Evaluation on the Quasi-Realistic Ionospheric Prediction Using an Ensemble Kalman Filter Data Assimilation Algorithm

Jianhui He1,2,3,4, Xinan Yue1,2,3,4, Huijun Le1,2,3,4, Zhipeng Ren1,2,3,4, and Weixing Wan1,2,3,4

1Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, 2Innovation Academy for Earth Science, CAS, Beijing, China, 3Beijing National Observatory of Space Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, 4College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China

Abstract In this work, we evaluated the quasi-realistic ionosphere forecasting capability by an ensemble Kalman filter (EnKF) ionosphere and thermosphere data assimilation algorithm. The National Center for Atmospheric Research Thermosphere Ionosphere Electrodynamics General Circulation Model is used as the background model in the system. The slant total electron contents (TECs) from global International Global Navigation Satellite Systems Service ground-based receivers and from the Constellation Observing System for Meteorology, Ionosphere and Climate are assimilated into the system, and the ionosphere is then predicted in advance during the quiet interval of 23 to 27 March 2010. The predicted ionosphere vertical TEC (VTEC) and the critical frequency foF2 are validated by the Massachusetts Institute of Technology VTEC and global ionosonde networks, respectively. We found that the ionosphere forecast quality could be enhanced by optimizing the thermospheric neutral components via the EnKF method. The ionosphere electron density forecast accuracy can be improved by at least 10% for 24 hr. Furthermore, the Thermosphere, Ionosphere, Mesosphere Energetics and Dynamics/Global Ultraviolet Imager (TIMED/GUVI) [O/N2] observations are used to validate the predicted thermospheric [O/N2]. The validation shows that the [O/N2] optimized by EnKF has better agreement with the TIMED/GUVI observation. This study further demonstrates the validity of EnKF in enhancing the ionospheric forecast capability in addition to our previous observing system simulation experiments by He et al. (2019, https://doi.org/10.1029/2019JA026554).

Plain Language Summary The significance of the coupled thermosphere and ionosphere data assimilation for ionosphere forecasting has been well proven recently. The neutral state variables can be optimized by assimilating ionosphere observations via their correlation represented by the ensemble based error covariance. In this study, the slant total electron contents from global ground-based International Global Navigation Satellite Systems Service receivers and space-based Constellation Observing System for Meteorology, Ionosphere and Climate are ingested into the data assimilation system to evaluate the quasi-realistic ionospheric forecasting capability during the geomagnetic quiet conditions. The Massachusetts Institute of Technology vertical total electron content and global ionosonde network foF2 observations are chosen to make independent validation. The results show that the ionosphere forecasting capability is enhanced via optimizing the background thermosphere and its effect could last more than 24 hr. In addition, the well-optimized neutral background [O/N2] by ensemble Kalman filter (EnKF) can be confirmed through the comparison with Thermosphere, Ionosphere, Mesosphere Energetics and Dynamics/Global Ultraviolet Imager observations. This study further demonstrates the validity of EnKF in enhancing the ionospheric forecast capability in additional to our previous observing system simulation experiments.

1. Introduction

The Earth's ionospheric conditions have a significant impact on positioning, navigation, and communication systems through refraction, scattering, and reflection of the radio wave. Due to the greatly increasing human space activities, an accurate specification of both the current and future ionospheric status has drawn more attention recently. In the past decade, lots of work about ionosphere nowcasting and forecasting have
been conducted by applying data assimilation technique (Bust et al., 2004; Galkin et al., 2012; Pi et al., 2003; Schunk et al., 2004; Wang et al., 2004). It has been proven that the effect of updating ionospheric electron density as initialization in the theoretical model prediction can only last about 2–3 hr (Jee et al., 2007), due to its variability resulting from a variety of drivers, such as solar extreme ultraviolet radiation, the lower atmospheric tides, and the high-latitude convection pattern. Therefore, a longer forecast timescale of ionosphere is still challenging.

As a closely coupled system, the state of ionosphere is greatly controlled by the conditions of thermospheric variables, such as the neutral compositions, wind, and temperature. Especially for neutral composition, it plays a key role in the production and loss of the ionosphere daytime F2 region plasma density. Meanwhile, the variation of neutral composition is relatively slow versus time in comparison with that of electron density. Thus, the coupled thermosphere and ionosphere system need to be taken into consideration for longer timescale of ionosphere forecasting.

Recently, it has been demonstrated that the longer timescale of ionosphere forecasting can be achieved by ingesting the ionosphere observations into the coupled thermosphere and ionosphere background model based on the ensemble Kalman filter (EnKF) data assimilation method (Chartier et al., 2016; Chen et al., 2017; He et al., 2019; Hsu et al., 2018; Lee et al., 2012; Matsuo et al., 2013; Yue et al., 2007). Theoretically, a variety of ionosphere drivers such as neutral composition and winds, and electric fields can be well adjusted through assimilating the ionosphere observations via EnKF data assimilation algorithm. Hsu et al. (2014) conducted Observing System Simulation Experiment (OSSE) during the quiet day to examine the capability of ionosphere forecasting through initializing the different combinations of ionosphere and thermosphere variables by assimilating Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) electron density profiles into the TIEGCM model based on Data Assimilation Research Testbed. They concluded that a longer ionosphere electron density forecast performance can be achieved by simultaneously initializing the thermosphere states, especially the neutral compositions. Chartier et al. (2013) also conducted the similar assimilation experiments but for the storm time. Their conclusion is that updating the neutral compositions is more effective in global ionospheric electron density specification and prediction. Pedatella et al. (2018) achieved reasonable ionospheric total electron content (TEC) forecast through applying EnKF to a whole atmosphere model during stratospheric sudden warming events.

Recently, He et al. (2019) constructed a new EnKF ionosphere and thermosphere data assimilation system based on National Center for Atmospheric Research TIEGCM and a sparse matrix method on an ordinary workstation. Through a series of OSSE studies to evaluate the validity and reliability of the system, the results show that the short-term forecast capability of ionosphere electron density is enhanced significantly and can extend to last longer than 24 hr due to the efficiently optimized neutral compositions. In comparison with other similar studies, the main advantage of our algorithm is that it could be run on an ordinary workstation, due to the efficiency of sparse matrix in both computing and storage. However, He et al. (2019) only show some OSSE results. In this study, we will further demonstrate the validity and efficiency of our algorithm by assimilating real observations. Specifically, the global ground- and space-based slant TECs will be assimilated into the system and evaluated by independent data sources such as gridded vertical TEC processed by Millstone Hill Haystack observatory of Massachusetts Institute of Technology (MIT), ionosonde critical frequency (foF2), and [O/N2] ratio measured by Thermosphere, Ionosphere, Mesosphere Energetics and Dynamics/Global Ultraviolet Imager (TIMED/GUVI).

The remainder of the paper is organized as follows. In section 2, the data used in this study and the design of EnKF data assimilation experiments will be described in detail. In section 3, we will present the experiment results. Discussion and summary will be given in sections 4 and 5, respectively.

2. Data Assimilation Prediction Description

2.1. EnKF Data Assimilation Algorithm

The EnKF thermosphere and ionosphere data assimilation system used in this study is constructed by He et al. (2019). This EnKF system uses a sparse matrix method to mitigate the huge computation and storage problems, which is based on three reasonable assumptions: (1) The background error covariance (P) needs to
be localized in EnKF algorithm to attenuate spurious statistical correlations due to the limited ensemble numbers, enabling the sparsity of this covariance. (2) For the observational operator matrix (H), the realistic ionosphere observations (e.g., TEC) only occupies a small portion of the grids of background model, enabling the sparsity of this matrix. (3) For the observation error covariance, this covariance is a diagonal matrix under the assumption of the uncorrelated observations error and thereby is a sparse matrix. In order to avoid the huge computation challenges of the matrix inverse in the Kalman filter, an iteration solving approach of Kalman gain and the left preconditions generalized minimum residual method is utilized (Yue et al., 2014). Moreover, a background error covariance inflation method is used to cope with the filter divergence problem, and a cutoff covariance localization scheme is applied to this EnKF data assimilation system (Houtekamer & Mitchell, 1998). At that moment, this EnKF ionosphere and thermosphere data assimilation system can be run on a general workstation. The National Center for Atmospheric Research TIEGCM was chosen as the background model in this EnKF data assimilation system, which can represent physical coupling between thermosphere and ionosphere (Richmond et al., 1992). The model resolution is 5° × 5° in longitude and latitude, half scale height in altitude, and 180 s in time. The realistic solar and geomagnetic activity indexes are used to drive the model. The used convection model in the high latitude is Heelis empirical model (Heelis et al., 1982).

In general, the data can be ingested into the physical-based model through the forward model, which is referred to within the data assimilation community as the observation operator. The ionosphere observations can be related to the background model variables through the forward operator. Mostly, the observation operator is a simple interpolative routine that interpolates the background model states to the observation location for the direct ionospheric measurement, such as electron density profile data from ionosonde. For the slant TEC, however, it is the integrating electron density along the raypath between Global Navigation Satellite Systems (GNSS) satellites and receivers. In this paper, the observation operator of slant TEC from ground- and space-based observations will be given as

\[ \text{stec} = \sum_{i=1}^{n} N_{ei} \Delta s_i, \quad (1) \]

where stec is the synthetic TEC of the TIEGCM along the raypath from satellite position to the receiver position, Nei is the electron density in the ith grid of TIEGCM, and the \( \Delta s_i \) is the raypath length within the ith grid, which is also the corresponding observation operator of slant TEC.

The best assimilation results could be obtained from the weighted average of the model state and the observations. The weights of model state and observations are related with their error covariance closely. For EnKF algorithm, the background error covariance is estimated from the ensemble of model forecasts at each assimilation cycle, which is called the flow-dependent background error covariance (P; Evensen, 1994). Regarding to the observational error covariance, the observation error is usually assumed to be independent to make the observation error covariance matrix diagonal (Bust et al., 2004), which was also the case in this study. Thus, the observation error covariance between ith and jth of the observations can be represented as

\[ R_{ij} = \begin{cases} \alpha \times H_i P_f H_i^T, & i = j \\ 0, & i \neq j \end{cases}, \quad (2) \]

where \( \alpha \) represents the ratio of the diagonal value of the observational error covariance to the diagonal value of \( HP_f H^T \); this ratio can be adjusted in our EnKF data assimilation system. According to formula (2), this ratio can significantly determine the weight of the real ionospheric observations in our assimilation system. If this ratio enlarges, the weights of the observations will be naturally small because the corresponding error covariance will become large and vice versa. Meanwhile, given that we believe in the observations much more than the model state, the ratio is typically small. In this paper, the value of \( \alpha \) is set to be 0.01, which is same as the value we used in the previous reasonable ensemble experiment results (He et al., 2019). It is clear that the observations can be given much more weights than the model state when the value is equal to 0.01. Although the ratio is a constant value, the observational error covariance applied in our EnKF assimilation system will be correspondingly adjusted in terms of time and space dimensions owing to the...
estimated flow-dependent background error covariance as indicated above. Thus, the observations can be well ingested into the background model at each assimilation cycle.

2.2. Data Description

The slant TECs from global ground-based International GNSS Service (IGS) receivers and the space-based COSMIC are ingested into the EnKF data assimilation system during the quiet day from 23 to 27 March 2010 in this study. The ground-based GNSS observations have a high time resolution and can provide a continuous global monitoring of the ionosphere. The slant TEC is derived from about 400 of the IGS ground-based receivers, which were calculated according to the global ionospheric map-aided method described by She et al. (2019). The IGS-leveled TEC is calibrated with the satellites differential code biases (DCBs) from Deutsches Zentrum für Luft- und Raumfahrt. The IGS receiver DCBs are estimated by the daily average deviation between the observed vertical TEC and the vertical TEC interpolated from IGS global ionospheric map at the same ionospheric pierce point. Moreover, the observations with elevation angles less than 10° are discarded to avoid the multipath effects. The space-based slant TEC data from COSMIC are provided by the University Corporation for Atmospheric Research/COSMIC Data Analysis and Archive Center. This slant TEC processing includes the cycle slip detection, multipath calibration, leveling of phase TEC to the pseudo-range TEC, and DCB calibration. The GNSS satellite DCB is calibrated by the Center for Orbit Determination in Europe. The COSMIC receiver DCB is estimated by a least squares fitting methods based on spherical symmetry of the overhead ionosphere, which is assumed to be a constant during 1 day. The accuracy of this slant TEC is about 1–3 TECU (Yue et al., 2011).

In this paper, the MIT vertical TEC (VTEC) and ionosonde data are used as independent data sources to evaluate the quasi-realistic ionosphere forecasting capability. The MIT VTEC is calculated by the software package MIT Automated Processing of GPS to automate the process of downloading and processing GPS data to produce global TEC maps (Rideout & Coster, 2006). The TEC maps are provided globally in 1° × 1° bins in latitude and longitude and have a 5 min time resolution. The standard archiving output format Digisonde ionogram data, provided by the National Oceanic and Atmospheric Administration’s National Geophysics Data Center, have been autoscaled by the Automatic Real-Time Ionogram Scaler with True height analysis software to get the ionospheric F2 critical frequency (Gardner et al., 2014).

The TIMED/GUVI [O/N2] observations are used to make a comparison with the [O/N2] updated by EnKF. TIMED/GUVI was launched on 7 December 2001 into a 625 km circular 74.1° inclination orbit with a 97.8 min period. GUVI samples all local solar times every 60 days. This means that the satellite orbit is near the same local time throughout a given day and the global maps can be created by compiling measurements from successive orbits (Stephan et al., 2008). The [O/N2] is formally defined as the ratio of the vertical column density of the atomic oxygen to that of the molecular nitrogen. The base of that column is the altitude where the N2 column density is equal to 1017 cm−2, typically around 135–140 km (Meier et al., 2005).

As an example, Figure 1 shows the distribution of the ground IGS stations associated with the ionospheric pierce point coverage and ionosonde stations (Figure 1a) and also gives a typical daily coverage by occultation tangent points (Figure 1b) during the day of 24 March 2010.

2.3. Prediction Setup

In order to evaluate the prediction performance of the system, three data assimilation experiments were conducted as listed in Table 1. The first one only updates the ionosphere parameters (Ne and O+). The second one is performed with simultaneously updating ionosphere and thermosphere state vectors, including the electron density (Ne), atomic oxygen ion density (O+), neutral temperature (Tn), and neutral compositions (O and O2). The third one is a TIEGCM default run without data assimilation, referred to as the control experiment henceforth. For each experiment, the assimilation window is set to be 1 hr, and a 24 h forecast for each assimilation cycle is conducted from the day of 23 to 27 March 2010. The ground-based slant TEC and COSMIC slant TEC within 1 hr are assimilated into TIEGCM for each assimilation cycle. It is worthwhile to note that the upper boundary of the TIEGCM located around 500–700 km in altitude, depending on the solar activity level. However, the slant TEC is the integration of electron density from GPS receivers to the height of GPS satellites, which are located around 20,000 km. There is a significant contribution to GPS slant TEC from the electron density above the upper boundary of the TIEGCM. Same as method used
in the other research (Chen et al., 2016), the electron density above the upper boundary of TIEGCM was extrapolated as

\[
N_{e,h} = N_{e,h} \exp \left( \frac{h - h'}{H} \right),
\]

where \( h \) is the altitude above the upper boundary of TIEGCM (\( h' \)) and \( H \) is the scale height of electron density \( (N_e) \) in the ith grid of the upper boundary of TIEGCM. Using this vertical extrapolation method, the electron density up to around 20,000 km can be obtained at each latitudinal and longitudinal grid of TIEGCM. Note that the grid above the upper boundary of TIEGCM was set to be 5° × 5° in longitude and latitude and 10 km in altitude. The raypath length at each grid of the upper boundary of TIEGCM can be obtained along each line of sight of GPS signal path. Then, according to formula (1), we can remove the topside slant TEC above the upper boundary of TIEGCM for each assimilated slant TEC observation from IGS ground-based receivers and the COSMIC. According to the example of the extrapolation from Chartier et al. (2016), the topside TEC can account for 12% of the total TEC on average. Although there are several problems with this approach, the extrapolation is an improvement over the assumption that there is no electron density above the upper boundary of TIEGCM. The other method calculating the TEC ratio of the upper boundary to the whole TIEGCM range and removing the topside slant TEC by the TEC ratio instead of the modeled TEC value can be also taken into consideration in the future.

The number of ensemble members is chosen to be 90, which is similar to the previous studies (Chartier et al., 2016; Chen et al., 2017; Hsu et al., 2018; Lee et al., 2012; Matsuo et al., 2013). The model ensemble is generated by perturbing the main TIEGCM drivers via centered Gaussian distributions, including solar 10.7 cm radio flux \( (F_{10.7}) \) and Kp indices. Note that the mean values of \( F_{10.7} \) and Kp are from realistic value with a standard deviation of 15% and 10%, respectively. A cutoff covariance localization scheme associated with 5° correlation distance in meridional and 10° in zonal direction and 30 km in height is applied to localizing the background error covariance to attenuate spurious correlations over long spatial distances due to limited ensemble number. To avoid the effect of initial conditions on the ensemble assimilation and control experiment, the model has been running for about 2 weeks before starting assimilation.

### 3. Results

To get a sense of the difference of ionospheric predictable timescale through updating different thermosphere and ionosphere state vectors,

| Table 1 |
| List of the EnKF Data Assimilation Experiments and Corresponding Updated Parameters in the Assimilation |
| Number of experiments | Updated variables |
|------------------------|-------------------|
| 1                      | \( \text{Ne, O}^+ \) |
| 2                      | \( \text{Ne, O}^+, Tn, O, O_2 \) |
| 3                      | TIEGCM run (without data assimilation) |

Note. \( \text{Ne, O}^+, Tn, O, \text{and O}_2 \) represent electron density, oxygen ion, neutral temperature, atomic oxygen mass mixing ratio, and molecular oxygen mass mixing ratio, respectively.
Figure 2 shows a TEC comparison among the control prediction (the first column) and forecast results of experiments 1 (the third column) and experiment 2 (the fourth column) with MIT VTEC observations (the second column) during the day of 24 March 2010 for different advanced hours. It can be seen that the values of VTEC are overall underestimated by TIEGCM control prediction in comparison with realistic observations. But this feature is corrected after both EnKF data assimilation experiments. For experiment 1, at 0600 UT (1 hr forecast), the value of forecast VTEC is closer to that of the MIT VTEC observation, especially in the equatorial ionization anomaly (EIA) area. At 0800 UT (3 hr forecast), the EnKF forecasting VTEC still has better performance in terms of EIA amplitude than the control one. However, the forecasting improvements of VTEC of experiment 1 become less obvious at 1000 UT (5 hr forecast) and 1200 UT (7 hr forecast). This means that the updated O+ modified by data assimilation process has approximately returned to the control state. Therefore, it demonstrates that the ionosphere forecasting timescale can only last for about 3 hr when only the ionosphere state variable O+ is initialized. For experiment 2, at 0600 and 0800 UT, the EnKF forecasting VTEC is closer to observations than the TIEGCM results. However, unlike experiment 1 at 1000 UT and 1200 UT, there exists relatively a closer value between the experiment 2 and MIT VTEC at the corresponding UTs, especially in the EIA area. In other words, the ionosphere forecasting capability is enhanced when the ionosphere and thermosphere variables are simultaneously initialized.

In order to study how long the predicted ionosphere by experiment 2 return to the control state, Figure 3 further gives the local time and magnetic latitude variation of the deviation of forecasting VTEC in experiment 2 and the control results from MIT VTEC observations during the whole day of 24 March 2010. There is a relatively larger mean deviation and root-mean-square error (RMSE) of TIEGCM predicted VTEC from observations during the daytime. Note that the daytime is referred to as the time between 0600 and 1800 local time, and the other time is defined as nighttime in this paper. The mean deviation and RMSE of VTEC at daytime is 9.57 TECU and 11.76 TECU, respectively, for TIEGCM. However, there is an obvious decrease in mean deviation and RMSE for the 1 hr EnKF VTEC forecast in comparison with
TIEGCM prediction. The mean deviation of VTEC is 5.13 TECU, and the RMSE is 8.09 TECU. In comparison with the 1 hr forecasting results, the value of mean deviation and RMSE of VTEC increases at 3 and 12 hr forecasts but still smaller than that of the control one. Even at the 24 hr forecasting, the value of mean deviation and RMSE of VTEC is 8.70 TECU and 10.73 TECU, respectively, which is still a little smaller than control results. It indicates that the effect of initializing neutral atmosphere on forecast could last 24 hr or even longer. Meanwhile, we can find a tiny improvement of VTEC forecasting by EnKF during the nighttime. The mean deviations and RMSE of TIEGCM prediction VTEC from observation are 3.10 TECU and 6.46 TECU, which are 3.04 TECU and 6.35 TECU for the 1 hr forecasting results. To get a clear depiction of how the forecast skill evolves over the 24 hour forecast period for different local time and different latitudes, Figure 4 presents the mean deviation ratio of forecasting VTEC of experiment 2 from MIT VTEC values for 24 h forecast period in three different magnetic latitude regions for experiment 2 during the whole day of 24 March 2010. Note that the mean deviation ratio is defined as

**Figure 3.** The mean deviation (the left panels) and RMSE of the deviations (the right panels) of global VTEC versus magnetic latitude—local time for TIEGCM results without data assimilation (the first row) and the forecast results of experiment 2 at different hours (from the second to fifth rows) during the whole day of 24 March 2010.
where MIT VTEC represents the global VTEC observation data and the forecast VTEC denotes the forecast results from experiment 2, respectively. From this figure, it is clear that there exists a smaller mean deviation ratio at the daytime in all latitude regions even at 24 hr forecast. Meanwhile, we can find that there exists an obvious larger mean deviation ratio at the nighttime than that at the daytime in all latitude regions, even for 1 hr forecast. It may be due to the fact that the assimilation is less effective at night and further makes a negative impact on the short-term forecasting results at night. Furthermore, we assimilate the global slant TEC simultaneously for both daytime and nighttime in our EnKF algorithm (He et al., 2019). Given that the nighttime TEC amplitude is much smaller than daytime, it is reasonable to expect that the assimilation effect during nighttime is less significant than daytime in relative sense.

Figure 5 further gives the statistical VTEC RMSE of the deviation of the control and assimilation forecast results of experiment 2 from MIT VTEC observations at 1300 LT and 2100 LT during 24–27 March 2010. According to this figure, at 1300 LT, there exists a large VTEC difference between the control results and the MIT VTEC observations especially at the midlatitude and low-latitude region, with the average VTEC RMSE of 15.21 TECU. Such deviation has an obvious reduction for 1 hr forecasting, which is 10.1 TECU. For the 3, 12, and 24 hr forecasting, the value of VTEC RMSE has an increase, but still much smaller than the control one, which is 11.90 TECU, 13.4 TECU, and 13.6 TECU, respectively. It again demonstrates that the daytime ionosphere forecasting timescale can be enhanced through updating thermosphere state.
variables. In addition, the comparison results at 2100 LT were also presented here for comparison. We find no improvement of forecasting VTEC during nighttime as indicated in Figures 3 and 4. The overall RMSE is 5.30 TECU for the forecasting VTEC during the nighttime even at the 1 hr forecast in comparison with 5.21 TECU for the TIEGCM control run.

The ionosonde observed foF2 from National Oceanic and Atmospheric Administration’s National Geophysics Data Center is also used to make an independent validation of the forecasting results by EnKF. Figure 6 gives an example of the foF2 comparison of the control and the forecast results of experiment 2 at different forecast hours with ionosonde observations at Brisbane station (152.92°E, −27.53°N) from the day of 24 to 27 March 2010. The overall control results underestimate the value of realistic foF2 observations. Meanwhile, it can be seen that the overall 1 hr EnKF forecasting results are obviously closer to the ionosonde observations. The overall average deviation of TIEGCM foF2 prediction from observations is 1.68 MHz, whereas that is 0.76 MHz for the 1 hr forecasting foF2. At 3 and 12 hr forecasts, the mean deviation is 0.83 and 1.22 MHz, respectively. For the 24 hr EnKF forecasting, the mean deviation is 1.38 MHz, which still is lower than the control one. This again demonstrates the enhanced forecasting timescale by updating the thermosphere state variables. Moreover, the global ionosonde network critical frequency observations from the corresponding days are also used to evaluate the forecasting results. Figure 7 shows the mean foF2 results of the TIEGCM control, EnKF forecasting results of experiment 2, and all available ionosonde observations during the quiet day from 24 to 27 March 2010. Note that the TIEGCM and forecasting results are interpolated to the locations of ionosondes. It can be seen that there is significantly underestimation of the foF2 predicted by TIEGCM in comparison with the ionosonde observations. The mean deviation of the
control case is 1.22 MHz. However, for the 1 and 3 hr $f_{o}F_2$ forecasting, there is a much closer value to the realistic $f_{o}F_2$ observations, and the mean deviation of which is 0.63 and 0.89 MHz, respectively. At 12 and 24 hr forecasts, the value of the EnKF $f_{o}F_2$ forecasting is still closer to the realistic $f_{o}F_2$ data than the control one, and the mean deviation of which is 0.98 and 1.09 MHz, respectively. This means that there exists still an $f_{o}F_2$ accuracy improvement even though the advanced forecasting time extends to 24 hr. Similar to Figure 3, it can be also found that the main improvement of forecasting $f_{o}F_2$ occurs during the daytime and there is little improvement relative to the ionosonde observations at nighttime.

The above results have demonstrated that the ionospheric forecasting capability is enhanced due to the optimization of neutral backgrounds via EnKF method. To evaluate independently, the comparison between the EnKF $[O/N_2]$ forecasting and TIMED/GUVI $[O/N_2]$ observations has been made. As indicated above, the TIMED/GUVI $[O/N_2]$ is derived by taking the ratio of O column density to N$_2$ column density above an altitude where the N$_2$ column density is up to 10$^{17}$ cm$^{-2}$ (Strickland et al., 1995). We also calculate $[O/N_2]$ from the model grid based on the same geometry and then interpolated to the location of observations. Figure 8 gives the comparison results of the EnKF forecasting $[O/N_2]$ of experiment 2 and the control one with TIMED/GUVI measurements at around 1000 LT. As indicated, there exists significant difference between the control $[O/N_2]$ and the TIMED/GUVI $[O/N_2]$ observations. The disagreement has been significantly reduced for the 1 hr $[O/N_2]$ forecasting results. Especially over high-latitude region, the $[O/N_2]$ depletion in the observation is well reproduced by the EnKF optimization, though it is not obvious in the TIEGCM control run. The overall average deviation of global $[O/N_2]$ produced by EnKF forecasting results from

---

**Figure 6.** The $f_{o}F_2$ comparison between ionosonde observations (red dots), TIEGCM (gray dashed line), and the different hour forecast results of experiment 2 (green dashed line) at Brisbane station (152.92°E, −27.53°N) from the day of 24 to 27 March 2010.
TIMED/GUVI observations is 0.13, which is 0.19 for TIEGCM results. However, it can be seen that degrading [O/N_2] forecasting results exist in the low-latitude region. The potential reason is that the coupling between the electron density and [O/N_2] in the low latitudes is less significant due to the complex electrodynamics and other processes. The overall amplitude and distribution feature of the global forecasting [O/N_2] is still in alignment with the TIMED/GUVI [O/N_2] observations at 3 and 12 hr forecasts. The corresponding average deviation of global mean [O/N_2] by forecasting result is 0.13 and 0.12, respectively. We can find that even at the 24 hr forecasting, the mean deviation of global forecasting [O/N_2] is 0.14. This well demonstrates that EnKF can optimize the thermosphere states (can be also called ionospheric drivers) through the correlative relationship between these states and electron densities through the physical mechanisms included in the theoretical model. Meanwhile, it suggests that the updated thermosphere state variables have a longer “memory” time to recovery to the control state and thereby contributes to the ionosphere long timescale forecasting capability.

4. Discussion

The results show that the ionosphere forecast capability can be enhanced by initializing both thermosphere and ionosphere state variables via EnKF data assimilation method. As was shown on the majority of previous studies in observation system simulation experiments (Chartier et al., 2013; He et al., 2019; Hsu et al., 2014), updating the thermosphere states (Tn, O, O_2) simultaneously has more impact on the ionosphere electron density prediction than the ionosphere state variables only (e.g., O^+). The reason is that

![Figure 7. The mean f_0F_2 versus LT and magnetic latitude between global ionosonde observations (the left top), TIEGCM (right top), and the different hour forecast results of experiment 2 (the second and third rows) during 24 to 27 March 2010.](image_url)
the thermosphere neutral compositions have a longer “memory” in comparison with the ionosphere states and that ionosphere electron density of the daytime F region is mainly controlled by the production and loss process of neutral compositions (Rishbeth, 1998).

The results concluded in previous researches that the thermosphere neutral compositions can be well optimized through assimilating the ionosphere observations were demonstrated through the comparison of the EnKF forecasting \([O/N_2]\) with the realistic TIMED/GUVI observations. The long recovery time to the control states can be found in section 3 and can give the reasonable explanation on the enhanced ionosphere forecast capability. However, based on Figure 8, the degraded thermosphere composition can be found in the low-latitude region. But the forecast skill has shown significant improvement in low-latitude region. The main reasons include (1) the forecast skill do improve more significant in low latitude when looking at the absolute amplitude since the low-latitude ionospheric electron density is much larger than other latitudes. However, as shown in Figure 4, when looking at the relative deviation, low-latitude improvement is not that significantly different from other latitudes. (2) In low-latitude ionosphere, electrodynamic process plays a significant role in ionospheric electron density generation and re-distribution. However, these parameters are not adjusted in our algorithm rather than neutral composition. Therefore, the adjusted \([O/N_2]\) has degraded performance in low-latitude region. (3) Although the optimized \([O/N_2]\) has degraded performance, the forecast skill in low latitude still has improvement in low latitude. This is due to the fact that

![Image](image_url)
there exist ambiguities between different ionospheric drivers. When only adjusting neutral composition in low latitude in EnKF, other drivers’ effect, such as ExB drift, will be considered to some extent by adjusting neutral composition. This will result in “worse” optimized \([O/N_2]\) but better electron density forecast, as is the situation in our paper.

Note that the realistic indices, such as \(F_{10.7}\) and Kp, were used as the TIEGCM drivers to produce each ensemble experiment in this paper. We refer to the EnKF experiment results as quasi-realistic ionosphere forecasts, which are due to the fact that a true realistic forecast would use predicted solar drivers as well as only observations that are available in near real time. Meanwhile, although the realistic geomagnetic indices are used in this paper, the EnKF ionosphere forecast results cannot be completely consistent with the observation even for 1 hr forecasting results, which can be found in the comparison results of ionosphere VTEC and \(f_0F_2\). Realistic thermosphere and ionosphere system is also controlled by other boundary conditions, such as high-latitude convection pattern and auroral particles precipitation, and lower atmospheric tides. This system in response to such boundary conditions is far more complex than what can be represented by the two parameters \((F_{10.7}\) and Kp). In the EnKF, the model error will increase during the assimilation process due to the misrepresentative of the model uncertainty from the inaccuracy of boundary conditions and model forcing (Matsuo et al., 2013). Thus, in comparison with previous OSSE results, there is a less effective improvement of ionosphere forecasting results when the forecasting timescale extends to be 24 hr. In order to prepare the purpose of the near real-time operation and obtaining more accurate ionosphere forecasting results in the future, more external driving forces of the TIEGCM need to be estimated.

5. Summary

In this study, both the ground- and space-based slant TECs have been assimilated into physical model TIEGCM through EnKF method to evaluate the quasi-realistic ionosphere forecasting capability during the geomagnetic quiet conditions. The MIT VTEC observations and ionosonde-based critical frequency have been used to make independent validation. All the comparison shows that the short-term ionosphere forecasting capability can be enhanced when simultaneously initializing the ionosphere and thermosphere states. In addition, the thermosphere variables \([O/N_2]\) optimized by EnKF forecasting has also been made a cross comparison with TIMED/GUVI data. Our validation shows that the EnKF forecasting \([O/N_2]\) is more consistent with real observations in comparison with the TIEGCM control case especially in high latitudes. The optimized \([O/N_2]\) can be used to explain the enhanced ionosphere short-term forecasting timescale.

References

Bust, G. S., Garner, T. W., & Gaussiran, T. L. J. II (2004). Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density simulation algorithm. *Journal of Geophysical Research, 109*, A11312. https://doi.org/10.1029/2003JA010234

Chartier, A. T., Jackson, D. R., & Mitchell, C. N. (2013). A comparison of the effects of initializing different thermosphere-ionosphere model fields on storm time plasma density forecasts. *Journal of Geophysical Research: Space Physics, 118*, 7329–7337. https://doi.org/10.1002/2013JA019034

Chartier, A. T., Matsuo, T., Anderson, J. L., Collins, N., Hoar, T. J., Lu, G., et al. (2016). Ionospheric data assimilation and forecasting during storms. *Journal of Geophysical Research: Space Physics, 121*, 764–778. https://doi.org/10.1002/2014JA020799

Chen, C., Lin, C., Chen, W., & Matsuo, T. (2017). Modeling the ionospheric prereversal enhancement by using coupled thermosphere-ionosphere data assimilation. *Geophysical Research Letters, 44*, 1652–1659. https://doi.org/10.1002/2016GL071812

Chen, C. H., Lin, C. H., Matsuo, T., Chen, W. H., Lee, I. T., Liu, J. Y., et al. (2016). Ionospheric data assimilation with thermosphere-ionosphere electrodynamics general circulation model and GPS-TEC during geomagnetic storm conditions. *Journal of Geophysical Research: Space Physics, 121*, 5708–5722. https://doi.org/10.1002/2015JA021787

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research, 99*(CS), 10,143–10,162. https://doi.org/10.1029/94JC00572

Galkin, I. A., Reinisch, B. W., Huang, X., & Bilitza, D. (2012). Assimilation of GIRO data into a real-time thermosphere-ionosphere-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Space Physics, 117*, A11315. https://doi.org/10.1029/2012JA017725

Gardner, L. C., Schunk, R. W., Scherliss, L., Sjoka, J. J., & Zhu, L. (2014). Global Assimilation of Ionospheric Measurements-Gauss Markov model: Improved specifications with multiple data types. *Space Weather, 12*, 675–688. https://doi.org/10.1002/2014SW001104

He, J., Yue, X., Wang, W., & Wan, W. (2019). EnKF ionosphere and thermosphere data assimilation algorithm through a sparse matrix method. *Journal of Geophysical Research: Space Physics, 124*, 7356–7365. https://doi.org/10.1029/2019JA026554

Heeles, R. A., Lowell, J. K., & Spiri, R. W. (1982). A model of the high-latitude ionospheric convection pattern. *Journal of Geophysical Research, 87*(A8), 6339–6345. https://doi.org/10.1029/JA087iA08p06339

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review, 126*(3), 796–811. https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAE3.0.CO;2

Hsu, C., Matsuo, T., Wang, W., & Liu, J. (2014). Effects of inferring unobserved thermospheric and ionospheric state variables by using an Ensemble Kalman Filter on global ionospheric specification and forecasting. *Journal of Geophysical Research: Space Physics, 119*, 9256–9267. https://doi.org/10.1002/2014JA020390
Hsu, C., Matsuo, T., Yue, X., Fang, T., Fuller-Rowell, T., Ide, K., & Liu, J. Y. (2018). Assessment of the impact of FORMOSAT-7/COSMIC-2 GNSS RO observations on mid- and low-latitude ionosphere specification: Observing system simulation experiments using ensemble square root filter. Journal of Geophysical Research: Space Physics, 123, 2296–2314. https://doi.org/10.1002/2017JA025109

Lee, I. T., Matsuo, T., Richmond, A. D., Liu, J. Y., Wang, W., Lin, C. H., et al. (2012). Assimilation of FORMOSAT-3/COSMIC electron density profiles into a coupled thermosphere/ionosphere model using ensemble Kalman filtering. Journal of Geophysical Research, 117, A10318. https://doi.org/10.1029/2012JA017790

Matsuo, T., Lee, I., & Anderson, J. (2013). Thermospheric mass density specification using an ensemble Kalman filter. Journal of Geophysical Research: Space Physics, 118, 1339–1350. https://doi.org/10.1002/jgra.50162

Meier, R., Crowley, G., Strickland, D. J., Christensen, A. B., Paxton, L. J., Morrison, D., & Hackert, C. L. (2005). First look at the 20 November 2003 superstorm with TIMED/GUVI: Comparisons with a thermospheric global circulation model. Journal of Geophysical Research, 110, A09S41. https://doi.org/10.1029/2004JA010990

Pedatella, N. M., Liu, H.-L., Marsh, D. R., Raeder, K., Anderson, J. L., Chau, J. L., et al. (2018). Analysis and hindcast experiments of the 2009 sudden stratospheric warming in WACCMX+DART. Journal of Geophysical Research: Space Physics, 123, 3131–3153. https://doi.org/10.1002/2017JA025107

Pi, X., Wang, C., Hajj, G. A., Rosen, G., Wilson, B. D., & Bailey, G. J. (2003). Estimation of ExB drift using a global assimilative ionospheric model: An observation system simulation experiment. Journal of Geophysical Research, 108(A2), 1075. https://doi.org/10.1029/2001JA009235

Richmond, A. D., Ridley, E. C., & Roble, R. G. (1992). A thermosphere/ionosphere general circulation model with coupled electrodynamics. Geophysical Research Letters, 19(6), 601–604. https://doi.org/10.1029/GL019i006p00601

Rishbeth, H. (1998). How the thermospheric circulation affects the ionospheric F2 layer. Journal of Atmospheric and Solar-Terrestrial Physics, 60, 1385–1402. https://doi.org/10.1016/S1364-6828(98)00062-5

Schunk, R. W., Scherliess, L., Sejka, J. J., Thompson, D. C., Anderson, D. N., Codrescu, M., et al. (2004). Global Assimilation of Ionospheric Measurements (GAIM). Radio Science, 39, RS1S02. https://doi.org/10.1029/2002RS002794

She, C., Yue, X., Hu, L., & Zhang, F. (2019). Estimation of ionospheric total electron content from a multi-GNSS station in China. IEEE Transactions on Geoscience & Remote Sensing, 57(2), 852–860. https://doi.org/10.1109/TGRS.2019.2941049

Stephan, A. W., Meier, R. R., & Paxton, L. J. (2008). Comparison of Global Ultraviolet Imager limb and disk observations of column O/N₂ during a geomagnetic storm. Journal of Geophysical Research, 113, A01301. https://doi.org/10.1029/2007JA012599

Strickland, D. J., Evans, J. S., & Paxton, L. J. (1995). Satellite remote sensing of thermospheric O/N₂ and solar EUV. Journal of Geophysical Research, 100, 12, 217–12. 226. https://doi.org/10.1029/95JA00574

Wang, C., Hajj, G., Pi, X., Rosen, I. G., & Wilson, B. (2004). Development of the global assimilative ionospheric model. Radio Science, 39, RS1S06. https://doi.org/10.1029/2002RS002854

Yue, X., Schreiner, W. S., Hunt, D. C., Rocken, C., & Kuo, Y.-H. (2011). Quantitative evaluation of the low Earth orbit satellite based slant total electron content determination. Space Weather, 9, S09001. https://doi.org/10.1029/2011SW000687

Yue, X., Schreiner, W. S., Kuo, Y. H., Braun, J. J., Lin, Y., & Wan, W. (2014). Observing system simulation experiment study on imaging the ionosphere by assimilating observations from ground GNSS, LEO-based radio occultation and ocean reflection, and cross link. IEEE Transactions on Geoscience & Remote Sensing, 52(7), 3759–3773. https://doi.org/10.1109/TGRS.2013.2275753

Yue, X., Wan, W., Liu, L., Zheng, F., Lei, J., Zhao, B., et al. (2007). Data assimilation of incoherent scatter radar observation into a one-dimensional midlatitude ionospheric model by applying ensemble Kalman filter. Radio Science, 42, RS6006. https://doi.org/10.1029/2007RS003631