remote sensing

MDPI

Article

Error Analysis of the "Bottom-Up" Retrieval Algorithm Using Simulated First-Order Radio Occultation E-Region Excess Phase Profiles

Cornelius Csar Jude H. Salinas [1,2,*], Dong L. Wu [2], Madison Stanford [3], Liying Qian [4], Nimalan Swarnalingam [2,5], Daniel J. Emmons [6], Yosuke Yamazaki [7] and Jorge Chau [7]

1 Goddard Earth Sciences Technology and Research II, University of Maryland Baltimore County, Baltimore, MD 21250, USA
2 NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA; dong.l.wu@nasa.gov (D.L.W.); nimalan.swarnalingam@nasa.gov (N.S.)
3 Jackson School of Geosciences, University of Texas at Austin, Austin, TX 78712, USA; madison.stanford@jsg.utexas.edu
4 National Center for Atmospheric Research High Altitude Observatory, Boulder, CO 80301, USA; lqian@ucar.edu
5 Department of Physics, The Catholic University of America, Washington, DC 20064, USA
6 Department of Engineering Physics, Air Force Institute of Technology, Wright-Patterson Air Force Base, Dayton, OH 45433, USA; daniel.emmons.1@us.af.mil
7 Leibniz Institute for Atmospheric Physics, 18225 Kühlungsborn, Germany; yamazaki@iap-kborn.de (Y.Y.); chau@iap-kborn.org (J.C.)
* Correspondence: cornelius.c.salinas@nasa.gov

Highlights

What are the main findings?

- Under ideal conditions, the current version of the "bottom-up" retrieval algorithm performs best when E-region Ne values are enhanced while F-region Ne values are suppressed.
- The algorithm's method of removing F-region contributions can produce artifacts when true E-region Ne values are low.

What are the implications of the main findings?

- The current version of the "bottom-up" retrieval algorithm performs best over the night-time high latitudes.
- The method to remove F-region contributions needs to be improved.

Abstract

This paper first quantifies the total errors of the "bottom-up" retrieval algorithm that is applied to synthetic excess phase profiles that are only a function of electron density (Ne). These profiles are constructed using Ne outputs from the Specified Dynamics–Whole Atmosphere Community Climate Model with ionosphere/thermosphere eXtension (SD-WACCM-X). Then, this paper quantifies the general sensitivity of these errors to the true E-region Ne and the peak F-region Ne. Results indicate that, under ideal conditions, errors are generally low over the night-time high latitudes above 100 km but are extremely high over the night-time low to middle latitudes. This paper generalizes that the errors are inversely proportional to true E-region Ne while directly proportional to peak F-region Ne. The retrieved values also only have adequate correlation (>0.7) to true E-region Ne values greater than $0.1 \times 10^{11} \frac{e^-}{m^3}$. Based on these results, we conclude that, under ideal conditions, the current "bottom-up" retrieval algorithm performs best when the true E-region Ne is relatively high while peak F-region Ne is relatively low, as is the case over the night-time

Check for updates

Academic Editor: Kan Wang

Received: 30 October 2026
Revised: 20 November 2026
Accepted: 24 November 2026
Published: 26 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

high latitudes. We also conclude that the method to remove F-region contributions still induces significant error and even forms artifacts in the retrievals.

Keywords: GNSS radio occultation; ionosphere; electron density

1. Introduction

The ionospheric E-region is located between 90 and 150 km, where electrons are primarily produced through the day-time photodissociation of molecular oxygen and molecular nitrogen by soft X-ray and solar extreme ultraviolet (EUV) radiation [1,2]. This region is important to a wide range of navigation and communication systems that rely on high-frequency radio waves [3–5]. This is because the refractive index of the E-region is conducive to reflection and refraction of these waves [6,7]. Since the E-region's refractive index is a function of electron density (Ne), operators of these systems are reliant on accurate knowledge and prediction of E-region Ne.

For a long time, our knowledge of E-region Ne has been limited due to constraints in ground-based and space-based instruments. The most common ground-based instruments to observe the E-region are ionosondes and incoherent scatter radar (ISR). However, ionosondes can only perform day-time observations. ISR's can observe both day- and night-time but their spatial coverage is localized. Space-based in situ instruments cannot perform measurements in the E-region because the altitude is too low. A potential space-based remote sensing technique is radio occultation (RO). RO measures the changes in global navigation satellite system (GNSS) signals as they propagate through the atmosphere/ionosphere and are captured by satellite receivers [8]. The properties of these changes can then be used to estimate essential meteorological and space weather parameters [9,10]. One such property is the signal phase of global navigation satellite system (GNSS) radio wave signals when they are received by a low-Earth orbit (LEO) satellite. Without an atmosphere and ionosphere (free space), the signal phase should not change as the signal propagates from the GNSS satellite to the LEO receiver. With an atmosphere and ionosphere, though, signal phase changes primarily as a result of its propagation through a varying medium. When the receiver onboard the LEO satellite compares the phase of the received signal $v^{'}$ with the phase of a version of the signal in free space v_{free} (hereafter referred to as the free-space signal), there will be a difference. The difference between them is referred to as the total excess phase $\phi_{ex} = v^{'} - v_{free}$. The main RO measurement is total excess phase $\phi_{ex}(h_T)$ as a function of tangent height h_T.

Retrieval methods aim to estimate electron density vertical profiles $N_e(z)$ that are functions of altitude given a measured excess phase profile $\phi_{ex}(h_T)$. This is based on the principle that, assuming the GNSS RO signal solely propagates through a plasma layer mainly comprising electrons, the total excess phase is primarily a function of the horizontal total electron content (hTEC):

$$\phi_{ex}(h_T) = \frac{40.3 \times 10^{16}}{f^2} hTEC(h_T) = \frac{40.3 \times 10^{16}}{f^2} \left[\int N_e(s) \cdot ds \right] \quad (1)$$

In other words, total excess phase is proportional to the line integral of the electron density profile $N_e(s)$ along the curved ray path specified by s. The ray path is curved due to ionospheric bending. Note here that, strictly speaking, it is a function of $N_e(s)$ and not $N_e(z)$. However, retrieval algorithms estimate $N_e(z)$. This is because these retrieval algorithms commonly assume spherical symmetry of electron density along that shell of the ionosphere where the RO signal propagates. This assumption of spherical symmetry is one of the major sources of errors in RO's ability to resolve finer-scale ionospheric structures. Hence, current

retrieval algorithms only enable RO data to resolve very slowly evolving planetary-scale ionospheric structures. In addition, the assumption of spherical symmetry is not the only source of error. A more appropriate expression for total excess phase $\phi_{ex}(h_T)$ or $hTEC(h_T)$ is that it is the sum of two main components:

$$hTEC(h_T) \approx \int N_e(s) \cdot ds + X'(h_T) \quad (2)$$

$X'(h_T)$ represents numerous other phenomena, such as instrument noise, that are not purely a function of electron density.

RO retrieval algorithms frequently assume spherical symmetry, and they are also tailored to remove as much of $X'(h_T)$. For the ionospheric F-region, the "onion-peeling" method has been the most frequently used retrieval algorithm to retrieve total electron content and ionospheric F2-region peak electron density values [11,12]. However, this method is unstable as you go into much lower ionospheric regions.

Recently, a relatively more stable inversion algorithm [13] was developed that retrieves E-region Ne profiles from radio occultation (RO) total excess phase measurements. This is called the "bottom-up" retrieval algorithm. Since the E-region is much deeper into the ionosphere, a key capability of any lower ionospheric retrieval algorithm is that it should first isolate the component of the excess phase that is solely due to the signal's propagation through E-region plasma $\phi_E(h_T)$. After processing the total excess phase such that only the E-region and F-region contributions (converted to hTEC) $hTEC_{E,F}(h_T) = hTEC_E(h_T) + hTEC_F(h_T)$ remain, the "Bottom-Up" method utilizes a key approximation for $hTEC_F(h_T)$: the F-region component of the bending and phase advances at these lower ionospheric regions is a linear function of tangent height. This linear function can be estimated by fitting $hTEC_F(h_T)$ values between 30 and 50 km with a linear function to yield coefficients a_1 and a_0. This linear function is then subtracted from $hTEC_{E,F}(h_T)$ to yield:

$$hTEC_E(h_T) = hTEC_{E,F}(h_T) - [a_1 h_T + a_0] \quad (3)$$

$hTEC_E(h_T)$ is then subject to an optimal estimation inversion. A more complete description of this algorithm is in the methodology section.

The inversion algorithm has already been used to retrieve E-region Ne profiles from RO measurements performed by numerous RO constellation missions [13,14]. Validation efforts of the algorithm have been through comparisons of the RO dataset with observations from ground-based instruments as well as with simulations from empirical and first-principles physics-based models [14–16]. These comparisons were all made under geomagnetically quiet conditions. Comparisons against ground-based instruments have shown that, while the RO dataset and ground-based measurements do exhibit similar relative variations, the RO dataset's absolute values are consistently larger than values from the available ground-based measurements [14–16]. Comparisons with model data have also shown that the RO dataset exhibits spatial structures that are not simulated by the models. One unique feature in RO is a more sustained enhancement in low-latitude E-region Ne during afternoon hours that even extends into evening hours. Overlaid on this enhancement is the presence of an equatorial trough.

In this paper, we further our assessment of this algorithm by analyzing the performance of the current version of this E-region "bottom-up" retrieval algorithm when it is applied on $hTEC_E(h_T)$ under simulated idealized conditions; that is, $hTEC_E(h_T)$ is only a function of electron density along the signal path and $X'(h_T) \approx 0$. At the same time, we know the corresponding tangent Ne profile $N_{e,tan}(z)$ of these $hTEC_E(h_T)$ profiles. Obviously, this work cannot be performed using real measured excess phase profiles because we cannot isolate $\int N_e(s) \cdot ds$ from these measurements. To do that, we need an instrument capable of measuring electron density along a GNSS signal's ray path from the GNSS satellite to

the tangent point and then finally to the low-Earth orbit (LEO) satellite receiver. This is currently impossible. This paper, though, simulates this setup using electron density simulations from the Specified Dynamics–Whole Atmosphere Community Climate Model with ionosphere/thermosphere eXtension. The approximate resolution of this model is around ~200 km longitude by ~200 km latitude by a few kilometers in altitude. Hence, there are no fine-scale ionospheric phenomena that produce very sharp gradients like Sporadic-E. This is important to note because the algorithm itself was never designed for these finer-scale phenomena [13].

This paper first quantifies the total errors, under ideal conditions, in the retrieved E-region Ne profiles $N_{e,retr}(z)$. The errors in $N_{e,retr}(z)$ is defined as the difference with its corresponding tangent point electron density profile $N_{e,tan}(z)$. Then, this paper quantifies the general sensitivity of these errors toward the simulated true E-region peak Ne and the peak F-region Ne. The sensitivity toward the true E-region peak Ne provides an assessment of the algorithm's sensitivity. The sensitivity toward the peak F-region Ne provides an assessment of the algorithm's method in removing F-region contributions. These will provide guidance on future improvements to the algorithm. We emphasize that this paper does not aim to quantify the different components of the total uncertainty in the simulated and, even more so, real RO data.

2. Data and Methods

In this section, we first describe the model outputs as well as data on spacecraft location. Then, we describe how we constructed first-order synthetic radio occultation excess phase profiles. We also describe the "bottom-up" retrieval algorithm and its current configuration. Finally, we describe our main metrics for quantifying the systemic errors of the "Bottom-Up" retrieval algorithm.

2.1. Model: Specified Dynamics–Whole Atmosphere Community Climate Model with Ionosphere/Thermosphere eXtension v2.1 (SD-WACCM-X)

This work utilizes electron density outputs from the Specified Dynamics–Whole Atmosphere Community Climate Model with ionosphere/thermosphere eXtension (SD-WACCMX) [17,18]. This model is a first-principles physics-based model that simulates the whole atmosphere from the surface to the ionosphere/thermosphere while accounting for the coupling of the atmosphere with the ocean, sea ice and land. SD-WACCM-X is nudged by the Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis dataset from the surface to around 50 km [19,20]. This ensures that the atmospheric temperatures and winds from the surface of the Earth to the stratosphere are realistic in all simulations. This work utilizes model outputs with a horizontal resolution of 1.9° in latitude and 2.5° in longitude and a vertical resolution of 2 points per scale height below ~50 km and 4 points per scale height above ~50 km. This corresponds to roughly 2 to 3 km vertical resolution in the E-region. The model outputs also have 3-hourly temporal resolution. This work only utilized model outputs for 21 March 2022 and 21 June 2022. These specific months were chosen because, being representatives of equinox and solstice seasons ensures they adequately include a wide range of E-region electron density variabilities.

2.2. Spire Radio Occultation Spacecraft Location Data

This work utilizes location data on GNSS satellite locations and RO receivers from the Spire RO constellation. Spire RO is a constellation of around 100 CubeSats known as Low-Earth Multi-Use Receivers (LEMURs) [21]. Their orbital inclinations include equatorial, sun-synchronous, 51.6-degree, 83/85-degree and 37-degree orbits. Their orbital altitudes range from 400 km to 600 km. Between 2022 and 2023, Spire RO provided ~15,000 occultations

throughout the globe in one day. The massive amount of occultations enables complete global local-time sampling in just 3 universal time hours. Hence, we did not need to use multiple days of data to cover all local times. To match the model outputs, this work only utilized location data for 21 March 2022 and 21 June 2022.

2.3. First-Order Synthetic Radio Occultation Excess Phase Profiles

To produce first-order synthetic radio occultation excess phase profiles (hereafter just referred to as "synthetic" RO profiles), we need to produce hTEC profiles as a function of tangent height from a 3D model of E-region electron density as well as the coordinates of the LEO receiver satellite, the GNSS satellite and the tangent point. Given the coordinates of the LEO satellite, GNSS satellite and tangent point, we first calculate the longitude, latitude and altitude along a GNSS signal's ray path from the GNSS satellite to the tangent point and finally to the LEO receiver. The coordinates of the GNSS satellite and the LEO receiver are provided by the level 1 files.

These coordinates are used to construct the LEO receiver location vector $\vec{r_L}$, the GNSS satellite location vector $\vec{r_G}$ and the tangent point location vector $\vec{r_T}$. Figure 1 shows an illustration of these important location vectors as well as other important vectors for the calculations. The steps to obtain the coordinates along the ray path are:

1. Convert $\vec{r_L}$, $\vec{r_G}$ and $\vec{r_T}$ into ECEF coordinates.
2. Subtract the position vector of the tangent point and the GNSS satellite: $\vec{r_T} - \vec{r_G}$.
3. Obtain the length of $\vec{r_G} - \vec{r_T}$ which is $L_{TG} = |\vec{r_T} - \vec{r_G}|$.
4. Divide this length by the step size along the line of sight, Δs. We use $\Delta s = 10$ km. The total number of steps between $\vec{r_T}$ and $\vec{r_G}$ is $\frac{L_{GT}}{\Delta s}$.
5. Position vectors along the line of sight, which are then given by $\vec{r}_{GT} = \vec{r_G} + (k\Delta s)(\vec{r_T} - \vec{r_G})$ where $k = 1, 2, \ldots, \frac{L_{TG}}{\Delta s}$.
6. Convert position vectors into longitude, latitude and height above sea-level coordinates.
7. To obtain points between the tangent point and LEO, repeat 2 to 5 but for $\vec{r_L} - \vec{r_T}$.

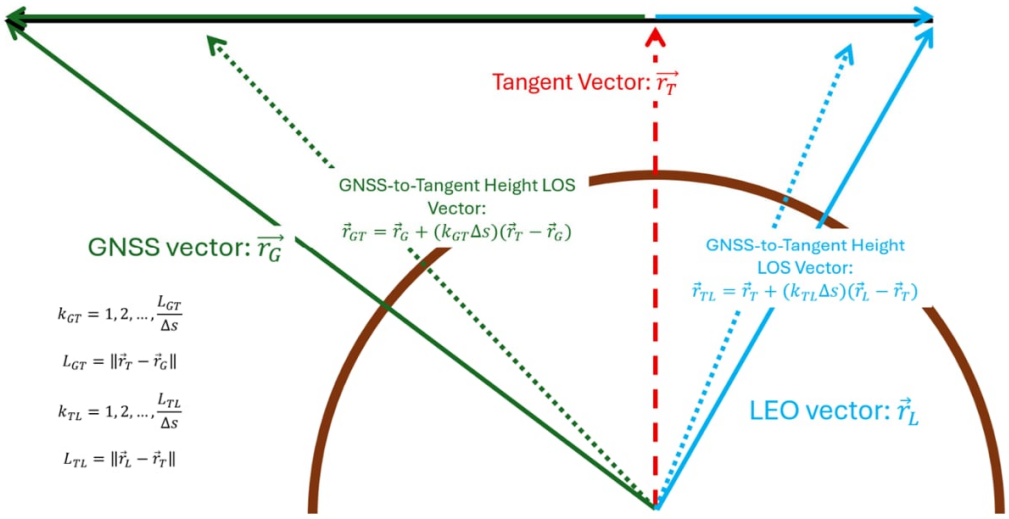

Figure 1. Location vectors of the GNSS satellite $\vec{r}_G$, LEO receiver $\vec{r}_L$ and tangent point $\vec{r}_T$. See main text for more information.

Figure 2a shows an illustration of calculated coordinates along the ray path in ECEF coordinates. In ECEF coordinates, this manifests as a straight line from the GNSS satellite to the tangent point and from the tangent point to the LEO receiver. Figure 2b–d show illustrations of calculated coordinates along the ray path in longitude–latitude–altitude

coordinates. In this coordinate system, the ray path manifests as a parabolic curve from the GNSS satellite to the tangent point and from the tangent point to the LEO receiver.

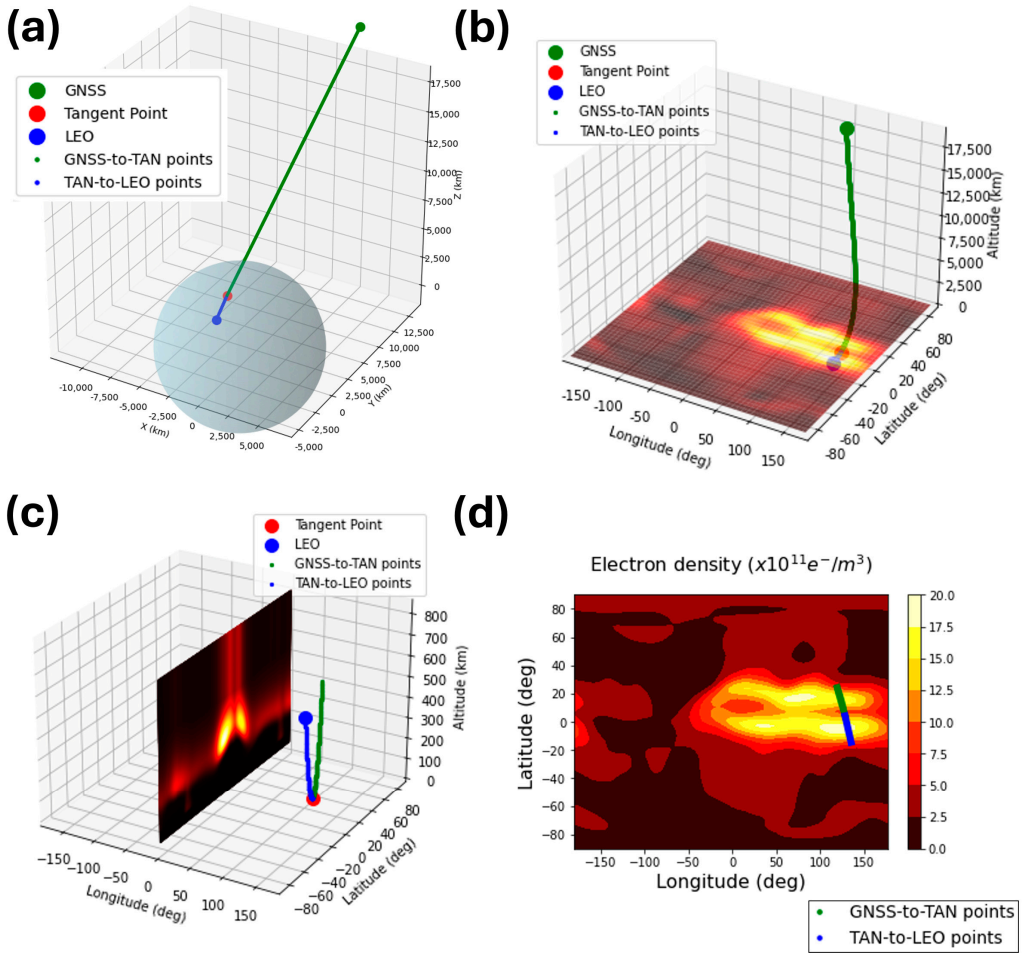

Figure 2. Subplot (**a**) is an illustration showing the calculated coordinates as green (GNSS to tangent point) and blue (tangent point to LEO receiver satellite) points in the ECEF coordinate system. Subplots (**b**–**d**) are illustrations showing the calculated coordinates as green (GNSS to tangent point) and blue (tangent point to LEO receiver satellite) points in the longitude–latitude–altitude coordinate system but at varying perspectives. The translucent sphere roughly represents planet Earth. The large green, red and blue dots indicate geographical coordinates of the GNSS satellite, tangent-point and LEO receiver satellite, respectively.

Given these coordinates along the ray path, we simply do linear interpolation on our 3D E-region Ne model from SD-WACCM-X. This provides us with the vertical Ne profiles $N_{e,GNSS}(z)$ and $N_{e,LEO}(z)$ constructed using the ray path coordinates from the GNSS satellite to the tangent point as well as using the ray path coordinates from the tangent point to the LEO receiver. Then, $hTEC(h_T)$ is calculated using the formula:

$$hTEC(h_T) = \int_{tan}^{GNSS} N_{e,GNSS}(z) \frac{zdz}{\sqrt{(R_E+z)^2 - (R_E+h_T)^2}} + \int_{tan}^{LEO} N_{e,LEO}(z) \frac{zdz}{\sqrt{(R_E+z)^2 - (R_E+h_T)^2}} \quad (4)$$

Note though, that although $hTEC(h_T)$ is constructed using $N_{e,GNSS}(z)$ and $N_{e,LEO}(z)$, the "truth" electron density corresponding to this $hTEC(h_T)$ profile is $N_{e,TAN}$ which is the electron density profile directly over the longitude and latitude coordinates specified by $\vec{r_T}$. This is the RO assumption of "spherical symmetry." Finally, since the "bottom-up" retrieval algorithm requires both L1 and L2 excess phase profiles, these $hTEC(h_T)$ profiles are con-

verted into L1 and L2 excess phase profiles using the formulae: $\phi_{L1}(h_T) = \frac{40.3\times10^{16}}{f_1^2} hTEC(h_T)$ and $\phi_{L2}(h_T) = \frac{40.3\times10^{16}}{f_2^2} hTEC(h_T)$ where $f_1 = 1.57542 \times 10^9$ Hz and $f_2 = 1.2276 \times 10^9$ Hz.

Bending of the signal path is not considered in this paper because previous studies have indicated that the differences from a straight-line scenario are only in the order of ~1 km [22]. As indicated earlier, SD-WACCM-X has a coarse resolution, so these bending impacts will not even be simulated. They will only increase computational load. Furthermore, this is considered a higher-order error which is not a focus of this paper.

2.4. "Bottom-Up" Retrieval Algorithm

Here, we provide a brief review of the "bottom-up" retrieval algorithm. Both the "Bottom-Up" retrieval algorithm and the "onion-peeling" method tap into how the total excess phase measured by an RO receiver is related to the electron density profile $N_e(z)$ through the equation:

$$\phi_{ex}(z) = \frac{40.3 \times 10^{16}}{f^2} hTEC = \frac{40.3 \times 10^{16}}{f^2}\left[\int N_e(z) \cdot ds\right] \quad (5)$$

Here, the integration is along the straight-line path ds of the signal. f is the signal frequency and z is a distance coordinate between the GNSS satellite and the LEO receiver. However, in the real atmosphere and ionosphere, the total excess phase of a GNSS signal as it propagates from the GNSS satellite to the LEO receiver is controlled by the phenomena of atmospheric $\phi_{bend,A}$ and ionospheric bending $\phi_{bend,I}(f_i)$ as well as ionospheric electron density values $\phi_p(f_i)$:

$$\phi_{ex}(f_i) = \phi_{bend,A} + \phi_{bend,I}(f_i) - \phi_p(f_i) \quad (6)$$

where f_i is the signal frequency. Atmospheric bending $\phi_{bend,A}$ does not substantially vary with frequency in the L-band. The signs of each component are meant to clearly signify that atmosphere and ionosphere bending are known to induce a phase delay $\phi_{bend,I}(f_i) > 0$ while simply propagating through plasma is known to induce a phase advance $\phi_p(f_i) < 0$ [13]. When $\phi < 0$, this is called phase advance because the path conditions modified the signal's phase such that the signal has 'advanced' ahead of the free-space signal. When $\phi > 0$, this is called phase delay because the path conditions modified the received signal's phase such that the signal was 'delayed.'

For the "Bottom-Up" method, we utilize the total excess phase of the L1 GNSS signal $\phi_{L1}(f_{L1}, h_T)$ and the L2 GNSS signal $\phi_{L2}(f_{L2}, h_T)$:

$$\phi_{L1}(f_{L1}, h_T) = \phi_{bend,A,L1}(h_T) + \phi_{bend,I,L1}(f_{L1}, h_T) - \phi_{p,L1}(f_{L1}, h_T) \quad (7)$$

$$\phi_{L2}(f_{L2}, h_T) = \phi_{bend,A,L2}(h_T) + \phi_{bend,I,L2}(f_{L1}, h_T) - \phi_{p,L2}(f_{L2}, h_T) \quad (8)$$

$f_{L1} = 1.57542$ GHz and $f_{L2} = 1.22760$ GHz. The trick to the "Bottom-Up" method lies in how to isolate the E-region components of $\phi_{p,L1}$ and $\phi_{p,L2}$.

First, to remove the atmospheric bending, we note that $\phi_{bend,A,L1}(h_T) = \phi_{bend,A,L2}(h_T)$ since bending is not a function of signal wavelength. Thus, subtracting both excess phases produces a new excess profile that only contains ionospheric bending and phase advance:

$$\phi_{L1L2}(f_{L1}, f_{L2}, h_T) = \phi_{bend,I,L1L2}(f_{L1}, f_{L2}, h_T) - \phi_{p,L1L2}(f_{L1}, f_{L2}, h_T) \quad (9)$$

The bending and phase advance terms can be further expanded into their F-region and E-region components:

$$\phi_{bend,I,L1L2}(f_{L1}, f_{L2}, h_T) = \phi_{bend,F,L1L2}(f_{L1}, f_{L2}, h_T) + \phi_{bend,E,L1L2}(f_{L1}, f_{L2}, h_T) \quad (10)$$

$$\phi_{p,L1L2}(f_{L1}, f_{L2}, h_T) = \phi_{p,F,L1L2}(f_{L1}, f_{L2}, h_T) + \phi_{p,E,L1L2}(f_{L1}, f_{L2}, h_T) \quad (11)$$

Owing to the short path length through the E-region, $\phi_{bend,E,L1L2}(f_{L1},f_{L2},h_T) \approx 0$. At this point, we only need to remove the contributions of $\phi_{bend,F,L1L2}(f_{L1},f_{L2},h_T) - \phi_{p,F,L1L2}(f_{L1},f_{L2},h_T)$ at E-region altitudes:

$$\phi_{L1L2}(f_{L1},f_{L2},h_T) = \phi_{bend,F,L1L2}(f_{L1},f_{L2},h_T) - \left[\phi_{p,F,L1L2}(f_{L1},f_{L2},h_T) + \phi_{p,E,L1L2}(f_{L1},f_{L2},h_T)\right] \quad (12)$$

Prior to removing them, we can already convert these excess phase profiles into horizontal total electron content (hTEC) using the formula: $hTEC_{L1L2}(h_T) = \frac{f_{L1}^2 f_{L2}^2}{40.3\times10^{16}(f_{L1}^2-f_{L2}^2)}\,\phi_{L1L2}(f_{L1},f_{L2},h_T)$.

To remove these remaining F-region contributions, the "Bottom-Up" method utilizes a key approximation of these contributions [13]: the F-region component of the bending and phase advances at these lower ionospheric regions is a linear function of tangent height. This linear function can be estimated by fitting $hTEC_{p,F,L1L2}$ values between 30 and 50 km with a linear function to yield coefficients a_1 and a_0. This linear function is then subtracted from $hTEC_{L1L2}(h_T)$ it to yield:

$$\Delta hTEC_E(h_T) = hTEC_{L1L2}(h_T) - [a_1 h_T + a_0] \quad (13)$$

This $\Delta hTEC_E$ can then be related to the E-region electron density profile using Equation (5):

$$\int N_e(z)\cdot ds = \Delta hTEC_E(h_T) \quad (14)$$

$N_e(z)$ is then solved using the optimal estimation inversion method. The measurement vector $\vec{y}$ is $\Delta hTEC_E(h_T)$. The state vector $\vec{x}$ is $N_e(z)$. The linearized forward model for the measurement vector is related to the state vector through the weighting function K as follows:

$$\vec{y} = \vec{y_0} + K\cdot(\vec{x} - \vec{x_0}) + \epsilon_y \quad (15)$$

The initial state vector $\vec{x_0}$ is a profile whose F-region comes from IRI and whose E-region comes from FIRI [13]. The optimal solution to this is:

$$\hat{x} = \left[S_a^{-1} + K^T S_y^{-1} K\right]^{-1}\left[S_a^{-1}\vec{a} + K^T S_y^{-1}\vec{y}\right] \quad (16)$$

$\vec{a}$ is the a priori of $\vec{x}$, which is set to be the same as $\vec{x_0}$. S_a and S_y are covariance matrices for $\vec{a}$ and $\vec{y}$, respectively. For more information on these inputs and how this was solved, see the primary retrieval algorithm paper [13].

2.5. Methods to Determine Errors, Identify Artifacts and Quantify Sensitivity

To quantify errors under our idealized conditions, this work first applies this bottom-up retrieval algorithm to synthetic $hTEC(h_T)$ profiles constructed using Equation (4). Then, the retrieved electron density profile $N_{e,retr}(z)$ is compared to the tangent point electron density profile $N_{e,tan}(z)$. Quantitative comparison is performed using statistical correlation and calculation of percent differences $100\% \times \frac{(N_{e,tan}(z)-N_{e,retr}(z))}{N_{e,retr}(z)}$.

To quantify the sensitivity of these errors toward the true E-region Ne values and true F-region conditions (marked by peak F-region Ne values), this work constructs a set of histograms. For sensitivity of the errors to the true E-region Ne values, we construct a histogram showing the distribution of errors with respect to the true E-region Ne values (hereafter referred to as "Error-Truth histogram"). For the sensitivity toward F-region conditions, we construct a histogram showing the distribution of errors with respect to peak F-region Ne values (hereafter referred to as "Error-F-region histogram"). For each type of histogram, we construct one version of this histogram for errors less than 100% and another version for errors greater than or equal to 100%.

3. Results

In this section, we first present the differences between the retrieved electron density profiles $N_{e,retr}(z)$ and the tangent point electron density profiles $N_{e,tan}(z)$ which we will refer to as 'error'. We present the local-time, latitude and seasonal dependencies (March equinox and June solstice) of these differences or errors. Then, we present the differences between the tangent point electron density profiles $N_{e,tan}(z)$ and the electron density profiles $N_{EregionOnly}(z)$ retrieved from $hTEC(h_T)$ profiles that did not have an F-region (Equation (17)). We will also equivalently refer to these differences as "errors due to F-region contributions".

3.1. March Equinox Error Analysis

Figure 3 shows scatter plots between $N_{e,retr}(z)$ and $N_{e,tan}(z)$ at select altitudes that are under day-time conditions (local times between 6 am and 18 pm). The color of the markers denotes the order of magnitude of the maximum value (calculated by $\log_{10}[\max(\Delta hTEC_E(h_T)])$ of the $\Delta hTEC_E(h_T)$ profile that yielded the $N_{e,retr}(z)$ profile. Showing this maximum value indicates how proportional the $\Delta hTEC_E(h_T)$ values are to the retrieved $N_{e,retr}(z)$ values. Ideally, $\log_{10}[\max(\Delta hTEC_E(h_T)]$ should be proportional to $N_{e,retr}(z)$ because this means that our method for estimating $\Delta hTEC_E(h_T)$ is reliable. Figure 3 shows that correlation coefficients between $N_{e,retr}(z)$ and $N_{e,tan}(z)$ are at least 0.78, with the better correlations at 100 km (r = 0.894) and at 110 km (r = 0.854). It also demonstrates that the retrieved values consistently overestimate the tangent-point values. This is consistent with our past validation studies indicating that our retrievals overestimate E-region electron density measured by ground-based measurements. Our findings suggest that, while the retrieved values may exhibit variations consistent with the tangent point variations, the exact values are higher than the true values.

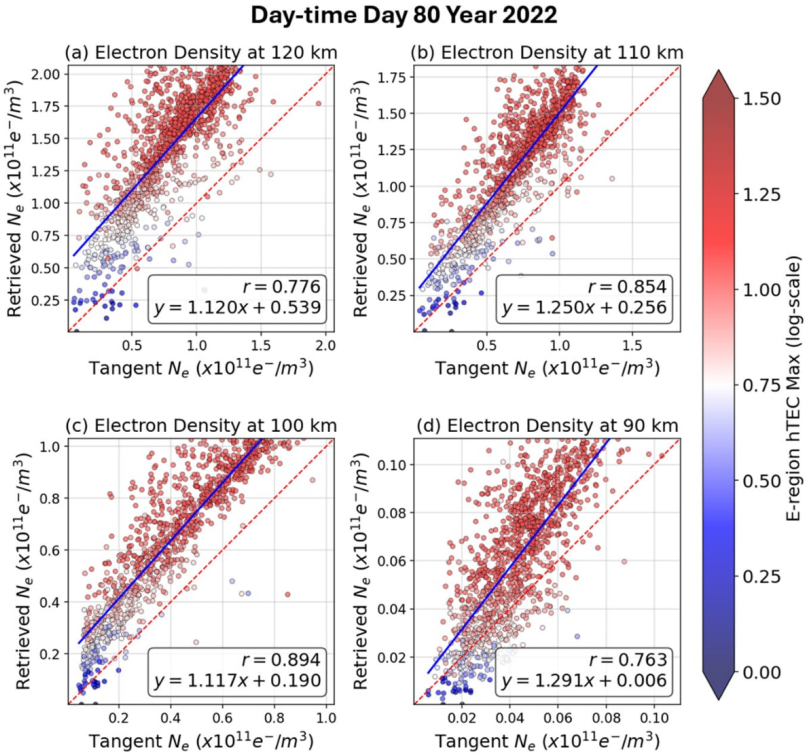

Figure 3. This figure shows scatterplots between all day-time (6 am to 18 pm) tangent-point and retrieved values at constant altitude for day-of-year 80. The color of the markers denotes the order of magnitude of the maximum value (calculated by $\log_{10}[\max(\Delta hTEC_E(h_T)])$ of the $\Delta hTEC_E(h_T)$

profile that yielded the $N_{e,retr}(z)$ profile. Showing this maximum value indicates how proportional the $\Delta hTEC_E(h_T)$ values are to the retrieved $N_{e,retr}(z)$ values. Subplot titles indicate the corresponding altitude of the scatterplot. Each subplot includes the correlation coefficients and linear regression fit. Each subplot also shows the linear fitted line (blue line) and the ideal line (red line) that would indicate that the retrieved values match the tangent-point values.

Figure 4 shows the scatter plots for night-time conditions (local times 18 pm to 6 am). The night-time figures demonstrate lower correlation coefficients than their day-time counterparts. However, a closer look shows that for altitudes of 100 km, 110 km and 120 km, when the tangent point values are greater than $0.1 \times 10^{11} \frac{e^-}{m^3}$, the correlation is better, though the uncertainties are still large. It is when the true values are less than $0.1 \times 10^{11} \frac{e^-}{m^3}$ that the retrieved values become very poorly correlated.

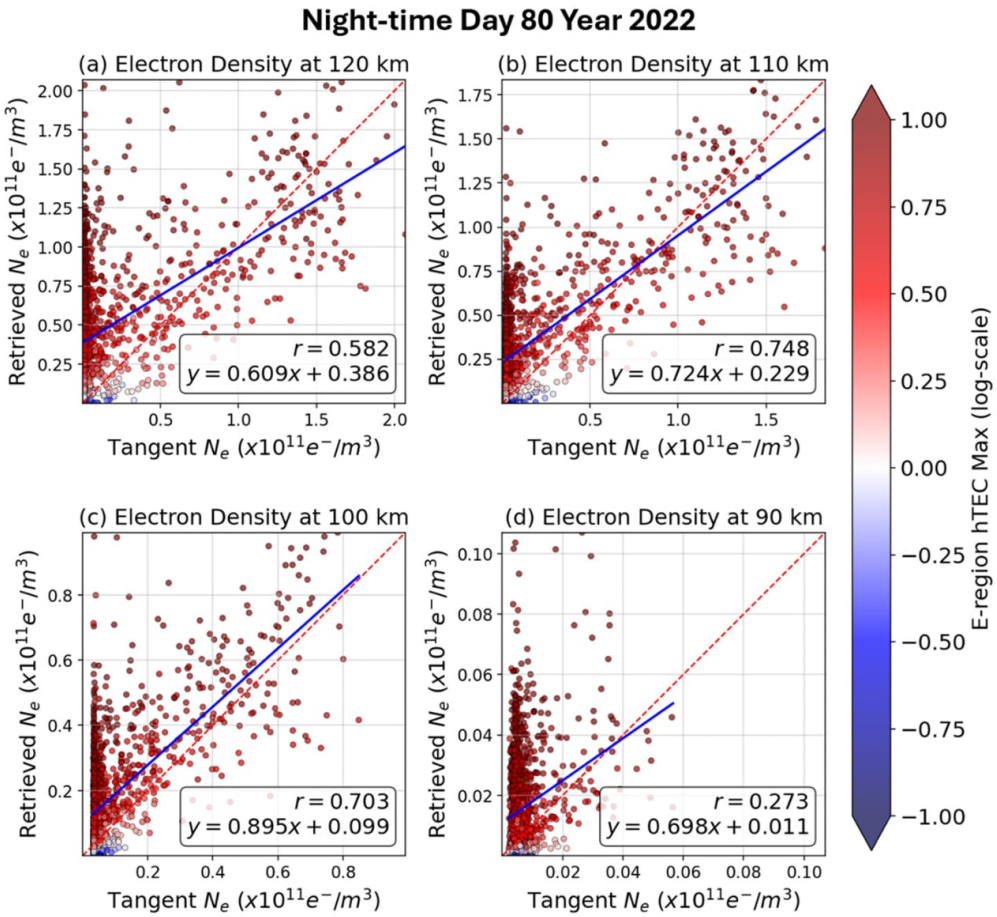

Figure 4. This figure shows scatterplots between all day-time (18 pm to 6 am) tangent-point and retrieved values at constant altitude for day-of-year 80. The color of the markers denotes the order of magnitude of the maximum value (calculated by $\log_{10}[\max(\Delta hTEC_E(h_T)])$ of the $\Delta hTEC_E(h_T)$ profile that yielded the $N_{e,retr}(z)$ profile. Showing this maximum value indicates how proportional the $\Delta hTEC_E(h_T)$ values are to the retrieved $N_{e,retr}(z)$ values. Subplot titles indicate the corresponding altitude of the scatterplot. Each subplot includes the correlation coefficients and linear regression fit. Each subplot also shows the linear fitted line (blue line) and the ideal line (red line) that would indicate that the retrieved values match the tangent-point values.

The marker colors in Figures 3 and 4 indicate that the maximum $\Delta hTEC_E(h_T)$ values are prominently at an order of magnitude close to 1. This includes night-time cases when the tangent point values are already less than $0.1 \times 10^{11} \frac{e^-}{m^3}$. Night-time tangent point values are at least an order of magnitude lower than day-time tangent point values. This indicates that the night-time $\Delta hTEC_E(h_T)$ values are not varying at a magnitude proportional to the

tangent point E-region electron density values. In addition, these figures show that the large night-time $\Delta hTEC_E(h_T)$ values are consistent with the inaccurate large $N_{e,retr}(z)$ values. Hence, the night-time inaccuracies may be tied to overestimated $\Delta hTEC_E(h_T)$ values.

To quantify the errors on average vertical profiles as a function of local time and latitude, Figure 5 shows the retrieved and tangent-point mean vertical profiles over 5 latitude bands split into day-time (Figure 5a–e) and night-time (Figure 5f–j). Note that for the night-time vertical profiles, the tangent-point profiles have been multiplied by 10 to help discern differences. The error bars on these vertical profiles are the standard deviation of all the available data points. Thus, together with the mean values, the profiles illustrate both the average values and the degree of variability around the mean. Complementary to this is Figure 6 which shows the calculations of the percent differences between the retrieved and tangent-point values as a zonal profile. For the day-time profiles, we can see that even with error bars accounted for, the retrieved profiles overestimate the tangent-point values. Figure 6a indicates that, above 100 km, retrieved values overestimate the tangent-point values by around 30% to 90%. Below 100 km, retrieved values are larger than tangent-point values by 100% over the middle to high latitudes. For the night-time profiles in Figure 5f–j and the night-time zonal profile in Figure 6b, we note that the retrieved profiles overestimate tangent-point values by at least 100%, but the disparity significantly drops over the high latitudes above 100 km. Figure 6b indicates that over the night-time high latitudes, the differences between the retrieved and tangent-point values are less than 25%. On the other hand, below 100 km, retrieved values over high latitudes are larger than tangent-point values by 100%. However, there is a narrow band below 100 km over latitude 60S-40S and latitude 40N-60N where their differences are relatively lower.

Looking closer, Figure 5f–j show that, unlike the night-time high latitudes and all day-time profiles, these tangent-point night-time profiles over the low to middle latitudes have peak E-region electron density values at around 100–105 km. The "bottom-up" retrieval algorithm currently uses an a priori profile that exponentially increases with altitude. Hence, it is currently tuned in such a way that the retrieved profiles will generally have peak E-region electron density values located above 110 km. Our results suggest that one of the reasons why the algorithm struggles at night is because it is currently tuned to only retrieve E-region electron density profiles whose vertical profiles have peak values above 110 km.

To quantify the errors on gridded latitude–local time profiles at constant altitude, Figure 7 shows latitude–local solar time profiles of retrieved and tangent-point electron density values at 90 km and 110 km as well as their percent differences. Figure 7c shows that the lowest day-time errors are located between ~9 am and ~15 pm over all latitudes where the retrieved values overestimate tangent-point values by around 50%. Figure 7c also demonstrates that the lowest night-time errors are located at high latitudes where the retrieved values still overestimate the tangent-point values but by less than 20%. In addition, Figure 7b shows a sustained enhancement in E-region Ne post-noon that extends into night. This has also been reported from the data [13,14] and is not apparent in Figure 7a.

While Figure 7c shows the lowest errors over the night-time high latitudes, Figure 7f indicates that the errors grow with decreasing altitude down to 90 km. This is consistent with Figure 6b, which specifically indicates that the performance begins to deteriorate at around 95 km. Figure 7f also indicates that unlike altitudes above 95 km, there are local-time and latitude bands where the retrieved values underestimate the tangent-point values. Retrieved values underestimate the tangent-point values the most by around 50% between midnight and 6 am over latitudes 60S to 40S and latitudes 40N to 60N. Looking at Figure 4d, we do see that there are data points whose order of magnitude is -1 (blue colored markers), and they are also data points indicating that they correspond to retrieved values that underestimate tangent-point values. Hence, these underestimated values may

be tied to cases when our $\Delta hTEC_E(h_T)$ are also lower than they should be. On the other hand, Figure 7c,f both indicate that the retrieved values at 90 km and 110 km between 6 am and 12 noon over the low latitudes overestimate tangent-point values by around 50%.

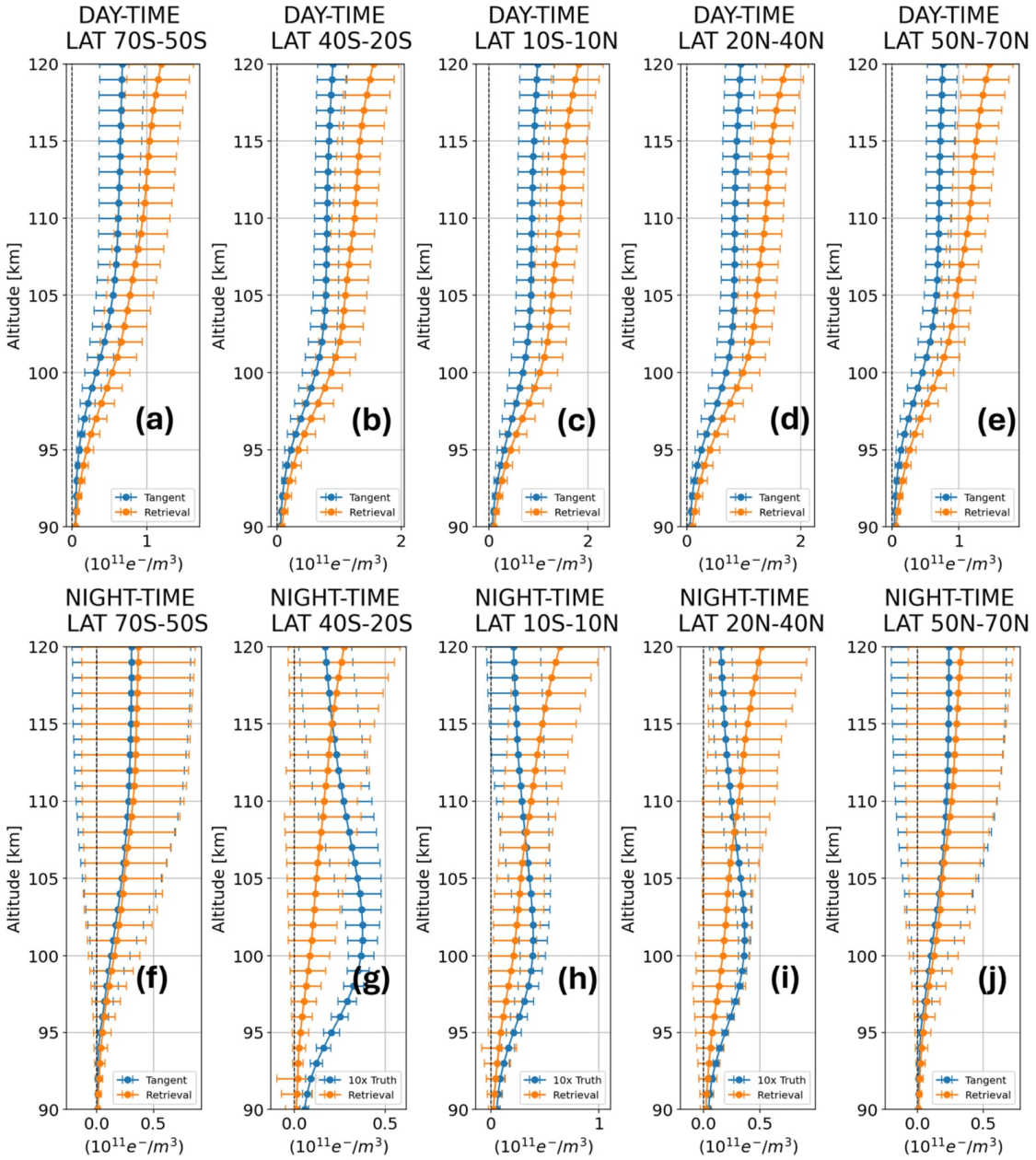

Figure 5. Each subplot shows the average of retrieved and tangent-point E-region Ne vertical profiles over a latitude band for day-of-year 80. The top panel shows day-time vertical profiles while the bottom panel shows night-time vertical profiles. The day-time vertical profiles are averaged over (**a**) latitudes 70S to 50S, (**b**) latitudes 40S to 20S, (**c**) latitudes 10S to 10N, (**d**) latitudes 20N to 40N and (**e**) latitudes 50N to 70N. Note that for subplots (**g**–**i**), the night-time truth values are multiplied by 10 to improve visibility. The night-time vertical profiles are averaged over (**f**) latitudes 70S to 50S, (**g**) latitudes 40S to 20S, (**h**) latitudes 10S to 10N, (**i**) latitudes 20N to 40N and (**j**) latitudes 50N to 70N. The error bars are the standard deviation. Subplot titles indicate the corresponding latitude band for the vertical profiles.

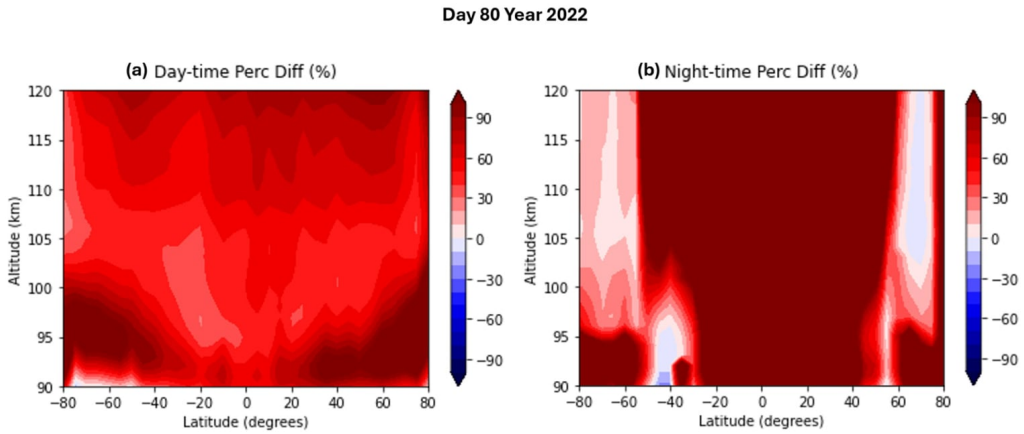

Figure 6. Zonal profile of percent difference between retrieved and tangent-point E-region Ne values for day-of-year 80 in year 2022 for (**a**) day-time and (**b**) night-time conditions. To clearly indicate locations with values within ±100%, all values beyond this range are capped at 100%. The color bars are capped at ±100%, as percent differences exceeding this threshold are collectively classified as extreme. Consequently, distinguishing between their exact magnitudes is not necessary.

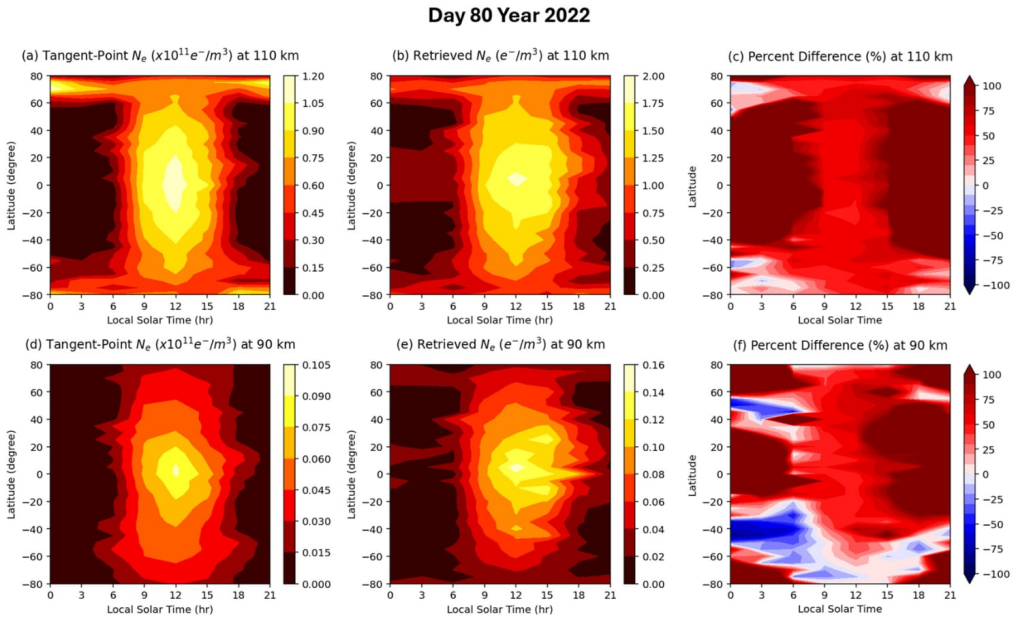

Figure 7. The top panel shows the latitude–local time profile of (**a**) tangent-point and (**b**) retrieved electron density values at 110 km, as well as their (**c**) percent differences for day-of-year 80 in year 2022. The bottom panel shows the same as the top panel but for values at 90 km. To clearly indicate locations with percent difference values within ±100%, all percent difference values beyond this range are capped at 100%. The color bars are capped at ±100%, as percent differences exceeding this threshold are collectively classified as extreme. Consequently, distinguishing between their exact magnitudes is not necessary.

3.2. June Solstice Error Analysis

Figure 8 shows scatter plots between $N_{e,retr}(z)$ and $N_{e,tan}(z)$ that are under day-time conditions (local times between 6 am and 18 pm) for the June solstice. For all altitudes, it indicates that correlation coefficients between the retrieved and tangent-point values are generally lower than the March equinox case in Figure 3. The maximum correlation is 0.79. This low correlation is evident in how large the spread of the values is in the plot. For altitudes of 100 km, 110 km and 120 km, it also shows a consistent overestimation of true

values. The marker colors though, add that day-time maximum $\Delta hTEC_E(h_T)$ values are consistently at an order of magnitude close to 1 even when the tangent-point values are less than $1.0 \times 10^{11} \frac{e^-}{m^3}$. This suggests that the inaccuracies may be tied to inaccuracies in $\Delta hTEC_E(h_T)$. At 90 km though, retrieved values underestimate tangent-point values greater than $0.1 \times 10^{11} \frac{e^-}{m^3}$. Note though that, unlike the March equinox conditions in Figure 3d, these day-time June solstice conditions at 90 km have Ne values up to $0.3 \times 10^{11} \frac{e^-}{m^3}$. Day-time March equinox tangent-point values at 90 km never went above $0.1 \times 10^{11} \frac{e^-}{m^3}$.

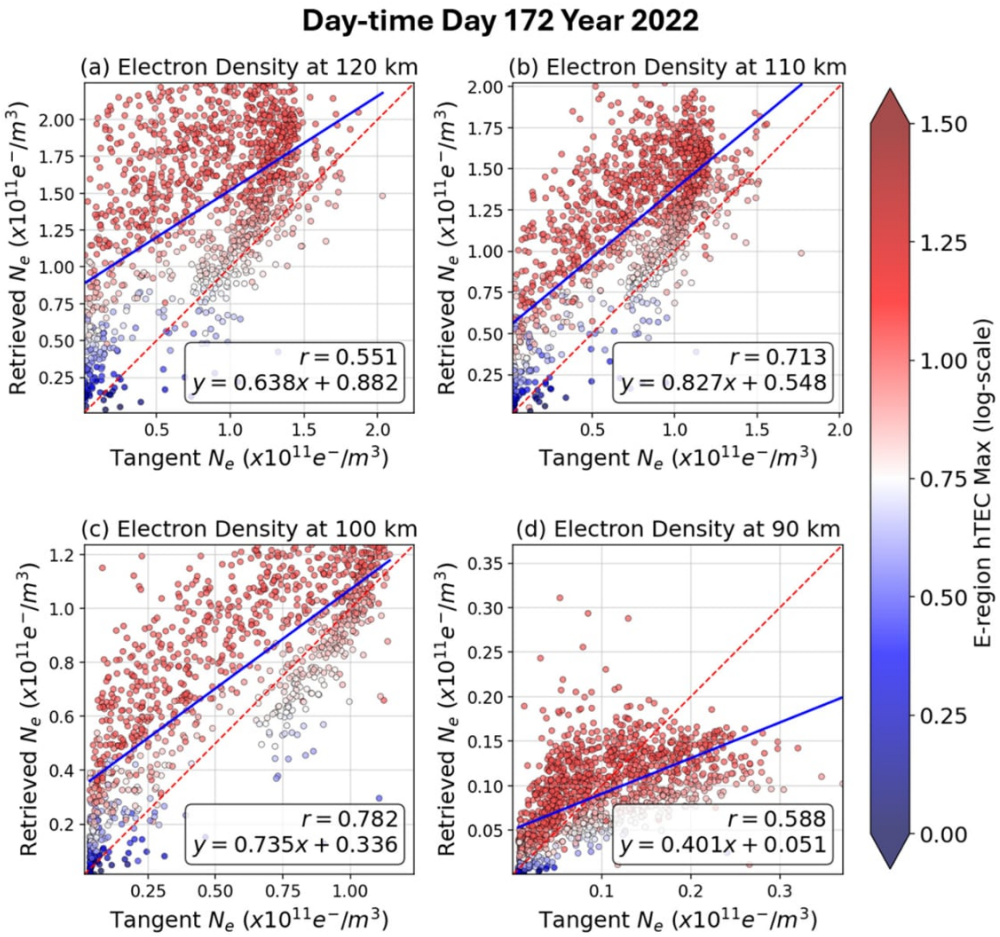

Figure 8. Scatterplots between all day-time (6 am to 18 pm) tangent-point and retrieved values at constant altitude for day-of-year 172. The color of the markers denotes the order of magnitude of the maximum value (calculated by $\log_{10}[\max(\Delta hTEC_E(h_T)])$ of the $\Delta hTEC_E(h_T)$ profile that yielded the $N_{e,retr}(z)$ profile. Showing this maximum value indicates how proportional the $\Delta hTEC_E(h_T)$ values are to the retrieved $N_{e,retr}(z)$ values. Subplot titles indicate the corresponding altitude of the scatterplot. Each subplot includes the correlation coefficient and linear regression fit. Each subplot also shows the linear fitted line (blue line) and the ideal line (red line) that would indicate that the retrieved values match the tangent-point values.

Figure 9 shows the scatter plots for night-time conditions (local times 18 pm to 6 am). Unlike the March equinox case, the correlation coefficients for these night-time conditions are larger than the day-time conditions at altitudes 100 km, 110 km and 120 km. On the other hand, similar to the March equinox case, when the tangent-point values are greater than $0.1 \times 10^{11} \frac{e^-}{m^3}$, the correlation between the retrieved values and tangent-point values is better. In addition, as in the March equinox case, the night-time maximum $\Delta hTEC_E(h_T)$ values in June solstice conditions are also still consistently at an order of magnitude close to 1 even for true values less than $0.1 \times 10^{11} \frac{e^-}{m^3}$. Thus, Figures 4 and 9 indicate that the night-time

$\Delta hTEC_E(h_T)$ values may not vary at a magnitude proportional to the true E-region electron density values.

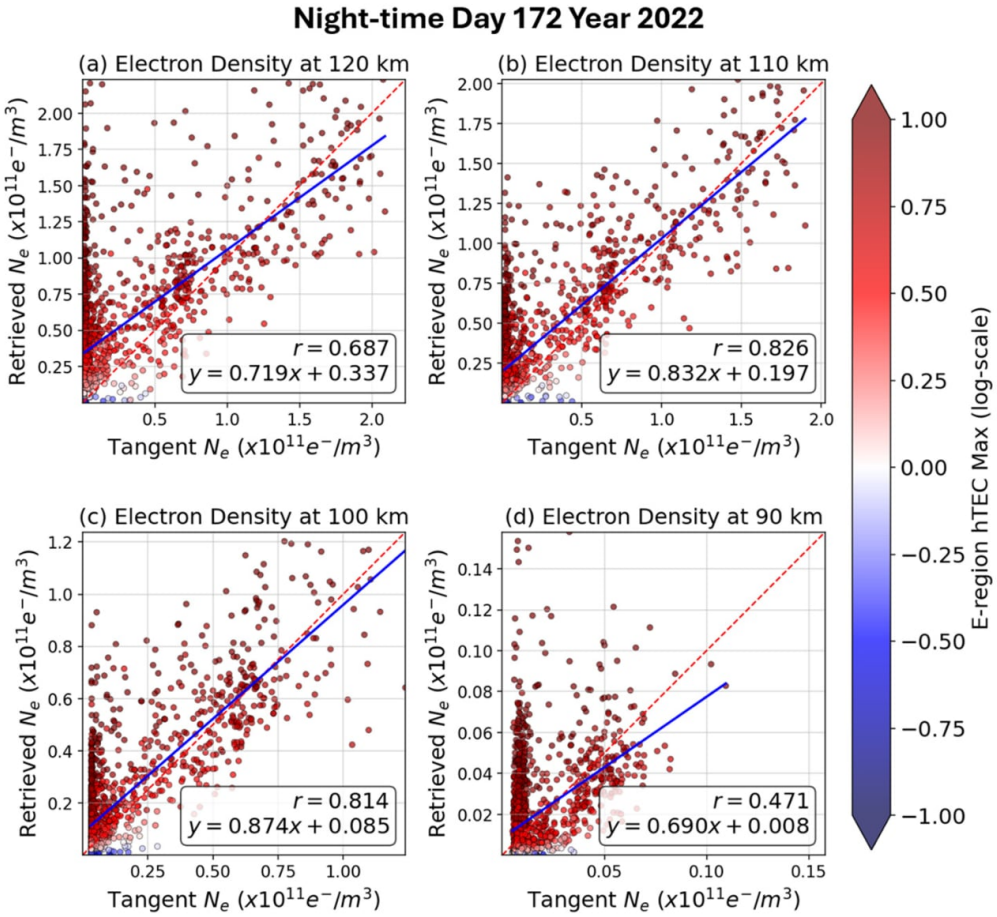

Figure 9. Scatterplots between all day-time (18 pm to 6 am) tangent-point and retrieved values at constant altitude for day-of-year 172. The color of the markers denotes the order of magnitude of the maximum value (calculated by $\log_{10}[\max(\Delta hTEC_E(h_T)])$ of the $\Delta hTEC_E(h_T)$ profile that yielded the $N_{e,retr}(z)$ profile. Showing this maximum value indicates how proportional the $\Delta hTEC_E(h_T)$ values are to the retrieved $N_{e,retr}(z)$ values. Subplot titles indicate the corresponding altitude of the scatterplot. Each subplot includes the correlation coefficient and linear regression fit. Each subplot also shows the linear fitted line (blue line) and the ideal line (red line) that would indicate that the retrieved values match the tangent-point values.

To quantify these errors on averaged vertical profiles, Figure 10 also indicates the retrieved and tangent-point mean vertical profiles but for day-of-year 172. Complementary to this is Figure 11 which shows the calculations of the percent differences between the truth and retrieved values as a zonal profile. Figure 10a–e illustrate that, even with error bars accounted for, day-time retrieved values above 100 km consistently overestimate tangent-point values over all latitudes, but the differences are lowest over the northern high latitudes coinciding with the summer high latitudes. Below 100 km, retrieved values also overestimate the tangent-point values except over the northern middle and high latitude bands where they now underestimate the tangent-point values. Figure 11a shows the very clear hemispheric and seasonal dependence of these errors. Above 100 km, retrieved values overestimate tangent-point values over the summer hemisphere by less than 50% which improves to less than 20% over the northern high latitudes. Below 100 km, retrieved values underestimate tangent-point values at around 20% over the middle latitudes but increase to around 50% over the high latitudes. On the other hand, retrieved values overestimate tangent-point

values between 90 km and 120 km throughout the southern hemisphere which also coincides with the winter hemisphere.

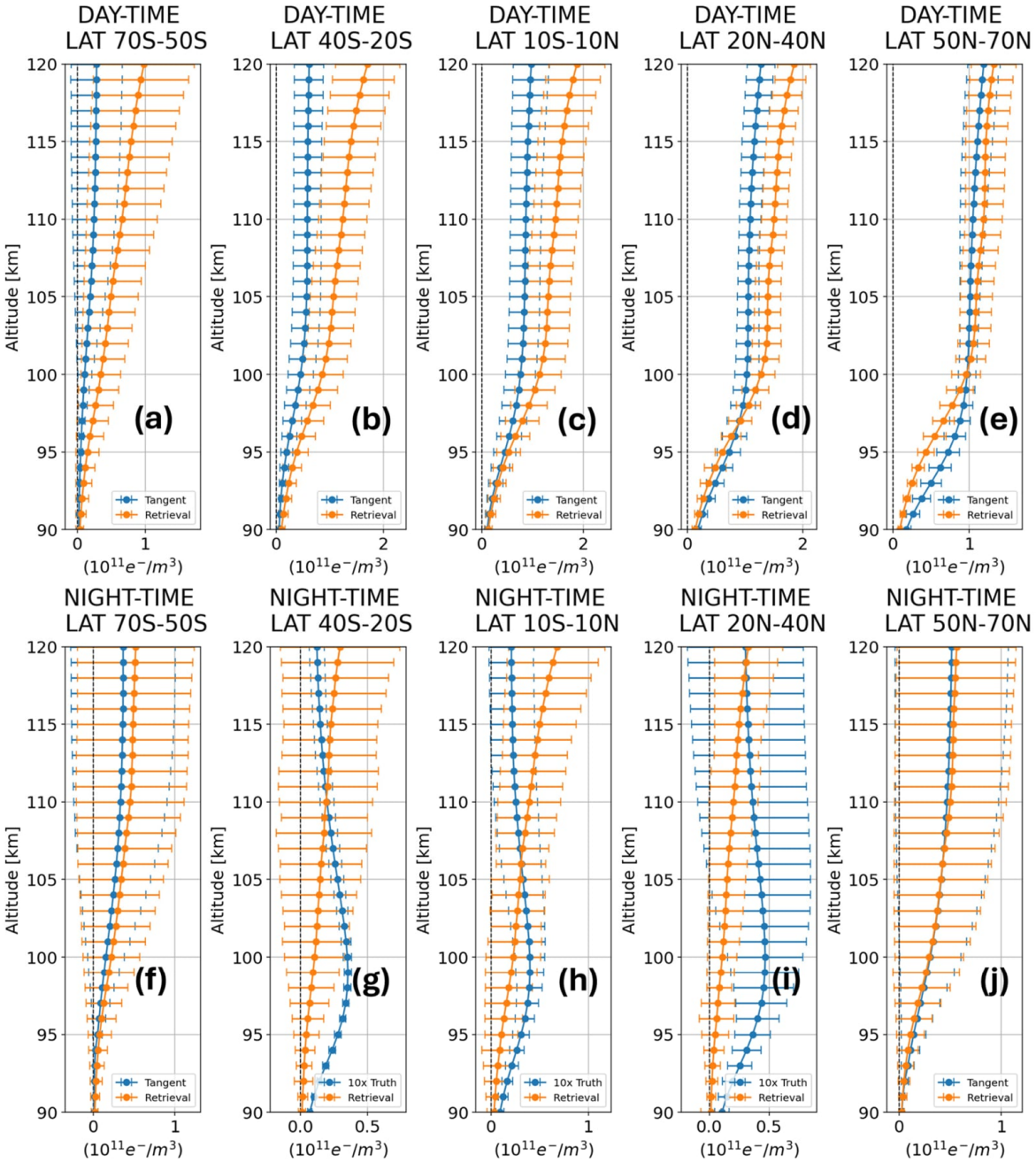

Figure 10. Each subplot shows the average of retrieved and tangent-point E-region Ne vertical profiles over a latitude band for day-of-year 172. The top panel shows day-time vertical profiles while the bottom panel shows night-time vertical profiles. The day-time vertical profiles are averaged over (**a**) latitudes 70S to 50S, (**b**) latitudes 40S to 20S, (**c**) latitudes 10S to 10N, (**d**) latitudes 20N to 40N and (**e**) latitudes 50N to 70N. Note that for subplots (**g**–**i**), the night-time truth values are multiplied by 10 to improve visibility. The night-time vertical profiles are averaged over (**f**) latitudes 70S to 50S, (**g**) latitudes 40S to 20S, (**h**) latitudes 10S to 10N, (**i**) latitudes 20N to 40N and (**j**) latitudes 50N to 70N. The error bars are the standard deviation. Subplot titles indicate the corresponding latitude band for the vertical profiles.

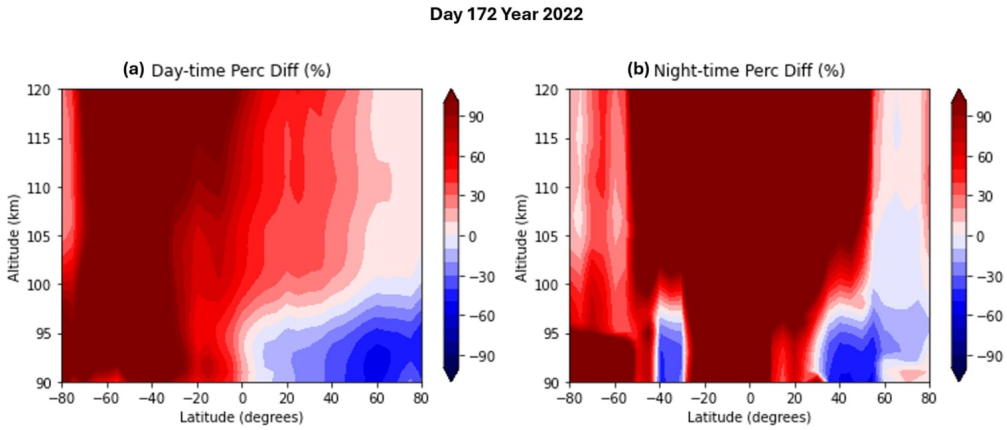

Figure 11. Zonal profile of percent difference between retrieved and tangent-point E-region Ne values for day-of-year 172 in year 2022 for (**a**) day-time and (**b**) night-time conditions. To clearly indicate locations with values within ±100%, all values beyond this range are capped at 100%. The color bars are capped at ±100%, as percent differences exceeding this threshold are collectively classified as extreme. Consequently, distinguishing between their exact magnitudes is not necessary.

For night-time profiles, Figure 10f–j show that the retrieved values overestimate the tangent-point values everywhere except over both high latitudes. Figure 11b indicates that the errors over both high latitudes are less than 50%, with errors over the northern high latitudes even lower at less than 20%. Figure 10g–i show that, for the June solstice, these tangent-point night-time profiles over the low to middle latitudes also have peak E-region electron density values at around 100–105 km. Below 100 km, retrieved values underestimate tangent-point values at around 20% over the northern high latitudes but increase to around 50% over the mid latitudes. On the other hand, retrieved values over the summer high latitudes substantially overestimate tangent-point values. More on this when we discuss Figure 12.

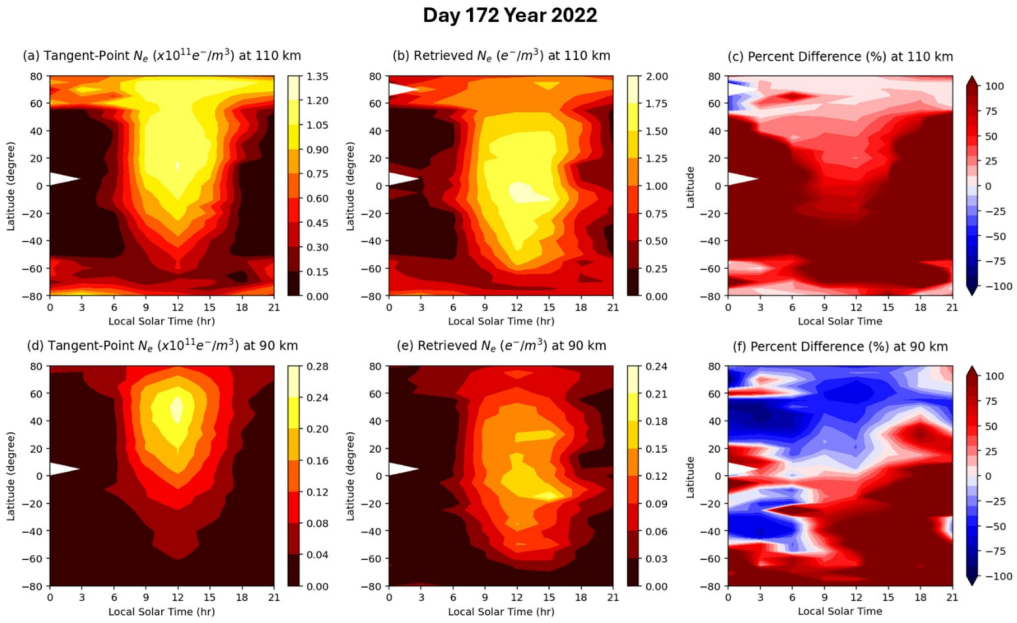

Figure 12. The top panel shows the latitude–local time profile of (**a**) tangent-point and (**b**) retrieved electron density values at 110 km as well as their (**c**) percent differences for day-of-year 172 in year 2022. The bottom panel shows the same as the top panel but for values at 90 km. To clearly indicate

locations with percent difference values within ±100%, all percent difference values beyond this range are capped at 100%. The color bars are capped at ±100%, as percent differences exceeding this threshold are collectively classified as extreme. Consequently, distinguishing between their exact magnitudes is not necessary.

To quantify the errors on gridded latitude–local time profiles at constant altitude, Figure 12 shows latitude–local solar time profiles of retrieved and tangent-point electron density values at 90 km and 110 km as well as their percent differences for day-of-year 172. Figure 12c shows that, at 110 km, the lowest errors are less than 10%, and they are located over the northern high latitudes and for all local times as well as over the southern night-time high latitudes. Over the day-time low to middle latitudes, errors increase to about 50%. Elsewhere, errors are greater than 100%.

At 90 km, Figure 12f shows that the main characteristic of the errors is that retrieved values over the summer hemisphere generally underestimate tangent-point values, while retrieved values over the winter hemisphere generally overestimate tangent-point values. This means that the retrieved values are generally underestimated (overestimated) over the summer (winter) hemisphere. The lowest errors are over the day-time middle to high latitudes where retrieved values underestimate tangent-point values by around 50%. Figure 12f also clarifies Figure 11b. It indicates that the larger absolute error over the mid latitudes is because of the presence of a relatively larger error between midnight and 6 am local time over latitudes 30N and 50N. At the same time, during the same local time but over the high latitudes, retrieved values overestimate the tangent-point values by around 20%.

3.3. Sensitivity to True E-Region Ne and Peak F-Region Ne Values

In this section, we now assess the general sensitivity of these errors toward the simulated true E-region peak Ne and the peak F-region Ne. The sensitivity toward the simulated true E-region peak Ne is important because it provides an assessment of the algorithm's current sensitivity, albeit under ideal conditions. The sensitivity toward peak F-region Ne is important because it provides an assessment of the accuracy of how the algorithm removes F-region contributions. To provide a generalized statistical assessment of the sensitivity of these errors toward the true E-region Ne values and F-region conditions, this section presents the "Error-Truth" histograms and the "Error-F-region" histograms.

Figure 13 shows these histograms. Errors in these histograms correspond to the values found at an altitude of 110 km and include all values over all locations and local times for the March equinox and June solstice. The other altitudes exhibit similar behavior, so we only show this single altitude. Figure 13a,b indicate that errors less than 100% generally have true E-region values greater than $0.1 \times 10^{11} \frac{e^-}{m^3}$. This is consistent with Figures 3, 4, 8 and 9 which suggest the algorithm has good sensitivity when true values are greater than $0.2 \times 10^{11} \frac{e^-}{m^3}$. They also indicate that the errors are inversely proportional to the true E-region Ne values. Figure 13a indicates that errors less than 50% coincide with true E-region Ne values greater than $1 \times 10^{11} \frac{e^-}{m^3}$.

Figure 13c,d indicate that, unlike Figure 13a,b, there is not a clear demarcation between errors greater than 100% and errors less than 100% when looking at the peak F-region Ne values. Both sets of errors can fall under a similar range of peak F-region Ne values. While Figure 13c does indicate that for errors less than 100%, there is a clear direct proportionality between the errors and peak F-region Ne values. Figure 13c indicates that locations where errors are around 50% or less coincide with peak F-region Ne values less than $5 \times 10^{11} \frac{e^-}{m^3}$.

These generalized statistics are consistent with how the algorithm performs best over the night-time high latitudes. Over these latitudes, Figures 7 and 12 indicate true E-region Ne values greater than $1 \times 10^{11} \frac{e^-}{m^3}$. To show the exact peak F-region electron density values over these areas of low or high errors, Figure 14 shows the day-time and night-time zonal

profiles of F-region electron density. Figure 14c,d indicates that peak F-region Ne values over these night-time locations are less than $5 \times 10^{11} \frac{e^-}{m^3}$. Accounting for the error bars in Figure 13a,c, this would fall under the histogram bins of Figure 13a on the right-most side where the range of E-region Ne values is greater than $1 \times 10^{11} \frac{e^-}{m^3}$. This would also fall under the histogram bins of Figure 13c on the left-most side where peak F-region Ne values are less than $5 \times 10^{11} \frac{e^-}{m^3}$. The statistics also help explain the tendency for the errors to be lower over the summer night-time high latitudes than over the winter night-time high latitudes. Figure 14c,d indicate that although the F-region values over both hemispheres are low, the summer E-region high latitudes have higher E-region Ne values (greater than $1 \times 10^{11} \frac{e^-}{m^3}$) than the winter E-region. On the other hand, these statistics are also consistent with how the algorithm performs worse over the low- to middle latitudes at night-time. True E-region Ne values there are orders of magnitude lower than $0.1 \times 10^{11} \frac{e^-}{m^3}$.

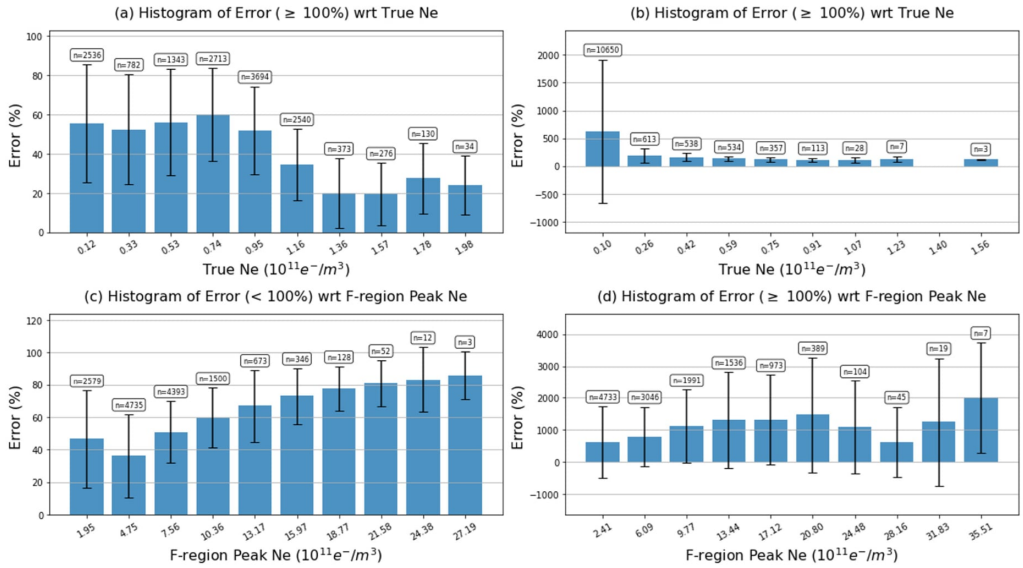

Figure 13. The top panels show a histogram of error (in percent difference), while the bottom panels show a histogram of peak E-region Ne density. Both histograms are with respect to the logarithm of peak F-region Ne density. The left column subplots only include percent errors less than 100%, while the right column subplots include percent errors greater than or equal to 100%. The error-bars denote the standard deviation of the values.

These generalized statistics also help explain the peculiar features only observed in the RO data; that is, the sustained enhancement in low-latitude E-region Ne during afternoon hours that even extends into evening hours, and overlaid on this enhancement is the presence of an equatorial trough. This specific feature is reproduced in Figures 7b and 12b. First, since the true values in Figures 7a and 12a do not show this feature, it already suggests that this feature is an artifact. Figure 14e,f add that this peculiar feature appears similar to the afternoon-to-evening features in the peak F-region Ne values. Over these areas, peak F-region Ne values are greater than $12 \times 10^{11} \frac{e^-}{m^3}$ while Figures 7a and 12a indicate that, over these areas, true E-region Ne values are less than $0.5 \times 10^{11} \frac{e^-}{m^3}$. The appearance of this feature in Figures 7b and 12b would be consistent with how the retrieval struggles when peak F-region Ne values are relatively high while true E-region Ne values are low. Overall, these suggest that the artifact may come from inaccuracies in our method to remove F-region contributions.

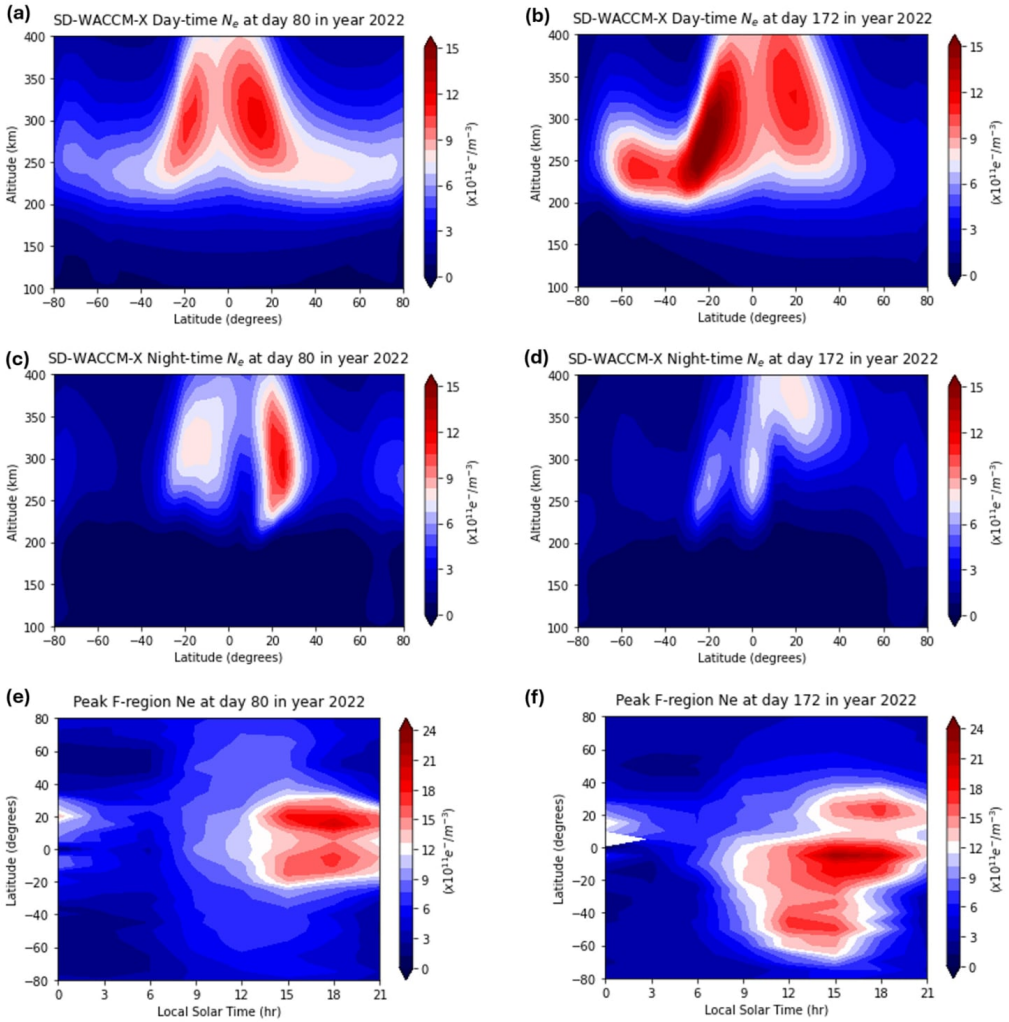

Figure 14. Latitude–altitude profiles of day-time (top panel) and night-time (middle panel) E-region to F-region electron density as well as latitude–local time profiles of peak F-region Ne values as simulated by SD-WACCM-X. The left column shows the profiles for day 80, while the right column shows the profiles for day 172.

4. Discussion

An ideal way to assess the performance of the "bottom-up" retrieval algorithm (or any other RO retrieval algorithm) would require an instrument to measure the electron density values along a GNSS signal's ray path from the GNSS satellite to the tangent point and then finally to the LEO satellite receiver. This can specifically determine how well the retrieved E-region electron density $N_{e,retr}(z)$ vertical profile compares to the E-region electron density profile located over an RO ray path's tangent point coordinate $N_{e,tan}(z)$. This instrument does not exist yet. This paper performs an alternative analysis that mimics this ideal observational setup but purely uses model simulations of electron density as well as information on the orbital locations of the GNSS satellite, the tangent point and the LEO receiver.

This paper first reported the local-time, latitude and seasonal dependencies behind the differences between $N_{e,retr}(z)$ and $N_{e,tan}(z)$ from 90 km and from 120 km. Over the low to middle latitudes, the local-time variations in their differences indicated that the $N_{e,retr}(z)$ generally overestimates $N_{e,tan}(z)$ and that the errors were lowest (less than 100%) around noontime, while the seasonal variations indicated that the errors are lowest over the

summer hemisphere. It is also revealed that the $N_{e,retr}(z)$ exhibits a sustained enhancement in low-latitude E-region Ne during afternoon hours that even extends into evening hours, and overlaid on this enhancement is the presence of an equatorial trough. This peculiar feature is not present in $N_{e,tan}(z)$ but has been observed in the RO data. On the other hand, over the high latitudes, the local-time variations in their differences reveal that the lowest errors are at night-time. Additionally, these low errors over the high latitudes persist regardless of season. Overall, the lowest differences between $N_{e,retr}(z)$ and $N_{e,tan}(z)$ (less than 25%) are located over the night-time high latitudes above 100 km, while the largest differences (larger than 100%) are located over the night-time low and middle latitudes.

Figures 3, 4, 8, 9 and 14 provide a generalization of this behavior. Figures 3, 4, 8 and 9 suggest that there is good correlation (greater than 0.7) when true values are at least $0.1 \times 10^{11} \frac{e^-}{m^3}$. Figure 14 further indicates that, regardless of location, local time and season, the demarcation between profiles with errors greater than 100% and those with errors less than 100% is a true E-region Ne value of around $0.1 \times 10^{11} \frac{e^-}{m^3}$. In addition, within these two categories, the errors become sensitive to peak F-region Ne values; that is, the errors are directly proportional to peak F-region Ne values. When true E-region Ne values are at least $1 \times 10^{11} \frac{e^-}{m^3}$ and peak F-region Ne values are lowest, errors can be less than 25%. On the other hand, when values are less than $0.1 \times 10^{11} \frac{e^-}{m^3}$, errors are extreme, and they can even form artifacts. It was found that the peculiar feature (sustained enhancement in low-latitude E-region Ne during afternoon hours that even extends into evening hours) is an artifact of these errors that stem from inadequacies in our method to remove F-region contributions.

Probing the night-time errors over low to middle latitudes further, this paper reveals that the vertical structure of night-time electron density over low to middle latitudes is different from the vertical structure of day-time electron density and even the night-time high latitudes. The peak night-time E-region electron density is located below 110 km, while it is located above 110 km during the day. It is possible for us to tune the algorithm through its covariance matrix such that the retrieved electron density profiles can also reflect this variation in vertical structure. However, this also assumes we believe the SD-WACCM-X simulation of the night-time vertical structure. We need more observations, preferably from high vertical resolution incoherent scatter radar, to prove that the night-time E-region electron density does indeed have this kind of vertical structure. Overall, until ground-based observations provide more support, these indicate that the retrieval algorithm will still have very large uncertainty over the night-time low to middle latitudes.

5. Conclusions and Future Work

These results showed that, under ideal conditions when $hTEC_E(h_T)$ is only a function of electron density along the signal path and $X^{'}(h_T) \approx 0$, the retrieved E-region Ne profiles have a correlation of at least 0.7 and errors less than 100% when true E-region Ne values are greater than $0.1 \times 10^{11} \frac{e^-}{m^3}$. Based on these results, we conclude that, under ideal conditions, the current "Bottom-Up" retrieval algorithm may retrieve E-region Ne profiles that have the highest correlation and lowest errors when true E-region Ne is relatively high while peak F-region Ne is relatively low, as is the case over the night-time high latitudes above 100 km. We also conclude that the method to remove F-region contributions still induces significant error and even forms artifacts in the retrievals when true E-region Ne values are relatively low.

Translating these into a guide for the real RO data is not straightforward. For now, the best we can do is refer to a validation paper that solely used observations and assess whether our findings in this paper are consistent with the validation paper. The validation paper did show that, under day-time conditions, the empirical model E-PROBED that utilizes RO data does have good correlation with ground-based ionosonde data when it comes to local

time (within 6–18 pm), seasonal and solar cycle variability. They also indicate that the data has a tendency to overestimate observations, but the errors are no more than 70%, with the highest values found during dusk hours. This 70% cap on error is indeed within our 100% cap in this paper. Likewise, the high correlation is consistent with our correlation analysis in this paper.

As for night-time validation, we refer to a previous paper [23] reporting on RO observations of night-time high latitudes. This paper has already shown that the RO data does observe peak enhancements over these locations during geomagnetic storms. The data indicate that E-region Ne at 110 km reached values greater than 3.0 $\times 10^{11} \frac{e^-}{m^3}$ and the spatial pattern shows how the enhanced areas remarkably follow the high oval. Retrieved values greater than 3.0 $\times 10^{11} \frac{e^-}{m^3}$ are far greater than our minimum 0.1 $\times 10^{11} \frac{e^-}{m^3}$ threshold in this paper. While we do not have adequate ground-based data to validate the exact accuracy of these values, the fact that the enhancements reach this magnitude while also following the high oval would be consistent with how our results in this paper indicate that, for true values greater than 0.1 $\times 10^{11} \frac{e^-}{m^3}$, retrieved values can be highly correlated with variabilities.

As for the suggested artificial structure over the low-latitude E-region that appeared to extend from afternoon to evening, note that the RO data indicated that this feature comes with E-region Ne values around 1.0 $\times 10^{11} \frac{e^-}{m^3}$. If this were not artificial, numerous low-latitude ionosonde data would have picked this up because, when converted to frequency, this would be within the range of ionosonde sensitivity. Since no ionosonde data has ever observed this, it would support our results in this paper, suggesting that this feature is an artifact.

Future work will aim to improve this retrieval algorithm by first determining the general vertical structure of night-time E-region electron density over low to middle latitudes. This will allow us to tune the retrieval algorithm such that it can yield retrieved vertical profiles with the appropriate night-time vertical structure. We plan to propose radar campaigns for this. Also, this work only analyzed first-order synthetic RO profiles. This hinders us from accurately translating our calculations into real data. Future work should redo this analysis using a more comprehensive RO model that introduces higher-order variabilities in these RO profiles.

Author Contributions: Conceptualization, C.C.J.H.S.; methodology, C.C.J.H.S. and D.L.W.; software, C.C.J.H.S. and D.L.W.; validation, C.C.J.H.S. and D.L.W.; formal analysis, C.C.J.H.S., M.S., D.L.W., N.S. and D.J.E.; investigation, C.C.J.H.S., M.S. and D.L.W.; resources, C.C.J.H.S., D.L.W. and L.Q.; data curation, C.C.J.H.S., D.L.W. and L.Q.; writing—original draft preparation, C.C.J.H.S.; writing—review and editing, C.C.J.H.S., D.L.W., M.S., L.Q., N.S., D.J.E., Y.Y. and J.C.; visualization, C.C.J.H.S. and D.L.W.; supervision, D.L.W., L.Q., D.J.E., Y.Y. and J.C.; project administration, D.L.W.; funding acquisition, C.C.J.H.S., D.L.W. and L.Q. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by NASA's Sun-Climate research project at GSFC (WBS 509496.02.03.01.17.04), by Living With a Star (LWS) program (WBS 936723.02.01.12.48), by NASA grant 80NSSC24K1114, by NASA grants 80NSSC20K1350 and 80NSSC20K0189, by a UMBC Strategic Awards for Research Transitions (START) grant, and by GESTAR-2 cooperative agreement with NASA Goddard Space Flight Center.

Data Availability Statement: Data can be accessed through [24].

Acknowledgments: The work is supported by the NASA Commercial Smallsat Data Acquisition (CSDA) programs. During the preparation of this manuscript/study, the author(s) used ChatGPT version GPT-5.6 Sol for the purposes of assisting in writing code. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GNSS	Global Navigation Satellite System
LEO	Low-Earth Orbit
SD-WACCM-X	Specified Dynamics–Whole Atmosphere Community Climate Model with Ionosphere/Thermosphere eXtension
EUV	Extreme Ultraviolet
RO	Radio Occultation
hTEC	Horizontal Total Electron Content
LEMURs	Low-Earth Multi-Use Receivers
MERRA	Modern-Era Retrospective Analysis for Research and Applications

References

1. Schunk, R.; Nagy, A. *Ionospheres*; Cambridge University Press: Cambridge, UK, 2000.
2. Banks, P.M.; Kockarts, G. *Aeronomy*; Elsevier: Amsterdam, The Netherlands, 2013; Volume A.
3. Feynman, J.; Gabriel, S.B. On Space Weather Consequences and Predictions. *J. Geophys. Res. Space Phys.* **2000**, *105*, 10543–10564. [CrossRef]
4. Board, S.S.; Council, N.R. Severe Space Weather Events–Understanding Societal and Economic Impacts: A Workshop Report. In *Severe Space Weather Events: Understanding Societal and Economic Impacts Workshop Report*; The National Academies Press: Washington, DC, USA, 2008.
5. Pisacane, V.L. *The Space Environment and Its Effects on Space Systems*; American Institute of Aeronautics and Astronautics: Reston, VA, USA, 2008.
6. Appleton, E.V.; Barnett, M.A.F. Local Reflection of Wireless Waves from the Upper Atmosphere. *Nature* **1925**, *115*, 333–334. [CrossRef]
7. Davies, K. *Ionospheric Radio Propagation*; US Department of Commerce, National Bureau of Standards: Gaithersburg, MD, USA, 1965; Volume 80.
8. Kursinski, E.R.; Hajj, G.A.; Schofield, J.T.; Hardy, K.R. Observing Earth's Atmosphere with Radio Occultation Measurements Using the Global Positioning System. *J. Geophys. Res. Atmos.* **1997**, *102*, 23429–23465. [CrossRef]
9. Wickert, J.; Reigber, C.; Beyerle, G.; König, R.; Marquardt, C.; Hocke, K. Atmosphere Sounding by GPS Radio Occultation: First Results from CHAMP. *Geophys. Res. Lett.* **2001**, *28*, 3263–3266. [CrossRef]
10. Jakowski, N. Radio occultation techniques for probing the ionosphere. *URSI Radio Sci. Bull.* **2005**, *2005*, 4–15. [CrossRef]
11. Jakowski, N.; Wehrenpfennig, A.; Heise, S.; Reigber, C.; Lühr, H.; Meehan, T.K. GPS Radio Occultation Measurements of the Ionosphere from CHAMP: Early Results. *Geophys. Res. Lett.* **2002**, *29*, 95–101. [CrossRef]
12. Syndergaard, S. A New Algorithm for Retrieving GPS Radio Occultation Total Electron Content. *Geophys. Res. Lett.* **2002**, *29*, 55–61. [CrossRef]
13. Wu, D.L.; Emmons, D.J.; Swarnalingam, N. Global GNSS-RO Electron Density in the Lower Ionosphere. *Remote Sens.* **2022**, *14*, 1577. [CrossRef]
14. Salinas, C.C.J.H.; Wu, D.L.; Swarnalingam, N.; Qian, L. Development of the Ionospheric E-region Prompt Radio Occultation Based Electron Density (E-PROBED) Model. *Space Weather* **2024**, *22*, e2024SW004037. [CrossRef]
15. Shaver, D.J.; Wu, D.L.; Swarnalingam, N.; Franz, A.L.; Emmons, D.J. Comparison of a Bottom-Up GNSS Radio Occultation Method to Measure D-and E-Region Electron Densities with Ionosondes and FIRI. *Remote Sens.* **2023**, *15*, 4363. [CrossRef]
16. Emmons, D.J.; Salinas, C.C.J.H.; Wu, D.L.; Swarnalingam, N.; Dao, E.V.; Forsythe, V.V. A Comparison of Modeled Daytime E Regions from E-PROBED and PyIRI with Ionosonde Observations. In *Annales Geophysicae*; Copernicus Publications: Göttingen, Germany, 2026; Volume 44, pp. 17–34.
17. Liu, H.-L.; Bardeen, C.G.; Foster, B.T.; Lauritzen, P.; Liu, J.; Lu, G.; Marsh, D.R.; Maute, A.; McInerney, J.M.; Pedatella, N.M.; et al. Development and Validation of the Whole Atmosphere Community Climate Model With Thermosphere and Ionosphere Extension (WACCM-X 2.0). *J. Adv. Model. Earth Syst.* **2018**, *10*, 381–402. [CrossRef]
18. Liu, J.; Liu, H.; Wang, W.; Burns, A.G.; Wu, Q.; Gan, Q.; Solomon, S.C.; Marsh, D.R.; Qian, L.; Lu, G.; et al. First Results From the Ionospheric Extension of WACCM-X During the Deep Solar Minimum Year of 2008. *J. Geophys. Res. Space Phys.* **2018**, *123*, 1534–1553. [CrossRef]
19. Rienecker, M.M.; Suarez, M.J.; Gelaro, R.; Todling, R.; Bacmeister, J.; Liu, E. Merra: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Clim.* **2011**, *24*, 3624–3648. [CrossRef]
20. Marsh, D.R.; Mills, M.J.; Kinnison, D.E.; Lamarque, J.F.; Polvani, L.M. Climate Change from 1850 to 2005 Simulated in CESM1 (WACCM). *J. Clim.* **2013**, *26*, 7372–7391. [CrossRef]

21. Angling, M.J.; Nogués-Correig, O.; Nguyen, V.; Vetra-Carvalho, S.; Bocquet, F.X.; Masters, D. Sensing the Ionosphere with the Spire Radio Occultation Constellation. *J. Space Weather Space Clim.* **2021**, *11*, 56. [CrossRef]
22. Hoque, M.M.; Jakowski, N. Ionospheric Bending Correction for GNSS Radio Occultation Signals. *Radio Sci.* **2011**, *46*, RS0D06. [CrossRef]
23. Salinas, C.C.J.H.; Qian, L. Quantifying Geomagnetic Activity's Contribution to the Global E-Region Electron Density's Day-to-Day Variability Using Spire Radio Occultation Observations. *Geophys. Res. Lett.* **2025**, *52*, e2024GL112874. [CrossRef]
24. Salinas, C.C.J. Data Used for Error Analysis Paper. *Zenodo* **2026**. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.