

The authors would like to thank the reviewer for their comprehensive suggestions, which have made the manuscript clearer and more reproducible. Responses to the individual points are included below. Line numbers refer to the new manuscript.

1) Specifically, the manuscript would benefit from: a clearer statement of novelty relative to existing assimilative models.

Added "NIMO was designed to be more adaptive than previous systems that couple first-principle and assimilative models." on lines 8–9 in the abstract.

2) Specifically, the manuscript would benefit from: greater transparency in validation design (distinguishing assimilated from independent data)

We added more details about the ingested (assimilated) data sets on lines 108–114. More details about the differences between assimilated and validation data sets are included around lines 165–170 and lines 221–224.

Lines 109–115:

"Data sets ingested for the validation runs include Global Positioning System (GPS) relative Total Electron Content (TEC) (Martire et al., 2024), Electron Density Profiles (EDPs) from the National Oceanic and Atmospheric Administration (NOAA) Mirrion 2 ionosondes (McNamara, 2005; National Centers for Environmental Information), and space-based radio occultation (RO) data. The RO data is obtained from the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) and COSMIC-2 constellations, which is publicly available at the University of Colorado (UCAR COSMIC Program, 2019, 2022). NIMO processes the slant TEC from the podTEC files. NIMO is also capable of ingesting commercial RO data (such as data sets provided by Spire and GeoOptics)."

Lines 229–232

"As was the case with the ionosonde data, TEC was also ingested into the NIMO data assimilation. However, in this instance the assimilated STEC was obtained from an alternate source (Martire et al., 2024) and does not require a receiver bias calibration. Thus, the validation data set will be distinct from the assimilated data set, even if the same satellite–receiver pairs are used."

3) Specifically, the manuscript would benefit from: some improve readability and reproducibility.

Addressed by adding more details into the model system, improving the figure quality, and improving the citations and acknowledgements for all the data sets.

4) The manuscript gives an overview of NIMO's structure but does not fully highlight how it differs from previous SAMI3/IDA-4D configurations or other assimilative systems. A concise comparison of capabilities and computational improvements would strengthen the case for novelty.

More information about NIMO and its development history is included on lines 32-36 and lines 100-107.

Lines 32-36:

"NIMO is built on a foundation of two well-known and sufficiently mature algorithms; both the Ionospheric Data Assimilation Four-Dimensional (IDA-4D) data assimilation scheme and the SAMI3 is also a Model of the Ionosphere (SAMI3) physics-based model have over 20 years of heritage. The predecessor to IDA-4D, IDA-3D, is comprehensively described in Bust et al. (2004). IDA-4D includes the temporal evolution of electron densities using a Gauss-Markov Kalman filter technique (Bust et al., 2007). More recently, IDA-4D evolved to solving for the log of the electron density (Bust and Datta-Barua, 2014)."

Lines 100-107:

"An early version of NIMO, called IDA-4D/SAMI3, ingested non-real time GPS data from ground stations available from the International GNSS Service with a 5-minute assimilation time step to study localized enhancements of electron density following geomagnetic storms (Chartier et al., 2021). In this study, IDA-4D/SAMI3 NmF2 was validated for two storm periods in November 2003 and August 2018 using in situ electron density data, autoscaled ionosonde NmF2 and reference GPS data. The assimilation model was found to reduce the Root Mean Squared Error (RMSE) of NmF2 in SAMI3 by up to 35 - 50%. This early version of the model was functionally similar to NIMO v1.0. The primary difference between IDA-4D/SAMI3 and NIMO v1.0 is that the pre-processor portions of IDA-4D were moved into a separate routine and the code was reconfigured to run in real-time."

5) Several of the validation datasets (ionosondes and GPS TEC) are also assimilated into NIMO. The authors acknowledge this but the distinction between assimilated and independent datasets should be made explicit.

The differences in the GPS TEC data sets are clarified on lines 229-232 (see point 2). More information about the differences between the ionosonde data sets are provided on lines 175-180.

Lines 175-180:

"The NOAA Mirrion 2 ionosonde database includes some of these ionosondes, but not all of them. This database also includes ionosondes that are not included in the validation. Although the ingestion process does remove

some of the provided data to prevent the dominance of one data set on the assimilation, this information is not available post-processing. Due to this complication, instances where ionosonde data is or is not assimilated are addressed on a case-by-case basis and the regional statistics do not make a distinction between these two possible states."

6) Validation against IRI-2016 is appropriate for a first version, but the authors should discuss briefly why other empirical or assimilative models were not included. If such comparisons are reserved for future work, this should be mentioned in the discussion.

We extended our reasoning for using IRI-2016 on lines 130-135. Our future plans involve comparing new versions of NIMO against older versions of NIMO and newer version of IRI, as well as new assimilative models that have reached out to us. As publication of those results will involve cooperation of a party not involved in this manuscript, we do not deem it wise to promise publication to the wider community.

Lines 130-135:

"We chose IRI-2016 to compare against because it was the most up-to-date version at the time this validation was performed. Since 2014, IRI has been recognized as the official standard for the Earth's ionosphere by the International Standardization Organization (ISO) (ISO 16457:2014; ISO 16457:2022). It is similarly endorsed by the International Union of Radio Science, the Committee on Space Research, and the European Cooperation for Space Standardization (Bilitza et al., 2022). Future version of NIMO and other ionospheric models (empirical, first-principles, or assimilative), may compare their performance against these results."

7) Figure 6 adds little beyond methodological illustration and could be moved to Supplementary Material

Removed figure altogether, as it was a bit elementary for the intended audience.

8) The discussion section could more explicitly quantify the improvement of NIMO over IRI-2016 (percentage reduction in RMSE or increase in correlation).

These are illustrated in the summary figures (Figures 8, 9, 12, 13, 14, 18, 19, 21, and 22). We found that choosing just one statistic to represent the model performance did not do a good job of highlighting the strengths and weaknesses of the two models, and so avoided doing this to not provide a misleading summary.

9) When describing storm-time performance (March 2015 run), it would be valuable to show a figure illustrating NIMO's dynamic response versus IRI.

Added RMSE improvements for the ionosonde statistics on lines 393–394, a new figure (Figure 10) that shows how the hmF2 varies over the storm when compared to IRI–2016, and a paragraph in the conclusions that presents meta–statistics across multiple instruments (lines 528–530).

Lines 393–394:

"with an improvement in the RMSE of 0.28 MHz for the foF2 and 2 km for the hmF2"

Lines 528–530:

"The storm time response from NIMO was also more representative of the ionosphere across all validation sets, when compared against IRI–2016. Examining just the RMSE from data containing all local times, PRMSE for the Mar 2015 storm was 60%. The ISR runs (which did not use RMSE) had Pr1 ,B = 75% against the unaltered NIMO output."

9) The conclusion could briefly describe operational readiness to substantiate the claim that NIMO is suitable for real–time space weather operations.

Added information on lines 37–39 to inform the reader that this model is currently operational.

Lines 37–39:

"This article provides an overview of NIMO version 1.0 (v1.0), the immediate precursor to NIMO v1.1.2 that began running operationally at the U.S. Fleet Numerical Meteorology and Oceanography Center."

10) #118: "knowledgable" → "knowledgeable" Fixed

11) #125: "hmF2 has an added complication that the conversion..." → "hmF2 estimation is complicated by the conversion...": This text was removed

12) "Data base" or "database" (ensure consistency): Fixed

13) Avoid repetitive phrasing such as "NIMO outperformed IRI–2016...": Went through the article and tried to use different phrasing in nearby text sections.

14) There are some acronyms not defined at first use: Fixed

15) Summary tables could be moved from the main text to a supplementary appendix: Kept things as they are, since the tables kept in the main text are the ones most directly referenced in the discussions.

16) Improve figure readability (axes labels, font size, color balance): Fixed

17) Provide DOI or URL references for the datasets used (COSMIC-2, Madrigal, JASON, ICON): Fixed

The authors would like to thank the reviewer for their suggestions, which have made the manuscript clearer and more reproducible. Responses to the individual points are included below, which have been broken up into numbered points so that the reviewer may more easily ensure each point has been addressed. Line numbers refer to the new manuscript.

1) The applications description in lines 22–30 is under referenced and there is no description in the introduction of other data assimilation systems that exist. As this is advocating the performance and novelty of an operational data assimilation system, there should be at least some acknowledgment or discussion of what methods and systems already exist. Is there a reason you have chosen IDA4D vs. other approaches? At least some discussion to that effect would be valuable to understanding the design trades involved in developing the model, particularly within an operational context, which has different challenges and restrictions than reanalysis-type or scientific data assimilation systems.

We added more description and references to the assimilation systems relevant to NIMO in lines 33–43, as well as lines 109–115.

2) Lines 49–50: This pre-processor requires significant further description. Data preprocessing, particularly in a real time scenario where phase leveling for GNSS, or ionosonde data filtering, can be much more challenging than in reanalysis configurations, ultimately setting the bounds on the available information content available from the measurements. While the study is validated in hindcast here, it is stated as an operational capacity, capable of near real time operation, so the pre-processing is a significant necessary part of such a system and must be adequately described. At the very least the processing applied to the hindcast ingested TEC data should be described and it should be caveated whether the approach is the same as that which would be applied in real time.

Details about the pre-processor have been added on lines 72–78.

3) Lines 50–51: How? The code hasn't been described in any detail, so it is hard to both corroborate this claim or understand the limits of the model.

Added on lines 60–67.

4) Lines 56–58: What is included in the assimilation state space? If you're updating the electron density at t_0 and propagating to t_1 without updating the external drivers of SAMI3 (EUV flux, thermosphere, winds, etc...) SAMI3 will largely just revert to determining an electron density self-consistent with the external drivers. The ionosphere has very little memory by virtue of the fast recombination rate and nearly instantaneous response to external

driving such that the prior electron density is only a small factor of the subsequent state. More information is needed here and some sort of demonstration of the forward propagation in the assimilation step to demonstrate that the model isn't just falling back to the state that is self-consistent with external drivers is essential to understanding how information from previous timesteps are being leveraged in the assimilation.

Added information on lines 72–84.

5) NIMO System: The details on how the assimilation is conducted are extremely limited. Such details are essential to interpret the validations later in this study. Details regarding the construction of the assimilation a-priori and measurement error covariances at a minimum should be provided, particularly given the apparent overfitting to ionosonde observations seen later in the manuscript.

Readers who wish to know more about the details of the assimilation can now go to the references provided in the new text that was added to address the similar comments made above (e.g., the papers that describe IDA-4D in detail).

6) Lines 91–93: This is not a recent change, so it's odd that the AMTB option was selected here, when the newer default option could have been changed to. The authors mention that this difference would be discussed in later sections, but it is not mentioned again after this point. Also, given that the authors are using IRI-2016, can they confirm that they have updated the IRI's internal `ig_rz.dat` and `apf107.dat` files to ensure that the model has been run in a nominal, rather than forecast, configuration in their study? For reference, if the pypi version by Mike Hersch is the one that has been used, without modification, then index files would have last been updated in February/May 2019.

According to corrections in the IRI FORTRAN files, the Shubin model was made the default on 10/03/2021. For our validation the code we used was frozen before that time, so we continued to use the AMTB default. Additionally, the Shubin model has caused problems for HF raytracers that we use so it still is not the operational standard that we use. The `iri.dat` files were all updated to include the correct timeframe. We were using the IRI FORTRAN directly, we were not using a python wrapper for this.

7) Lines 117 – 130, Ionosonde data Quality and Repeatability: All manually scaled data adhering to the URSI guidelines includes provision of qualifying letters to attest to the consistency and accuracy of the data. The reliability of those qualifying letters as a specification of manual scaling performance was assessed and validated in the 1970s and again in the 1980s as part of the URSI INAG endorsement process for the URSI handbook and guidelines.

Discussion of the limitations of manually scaled data has been removed as has the discussion of the Dandenault study. Text was changed to discuss that manually scaled data would likely not be available for real time analysis, which is what these tools were developed for.

8) Dandenault Study: That study involved having inexperienced scalers scale ionograms without any substantive training according to the URSI guidelines and does not represent the scaling performance of manual scaling as a whole. Any scaling that adheres to URSI guidelines is accurate and precise to within 0.05 MHz unless a corresponding qualifying letter is prescribed by the scaler, in which case the error threshold of the qualifying letter should be considered. In that study, many of the participants accidentally scaled the F1 peak as the F2 peak, a mistake one would expect of an autoscaling routine but not one I have ever seen made by a scaler adhering to the URSI guidelines and with corresponding training. The accuracy of ARTIST that you cite above this was determined against manually scaled data, so it is somewhat contradictory to imply that the accuracy of manually scaled data is not sufficient as to warrant getting the data manually scaled instead of using autoscaled data when it was sufficient to establish the performance of manual scaling in the first place.

See the response to (7), above.

9) Lines 127-130, Ionosonde data quality and processing: What efforts were made? You mention that the qualifying letters are not sufficient in and of themselves, but have you employed any filtering to your dataset or have you perhaps employed some preferential qualifying letters or confidence scores for certain parameters? It doesn't seem sufficient to just say that everything has errors so we didn't bother implementing anything as is currently implied.

The ionosonde data for validation is filtered to only include confidence scores of 70 or above. Although this seemed to work well in general, analysis after the fact showed that this did not remove all problematic scalings used in the validation. Added text say this on lines 164-165: "Based off this analysis, we use a cutoff of 70 for both foF2 and hmF2 to use the same times for both analyses."

10) Lines 132-134: This needs to be clarified: are data from all of the locations in Figure 4 passed to the assimilation or are some of them not included at all? If data is being rejected internally the criteria and methodology of that should be described somewhere. I would think that GNSS data would be often removed or down-sampled due to the over-correlated error covariance, but I would not think that the relatively infrequent ionosonde observations would ever be rejected on the basis of "dominating the assimilation".
Residuals vs. validation: In many instances the model is being compared to data that was assimilated either in part or in full. I all

such instances it should be made clear that this is the case and it should be caveated that such comparisons are residuals, not independent validation.

Not all of the ionosondes in Figure 4 are used in assimilation. The ionosondes used for assimilation came from the NOAA MIRRION 2 database. The validation database is larger than the assimilation database. Some of the ionosondes are included and additionally, there are assimilated ionosondes that are not part of the validation data base. Added text explaining this on lines 175–176: "The NOAA Mirrion 2 ionosonde database includes some of these ionosondes, but not all of them"

11) Ingested Data, Lines 177–178: If the sTEC was acquired through a different source and used a different bias estimation approach, it should be described here. The location of the assimilated GNSS stations should be added to figure 5 so that the reader can understand what amount and relative distribution of GNSS data was ingested into the model. It is also highly likely that your separate dataset is also included in the Madrigal one or is highly collocated, so comparison to TEC here is likely mainly an assessment of residual performance rather than independent validation.

Added a better description of the ingested TEC on lines 109–110. Added context about the difference between the ingested and validation data sets on lines 229–232.

Lines 109–110:

"Data sets ingested for the validation runs include Global Positioning System (GPS) relative Total Electron Content (TEC) (Martire et al., 2024),"

Lines 230–232:

"...the assimilated STEC was obtained from an alternate source (Martire et al., 2024) and does not require a receiver bias calibration. Thus, the validation data set will be distinct from the assimilated data set, even if the same satellite–receiver pairs are used."

12) Lines 199–201: Reference needed unless this was an assessment you have conducted, in which case it would benefit from being illustrated. Regardless, this does not mean that JASON2 is a correct reference, see for example: <https://doi.org/10.1007/s00190-021-01564-y>

This was an assessment that was performed in house, as clearly stated in Lines 249–256. No figure is needed to follow this analysis process.

13) Section 4.3.1: Ionosonde–based validation should likely be broken down by latitude region. The selected ionosonde dataset is highly

heterogeneous with a strong bias toward mid latitudes. Overall metrics using ionosonde data will thus be strongly biased toward the performance at mid latitudes. At the very least a comment should be added that the overall metric is likely strongly skewed toward mid latitudes.

Added a line stating the mid-latitude bias on 180-181.

14) Lines 293-294: Why do you believe that foF2 and hmF2 are more reliable than the parameters of the other layers? foE and foF1 do have some challenges in their scaling, but they are not as significant as those in the F2 peak and these layers are generally very stable and well represented by even climatology, so one would imagine that they are relatively easy goals. Given that the main stated objective is OTHR, which is highly sensitive to E and lower F-region plasma density, I would think that validation in that domain would be of critical importance to this work. In fact, one of the largest drawbacks of physics-based models compared to the IRI is their significant limitations in capturing the E-Region and F1-layer characteristics, so it is particularly important here, given your background model, to assess how it is doing below the F2 peak. It is, in fact, quite odd that despite having bottomside measurements either by ionosondes or ISRs, no illustration of the vertical structure of the assimilation and background model error statistics are provided. Given the importance for the stated application and the challenges experienced by many physics-based models in the bottomside, I think it is essential that the authors provide at least some illustration of the performance of the model in terms of its vertical structure, if only by comparison to ISR data.

We do not believe they are more reliable, and did not state that anywhere in the article. The F2 peak characteristics were chosen for the validation as they reflect the largest altitudinal feature in the ionosphere. Lines 538-540 were added to identify the bottomside and EIA as potential areas for validation expansion in the future.

Lines 538-540:

"Future validation efforts may also include a wider range of metrics that capture the variations in the bottomside ionosphere, as well as the morphology of key features like the Equatorial Ionization Anomaly."

15) Background model performance: It would be very valuable, if not essential, for the study to include comparison to the background model used for the assimilation system to better understand the innovation induced by the inclusion of the measurements (i.e. to understand how much the data moves the state away from the background).

The purpose of this paper is not to validate IDA-4D and SAMI3 independently, but to understand the performance of the entire NIMO

system. Independent validations of each of these models would require entirely separate papers and are out of scope for this manuscript.

References to prior development that led to NIMO has been added on lines 99-106, to add some context.

Lines 100-107:

"An early version of NIMO, called IDA-4D/SAMI3, ingested non-real time GPS data from ground stations available from the International GNSS Service with a 5-minute assimilation time step to study localized enhancements of electron density following geomagnetic storms (Chartier et al., 2021). In this study, IDA-4D/SAMI3 NmF2 was validated for two storm periods in November 2003 and August 2018 using in situ electron density data, autoscaled ionosonde NmF2 and reference GPS data. The assimilation model was found to reduce the Root Mean Squared Error (RMSE) of NmF2 in SAMI3 by up to 35 - 50%. This early version of the model was functionally similar to NIMO v1.0. The primary difference between IDA-4D/SAMI3 and NIMO v1.0 is that the pre-processor portions of IDA-4D were moved into a separate routine and the code was reconfigured to run in real-time."

16) Figure 8: This is a plot of residuals, where the ionosonde data was included in the assimilation itself. Can the author comment on what filtering was used in the ionosonde data during fitting? It appears like the assimilation ionosonde data is being overfit, given the significant outlier that is following the ionosonde observations, but the assimilation doesn't do the same for a subsequent outlier. Were the observations from the second outlier not assimilated while they were in the first case? How are you assigning uncertainty to the ionosonde observations in the assimilation? In this case at 0130 UT on April 10, 2020 the error is a second-hop trace scaled in error and with a very low assigned confidence value. This concern returns in Figure 11.

Added context on line 363-364: "At 01:30 UT the model fits a very high and unrealistic hmF2. The error in the EDP did match the true value of the error that was by determined by looking at the ionogram."

17) Figure 11: The assimilation is very clearly overfit to the collocated ionosonde observations here, where scaling errors are dominating the variability of the assimilation result. If that is somehow not the case, then the presence of the large anomalous swings in the assimilation must be explained. This figure is pointed to in the text as an example, but the contents of the figure and the behaviour of the assimilation demonstrated therein are not addressed or discussed anywhere in the manuscript. There needs to be some discussion of what is happening in the NIMO output in this figure. The variations seen look nothing like true variations seen in the ISR observations.

Figure 11 shows data from the JRO ISR and, as described in the text on 203, is not calibrated using an ionosonde. It is a completely independent data source.

18) Line 341: Why? The whole point of the ISRs over the ionosondes should be that the ISRs provide unambiguous vertical structure information. Reducing the comparison to just hmF2 and foF2, when the ISRs are themselves, in most cases, already calibrated against local ionosonde observations which you likely assimilate likely biases performance, since collocated hmF2 and foF2 observations were available at these locations from other instruments and it seems like a missed opportunity here to understand the vertical structure of model performance.

The ion line ISR data does not provide a direct measurement of the electron density and sometimes it is calibrated with ionosonde data to generate electron density. However, recent advances in the ISR detection of the plasma line are allowing to generate precise and independent electron density observations, especially at the F peak. This is specifically valid for the Millstone Hill and Arecibo ISRs. For the Jicamarca ISR, the electron density profiles are obtained using the Faraday rotation experiment (data presented in Figures 11, 12, and 13). Then the data presented in the paper is independent and unbiased from the ionosonde data.

As part of the validation, we performed comparisons the complete electron density vertical profile. However, following the lead of the ionosonde analysis, we included in the paper only the hmF2 and NmF2 comparisons, since they are easy to quantify and to compare.

19) Validation consistency: The authors repeatedly switch between what metrics they present for what comparisons. The authors should provide the same set of metrics for each comparison. RMSE should not be missing from Figures 12 and 13, just as correlation should not be missing from Figure 10, etc.... Given that the authors have spent considerable time establishing the importance of each of these metrics, they should be applying them equally to all comparisons. The same can be said for Figure 14 where RMSE returns but r disappears. The absence of particular metrics in certain validations could give the reader the false impression that the authors have been cherry picking metrics.

Our analysis found that no single metric accurately reflected each model's strength and weakness across the different validation data sets. A variety of statistics are provided to avoid providing a misleading summary of NIMO's performance.

20) Acknowledgements – Ionosonde data: The authors do not provide an acknowledgment for the ionosonde data used and do not adhere to the

rules of the road for use of ionosonde data. Rules of the Road:
<https://giro.uml.edu/didbase/RulesOfTheRoad.html> Acknowledgement List:
<http://giro.uml.edu/didbase/acknowledgements.html>

Fixed.

21) Acknowledgements – Madrigal TEC: Please adhere to the Madrigal TEC recommended practice described below for acknowledging use of Madrigal TEC products: https://cedar.openmadrigal.org/static/siteSpecific/tec_sources.html Also, you should provide a doi and reference to the relevant datasets used in this study. Madrigal provides a tool for composing doi's for sets of data if necessary. This can be done using Madrigal's `globalCitation.py` script in the Python API wrapper.

Fixed.

22) Line 53, UV data: Is there a reference for this or was it done as part of this study?

There is no reference for this, it is unpublished apart from its reference here.

23) Lines 65–66: This would imply that your forward propagation includes no storm-related behaviour of any sort, except, perhaps some minimal bleed-in from MSIS's storm response. While geomagnetic indices are used in your implementation, they would only end up being passed to MSIS, correct?

That is incorrect, the geomagnetic indices are also used by HWM, which has a disturbance component. The impact of the geomagnetic drivers is now illustrated in the new Figure 10.

24) Line 69: lowercase a (unless you mean the 24-hour average, in which case the capital is correct, but the three-hour is a bit confusing since formally A_p is calculated as a daily value at the end of each day and not as a sliding value).

Fixed

25) Line 78: Clarification is needed: Do you mean that you only ingest the slant TEC in the podTEC files or do you incorporate other processed R0 products as well?

Changed to "processes the slant TEC from the podTEC files"

26) Line 100 "validations" -> "validation"?

Fixed

27) Line 110: This is not strictly correct; scaling is the process of

isolating the complete ionospheric virtual height trace. The process you are referring to is trace inversion, which is a separate process conducted after scaling.

Changed to "scaled and inverted"

28) Line 113 version 4 or 5: The data also includes observations using v4.5. Despite it's subversioning, 4.5 is a distinct version of ARTIST using a different approach from both v4 and v5.

Added on line 155.

29) Line 114 scaling method unknown: It is not strictly unknown; it is just not reported by the quick char tool on the website. In most cases, this is either Autoscala for ionosondes operated by INGV or the Australian software suite if operated by the Australian Bureau of Meteorology. Russian ionosonde data is a mix of manually scaled and data scaled by Autoscala. That information is, however, contained in the ionosonde Standard Archiving Output files.

Reword this to say that it isn't included in the Fastchar data, which is what we used (line 156-157).

30) Lines 127-128: Worth citing the following to set some bounds on this: <https://doi.org/10.3390/rs12172671>

Removed the text where this would have been applicable.

31) ISR Data Calibration: Is this something you have done separately, or are you just using the data as it appears on Madrigal? If Madrigal, just cite the appropriate dois using their aggregate doi creation tool.

The ISR data was obtained directly from the PIs, not obtained from Madrigal.

32) Lines 172-173: This is out of date. The revised processing was reported in <https://amt.copernicus.org/articles/9/1303/2016/amt-9-1303-2016.html>

Added reference

33) Lines 173: While it is reported as an error, it is actually the "standard error" associated with the grid average (i.e. σ/\sqrt{N}). Also, I don't believe that it is true that they are typically on the order of a tenth of a TECU. Opening a random file from 2021, I get a global mean dtec of 0.85 TECU and median of 0.92 TECU with the distribution peaking at 1.4 TECU and appearing very multi-modal, with only 0.02% of all dtec values being less than 0.1 TECU. The randomly chosen file is that from May 26th, 2021, if you'd

like to verify. Regardless, this is not indicative of the error in the measurement. In relative TEC perhaps, but there is not a bias determination method that exists that can claim precision of this level. Even the best approaches settle in around 1 TECU if only because of the uncertainty in the residual error from phase-leveling, amongst other geometric limitations. The authors are again directed to <https://amt.copernicus.org/articles/9/1303/2016/amt-9-1303-2016.html> for a more up-to-date assessment of Madrigal TEC bias accuracy/precision or to <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2023SW003611> for an external assessment of their performance.

"On the order of a tenth" means between 0.1 and 0.99. As you have demonstrated to yourself, this is correct. However, to avoid confusion with readers, line 218 was rephrased to use simpler language.

Line 227:

"These errors typically range between 0.1 and 1.0 TEC Unit"

34) Line 327: To an extent this is true, but previous extensive assessments of the performance of autoscaling, which you cite in your introduction, can be used to set some bounds on our expectations of these errors. I would recommend revisiting your previous autoscaling performance numbers for the Galkin and other papers here for context, particular in foF2.

Added more context on lines 167–171, to better communicate the point that the expected errors from Autoscaling are not the only source of errors.

Lines 167–171:

"The hand scaled profiles also have errors associated with them that are harder to quantify. Only auto-scaled ionosondes were used in this study to replicate what would done in a real-time verification. Since the autoscale confidencescore does not catch all issues in scaling, it is likely that poorly scaled ionosonde data were included in the data validation, despite efforts to ensure a clean data set. The ingested data uses a different error analysis and may also include poorly scaled data."

35) Figure 11: The dates and times of this example must be provided so that the measurements can be corroborated and the dataset can be verified.

Added date to the caption.

36) Lines 435–436: missing "is"?

Fixed

37) Figure 19 and 20: Please add units to the axes or where appropriate.

Fixed