ANCHOR: Global Parametrized Ionospheric Data Assimilation

Victoriya V. Forsythe[1], Sarah E. McDonald[1], Kenneth F. Dymond[1], Bruce A. Fritz[1], Angeline G. Burrell[1], Katherine A. Zawdie[1], Douglas P. Drob[1], Meghan R. Burleigh[1], Dustin A. Hickey[1], Christopher A. Metzler[1], David D. Kuhl[1], Daniel Hodyss[1], Joseph H. Hughes[2]

[1]U.S. Naval Research Laboratory, Washington, DC, USA
[2]Orion Space Solutions, Louisville, CO, USA

Distribution Statement

Key Points:

- Novel approach for rapid ionospheric data assimilation using anchor points
- Global ionospheric data assimilation for plasma density parameters
- Extraction of the anchor points from the ionospheric data

Corresponding author: Victoriya V. Forsythe, victoriya.makarevich@nrl.navy.mil

Abstract

ANCHOR is a novel assimilative model developed at the U.S. Naval Research Laboratory. It extracts ionospheric parameters from RO and ionosonde data and assimilates them as point measurements into the maps of the background parameters using Kalman Filter approach. This paper introduces the ANCHOR algorithm, discusses its coordinate system and background, explains the background covariance formation, discusses the extraction of the ionospheric parameters from the data and the assimilation process, and, finally, shows the results of the observing system simulation experiment.

1 Introduction

The ionosphere starts ∼80 km from the Earth's surface and extends all of the way to the exosphere, which is thousands of kilometers above the Earth's surface. The free electrons that inhabit the ionosphere refract the electromagnetic waves that traverse through the ionosphere. Since high frequency (HF) communication is made possible by the propagation of electromagnetic waves through the ionosphere, it is crucial to know the amount of electrons along the communication path to establish the HF communication link by choosing the best transmission frequency. Climatological models, such as the International Reference Ionosphere (IRI) (Bilitza et al., 2022; Forsythe et al., 2023), provide good estimates of the electron density for a given time and solar activity level. However, the ionospheric state often deviates from the climatological prediction because it is a very active environment with localized structures and short-lived events. One way around this problem is to assimilate the ionospheric observations into the model background in order to obtain a realistic specification.

There are many ionospheric data assimilation (DA) models that use a Kalman filtering (KF) approach, including Global Assimilation of Ionospheric Measurements (GAIM) (Schunk et al., 2004), the Electron Density Assimilative Model (EDAM) (Angling & Cannon, 2004), the Ionospheric Data Assimilation Four-Dimensional (IDA4D) model (Bust & Crowley, 2007), and a Three-Dimensional Regional Assimilative Model of the Ionospheric Electron Density (Pignalberi, 2019). As a separate example, the IRI-Based Real-Time Assimilative Mapping (IRTAM) (Galkin et al., 2020) DA model uses KF on the time series of the ionospheric parameters, and is further discussed in Section 4.2.

There are several known disadvantages related to the non-parametrized DA models. First, a high computational load is required when performing the calculations for the 3-D electron density grid. When a global regularly-spaced grid with 2° horizontal resolution (or 16200 horizontal grid points) and 10 km vertical resolution (or 100 vertical grid points) is used, the total number of grid points becomes 1,620,000. The background covariance matrix for this grid should then be as large as 1,620,000 × 1,620,000. This matrix size slows down the calculations significantly.

The second challenge relates to the ingestion of non-local data types, such as slant total electron content (sTEC) from radio occultation (RO) data, sTEC from ground-based Global Positioning System (GPS) receivers, and ultraviolet (UV) radiance satellite imaging data. In the case of RO data, the sTEC measurement indicates the amount of electrons along a straight line between Low-Earth Orbit (LEO) and GPS satellites. However, the DA system has no information on how the data should be distributed along this path, or at which points along this straight line the modeled ionosphere should be corrected to match the observations. This makes the sTEC data non-localizable, with a very low information content on a per-measurement basis (Reid et al., 2023). One consequence of the ingestion of non-local data is difficulty resolving the horizontal variation of the electron density peak height (or hmF2) parameter. For example, the ingestion of RO sTEC into IDA4D mainly influences the peak density (NmF2), and infrequently adjusts the hmF2 (Forsythe & McDonald, 2022).

The third challenge relates to the ingestion of sTEC measurements in data-dense regions. As was previously noted by Forsythe et al. (2021), the intersection of sTEC rays can cause localized changes to electron density profiles (EDPs), creating unrealistic EDPs with fictitious layers or bite-outs. In other words, when sTEC rays intersect in an undetermined way, it is possible for the assimilation to produce the analysis that matches the measurements well, but contains non-physical EDPs. This last challenge is of particular importance for the DA applications, such as ray tracing, because an unrealistic bottom-side electron density will result in incorrect HF signal propagation path that lead to unusable communication recommendations.

These challenges provided the motivation to rethink the current approach to ionospheric DA. This led to the development of ANCHOR, a novel DA approach that assimilates anchor points extracted from data into horizontal maps of the ionospheric param-

eters. Naval Research Laboratory (NRL) currently implements this DA method in Python, as it has low computational requirements.

This paper starts by discussing the parametrization method used in ANCHOR. It then describes the coordinate system, background model (PyIRI), derivation of the background covariance, and DA scheme used in ANCHOR. Finally, it shows the results of an observing system simulation experiment (OSSE) that evaluates ANCHOR's performance.

2 ANCHOR DA

2.1 Parametrization

According to the IRI model (Bilitza et al., 2022), a typical vertical ionospheric profile consists of a main layer, called the F2 layer, that peaks at approximately 350 km. Figure 1 demonstrates a typical EDP. During the daytime, the photochemical layers (F1 and E) appear below the F2 peak. The D layer, which overlaps with the upper mesosphere is neglected in PyIRI due to its low plasma density. The D region is also ignored by ANCHOR due to the lack of D-region data to assimilate.

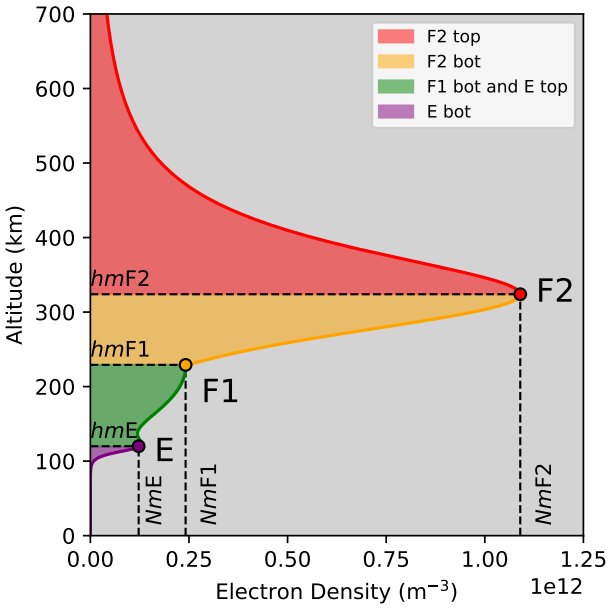


Figure 1. Example of an EDP and its main parameters.

It is possible to describe the EDP in terms of 11 parameters, where 6 parameters NmF2, NmF1, NmE, hmF2, hmF1, and hmE indicate the peaks and heights of the three layers, that are shown with red, orange and purple circles, respectively, in Figure 1. The remaining, five parameters are needed to describe the thicknesses of the top and the bottom sides of F2 layer (B_{top}^{F2}, B_{bot}^{F2}), the bottom side of F1 layer (B_{bot}^{F1}) and the bottom side of E layer (B_{bot}^{E}), as shown with colors in Figure 1. The 11 parameters can be reduced to five independent anchor points (hence the name ANCHOR): NmF2, hmF2, B_{top}^{F2}, B_{bot}^{F2}, and NmE, as the rest of the parameters can be either fixed at a constant values, such as hmE, B_{bot}^{E}, and B_{top}^{E}, or can be derived from the combination of the independent parameters, as it is done in PyIRI for the NmF1, hmF1, and B_{bot}^{F1} (Forsythe et al., 2023). Although, ANCHOR currently uses five independent anchor parameters, it is possible to extend this number in the future (e.g. to include the hmE correction, the D-region parameters, and the sporadic E layer).

2.2 Reference Frame and The Background

A reference frame selection is a very important step in the DA set-up process. Most ionospheric DA models operate in geographic coordinates. In this frame the subsolar point changes longitude position as the Earth rotates, whereas the position of the geomagnetic equator stays unchanged with time. As a result, the position of the ionospheric regions (in relation to the Earth's surface) also change with time. Figure 2a shows a snapshot of the NmF2 according to PyIRI for 02:30 Universal Time (UT) of 1 April 2020. The yellow circle shows the position of the subsolar point. The two crests of the Equatorial Ionization Anomaly (EIA) flank the geomagnetic equator shown with white dashed line.

A collection of the 24-hour snapshots in geographic coordinates describes the complete cycle of "motion" of the ionosphere from the right to the left (as the Earth rotates). Therefore, the daily mean μ and the standard deviation σ, shown in Figures 2c and 2e, respectively, among these snapshots determined in the geographic coordinates look as equally distributed in longitude enhancements that follow the curvature of the geomagnetic equator.

However, μ and σ parameters look completely different if calculated in another coordinate system. Figure 2b shows the NmF2 snapshot (same as in Figure 2a) organized in the Quasi-Dipole Latitude (QDLat) and Magnetic Local Time (MLT) coordinate sys-

tem. The subsolat point is now centered at 12 MLT and the geomagnetic equator is a straight line. The two crests of the EIA flank the straight geomagnetic equator.

Importantly, the other UT snapshots in QDLat-MLT coordinates look very similar to Figure 2b. Therefore, the daily mean μ, shown in Figure 2d, also looks very similar to each UT snapshot. Since the variance between each UT snapshot is very small, the standard deviation (σ) in QDLat-MLT coordinates is very low, especially in comparison to σ from Figure 2e.

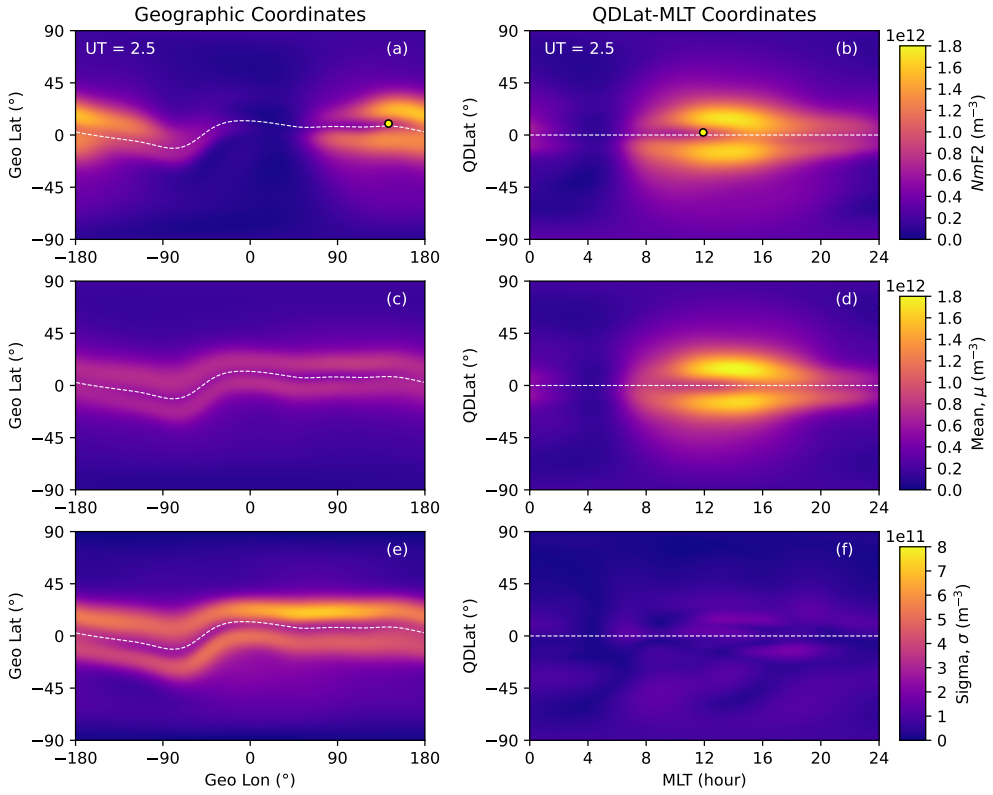


Figure 2. Comparison of the $NmF2$ and its daily variability expressed in geographic and QDLat-MLT coordinate systems for 1 April 2020. Panels (a) and (b) show the snapshots at 02:30 UT. Panels (c) and (d) show the daily mean, and panels (e) and (f) show the daily standard deviation. Geomagnetic equator is shown with white dashed line. Yellow circles in panels (a) and (b) show the locations of the subsolar points.

ANCHOR performs DA in the QDLat-MLT coordinate system. Unlike other DA models that use geographic coordinate systems that are agnostic to the subsolar location, the QDLat-MLT coordinate system considers both the orientation of the magnetic

field and the subsolar point location. Since this information is essential in the organization of ionospheric density (through photoionization and plasma transport), this reference frame is a more natural fit for ionospheric specification. This is reflected in the consistency of μ over the course of a day in the QDLat-MLT frame. The model variance and covariance also need to be defined specifically for QDLat-MLT system and are described below.

At the current stage of development, the PyIRI model (Forsythe et al., 2023) is employed as a background for ANCHOR. The background ionospheric parameters are determined for 15-min time frames because the regularly-spaced in QDLat-MLT coordinates grid points have different geographic locations in a geographic coordinate system (the PyIRI input needs to be in the geographic coordinate system). The μ and σ for all PyIRI background parameters for 1 April 2020 are shown in Figure 3. σ is expressed as percent difference from the μ. ANCHOR operates in $\log_{10}$ space for the electron density peaks to guarantee the positivity of the outputs. Therefore, NmF2 and NmE parameters in Figure 3a and 3b are expressed in $\log_{10}$ scale. The NmE parameter is strongly controlled by the solar irradiance, while the other parameters show spatial variations that reveal the importance of plasma transport. All of the parameters have very low diurnal standard deviations. When expressed in percentage of the mean, the density anchor parameters NmF2 and NmE have standard deviations that do not exceed two percent, with the largest standard deviations in the Antarctic region, where observations are sparse. The percent standard deviation is most significant for the B_{bot}^{F2} parameter.

The model daily σ, as is, cannot be used as model covariance σ, for the following reasons. First, it is so low that it would not allow the background to deviate far enough to match the observations. Second, the horizontal structure of σ should be considered with caution. For instance, σ for the NmE is equal to zero around noon and midnight, as shown in Figure 3g, because the E-region model takes into account only the position of the sun. These zero values for σ will prevent the propagation of the information from the data to the model, even if the variance is inflated using some multiplier. Another example can be made about the NmF2 σ, shown in Figure 3f. During the day time, it has zero values in the Northern hemisphere. This means that the data observed in the Northern hemisphere will have minimum DA influence compared to the data observed in the Southern hemisphere, just because the background will be restricted to move away from the μ.

The horizontal distribution of μ is more important in the QDLat-MLT coordinate system, than the horizontal distribution of σ. For each parameter it clearly indicates where the values should be high, and where they should be low. Therefore, ANCHOR models the variance, setting it to 30% of the daily background mean. It is important to mention that 30% is not important by itself as a number, because for the DA the ratio between the variance of the model to the variance of the data is more important. For our OSSE (Section 3.1), the data error will be set to 10%, meaning that the DA should trust the data three times more than it trusts the background.

The background vector $\vec{x}_b$ is formed for each time frame as a flattened 1-D array that contains all the background parameters. For example, in case of 15 min MLT and 2° QDLat resolutions, the grid for each map of the parameters has the size of 8827 elements, and the $\vec{x}_b$ vector has 44135 elements. A regular DA scheme (with similar horizontal resolution and 10 km vertical resolution) would have a state vector 183 times as large. This reduction in size allows much faster computation and lower hardware requirements.

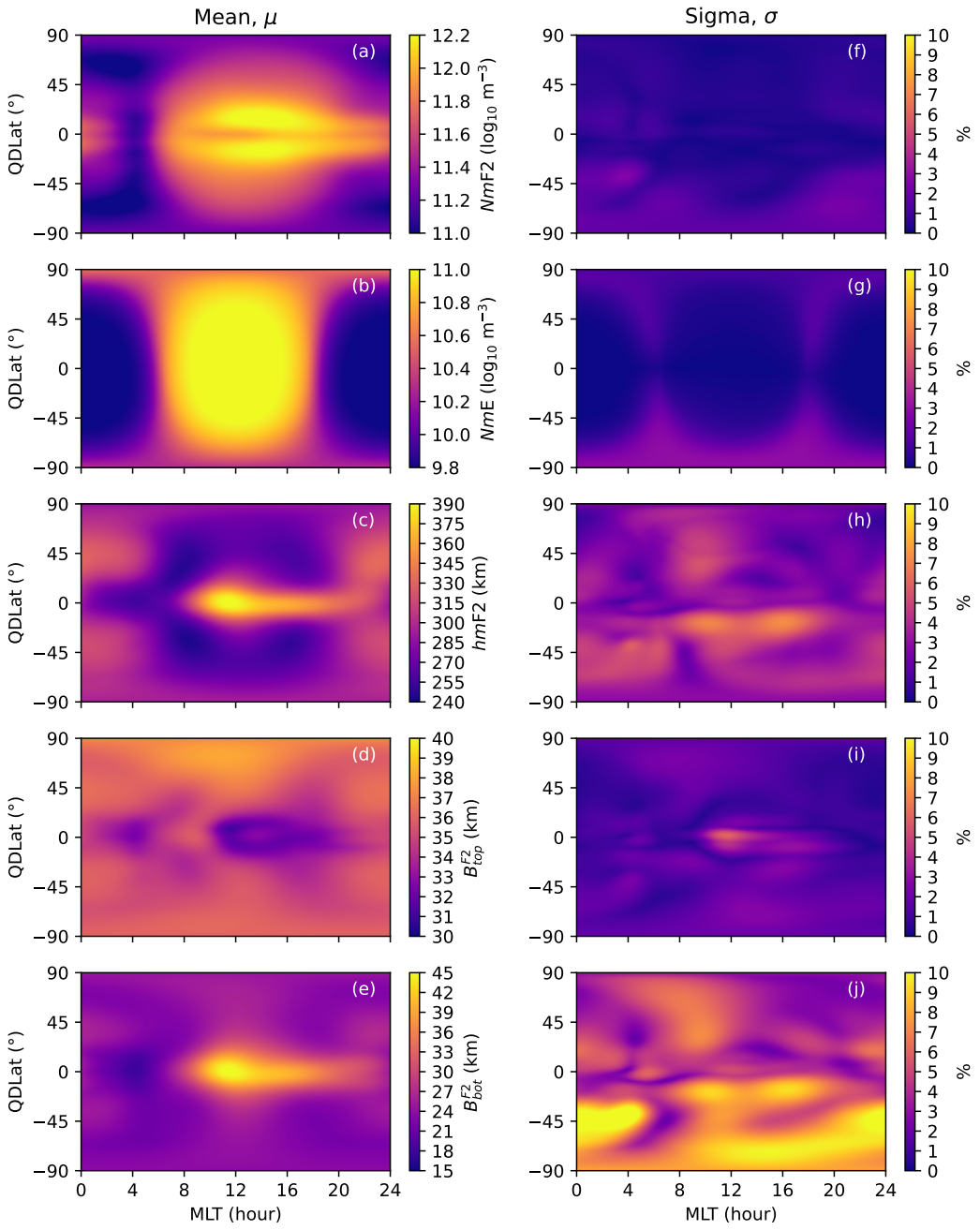


Figure 3. Daily mean μ and standard deviation σ of the anchor parameters in QDLat-MLT coordinates for 1 April 2020. σ is expressed as percent from the μ.

2.3 Background Covariance Matrix

Now that the model variance is established, it is necessary to develop the model covariance. A novel element here is the coordinate frame. ANCHOR calculates the back-

ground covariance, $\tilde{P}_b$, for each day of assimilation. First, it finds the deviation of each time frame from the daily mean for each anchor parameter. Then it finds the correlations of these deviations for each grid location with other locations. Since the correlations are calculated and not modeled as Gaussian distributions, it is necessary to localize them. Without localization one observation would transport the information all the way to the opposite side of the globe. Therefore, the correlations are multiplied with Gaspari-Cohn localization (Gaspari & Cohn, 1999) with the radius of 20° of great circle distance (GCD). This localization guarantees that at 40° GCD the correlation fully drops to zero. The localization radius of 20° GCD was chosen based on the previous work by Forsythe, Azeem, and Crowley (2020), where according to the GPS data the maximum of the mean horizontal ionospheric correlation length was found to be approximate 20° GCD.

The localized correlations for the NmF2 parameter around six reference points located at 12° QDLat and separated by 4 MLT are shown in Figure 4. It is obvious that in QDLat-MLT coordinates the mean MLT position of the equatorial crests determines the unique shape of the correlation ellipses. Whereas three reference points at 12, 16, and 20 MLT have secondary correlation peaks that indicate the geomagnetically conjugate points on the opposite side of the equator that are present only for the NmF2 parameter. This happens due to the latitudinal symmetry in the diurnal variance of the EIA. Importantly, the applied localization radius is large enough and does not suppress the horizontal structure of the correlations.

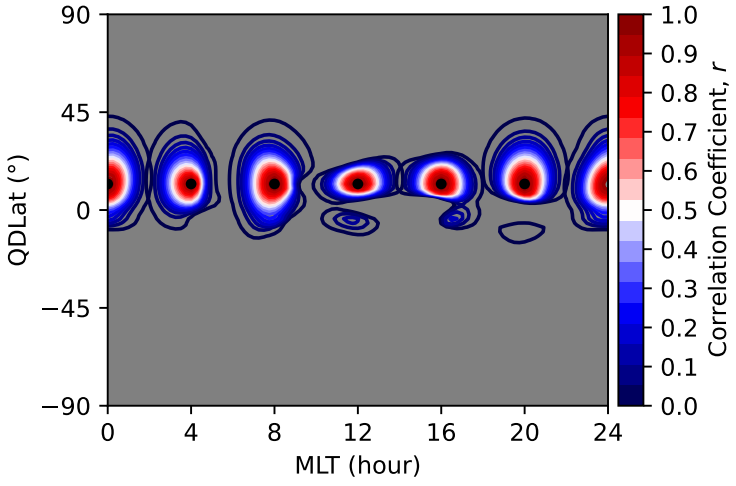


Figure 4. Correlations for NmF2 around six reference points for 1 April 2020.

The background covariance matrix is formed for all of the anchor parameters, taking into account the localized correlations for the off-diagonal elements and the allowed 30% standard deviation for the diagonal elements. It is important to note that because each parameter varies within this limit, it is impossible to obtain an unrealistic EDP. For example, it is impossible to obtain a fictitious ionospheric layer or a bite out, because it would require the addition of more parameters to the EDP description formalism. This fully addresses the third challenge of the non-parametrized ionospheric DA listed in the Section 1.

The covariance matrix is, however, further reduced to contain only the covariance between the same parameters, excluding the cross-parameter correlations. Therefore, at the current state of development, ANCHOR treats all five core parameters as independent of each other. The cross-parameter correlation requires further investigation and is beyond the scope of this study.

2.4 Observation Vector

Ionospheric parameters are extracted from RO and ionosonde data. This process is described in the upcoming Section 3.1. The observation vector, $\vec{y}$, is formed as a 1-D array of all the measured parameters, combining RO and ionosonde measurements, following the same order of the parameters as in the $\vec{x}_b$ array. The observation error covariance matrix, $\tilde{R}$, is formed assuming that the parameters are independent and setting the observation error to 10% of the measurement. As was mentioned before, in combination with the given 30% of the background variance, the DA trusts the data three times more than it trusts the background.

2.5 The Observation Operator

The observation operator, $\tilde{H}$, is formed as a matrix with shape $[N_{obs}, N_b]$, where N_{obs} is the number of observations in array $\vec{y}$, and N_b is the number of elements in the $\vec{x}_b$ array. The nearest-neighbor interpolation is used as a convenient simplification for the localized point measurements. For each observation, $\tilde{H}$ will have an element equal to unity at the closest grid point to the occurred observation, and zeroes everywhere else. Thus, the following matrix multiplication

$$\vec{y}_b = \tilde{H}\vec{x}_b \tag{1}$$

gives the expected observations $\vec{y}_b$ according to the given background.

The $\vec{y}_b$ is further used to exclude observations that differ significantly from the background. For the first three parameters (NmF2, NmE, and hmF2) the data points are excluded in the case where the residuals $|\vec{y}-\vec{y}_b|$ are larger than three times the background standard deviation. This is a so-called 3σ rule. However, this multiplier is increased to 10 for the B^{F2}_{top} and B^{F2}_{bot} since the residuals are significantly higher for those parameters (as expected given the empirical model variation and/or worse performance of the model for these parameters).

Additionally, it was found useful to combine observations that are close (located within 10° GCD) and differ by more that 10%. The presence of such contradicting observations often creates an unrealistically sharp gradient in the analysis. Therefore, such observations are averaged to reduce the number of contradicting measurements. The number of these exclusions is close to zero for the simulated data, but this reduction becomes important for the real data (Sakov & Sandery, 2017).

2.6 Kalman Filter

The Kalman Filter formalism is used for the ANCHOR DA scheme, where the analysis $\vec{x}_a$ represents the background corrected by the ingested observations and is defined as

$$\vec{x}_a = \vec{x}_b + \tilde{K}\left[\vec{y} - \vec{y}_b\right], \tag{2}$$

where $\tilde{K}$ is the Kalman gain defined as

$$\tilde{K} = \tilde{P}_b\tilde{H}^T\left[\tilde{H}\tilde{P}_b\tilde{H}^T + \tilde{R}\right]^{-1}, \tag{3}$$

In such a configuration of the filter, the information from the previous time frame does not propagate to the next time frame. For ANCHOR this is kept on purpose, because in the future ANCHOR will be coupled with full physics-based SAMI3 (Huba et al., 2000) model. This model coupling will take care of the information propagation from the current DA time frame to the next one.

However, for this paper ANCHOR will also be run in a different filter setting known as Gauss-Markov model, where the $\vec{x}_a$ is defined as

$$\vec{x}_a = \vec{x}_f + \tilde{K}\left[\vec{y} - \vec{y}_f\right], \tag{4}$$

where $\vec{x}_f$ is the forecast calculated as

$$\vec{x}_f = \vec{x}_{a'} + (\vec{x}_{a'} - \vec{x}_b)e^{-dt/\tau}, \tag{5}$$

where $\vec{x}_{a'}$ is the analysis from the previous assimilation time frame, dt is the DA time resolution (0.25 hours), and τ is decorrelation time that defines how soon the data becomes old (Forsythe, I. Azeem, et al., 2020), which is set to 2 hours. Appendix A has more information on the decorrelation time for ANCHOR.

2.7 3-D Density

After each time frame ANCHOR finds other dependent parameters, such as NmF1, hmF1, and B_{bot}^{F1}, utilizing the analysis of the main anchor parameters. Then, the maps of all parameters are converted to 3-D electron density. ANCHOR uses the same formalism as PyIRI to calculate the dependent parameters and to perform the computationally efficient 3-D density derivation, as described in details in Forsythe et al. (2023).

3 Results and Performance Evaluation

3.1 OSSE Truth and Simulated Observations

The evaluation of the ANCHOR algorithm was performed using an OSSE, set up with the full physics-based SAMI3 model outputs as *truth*. Observations were simulated based on the SAMI3 electron density output resampled into a regularly spaced geographic grid. Figure 5 compares the PyIRI background parameters to the SAMI3-derived parameters. All the parameters besides the NmF2 show a significant deviation from the background, which is desirable for the DA evaluation.

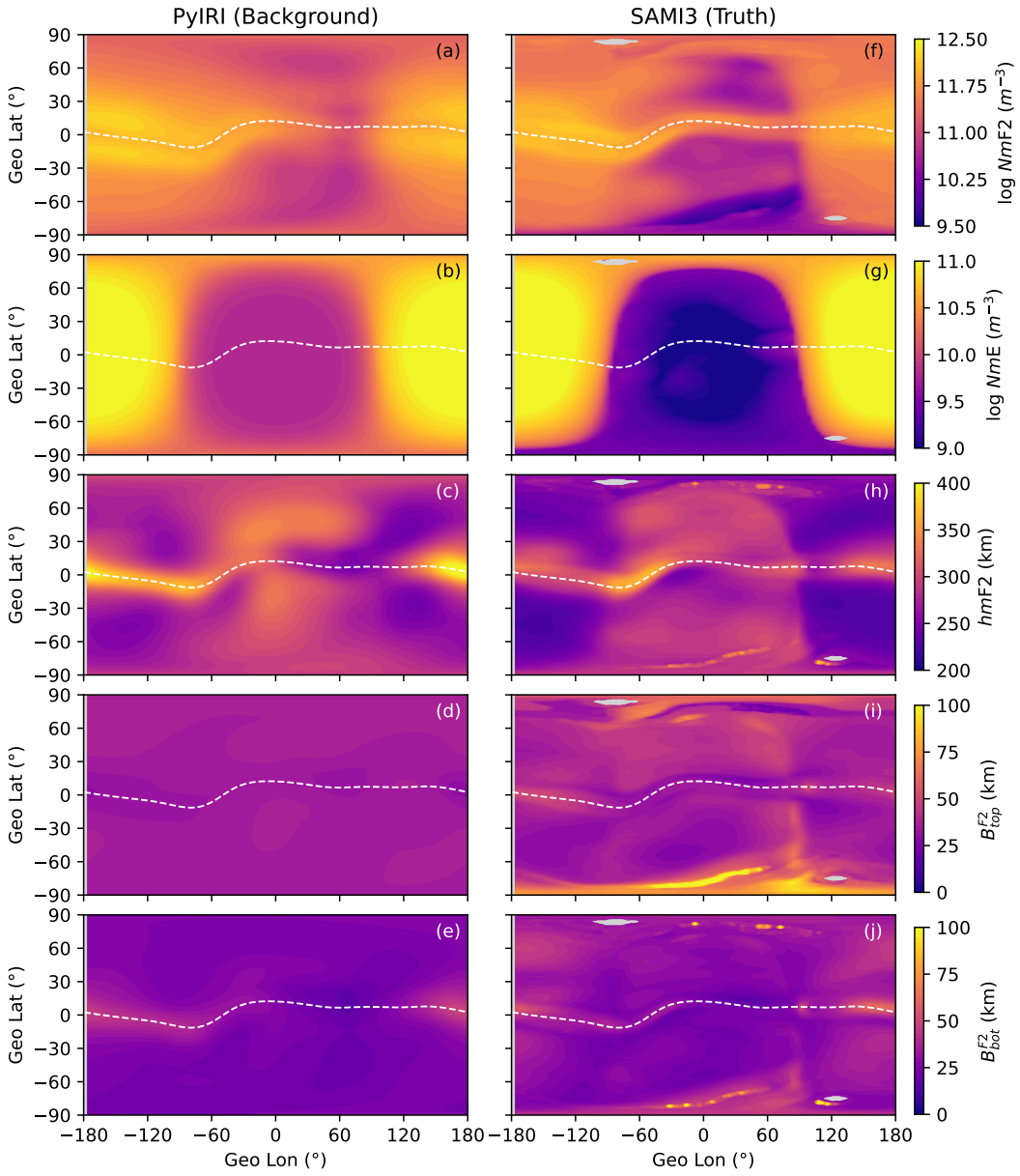


Figure 5. Comparison between background parameters and SAMI3 parameters used as known truth for 1 April 2020, 00:00:00 UT. Geomagnetic equator is shown with white dashed line.

COSMIC-2 (Lin et al., 2020) satellite EDPs along the tangent points (TP) were simulated using SAMI3 as a known truth for 1 April 2020 (by finding electron density along the TP locations from SAMI3). It is important to mention that this approach simplifies the OSSE, making the results look better than the real-life DA performance, where the RO sTEC measurements are first inverted to EDP using the Abel inversion technique. This choice is discussed in the Section 4. Figure 6a shows the locations of the COSMIC-

2 ROs between 2 and 3 UT. The white circles show the location of the observed NmF2, with red and blue lines indicating the locations of top- and bottom-sides of the F2 regions along the TPs, respectively. A black circle in Figure 6a shows the EDP shown in Figure 6b. The black line in Figure 6b shows the simulated profile, which is not smooth because the vertical resolution of the TP is higher than the SAMI3 output file vertical resolution. The NmF2 and hmF2 are first found. Then, using least square fitting, the thicknesses parameters B_{top}^{F2} and B_{bot}^{F2} are found using Epstein function formalism that is explained in details in Section 4 of the PyIRI paper (Forsythe et al., 2023). Red and blue curves in Figure 6b demonstrate the reconstruction of the topside and bottomside using the B_{top}^{F2} and B_{bot}^{F2} parameters, respectively. Next, the blue curve is subtracted from the black curve and a location of the NmE is found as a local maxima around 110 km with a 10 km allowed deviation.

The observation operator that was discussed in Section 2.5 localizes the RO measurements to the location of the observed NmF2, or the white circles in Figure 6a. However, given the data influence radius defined by the background covariance model (shown in Figure 4), the data influence will be distributed around the points, fully enclosing most of the TP tracks in Figure 6a, especially the bottom side portions, shown with red color, since they are usually shorter than the top-side portions.

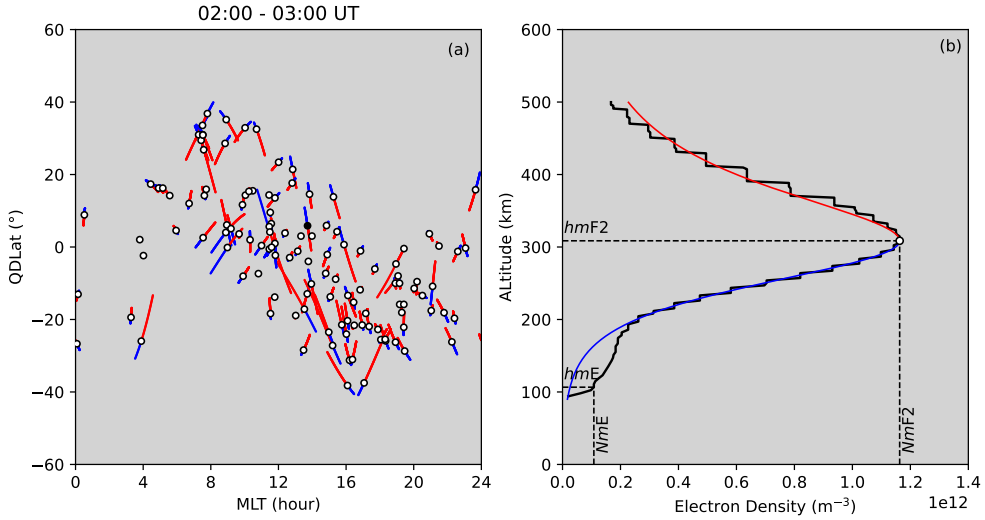


Figure 6. Locations of COSMIC-2 ROs and an example of the simulated data EDP.

A similar fitting procedure was applied to the vertical ionosonde data collected for 1 April 2020 for 20 ionosondes. Additionally, the locations of the ionosondes were randomly perturbed to mask the operational data set. The parametrization differed from RO only in excluding the determination of the B_{top}^{F2}, since ionosondes can not measure the ionospheric density above the peak density. In this OSSE the simulated data represent the ideal synthetic observation, without adding any additional instrumentation error.

3.2 DA Analysis

Figure 7 shows the results of the assimilation for one 15-min time frame for 1 April 2020, 00:30:00 UT. The first column in Figure 7 shows the background parameters, the second column shows the analysis with the ingested point measurements shown in circles, and the third column shows the percent difference between the background and the analysis together with the difference in the expected and observed parameters shown with circles. The largest modification to the background is made for the B_{top}^{F2} and B_{bot}^{F2} parameters, which use expanded color bars to ensure all % differences are visible.

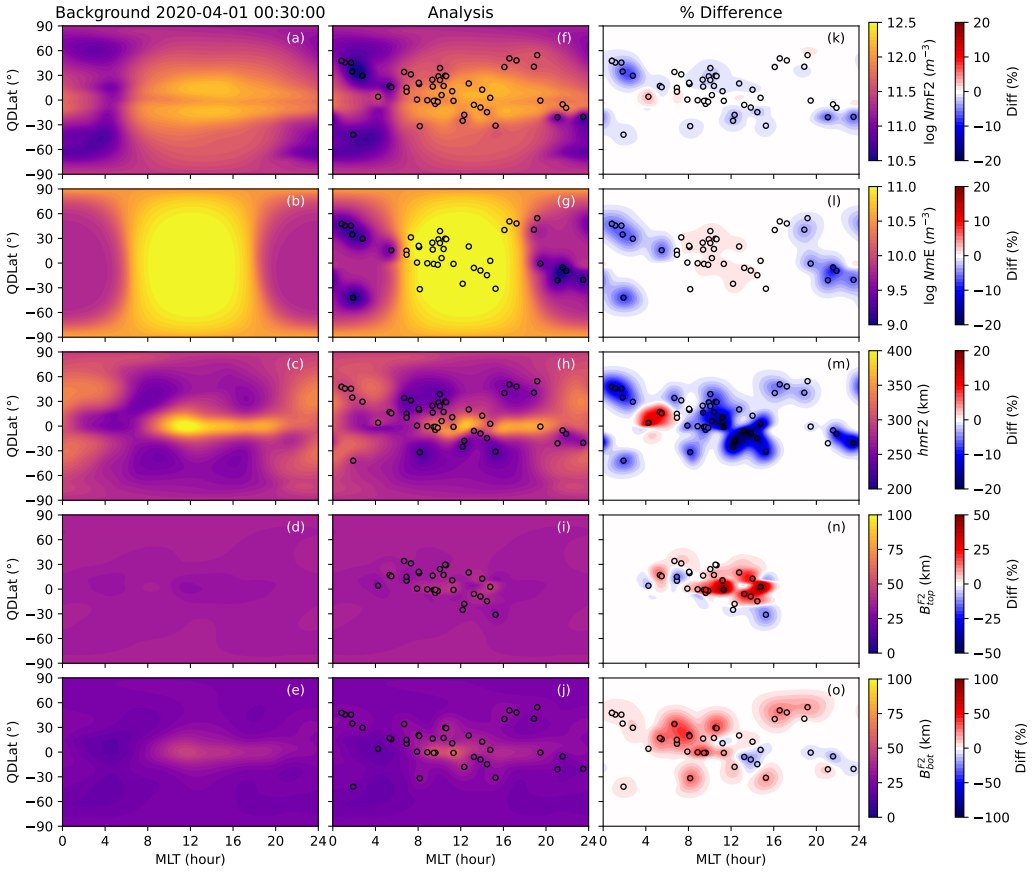


Figure 7. ANCHOR results for 1 April 2020, 00:30:00 UT.

3.3 Root Mean Square Errors

To analyze the ANCHOR results, the root mean square errors (RMSEs) between the SAMI3 truth and both the background and the analysis are found for the locations that have non-zero percent difference shown in the third column of Figure 7. Five panels in Figure 8 show the RMSEs for the background and analysis with black and red lines, respectively, for the five anchor parameters. In the case of NmF2 and NmE, the RMSEs are reduced by half in the analysis. In case of hmF2 the RMSEs are reduced by about 35%. Due to the large difference between the truth and the background, the B^{F2}_{top} and B^{F2}_{bot} show slightly lower reduction of RMSEs in the analysis.

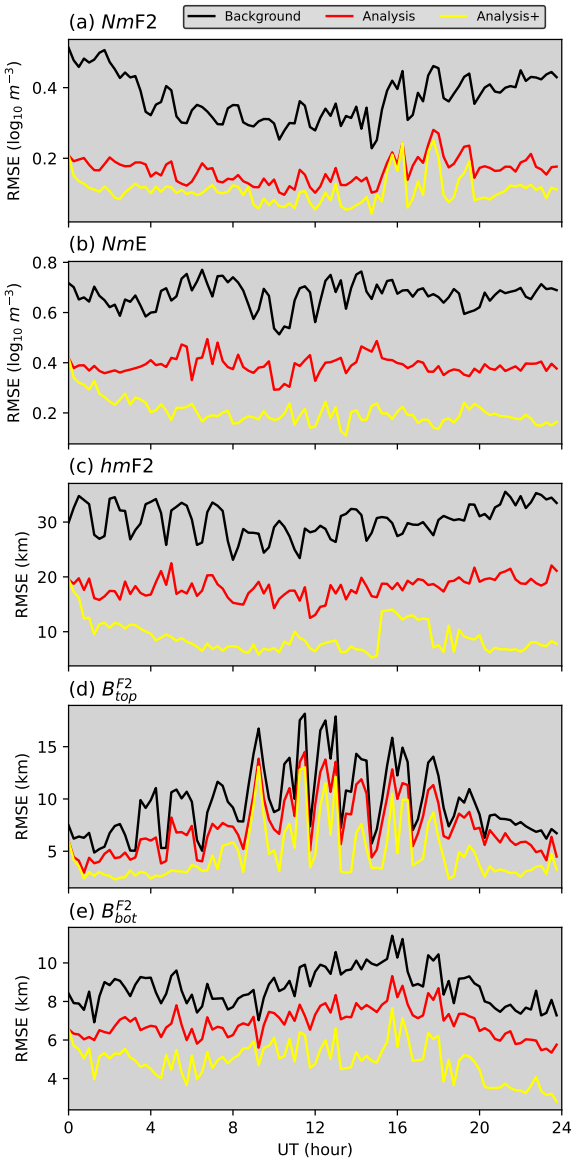


Figure 8. RMSEs for two ANCHOR runs.

In case one wonders about the 1.5-hour oscillations in RMSEs, they arise due to the clustering of the RO data. Figure 7i showed only RO data locations (since ionosonde data does not have B_{top}^{F2} parameter). For this 15-min time window the RO points are clustered between 4 and 16 MLT. Whereas the MLT sector moves by 4 MLT hours to the right every 15 minutes, making the full loop in 1.5 hours. Since the RMSE calculation includes only data-rich regions, the locations of the RO data influences these values.

So far, this study has focused on nowcast analysis results using only data available at the time of the assimulation. It is possible to use methods developed for forecasting to further reduce the analysis RMSE by propagating the information from the previous assimilation window to the next one by using Gauss-Markov model defined in Equation 4, because the state starts closer to a realistic ionospheric state. The yellow line in Figure 8 shows the RMSEs for the case when the analysis from the previous assimilation window was used to update a background for the current window.

4 Discussion

4.1 Challenges

In the beginning of this paper, the current challenges of the non-parametrized ionospheric DA models were presented. This section discusses how those challenges were addressed by ANCHOR.

The first challenge was related to the high computational intensity of the non-parametrized models, such as IDA4D. Due to the anchor parametrization methods, ANCHOR's covariance matrix is 1347 times smaller than a typical covariance matrix of IDA4D, which makes ANCHOR very rapid. On average it takes about 10 sec to produce the analysis for one assimilation time frame on a regular PC, while IDA4D takes several minutes when configured to use 44 openMP threads on a system with 200 GB of memory.

The second challenge was related to information propagation during the assimilation of non-local data types, such as sTEC from RO data. ANCHOR directly addresses this problem through the extraction of the EDP parameters from post-processed RO data, thus converting the non-local RO data type into a local measurement. Some may argue that an additional data pre-processing, such as Abel inversion for the RO data, introduces additional errors for the NmF2 parameter. However, as it was demonstrated in Figure 7, this approach enables the propagation of the information to all EDP parameters. Additionally, since most DA systems ingest RO sTEC as a relative measurement (by the assimilation of the difference between sTEC from a chosen base ray and the other rays), it does not guarantee the precise determination of sTEC, nor of NmF2 in addition to leaving the other parameters unchanged from the background. In the future, ANCHOR will introduce an advanced RO data pre-processing, by using the information about the horizontal density gradients derived from the analysis to improve the currently used

Abel inversion technique (by avoiding the symmetry assumption that is used in a standard Abel inversion).

The third challenge described the presence of unrealistic profiles in IDA4D analysis. Since ANCHOR is a parametrized DA model, this issue is fully addressed. Because each parameter varies within a physically realistic limit (set by the given model variance) it is impossible to obtain a discontinuity, a bite-out, or a fictitious F3 layer in the assimilation results. This makes ANCHOR a very reliable algorithm, where the output is always compatible with the ray tracing applications.

4.2 ANCHOR in Comparison to Other Parametrized DA Models

The parametrization for the ionospheric DA models is a rational approach. Examples of two other models that applied the parametrization are IRTAM (Galkin et al., 2020) and The Assimilative Canadian High Arctic Ionospheric Model (A-CHAIM) (Reid et al., 2023).

IRTAM is an operational ionospheric weather model based on ionosonde data from the Global Ionosphere Radio Observatory (GIRO) (Reinisch & Galkin, 2011). IRTAM adjusts IRI diurnal coefficients by looking at the diurnal time series of the NmF2, hmF2, B0 and B1 (the last two are the IRI parameters that describe the F2 bottomside thickness) parameters at each ionosonde station. Next, IRTAM adjusts global IRI coefficients to connect the ionosonde stations and to produce a smooth global distribution of the parameters. Since the diurnal distribution of the parameters at fixed locations (like ionosonde stations) is the key to the IRTAM, it is challenging to add RO data into the same IRTAM DA formalism, because the locations of RO measurements also change with time. Unlike IRTAM, ANCHOR does not perform DA on time series, and therefore is able to combine ionosonde and RO data, treating it in a similar way. Comparing the two DA systems, IRTAM provides smoothed in time and space global ionospheric background, whereas ANCHOR focuses on the nowcasting for each time frame and reveals higher spacial resolution features in the distribution of the parameters.

A-CHAIM applies particle filter DA approach to the series of spherical cap harmonic perturbations on the Empirical Canadian High Arctic Ionospheric Model (E-CHAIM). Unlike ANCHOR, A-CHAIM parametrizes the ionosphere not only vertically, but also horizontally, by finding a set of spherical harmonic coefficients that describe the hori-

zontal distribution of the parameters at each time frame. The particle filter DA approach makes A-CHAIM somewhat similar to Ensemble KF approach, where a particular realization of the model is chosen as a combination of all other possible model realizations. A-CHAIM is a regional DA model, unlike ANCHOR that produces the analysis on the global scale.

5 Conclusion and Future Directions

A new nowcast data assimilation algorithm, ANCHOR, was developed at the U.S. Naval Research Laboratory. It extracts ionospheric parameters from RO and ionosonde data and assimilates them as point measurements into the maps of the background parameters using Kalman Filter approach. ANCHOR is written in Python. It has a high computational efficiency, and is guaranteed to provide a reliable analysis suitable for the ray tracing applications.

There are many avenues for the future development of this DA model. Ground based GPS, UV radiance, and HF radar data will be added to the assimilation process. In the future ANCHOR will use SAMI3 as a background and will have a sophisticated forecast model utilizing the coupling between physics-based SAMI3 model and the ANCHOR nowcast. The possibility to assimilate sporadic E data into ANCHOR will be investigated. ANCHOR will undergo thorough optimization, stress testing, and validation in future efforts, with the goal of providing robust, operational DA support for space weather models.

6 Open Research

- PyIRI software used as ANCHOR's background model is available to the community at GitHub (Forsythe & Burrell, 2023).
- Raw COSMIC-2 data is available at `https://data.cosmic.ucar.edu/gnss-ro/cosmic2/provisional/spaceWeather/`.
- SAMI3 density outputs used as truth here are available at `https://doi.org/10.5281/zenodo.10196102`.
- Simulated RO and ionosonde parameters are available at `https://doi.org/10.5281/zenodo.10196588`.

- ANCHOR assimilation outputs are available at `https://doi.org/10.5281/zenodo.10196650`.

Acknowledgments

This work is sponsored by the Office of Naval Research.

Appendix A Decorrelation Time

Even though, ANCHOR is developed to be a nowcast assimilation scheme, the question about the decorrelation time was investigated. The decorrelation time should indicate the appropriate aging of the data in relation to the forecast background. A long decorrelation time would indicate slower ionospheric processes and therefore long-lasting data influence, whereas a short decorrelation time should point to the dynamic ionospheric region where the data becomes old faster. Figure A1 compares decorrelation time between geographic and QDLat-MLT coordinate systems. The decorrelation time, τ was calculated from the PyIRI model following the recipe from Forsythe, I. Azeem, et al. (2020). It is rather counterintuitive, but in the QDLat-MLT system τ is shorter in comparison with the geographic coordinate system, even though the variance is much smaller than in the geographical system, as was demonstrated in Figure 2. This indicates the presence of small-amplitude pulsations of the equatorial crests around its MLT and QDLat position. However, it is questionable whether these pulsations have any physical origin or whether they arise as an artificial effect that is coming from the URSI coefficients (a model that determines the global and diurnal distribution of $NmF2$ in PyIRI) that were derived for the geographic coordinate system. Further investigation, or even further improvement of the climatological coefficients, is required to address this issue.

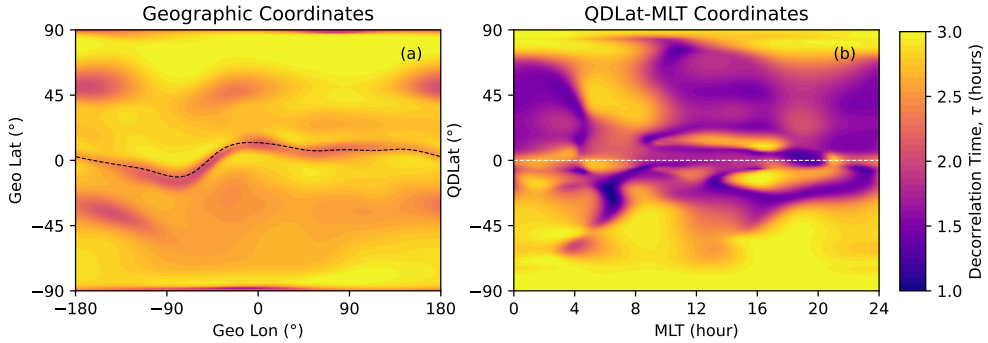


Figure A1. Comparison of the decorrelation time in geographic and QDLat-MLT systems. Geomagnetic equators are shown with dashed lines.

References

Angling, M. J., & Cannon, P. S. (2004). Assimilation of radio occultation measurements into background ionospheric models. *Radio Science*, *39*(1). doi: 10.1029/2002RS002819

Bilitza, D., Pezzopane, M., Truhlik, V., Altadill, D., Reinisch, B. W., & Pignalberi, A. (2022). The International Reference Ionosphere model: a review and description of an ionospheric benchmark. *Rev. Geophys.*, *60*(4). doi: 10.1029/2022RG000792

Bust, G. S., & Crowley, G. (2007). Tracking of polar cap ionospheric patches using data assimilation. *J. Geophys. Res. Space Physics*, *112*(A5). doi: 10.1029/2005JA011597

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*. doi: 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*. doi: 10.1029/2020RS007159

Forsythe, V. V., Bilitza, D., Burrell, A. G., Dymond, K. F., Fritz, B. A., & McDonald, S. E. (2023). PyIRI: Whole-Globe Approach to the International Reference Ionosphere Modeling Implemented in Python. *Space Weather*. doi: 10.22541/essoar.169592556.61105365/v1

Forsythe, V. V., & Burrell, A. G. (2023). victoriyaforsythe/PyIRI: v0.0.1. *GitHub*.

doi: 10.5281/zenodo.8235173

Forsythe, V. V., I. Azeem, G. C., Makarevich, R. A., & Wang, C. (2020). The global analysis of the ionospheric correlation time and its implications for ionospheric data assimilation. *Radio Science*, *55*. doi: 10.1029/2020RS007181

Forsythe, V. V., & McDonald, S. E. (2022). Ingestion of radio occultation data for ionospheric data assimilation. *Poster presented at CEDAR*.

Galkin, I. A., Reinisch, B. W., Vesnin, A. M., Bilitza, D., Fridman, S., Habarulema, J. B., & Veliz, O. (2020). Assimilation of sparse continuous near-Earth weather measurements by NECTAR model morphing. *Space Weather*, *18*. doi: 10.1029/2020SW002463

Gaspari, G., & Cohn, S. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*.

Huba, J. D., Joyce, G., & Fedder, J. A. (2000). Sami2 is Another Model of the Ionosphere (Sami2): a new low-latitude ionosphere model. *J. Geophys. Res.: Space Phys.*, *105*. doi: 10.1029/2000JA000035

Lin, C.-Y., Lin, C. C.-H., Liu, J.-Y., Rajesh, P. K., Matsuo, T., Chou, M.-Y., ... Yeh, W.-H. (2020). The Early Results and Validation of FORMOSAT-7/COSMIC-2 Space Weather Products: Global Ionospheric Specification and Ne-Aided Abel Electron Density Profile. *Journal of Geophysical Research: Space Physics*, *125*(10). doi: 10.1029/2020JA028028

Pignalberi, A. (2019). A three-dimensional regional assimilative model of the ionospheric electron density. *Ph.D. Thesis*.

Reid, B., Themens, D. R., McCaffrey, A., Jayachandran, P. T., Johnsen, M. G., & Ulich, T. (2023). A-CHAIM: Near-real-time data assimilation of the high latitude ionosphere with a particle filter. *Space Weather*, *21*. doi: 10.1029/2022SW003185

Reinisch, B. W., & Galkin, I. A. (2011). Global ionospheric radio observatory (GIRO). *Earth, Planets and Space*, *63*, 377–381. doi: 10.5047/eps.2011.03.001

Sakov, P., & Sandery, P. (2017). An adaptive quality control procedure for data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, *69*. doi: 10.1080/16000870.2017.1318031

Schunk, R. W., Scherliess, L., Sojka, J. J., Thompson, D. C., Anderson, D. N.,

500 Codrescu, M., ... Howe, B. M. (2004). Global Assimilation of Ionospheric
501 Measurements (GAIM). *Radio Science*, *39*. doi: 10.1029/2002RS002794

Figure 1.

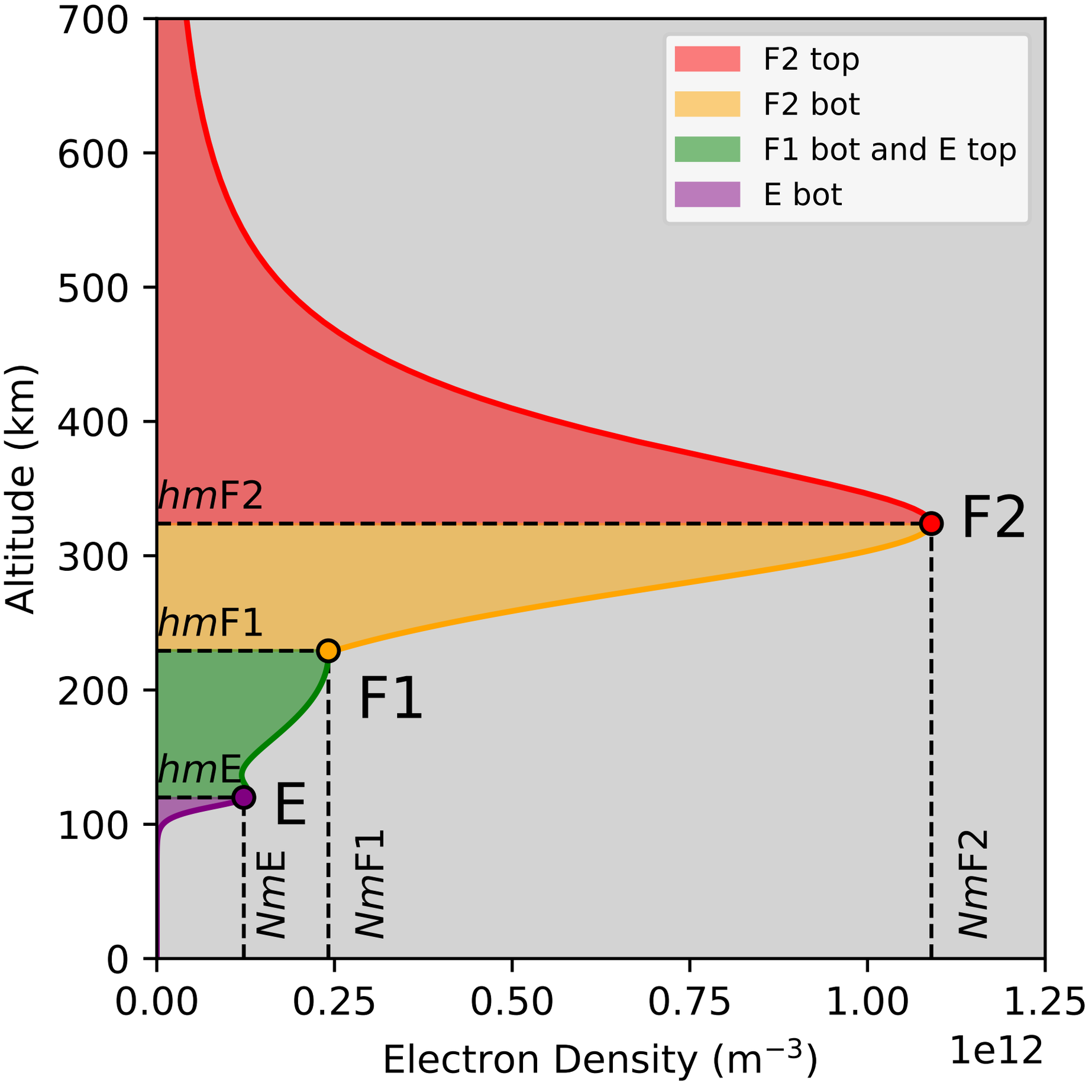

F2 top
F2 bot
F1 bot and E top
E bot
hmF2
hmF1
hmE
F2
F1
E
NmE
NmF1
NmF2
Altitude (km)
Electron Density (m^{-3})
1e12
0
100
200
300
400
500
600
700
0.00
0.25
0.50
0.75
1.00
1.25

Figure 2.

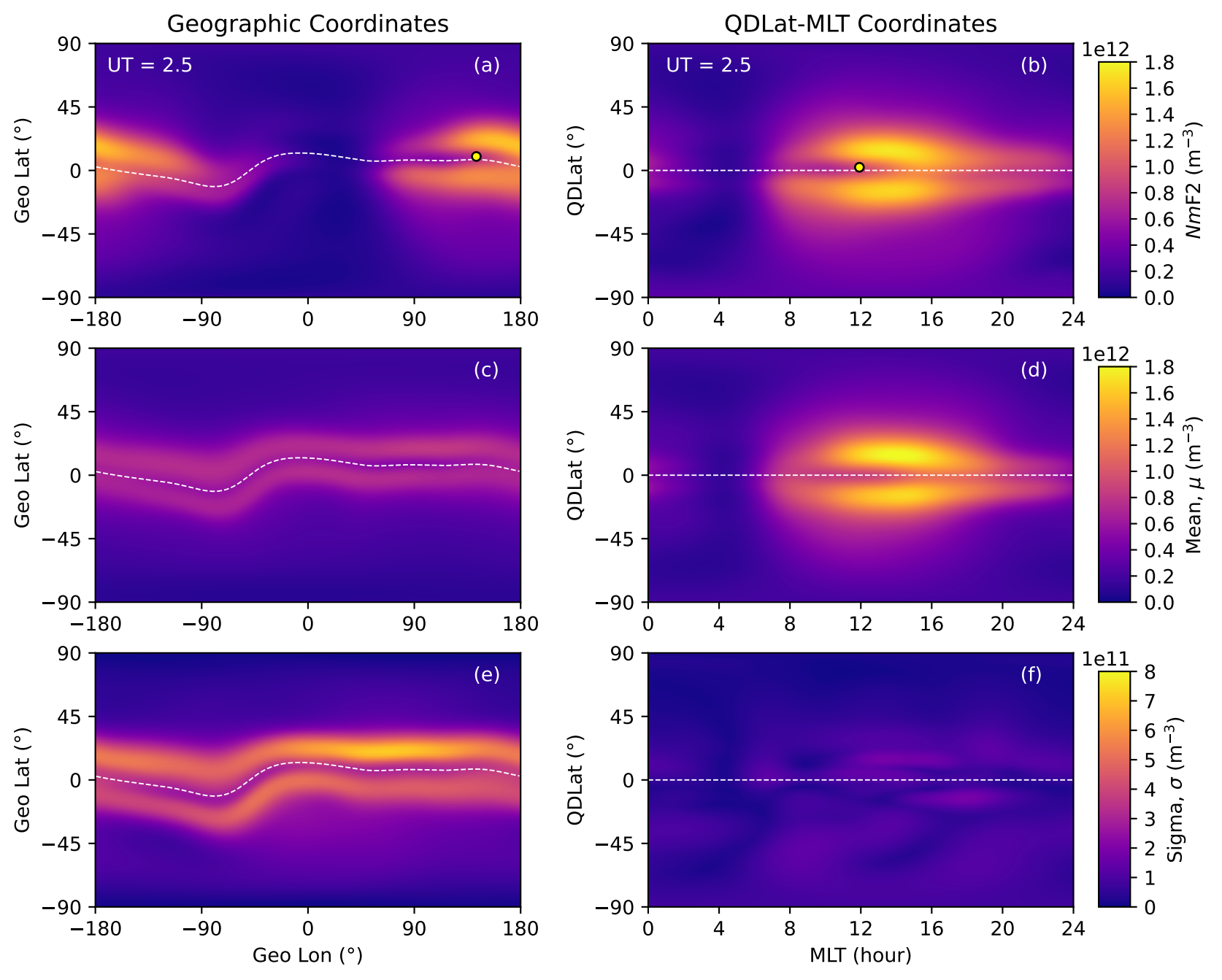

Geographic Coordinates
QDLat-MLT Coordinates
UT = 2.5
(a)
(b)
(c)
(d)
(e)
(f)
Geo Lat (°)
QDLat (°)
Geo Lon (°)
MLT (hour)
1e12
1e11
NmF2 (m^{-3})
Mean, μ (m^{-3})
Sigma, σ (m^{-3})

Figure 3.

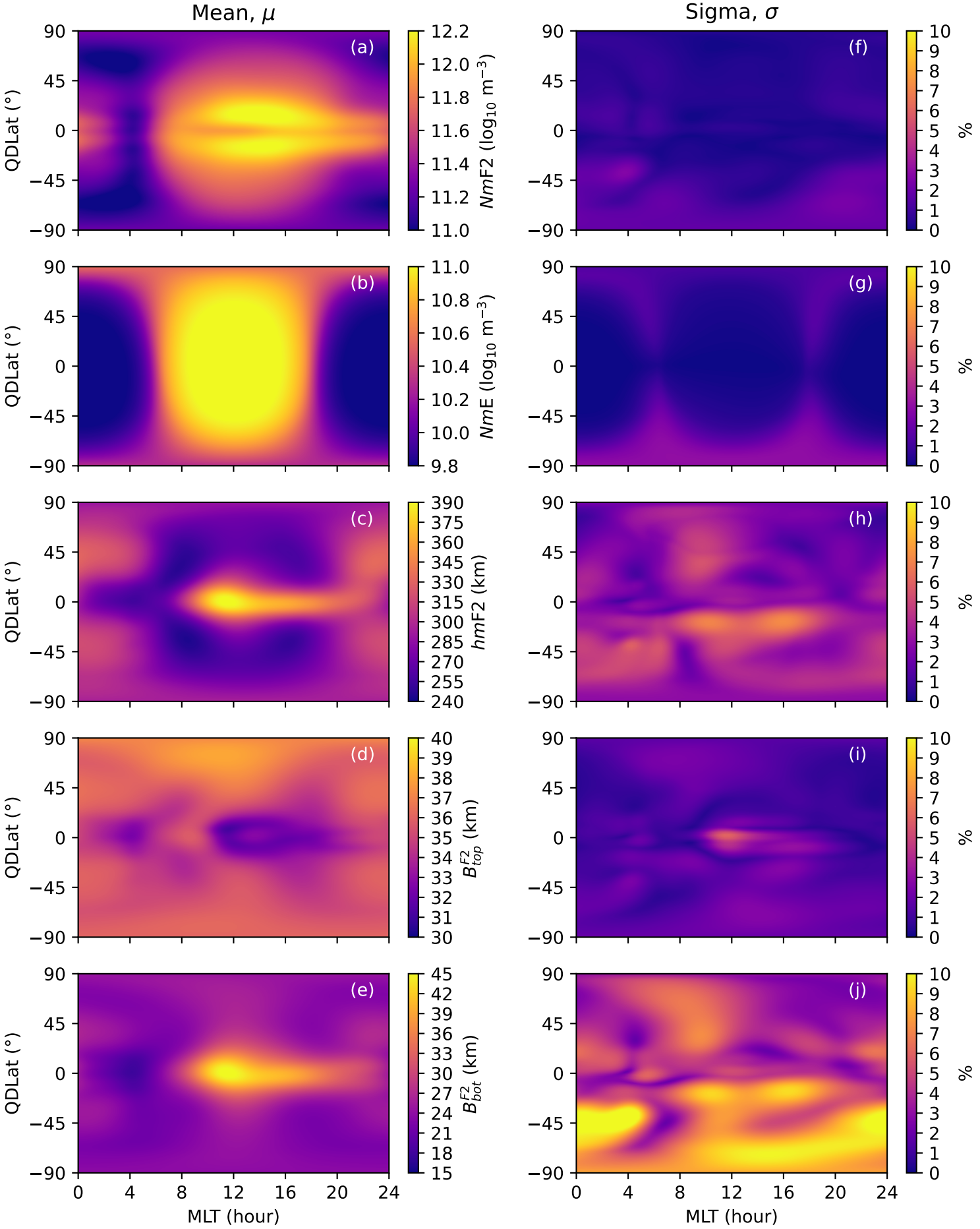

Mean, μ
Sigma, σ
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
QDLat (°)
MLT (hour)
$NmF2$ ($\log_{10}$ m^{-3})
NmE ($\log_{10}$ m^{-3})
$hmF2$ (km)
B_{top}^{F2} (km)
B_{bot}^{F2} (km)
%

Figure 4.

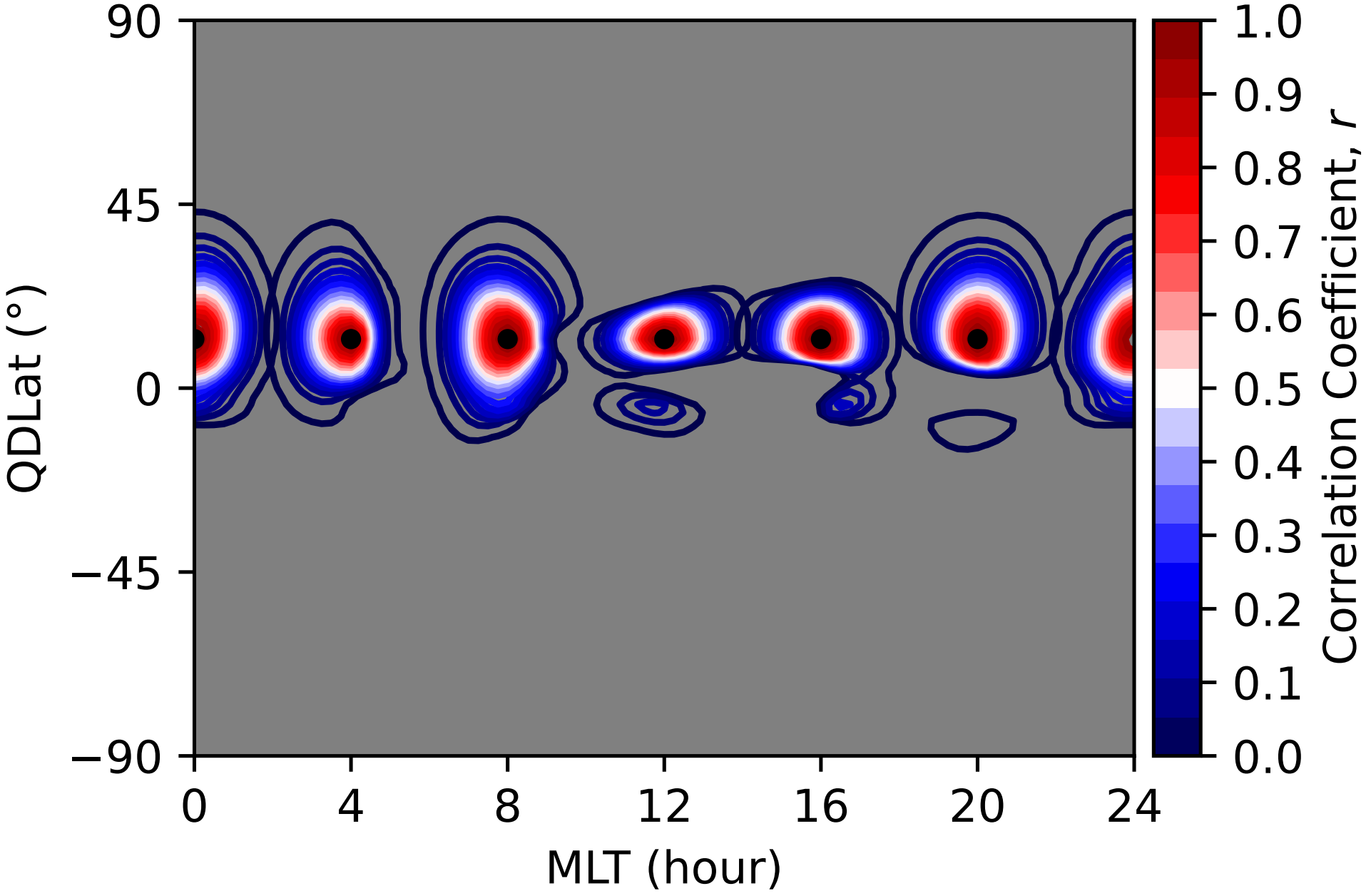

90
45
0
−45
−90
QDLat (°)
0
4
8
12
16
20
24
MLT (hour)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Correlation Coefficient, r

Figure 5.

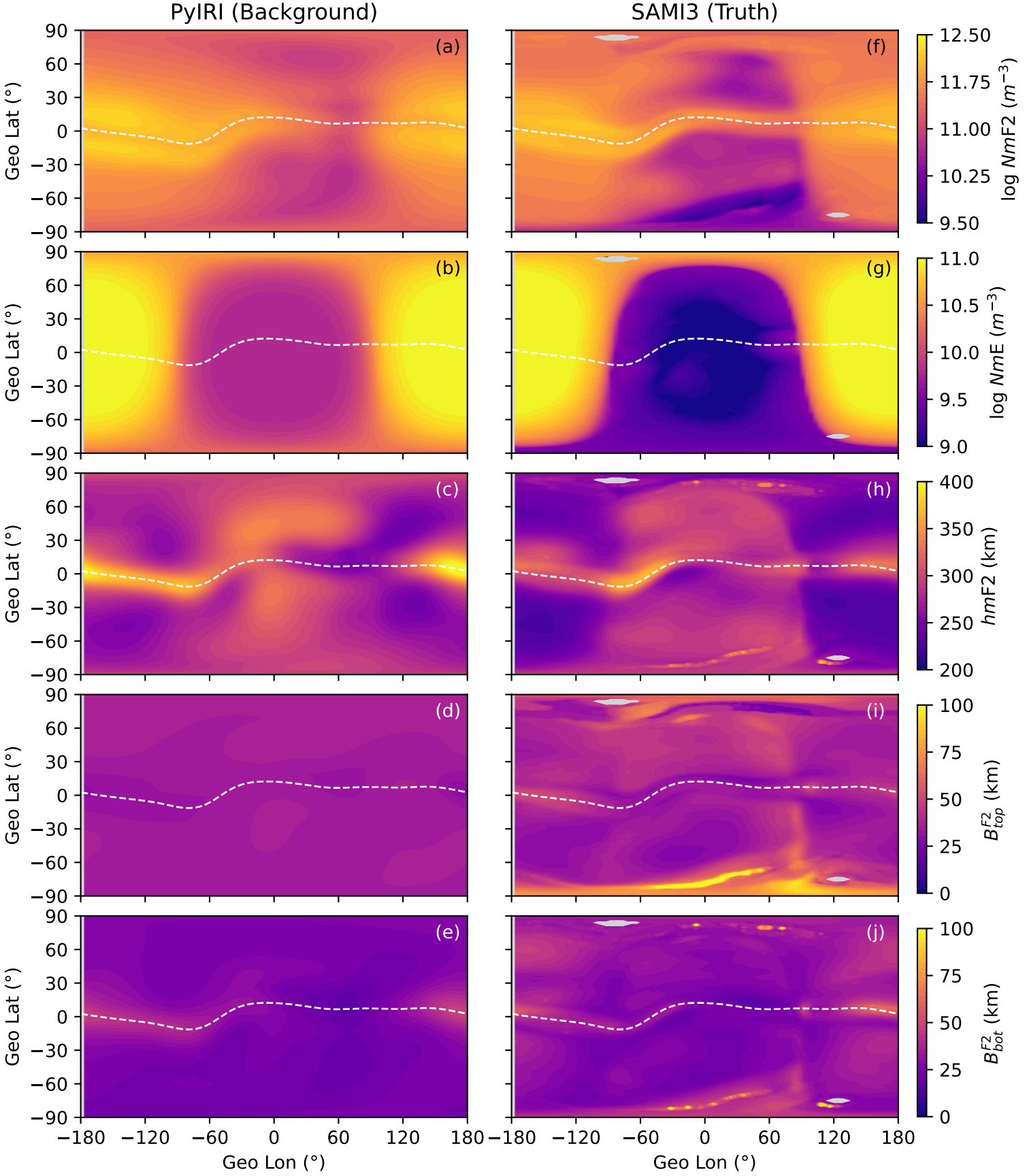

PyIRI (Background)
SAMI3 (Truth)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
Geo Lat (°)
Geo Lon (°)
90
60
30
0
−30
−60
−90
−180
−120
−60
0
60
120
180
12.50
11.75
11.00
10.25
9.50
log NmF2 (m^{-3})
11.0
10.5
10.0
9.5
9.0
log NmE (m^{-3})
400
350
300
250
200
hmF2 (km)
100
75
50
25
0
B_{top}^{F2} (km)
B_{bot}^{F2} (km)

Figure 6.

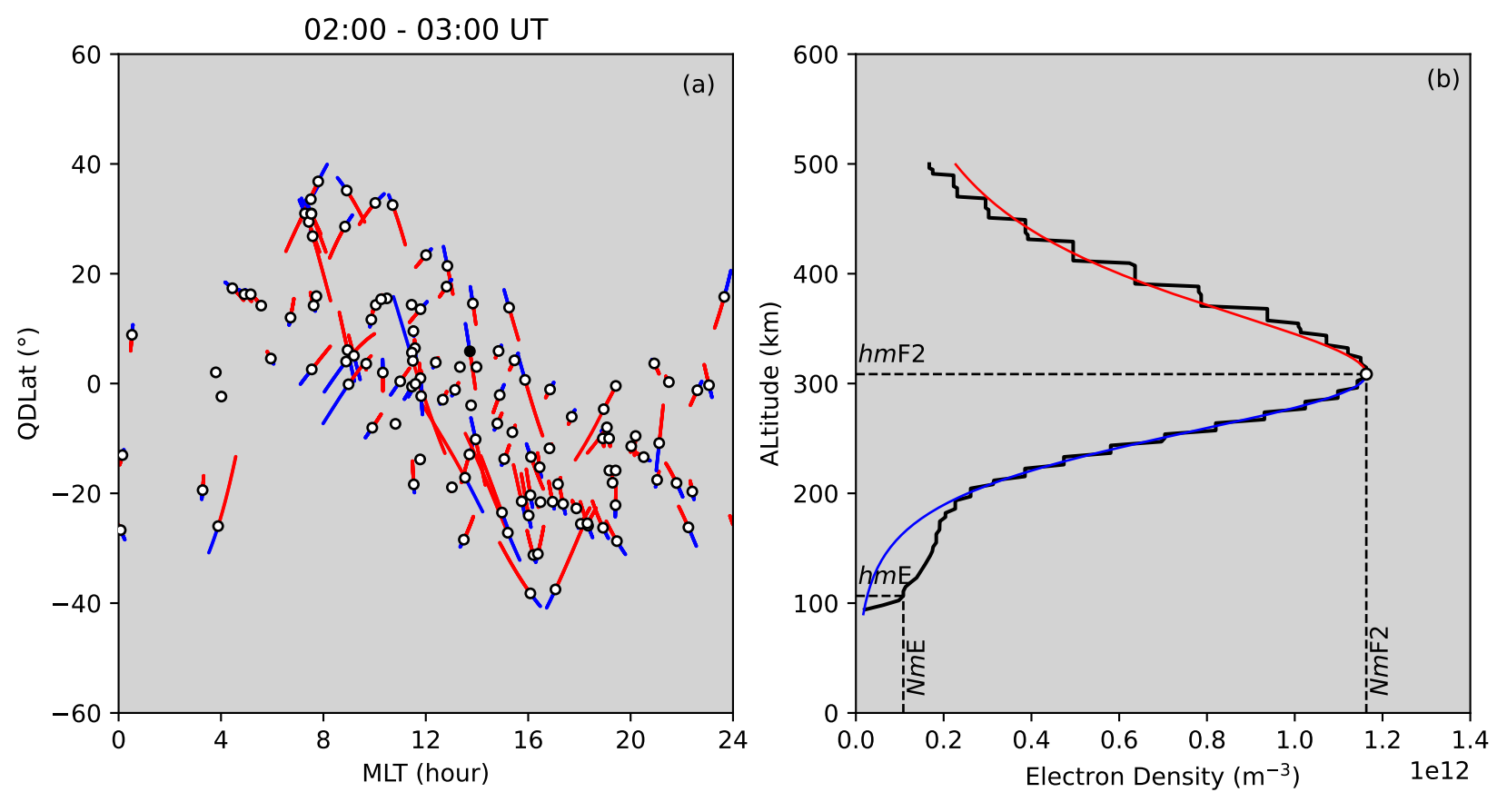

02:00 - 03:00 UT
(a)
QDLat (°)
MLT (hour)
(b)
ALtitude (km)
hmF2
hmE
NmE
NmF2
Electron Density (m^{-3})
1e12

Figure 7.

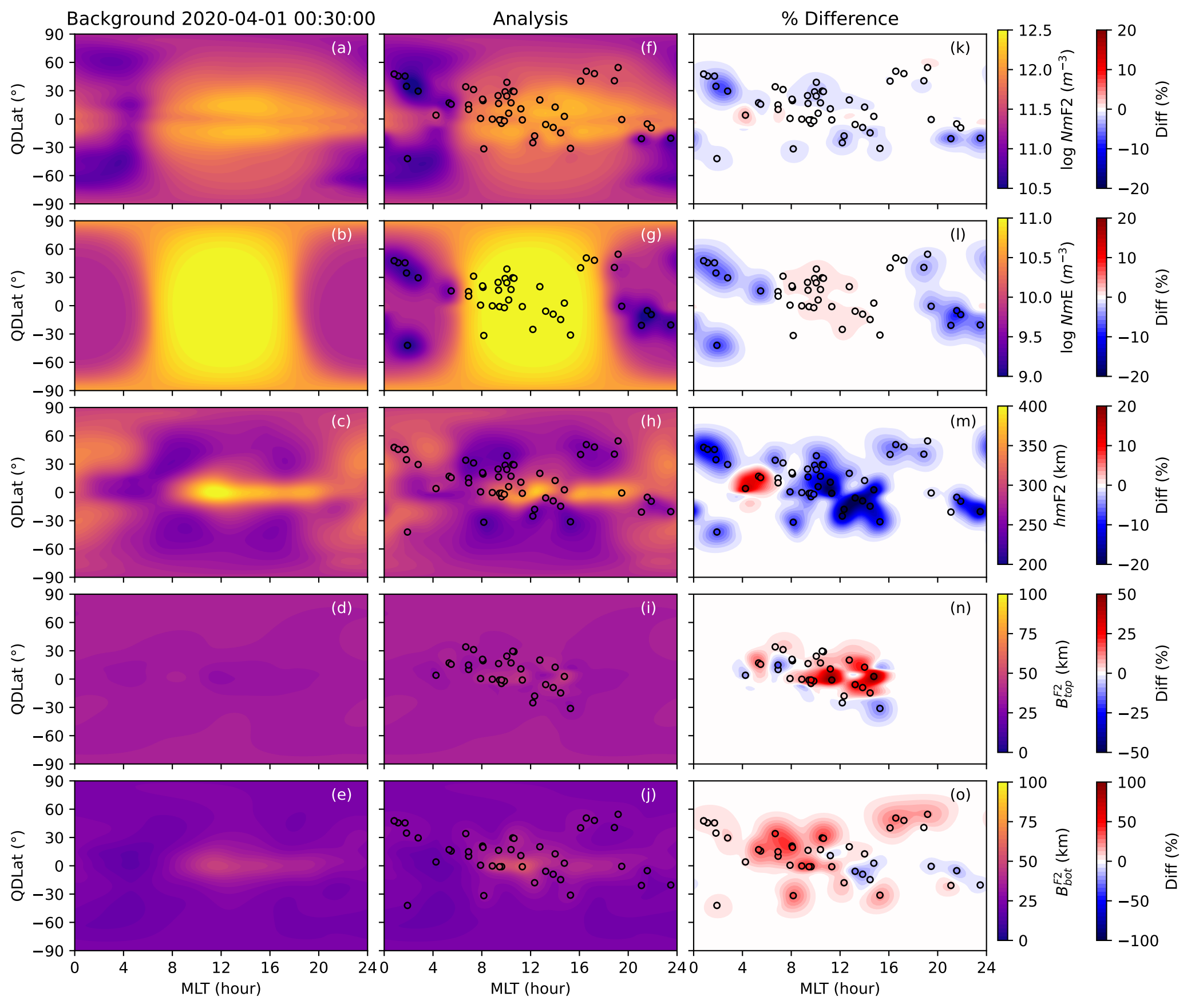

Background 2020-04-01 00:30:00
Analysis
% Difference
(a)
(f)
(k)
(b)
(g)
(l)
(c)
(h)
(m)
(d)
(i)
(n)
(e)
(j)
(o)
QDLat (°)
MLT (hour)
log NmF2 (m^{-3})
log NmE (m^{-3})
hmF2 (km)
B^{F2}_{top} (km)
B^{F2}_{bot} (km)
Diff (%)

Figure 8.

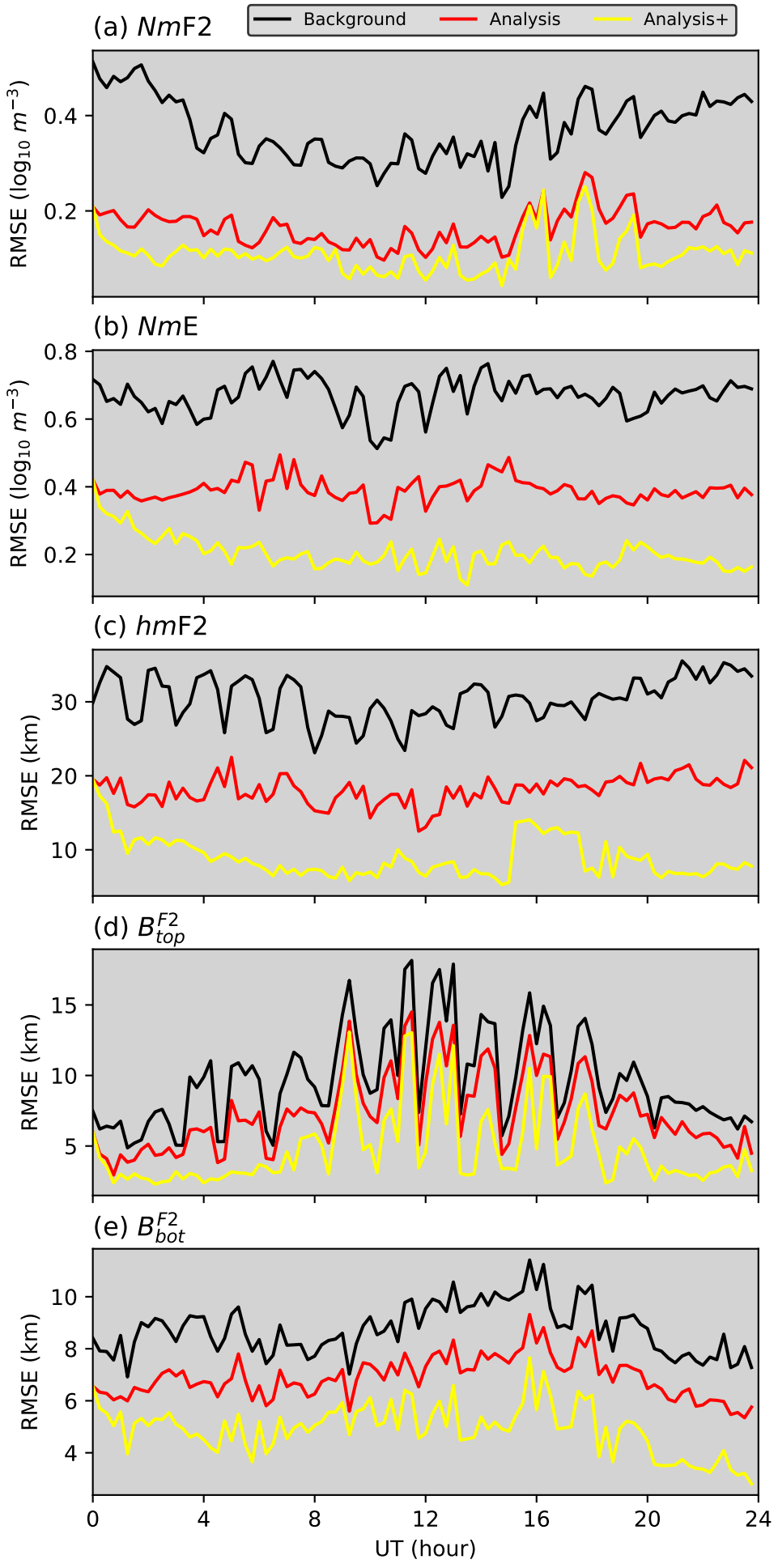
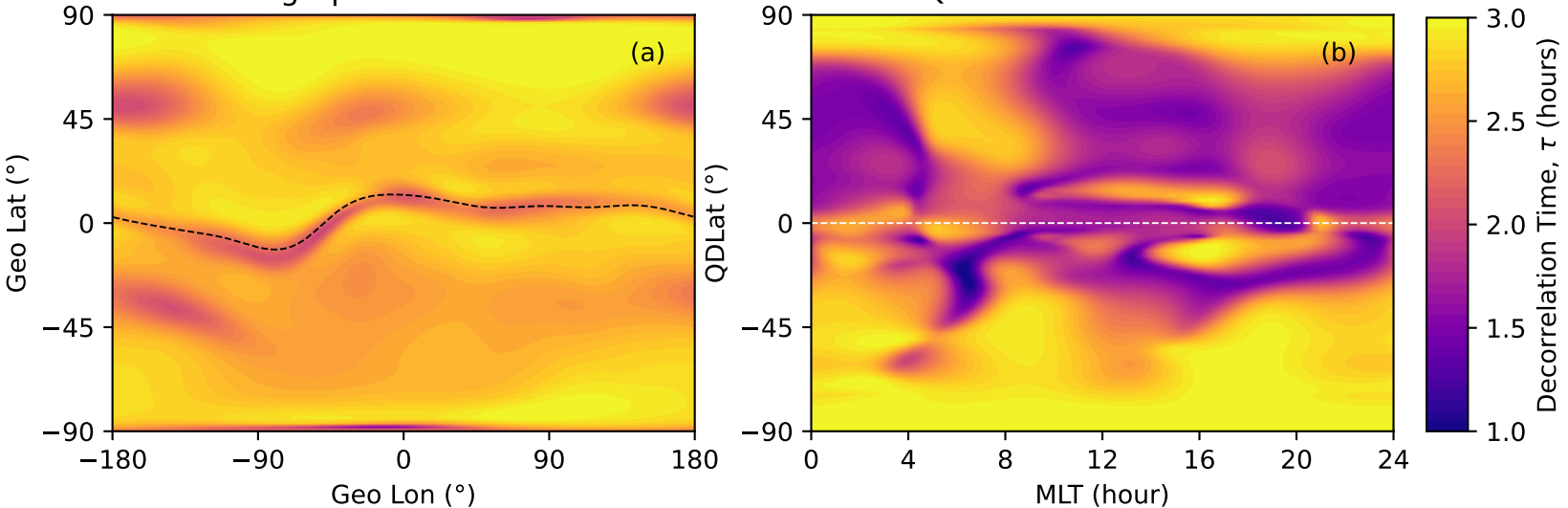

Background
Analysis
Analysis+
(a) NmF2
RMSE (log10 m−3)
0.4
0.2
(b) NmE
RMSE (log10 m−3)
0.8
0.6
0.4
0.2
(c) hmF2
RMSE (km)
30
20
10
(d) B_top^F2
RMSE (km)
15
10
5
(e) B_bot^F2
RMSE (km)
10
8
6
4
0
4
8
12
16
20
24
UT (hour)

Figure A1.


Geographic Coordinates
(a)
Geo Lat (°)
Geo Lon (°)
90
45
0
−45
−90
−180
−90
0
90
180
QDLat-MLT Coordinates
(b)
QDLat (°)
MLT (hour)
0
4
8
12
16
20
24
Decorrelation Time, τ (hours)
3.0
2.5
2.0
1.5
1.0