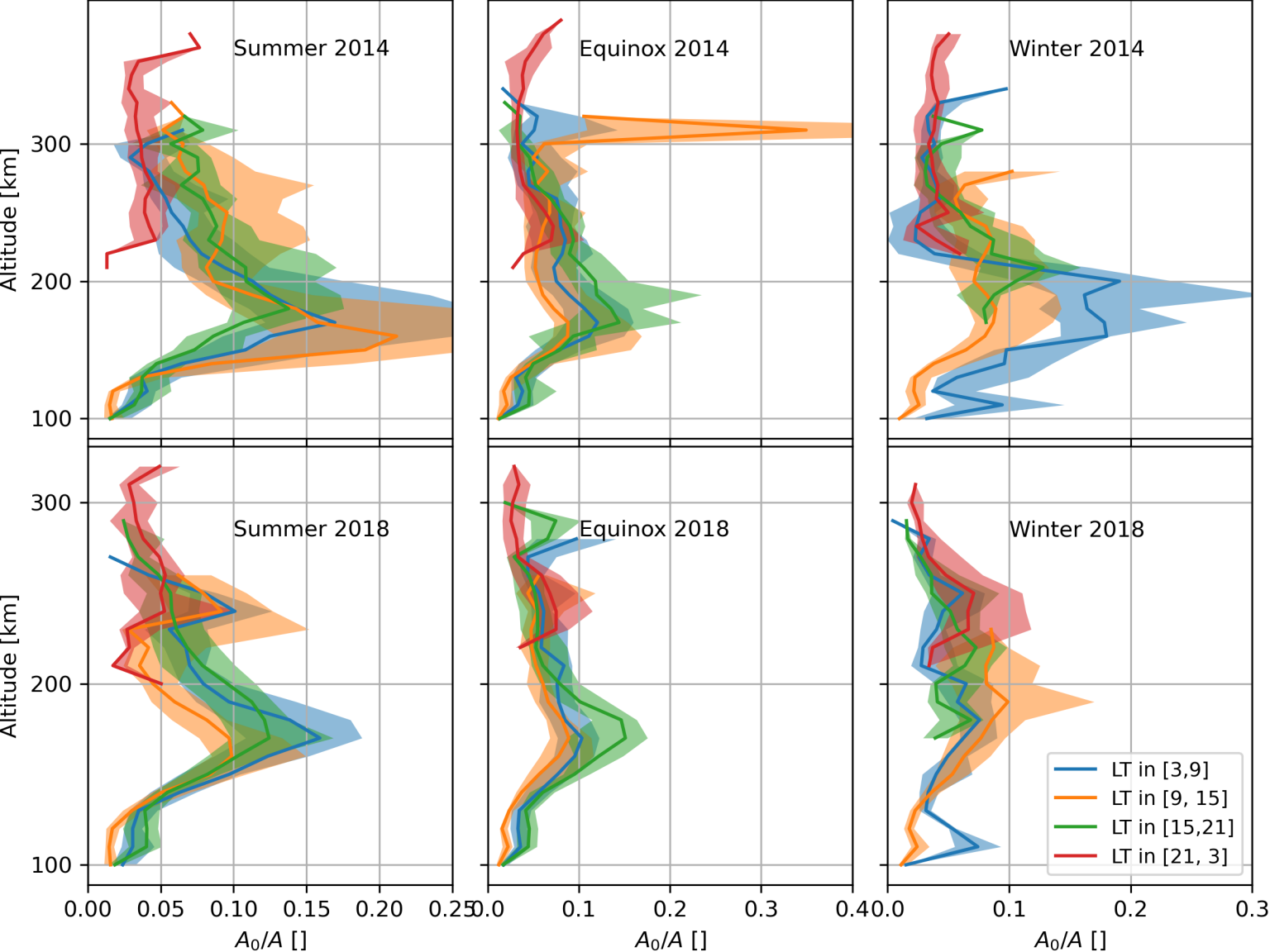

Summer 2014
Equinox 2014
Winter 2014
Summer 2018
Equinox 2018
Winter 2018
Altitude [km]
100
200
300
0.00
0.05
0.10
0.15
0.20
0.25
0.0
0.1
0.2
0.3
0.4
A_0/A []
LT in [3,9]
LT in [9, 15]
LT in [15,21]
LT in [21, 3]

Figure 11.

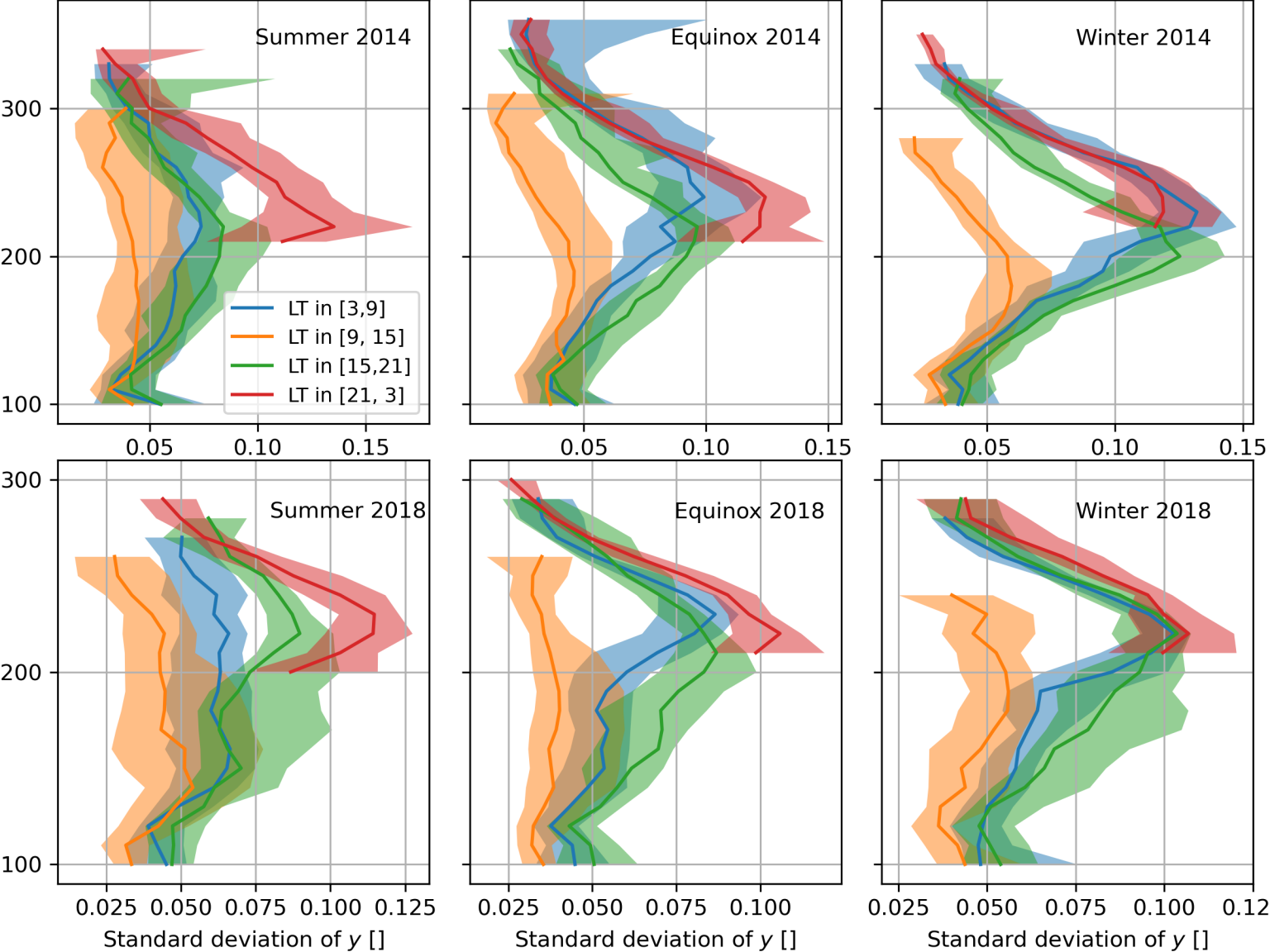

Summer 2014
Equinox 2014
Winter 2014
Summer 2018
Equinox 2018
Winter 2018
LT in [3,9]
LT in [9, 15]
LT in [15,21]
LT in [21, 3]
300
200
100
0.05
0.10
0.15
0.025
0.050
0.075
0.100
0.125
Standard deviation of y []

Figure 12.

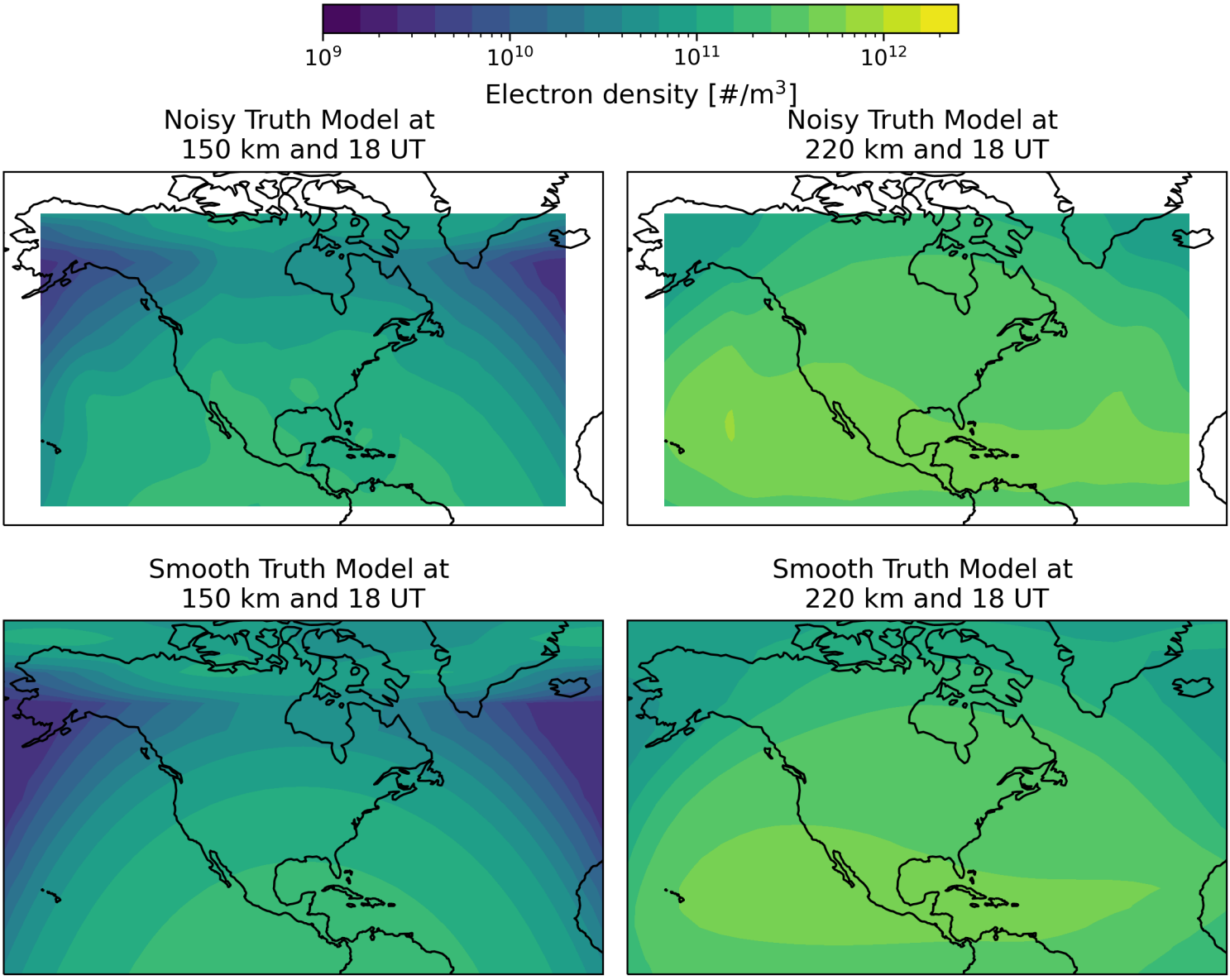

10^9
10^{10}
10^{11}
10^{12}
Electron density [#/m^3]
Noisy Truth Model at
150 km and 18 UT
Noisy Truth Model at
220 km and 18 UT
Smooth Truth Model at
150 km and 18 UT
Smooth Truth Model at
220 km and 18 UT

Figure 13.

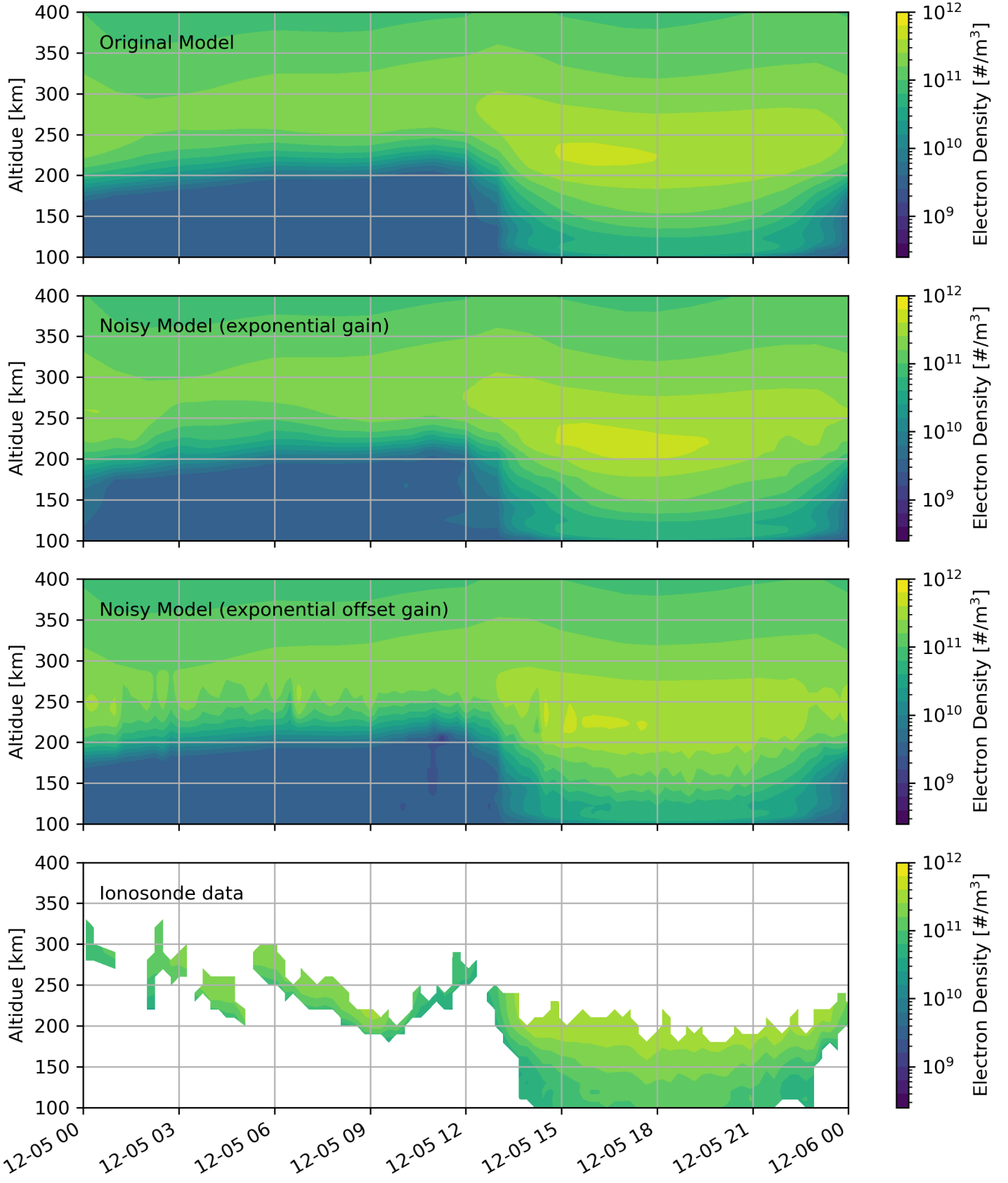
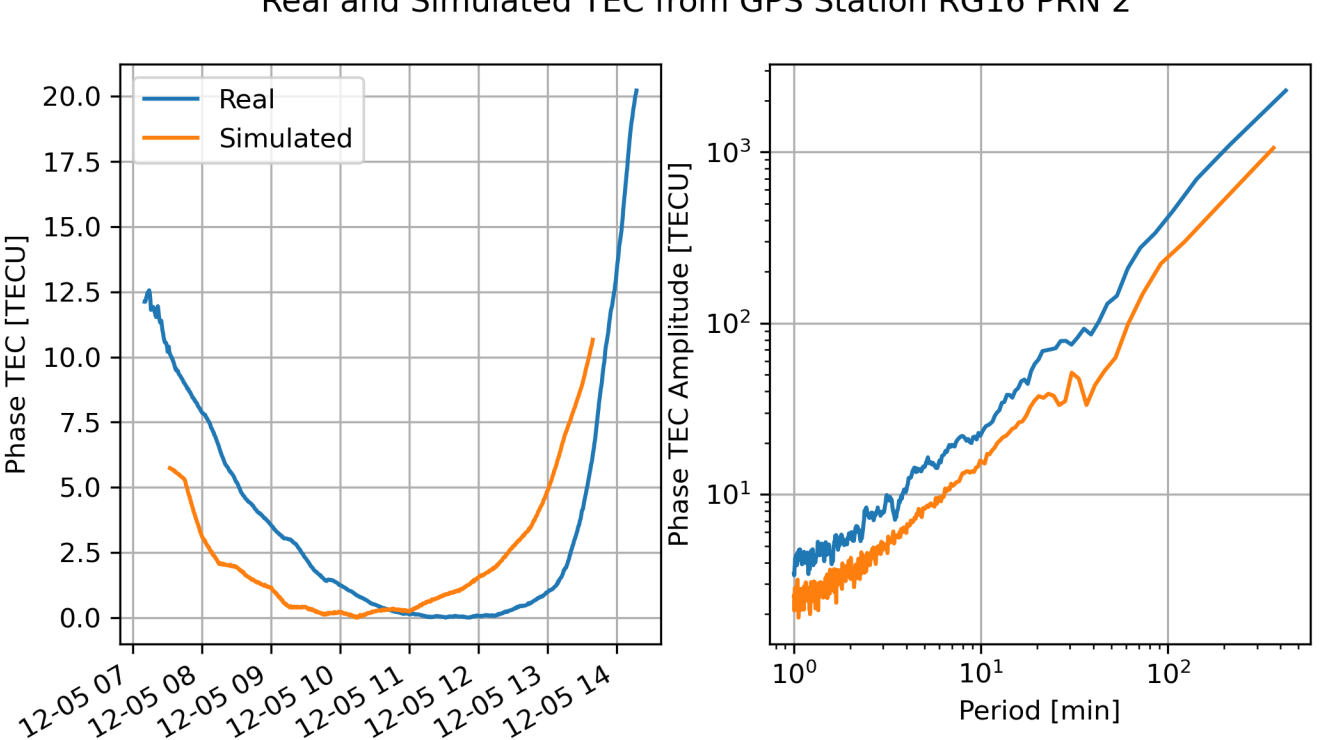

Original Model
Noisy Model (exponential gain)
Noisy Model (exponential offset gain)
Ionosonde data
Altidue [km]
400
350
300
250
200
150
100
Electron Density [#/m^3]
10^{12}
10^{11}
10^{10}
10^{9}
12-05 00
12-05 03
12-05 06
12-05 09
12-05 12
12-05 15
12-05 18
12-05 21
12-06 00

Figure 14.


Real and Simulated TEC from GPS Station RG16 PRN 2
Real
Simulated
Phase TEC [TECU]
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
12-05 07
12-05 08
12-05 09
12-05 10
12-05 11
12-05 12
12-05 13
12-05 14
Phase TEC Amplitude [TECU]
10^3
10^2
10^1
10^0
10^1
10^2
Period [min]