The Impact of Assimilating Ionosphere and Thermosphere Observations on Neutral Temperature Improvement: Observing System Simulation Experiments Using EnKF

Jianhui He¹,²,³, Xinan Yue¹,²,³, and Zhipeng Ren¹,²,³

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

Abstract Accurate specification of the thermosphere states is crucial to the low Earth orbit satellite operation. In this work, the impact of different ionosphere and thermosphere observing systems on the improvement of neutral temperature of the data assimilation model has been investigated by a series of observing system simulation experiments. The selected observations include the Global Navigation Satellite System total electron content (e.g., MIT vertical total electron content [VTEC]) and the daytime Global-scale Observations of the Limb and Disk (GOLD) level-2 disk temperature ($T_{\text{disk}}$). Such observations are ingested into the coupled ionosphere and thermosphere model based on our developed ensemble Kalman Filter data assimilation systems on the basis of the ensemble Kalman filter algorithm and the National Center for Atmospheric Research Thermosphere Ionosphere Electrodynamics General Circulation Model. The main findings are as follows: (a) A considerable improvement of the neutral temperature estimation of the physical-based model can be obtained in the global region by assimilating either the MIT VTEC or the GOLD $T_{\text{disk}}$ observations; (b) the assimilation of the GOLD can further contribute to temperature improvement in the lower thermosphere (<200 km), relative to the MIT VTEC assimilation; and (c) simultaneously assimilating both observation types can better improve the quality of neutral temperature estimation over the global area during the whole data assimilation process. The current results demonstrate that assimilating GOLD observations is important to improve the forecast capability of the physical-based model for the lower thermosphere states and can provide a possible reference for the joint assimilation of the ionosphere and thermosphere observations to better thermosphere specification.

Plain Language Summary In this study, the Global-scale Observations of the Limb and Disk (GOLD) $T_{\text{disk}}$ and the MIT vertical total electron content observations have been assimilated into the Thermosphere Ionosphere Electrodynamics General Circulation Model through the ensemble Kalman Filter data assimilation method. The impact of different ionosphere and thermosphere observations on the improvement of neutral temperature estimation of the physical-based model has been evaluated in detail. We found that the assimilation of the GOLD $T_{\text{disk}}$ is necessary for further temperature improvement in the lower thermosphere. Meanwhile, better thermosphere temperature estimation results can be obtained by simultaneously assimilating both observation types during the whole data assimilation period.

1. Introduction

The Earth's thermosphere refers to a region of the atmosphere from around 85–1,000 km. This region is crucial for human space exploration in the current space age, which has a nonnegligible drag effect on the low-orbit spacecraft operation (Emmert, 2015; Marcos, 2006; Sutton, 2018). The spatial-temporal variability of the thermosphere is primarily controlled by the process of solar and geomagnetic activity and lower atmosphere waves, etc. As a nonlinearly coupled complex system, its variability depends on not only the charged particles but also the neutral parameters, such as the neutral temperature, compositions, and density (Shim et al., 2014). Thus, it is challenging to characterize accurately the global or local variations in the thermosphere, especially during space weather events.
In the past decades, great efforts have been paid to get a clear understanding of the variability of the thermosphere parameters during the quiet day and geomagnetic disturbance period through the physical-based general circulation model (Liu et al., 2018; Millward et al., 1996; Qian et al., 2016; Richmond et al., 1992; Ridley et al., 2006) such as the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIEGCM) and Whole Atmosphere Community Climate Model with thermosphere ionosphere eXtension (WACCMx). These models can self-consistently give a spatial and temporal description of the thermosphere states using a finite difference solution to the conservation equations of momentum, mass, and energy. However, these general circulation models exhibit significant deviation from the realistic thermosphere data. Based on the assumption of static equilibrium, the temperature distribution results of those thermospheric theoretical models can be first calculated. Then, the atmospheric mass density is finally obtained according to the relationship between temperature and neutral composition. Therefore, it is a feasible solution to reduce the error of atmospheric density calculation by improving the accuracy of the simulated neutral temperature in those theoretical models.

Data assimilation has been proven to be a powerful tool in specifying and forecasting highly dynamic systems like Earth’s thermosphere. This technique can provide a relatively accurate initial field for a numerical forecast model combining the background value and observations (Daley, 1991). Currently, the ensemble Kalman Filter (EnKF) data assimilation system has been well used in the specification of the upper atmosphere states, including Data Assimilation Research Testbed and the Gridpoint Statistical Interpolation (Anderson et al., 2009; Hsu et al., 2018; Lee et al., 2012). The feature of the EnKF algorithm is the flow-dependent background error covariance estimation in each assimilation cycle (Evensen, 1994). These software packages have been well examined by assimilating the massive ground-based and space-based ionospheric observations. A good nowcast and forecast capability of the ionosphere states can be achieved in near real-time. However, those systems have not been directly used to assimilate thermosphere-related observations to specify the thermosphere states. This is due to the fact that the previous low-Earth orbiting satellite-based measurements can only obtain the limited thermosphere data at the fixed local time, such as the Challenging Mini-Satellite Payload (CHAMP), the Gravity Recovery And Climate Experiment (GRACE), and Thermosphere, Ionosphere, Mesosphere Energetics, and Dynamics/Global Ultraviolet Imager (TIMED/GUVI). For example, Matsuo et al. (2013) assimilated the neutral mass density from the CHAMP mission into the TIEGCM. They found that the improvement of temperature and major compositions (e.g., O and O2) is only in the vicinity of satellite orbits, but is short of making a global impact.

According to the relationship between the observed and unobserved state vectors captured in the background error covariance in the EnKF algorithm, some previous studies indirectly specify the thermosphere parameters, such as the neutral winds, mass density, and temperature, through assimilating the massive ionosphere observations (Chen et al., 2016, 2017; He, Yue, Le, et al., 2020; Matsuo et al., 2013). Chen et al. (2017) used the EnKF to assimilate ionosphere total electron content (TEC) observations into the TIEGCM. They found that the thermospheric winds, temperature, and compositions can be adjusted simultaneously resulting in improved specification and forecast of eastward pre-reversal enhanced electric field. Additionally, Chen et al. (2016) found that the thermospheric \([O/N_2]\) can be optimized to some extent by EnKF in comparison with TIMED/GUVI observations when assimilating ionosphere TEC observations. Meanwhile, the previous studies related to the longer forecast time of ionosphere electron density also suggest that there exists a good thermosphere state adjustment (e.g., neutral temperature and compositions) when just assimilating ionosphere observations (He et al., 2019; Hsu et al., 2014; Pedatella et al., 2020).

Recently, the National Aeronautics and Space Administration (NASA) Global-scale Observations of the Limb and Disk (GOLD) mission with a far-ultraviolet imager onboard the SES-14 satellite has been launched in the geostationary orbit at 47.5°W longitude. For the first time, a synoptic view of the thermospheric states in those theoretical models can be retrieved at a 30-min cadence from about 06:10 to 00:40 universal time each day (Eastes et al., 2020). Such unprecedented spatial and temporal coverage of thermosphere data can provide a rare opportunity for directly performing the thermosphere observations data assimilation through the EnKF combined with the physical-based model. Some researchers have already shown that GOLD thermospheric observations are an excellent data set used for thermospheric assimilation studies and operational purposes (Eastes et al., 2020; Laskar, Pedatella, et al., 2021). Cantrall et al. (2019) presented assimilation experiments of simulated GOLD LBH emission data using an ensemble filter. They found that assimilation of emission data can significantly
reduce the bias in model temperature specification under both geomagnetic quiet and disturbed conditions. Laskar, Pedatella, et al. (2021) investigated the improvement of forecast capability of the WACCMx by assimilating GOLD disk temperatures and other conventional lower atmospheric observations. They found that the forecast capability of the WACCMx can be improved by assimilating the GOLD temperatures.

The EnKF data assimilation system developed by He et al. (2019) has the capability of assimilating common ionosphere observations to specify and forecast ionosphere parameters, while thermosphere observations have not yet been included in this system. Therefore, in this study, we will augment this system with current thermosphere observations, for example, the GOLD disk temperature. Meanwhile, the impact of assimilating different ionosphere and thermosphere observations on the accuracy improvement of the simulated neutral temperature in the background theoretical model will be assessed in detail by a series of observing system simulation experiments (OSSEs). The ground-based ionospheric MIT vertical total electron content (VTEC) observations and the GOLD daytime disk temperature will be assimilated into the physical-based TIEGCM through our developed EnKF data assimilation system. The remainder of the study is organized as follows. In Section 2, the OSSE including simulated data and assimilation algorithm will be introduced 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 Method and Experiments Description

2.1. EnKF Data Assimilation System

The coupled thermosphere and ionosphere EnKF data assimilation system used in this study was developed by He et al. (2019), which was implemented with the first-principle model NCAR-TIEGCM as the numerical forecast model and the ensemble Kalman filter as the data assimilation algorithm. The background model has a default resolution of 5° × 5° in longitude and latitude, half-scale height in altitude, and 180 s in time step. The time evolution of the global distribution of the thermosphere and ionosphere parameters can be self-consistently produced by the specified solar/geomagnetic activity index and lower boundary tides. The feature of this data assimilation algorithm is that the Kalman gain is calculated by a sparse matrix method and iteration solving linear equations. The background error covariance is estimated from the finite ensemble members in each assimilation cycle. This EnKF data assimilation system can be conducted without using a supercomputer for a fast calculation. Currently, this system has been well used in specifying and forecasting global and regional ionosphere states (He, Yue, Hu, et al., 2020; He, Yue, Le, et al., 2020).

2.2. Observation Operator

The data assimilation requires an observation operator (H) that maps the background model states to the real-world observations. In the current work, the synthetic ionosphere VTEC observations and the GOLD daytime disk temperature will be assimilated into the TIEGCM using the EnKF method. The MIT VTEC are from the MIT Haystack Observatory, which can be publicly available at the Madrigal Database (http://cedar.openmadrigal.org/). It is provided globally in 1° by 1° bin in latitude and longitude and has a 5 min time resolution (Rideout & Coster, 2006). For this observation type, it is the integrated electron density along the vertical direction raypath between the GNSS satellites and the ground-based receivers. The observation operator (H) for VTEC can be built based on the following relationship,

\[
\text{VTEC} = \sum_{i=1}^{n} N_e \Delta S_i
\]

where VTEC is the synthetic vertical total electron content of the TIEGCM at the location of real-world observations. Ne stands for the electron density in the ith grid in the corresponding raypath. The \(\Delta S_i\) represents the vertical direction raypath length within the ith grid.

The level-2 disk temperatures (\(T_{disk}\)) provided by GOLD can be publicly obtained from the GOLD Science Data Center (https://gold.cs.ucf.edu/). This data can be available between about 06 and 23 UT each day, which covers about 1/4 globe over about ±70° latitude and 30°E–120°W longitude. It is retrieved by fitting the observed rotational structure of the \(N_\lambda\) Lyman-Birge-Hopfield (LBH) bands emission that are mainly from the lower thermosphere (100–200 km). Thus, the \(T_{disk}\) can be considered to be the integrated
temperature through the weight of the temperature at different heights, and the maximum weight is at a height of approximately 160 km (Eastes et al., 2020). According to the contribution function (weight) suggested by Laskar, Pedatella, et al. (2021), the observation operator of $T_{\text{disk}}$ can be represented as,

$$ T_{\text{disk}} = \sum_{i=1}^{n} T_{i} CF_i $$

(2)

$$ CF_i = \left( \frac{A}{H_i} \right) \times e^{\frac{-(H_i-\mu)^2}{2\sigma^2}} $$

(3)

For Equations 2 and 3, the $T_{\text{disk}}$ is the synthetic disk temperature of the TIEGCM at the location of the real GOLD observations. $T_i$ and $CF_i$ stand for the temperature and the contribution function at the $i$th grid, respectively. $A$, $\mu$, and $\sigma$ represent the amplitude, mean, and standard deviation for $H_i$. Note that our developed EnKBF data assimilation system uses the height as the vertical coordinate. Thus, $H_i$ represents the corresponding height of the $i$th grid.

2.3. Observing System Simulation Experiments Setup

To evaluate the impact of different ionosphere and thermosphere observing systems on the improvement of thermosphere temperature estimation of the physical-based model, three OSSEs are conducted in the current work, which are listed in Table 1, henceforth referred to as the OSSE 1, OSSE 2, and OSSE 3. The synthetic observations are used in all OSSEs, which include the ionosphere MIT VTEC observations and GOLD $T_{\text{disk}}$ neutral temperature as described in Section 2.2. These synthetic observations are interpolated from the simulated “truth” case based on their corresponding observation configuration. According to Table 1, the OSSE 1 assimilates the synthetic MIT VTEC observations. The OSSE 2 assimilates the synthetic GOLD $T_{\text{disk}}$ observations. The OSSE 3 is conducted by simultaneously assimilating both observation types. Note the synthetic MIT VTEC data without observation errors. The observation error of such synthetic data is considered to be small, for example, the ground-based GNSS slant TEC with an error of $\sim 1$–3 TECU (Mannucci et al., 2005). It means that the analysis and forecast quality of data assimilation results will not be a noticeable change even if such errors are added to such synthetic observations. Due to the higher signal-to-noise ratio of real GOLD $T_{\text{disk}}$ observations, the observational errors ($\varepsilon$) are added to such synthetic observations, which are based on a centered Gaussian distribution with the standard deviation of $\sigma$ the magnitude of a unit of observation data, namely $\varepsilon \sim N(0, \sigma^2)$. Compared with the random uncertainty indicated by Laskar, Eastes, et al. (2021), the value of $\sigma^2$ is relatively small, which is set to be 55 K in the current work. This will not make a noticeable influence on the GOLD $T_{\text{disk}}$ data assimilation results. The synthetic GOLD disk temperature observations used in the current work can be found in Figure 1.

Besides carrying out these OSSEs, two TIEGCM default run experiments without data assimilation are also conducted by different model drivers. The one TIEGCM default run, henceforth referred to as the control experiment, is driven by F10.7 = 76 sfu, CP = 30.8 kV, and HP = 18.30 GW at the December solstice. The other one TIEGCM run is referred to as the simulated “truth” experiment henceforth, which is driven by
F10.7 = 100 sfu, CP = 30.8 kV, and HP = 18.30 GW. For each OSSE, ensemble members are generated by perturbing the solar and geomagnetic index (e.g., F10.7, KP) using the centered Gaussian distributions method. The mean values of these model drivers are the same as that of the control one, with 15% standard deviations. The forcing parameters assigned to each member are held unchanged during the whole data assimilation period. Meanwhile, to avoid the effect of initial conditions, the spin-up time for the control and the simulated “truth” case and each data assimilation experiment is about two weeks.

For each OSSE, the data assimilation period is from 0000 UT to 0000 UT of next day. The assimilation time window is 1 hr. Note that the different thermosphere and ionosphere state vectors are updated in different OSSEs. The OSSE 1 is performed by simultaneously updating ionosphere and thermosphere state vectors, including the electron density (Ne) and neutral temperature (Tn). For OSSE 2, only the neutral temperature is updated. The updated state vectors of OSSE 3 are the same as that of OSSE 1.

The other key parameters configuration for the EnKF algorithm are summarized as follows. The number of ensemble members is chosen to be 60. In order to minimize spurious correlations over long spatial distances, both observation types are localized in the horizontal direction using a Gaspari-Cohn function with a half-width 10°. The vertical localization is applied to both observation types also using a Gaspari-Cohn function with a half-width 100 km.

3. Results

The impact of assimilating ionosphere VTEC observations on the improvement of temperature estimation of the physical-based model is first presented here. Before investigating the improvement level of thermosphere temperature in OSSE 1, the comparison of the data assimilation results of ionosphere analysis NmF2 with the control case are first shown in Figure 2. According to the figure, it can be found that the NmF2 deviations have an obvious reduction at all selected times, relative to that of the control case. It illustrates...
that VTEC observations have been well assimilated into the TIEGCM. Then, Figure 3 gives the adjusted thermosphere temperature results through assimilating ionosphere observations in OSSE 1. To demonstrate the neutral temperature can be adjusted by assimilating ionosphere data, the left panel of Figure 3 shows the longitude and latitude distribution of the absolute deviation of the control, estimation, and analysis thermosphere temperature for OSSE 1 from the simulated “truth” case at ZP = 2 (about 300–310 km) at 1600 UT. Note that the prior 1 hr forecast results of the TIEGCM at each data assimilation cycle are referred to as the estimation results in this study. To illustrate there still exist an improved accuracy of the simulated neutral temperature in background theoretical models, the right panel shows the time evolution of the global mean estimated temperature at ZP = 2 during the whole data assimilation period. From the left panel of this figure, it can be found that there exists the biggest global mean neutral temperature difference between the simulated “truth” and control one. It is due to the different F10.7 values used in driving the TIEGCM run. The difference of the estimated (analysis) temperature has an obvious decrease. The overall average absolute deviation is 90.62 K for the control one, which is 40.07 and 20.62 K for the estimation and analysis one, respectively. According to the right panel, the global mean values of the estimated neutral temperature at about 300–310 km have a rapid increase and tend to be stable near 12 UT, which is relatively closer to that of the simulated “truth” case at the end of the data assimilation period. It is due to the fact that the background model has been up to thermodynamic equilibrium at this stage. The overall mean value of the estimated neutral temperature, relative to that of the control one, improved by about 47%. This illustrates that the accuracy of the simulated neutral temperature in the physical-based model can be improved to some extent through assimilating electron densities. Meanwhile, the higher temperature differences at many locations can also be found, which could be due to the following two reasons. The first reason may be that there are lack of data in the ocean area during the daytime. The second one may be that the improvement of the estimated temperature is limited in some regions during the nighttime.

Similarly, Figure 4 shows the GOLD disk temperature comparison among the simulated “truth,” control, estimation, and analysis results at 1600 UT (the left panel), and the time evolution of the global mean estimated temperature at ZP = −2 (about 160–170 km) during the whole data assimilation period (the right panel). It is noteworthy that the GOLD disk temperature is calculated from the model grid based on the same geometry and then interpolated to the location of realistic observations. From the left panel of Figure 4, it can be found that there exists a smaller temperature value of the control one than that of the simulated “truth” case. The values of the estimated and analysis GOLD disk temperature have a great improvement, which is much more consistent with the simulated “truth” case. From the right panel, the values of the global mean estimated temperature at about 160–170 km are equal to that of the control one from the 0000 UT to 0600 UT, which is due to the fact that there are no available GOLD disk observations during this time interval.

Figure 2. The longitude-latitude distributions of the NmF2 deviation of the control case (the top panel) and analysis results for observing system simulation experiment 1 (the bottom panel) from the simulated “truth” case during the whole assimilation period. From the left to the right are the results at 0000 UT, 0400 UT, 0800 UT, 1200 UT, 1600 UT, and 2000 UT, respectively.
The mean values of the estimated neutral temperature have gradually increased from 0700 UT to 2000 UT, which are much closer to the simulated “truth” case. For OSSE 2, the biggest improvement of the global mean estimated neutral temperature is 42%, relative to that of the control one. It means that the GOLD observations have been well assimilated into the TIEGCM via our developed EnKF data assimilation system. The forecasting capability of the neutral temperature of the physical-based model can be improved by assimilating GOLD observations. Due to the less amount of observations available, it can be found that such temperature improvement has a slight decrease from 2100 UT to 2400 UT.

From the results presented in Figures 3 and 4, it is thus expected that there exists a better thermosphere neutral temperature improvement in OSSE 3 in the whole range of heights. To make a cross-comparison of the improvement level on temperature estimation through assimilating different ionosphere and thermosphere observations, Figure 5 shows the latitude-height distribution of the estimated thermosphere temperature of control results, OSSE 1–3, and their corresponding deviation from the simulated “truth” case along 50°W at 1600 UT. The selected longitude near the sunlit position at this UT. It can be seen that thermosphere temperature value of the control one is much smaller than that of other OSSEs in the whole range of heights and latitudes. The biggest temperature difference exists between the control one and the simulated “truth” case, the average difference of which is 79.2 K. For OSSE 1, which only assimilates the VTEC observations, such a difference has an obvious reduction in all latitudes and heights, relative to the control case. The mean difference for OSSE 1 is 39.7 K. For OSSE 2, although the assimilated daytime GOLD disk temperature observations have only coverage over about ±60° latitudes, the temperature improvement can be found in all latitude and height ranges. The mean difference for OSSE 2 is 36.9 K. Meanwhile, it can be also found that there exists a bigger temperature improvement of OSSE 2 than that of OSSE 1 in the lower thermosphere.

**Figure 3.** The longitude and latitude distributions of the absolute deviation of the control, estimation, and analysis thermosphere temperature of observing system simulation experiment 1 from the simulated “truth” case at ZP = 2 (about 300–310 km) at 1600 UT (the left panel), and the time evolution of the global mean estimated temperature at ZP = 2 during the whole data assimilation period (the right panel).
heights. For OSSE 3, the optimal improvement results of thermosphere temperature can be obtained, which has the smallest deviation than other cases. The mean difference for OSSE 3 is 26.1 K. It demonstrates that a considerable improvement of the simulated neutral temperature in the TIEGCM can be obtained in the whole range of the heights and latitudes by assimilating either the VTEC or the GOLD observations. Meanwhile, simultaneously assimilating both observation types can further improve the accuracy of neutral temperature estimation.

This section further represents the temperature improvement of each OSSE at different thermosphere heights. In Figure 6, we compare the improvement ratio of the estimated thermosphere temperature for

Figure 4. The Global-scale Observations of the Limb and Disk (GOLD) disk temperature comparison among the simulated “truth,” control, estimation, and analysis results at 1600 UT (the left panel), and the time evolution of the global mean estimated temperature at ZP = −2 (about 160–170 km) during the whole data assimilation period (the right panel).
the OSSE 1–3 at ZP = −2 (about 160–170 km) during some selected UTs of the assimilation period. The improvement ratio is calculated as follows.

$$\text{improvement ratio} = \left(1 - \frac{\text{estimation}}{\text{control}}\right) \times 100$$

(4)

It can be found that an evident improvement of the estimated neutral temperature exists in all OSSEs. For the OSSE 1, due to the global coverage of the assimilated VTEC observations, the temperature improvement occurs in the global area at each selected UT. The overall mean temperature improvement is about 34.4%, 37.4%, and 39.4% at 1300 UT, 1700 UT, and 2200 UT, respectively. For the OSSE 2, a large temperature improvement can be found in the region where there are GOLD $T_{\text{disk}}$ observations assimilated. For example, the biggest improvement ratio located in the sunlit position can be up to 75.9% at 1700 UT. Meanwhile, there also exists temperature improvement in the region where there are no GOLD observations available. For OSSE 3, which simultaneously assimilates the VTEC and GOLD $T_{\text{disk}}$, the biggest temperature improvement ratio exists in a global area during the whole selected UTs in comparison with other experiments. Similar to Figure 6, Figure 7 shows the cross-comparison of the improvement ratio of the estimated temperature for OSSE 1–3 at ZP = 2 (about 300–310 km) during some selected UTs of the assimilation period. According to the results shown above, for OSSE 1, the global temperature has a larger improvement at the higher thermosphere heights than that at the lower heights, the average value of which is about 52.8%, 53.9%, and 54.0% at 1300 UT, 1700 UT, and 2200 UT, respectively. With respect to OSSE 2, we found that an evident improvement also exists in the higher thermosphere heights. The optimal thermosphere temperature estimation can be obtained in OSSE 3. According to the results described above, it again demonstrates that there exists a larger temperature improvement for assimilating the GOLD $T_{\text{disk}}$ than that for VTEC at the lower thermosphere heights. Meanwhile, the simultaneously assimilating ionosphere and thermosphere can further enhance the forecast capability of the thermosphere temperature of the physical-based model in the whole range of the heights and over the global coverage.
Figure 8 further gives the longitude-universal time distribution of the root mean square errors (RMSEs) comparison of the estimated temperature for the control run and OSSEs 1–3 in the whole range of heights and latitudes. From this figure, it can be found that the biggest neutral temperature RMSEs exist between the simulated “truth” and control one. Such a difference has an obvious reduction in OSSE 1–3. For OSSE 1, the assimilated VTEC observations have enough spatial-temporal coverage, and lower RMSEs exist at all longitudes and UTs. For OSSE 2, the estimation temperature RMSEs have an obvious reduction in the regions of available GOLD $T_{\text{disk}}$ data. Meanwhile, a considerable temperature improvement can also be found where there are no GOLD observations assimilated, for example, within 60°E to 180°E longitudes from 1300 UT to 2400 UT. The optimal thermosphere temperature estimation can be found in OSSE 3 in the time interval of 0600 UT and 2400 UT. The mean RMSEs of the estimated temperature are 68.8 K, 34.4 K, 44.2 K, and 29.6 K for control one, OSSE 1–3, respectively. It means that simultaneously assimilating different observation types can further improve the quality of thermosphere temperature specification over the global area during the whole data assimilation process.

4. Discussion

In this study, the impact of different observations, including the VTEC from ground-based GNSS receivers and the $T_{\text{disk}}$ from the GOLD mission, on the improvement of thermosphere temperature of the physical-based model has been evaluated by performing a series of OSSEs. We found that assimilating the VTEC or the GOLD $T_{\text{disk}}$ observations can improve the accuracy of the simulated neutral temperature in the background theoretical model in a global scale. There exists a larger temperature improvement in the lower
thermosphere for assimilation of the GOLD $T_{\text{disk}}$. A better thermosphere temperature improvement can be obtained over the global area during the whole data assimilation process when simultaneously assimilating both observation types.

In the previous studies, it has been well proven that the thermosphere parameters (e.g., neutral temperature and compositions) can be estimated to some extent by only assimilating electron density observations into the theoretical background model using the EnKF method (Chen et al., 2016; Hsu et al., 2014; Matsuo et al., 2013). The main reason is that the ionosphere and thermosphere are considered to be an integral part of both the analysis and forecast steps when using a coupled thermosphere and ionosphere dynamics in the EnKF. In the analysis stage of the data assimilation, the state vectors, including the ionosphere and thermosphere parameters of the background model, will be simultaneously adjusted to some extent by assimilating ionosphere observations according to the statistical relationship captured in the background error covariance. In the forecast stage of data assimilation, both ionospheric and thermospheric variables are further constrained by self-consistent nonlinear dynamics of the physical-based model. According to the results presented in the current work, it has been again demonstrated that the estimated global thermosphere temperature at different heights can be improved to some extent by only assimilating the VTEC observations (e.g., Figure 3). This has an important implication for assimilating the massive and continuous ionospheric parameters to make an operational global monitoring of thermospheric states.

Due to the fact that the GOLD disk temperature observations are retrieved by fitting the observed rotational structure of the 135.6 nm LBH bands emission, these emissions emanate primarily from the lower thermosphere around 160–200 km at daytime (Aksnes et al., 2006; Krywonos et al., 2012). It is expected that assimilating the GOLD $T_{\text{disk}}$ observations will play a positive role in the improvement of the thermosphere temperature.
temperature estimation at the lower thermosphere heights. Meanwhile, due to the effect of thermodynamic process of the physical-based model, the estimated temperature improvement is not limited to the region where the GOLD observations are available, which also occurs in other regions (e.g., Figures 4 and 5). This suggests that assimilating the GOLD detect data is important to further improve the specification of the global three-dimensional temperature structure of the coupled ionosphere and thermosphere model, especially in the lower thermosphere. In current work, the contribution function with 0° solar zenith angle is used for the GOLD $T_{\text{disk}}$ assimilation. According to Laskar, Pedatella, et al. (2021), the contribution functions with different solar zenith angles have similar shapes, and the peak altitudes vary by only $\sim 10$ km until about 55° solar zenith angle. Thus, the contribution function adopted in regions of the lower solar zenith angles will not have an obvious impact on the GOLD $T_{\text{disk}}$ data assimilation results. However, for the higher solar zenith angle, the assimilation results may be different from the results presented in the current study, especially in the edge of the disk.

As indicated above, for a general circulation model, the atmospheric mass density is calculated using the relationship between temperature and neutral composition. Meanwhile, the temperature distribution results are first calculated under the assumption of static equilibrium. It means that the better mass density results will be modeled resulting from a better specification of the global three-dimensional temperature in the coupled ionosphere and thermosphere model. According to the results presented in this study, the accuracy of the simulated neutral temperature in background theoretical models can be improved by assimilating both ionosphere and thermosphere observations. The corresponding mass density forecast capability is thus expected to be enhanced, which will be investigated in detail in the future.

It is worth noting that the data assimilation experiments conducted in the current work are all observing system simulation experiments. The OSSEs have the advantage of accurately assessing the analysis errors and forecast skill of the data assimilation system under the condition that the assimilated synthetic observations are precisely known (Errico et al., 2013; Privé et al., 2013). For example, it is shown in Figure 5 that the improvement of thermosphere estimation temperature at different heights can be clearly found in different OSSEs. Meanwhile, it is apparent in Figures 6 and 7 that the forecast capability of the physical-based model can also be greatly improved for OSSE 2 in the region where there are no available GOLD observations. In general, for OSSEs, the different models should be used to reproduce the nature run field and to conduct the forecast experiments. In this study, just like the previous studies (Hsu et al., 2014; Matsuo et al., 2013), the

Figure 8. The longitude-universal time distribution of the estimation temperature root mean square errors comparison among the control, observing system simulation experiment (OSSE) 1–3 in the whole range of the heights and the latitudes.
same models are used to reproduce the nature run field and to conduct the forecast of atmospheric states, which is the TIEGCM. Though the synthetic observations are specified with observation errors, the model bias has not been taken into consideration for those cases. Thus, the improvement of the thermosphere estimation temperature results presented in the current work would have a degradation when there is another independent model used. However, currently, it is hard to make the full set of synthetic observations from another first-principle model in the space science community. Next, the impact of assimilation of the ionosphere and thermosphere observations in the real world on the thermosphere temperature estimation will also be investigated in detail in the future.

5. Summary

In this study, we evaluate the impact of different observation types from different observation systems, which includes the VTEC from global ground-based GNSS receivers and the neutral disk temperature from the GOLD mission, on the thermosphere temperature improvement by performing a series of OSSEs. It is found that a considerable improvement of the thermosphere temperature estimation can be obtained in the global region by assimilating either the VTEC or the GOLD observations. However, there exists a larger temperature improvement for the assimilation of the GOLD in the lower thermosphere (<200 km) in comparison with the VTEC assimilation. Also, simultaneously assimilating different observation types can further improve the quality of thermosphere temperature estimation over the global area during the whole data assimilation process.

Data Availability Statement

The assimilated GOLD data can be publicly obtained from the GOLD Science Data Center (https://gold.cs.ucf.edu/). The assimilated MIT VTEC data can be publicly available at the Madrigal Database (http://cedar.openmadrigal.org/). The simulated data and results presented in this study are publicly available at https://osf.io/sk6bw/.

References

Aksnes, A., Eastes, R., Budzien, S., & Dymond, K. (2006). Neutral temperatures in the lower thermosphere from N2 Lyman-Birge-Hopfield (LBH) band profiles. Geophysical Research Letters, 33, L15103. https://doi.org/10.1029/2006GL026255

Anderson, J. L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Arelano, A. (2009). The data assimilation research testbed: A community facility. Bulletin of the American Meteorological Society, 90, 1283–1296. https://doi.org/10.1175/2009BAMS2618.1

Cantrall, C. E., Matsuo, T., & Solomon, S. C. (2019). Upper atmosphere radiation data assimilation: A feasibility study for GOLD far ultraviolet observations. Journal of Geophysical Research: Space Physics, 124, 8154–8164. https://doi.org/10.1029/2019JA026910

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

Daley, R. (1991). Atmospheric data analysis. Atmospheric data analysis. Cambridge University Press.

Eastes, R. W., McClintock, W. E., Burns, A. G., Anderson, D. N., Andersson, L., Aryal, S., et al. (2020). Initial observations by the GOLD mission. Journal of Geophysical Research: Space Physics, 125(7), e2020JA027823. https://doi.org/10.1029/2020JA027823

Emmert, J. T. (2015). Thermospheric mass density: A review. Advances in Space Research, 56, 773–824. https://doi.org/10.1016/j.asr.2015.05.038

Errico, R. M., Yang, R., Privé, N., Tai, K.-S., Todling, R., Sienkiewicz, M., & Guo, J. (2013). Development and validation of observing-system simulation experiments at NASA’s Global Modeling and Assimilation Office. Quarterly Journal of the Royal Meteorological Society, 139, 1162–1178. https://doi.org/10.1002/qj.202710.1002/qj.2027

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(C5), 10143–10162. https://doi.org/10.1029/94JC00572

He, J., Yue, X., Hu, L., Wang, J., Li, M., Ning, B., et al. (2020). Observing system impact on ionospheric specification over China using EnKF assimilation. Space Weather, 18, e2020SW002527. https://doi.org/10.1029/2020SW002527

He, J., Yue, X., Le, H., Ren, Z., & Wan, W. (2020). Evaluation on the quasi realistic ionospheric prediction using an ensemble Kalman filter data assimilation algorithm. Space Weather, 18, e2019SW002410. https://doi.org/10.1029/2019SW002410

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

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(11), 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

Krywonos, A., Murray, D. J., Eastes, R. W., Aksnes, A., Budzien, S. A., & Daniell, R. E. (2012). Remote sensing of neutral temperatures in the Earth’s thermosphere using the Lyman-Birge-Hopfield bands of N2: Comparisons with satellite drag data. *Journal of Geophysical Research*, 117, A09311. https://doi.org/10.1029/2011ja017226

Laskar, F. I., Eastes, R. W., Codrescu, M. V., Evans, J. S., Burns, A. G., Wang, W., et al. (2021). Response of GOLD retrieved thermospheric temperatures to geomagnetic activities of varying magnitudes. *Geophysical Research Letters*, 48, e2021GL093905. https://doi.org/10.1029/2021GL093905

Laskar, F. I., Pedatella, N. M., Codrescu, M. V., Eastes, R. W., Burns, A. G., & McClintock, W. (2021). Impact of GOLD retrieved thermospheric temperatures on a whole atmosphere data assimilation model. *Journal of Geophysical Research: Space Physics*, 126, e2020JA028646. https://doi.org/10.1029/2020JA028646

Lee, J. 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, A01318. https://doi.org/10.1029/2011ja017700

Liu, H.-L., Bardeen, C. G., Foster, B. T., Lauritzen, P., Liu, J., Lu, G., et al. (2018). Development and validation of the Whole Atmosphere Community Climate Model with thermosphere and ionosphere extension (WACCM-X 2.0). *Journal of Advances in Modeling Earth Systems*, 10, 381–402. https://doi.org/10.1002/2017MS001232

Mannucci, A. J., Tsurutani, B. T., Iijima, B. A., Komjathy, A., Saito, A., Gonzalez, W. D., et al. (2005). Dayside global ionospheric response to the major interplanetary events of October 29–30, 2003 “Halloween Storms”. *Geophysical Research Letters*, 32, L12S02. https://doi.org/10.1029/2004GL021467

Marcos, F. A. (2006). New satellite drag modeling capabilities. 44th AIAA Aerospace Sciences Meeting and Exhibit.

Matsuo, T., Lee, I., & Anderson, J. (2013). Thermospheric mass density specification using an ensemble Kalman filter. *Journal of Geophysical Research: Space Physics*, 118(3), 1339–1350. https://doi.org/10.1002/jgra.50162

Millward, G. H., Moffett, R. J., Quegan, S., & Fuller-Rowell, T. J. (1996). A coupled thermosphere-ionosphere-plasmasphere model (CTIP). In R. W. Schunk (Ed.), *STEP handbook on ionospheric models* (pp. 239–279). Utah State University.

Pedatella, N. M., Anderson, J. L., Chen, C. H., Raeder, K., Liu, J., Liu, H.-L., & Lin, C. H. (2020). Assimilation of ionosphere observations in the Whole Atmosphere Community Climate Model with thermosphere-ionosphere eXtension (WACCMX). *Journal of Geophysical Research: Space Physics*, 125, e2020JA028251. https://doi.org/10.1029/2020JA028251

Privé, N., Errico, R. M., & Tai, K.-S. (2013). Validation of forecast skill of the Global Modeling and Assimilation Office observing system simulation experiment. *Quarterly Journal of the Royal Meteorological Society*, 139, 1354–1363. https://doi.org/10.1002/qj.2029

Qian, L., Burns, A. G., Wang, W., Solomon, S. C., Zhang, Y., & Hsu, V. (2016). Effects of the equatorial ionosphere anomaly on the interhemispheric circulation in the thermosphere. *Journal of Geophysical Research: Space Physics*, 121, 2522–2530. https://doi.org/10.1002/2015JA022169

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/10.1029/92GL00401

Rideout, W., & Coster, A. (2006). Automated GPS processing for global total electron content data. *GPS Solutions*, 10(3), 219–228. https://doi.org/10.1007/s10291-006-0019-8

Ridley, A. J., Deng, Y., & Toth, G. (2006). The global ionosphere-thermosphere model. *Journal of Atmospheric and Solar-Terrestrial Physics*, 68(8), 839–864. https://doi.org/10.1016/j.jastp.2006.01.008

Shim, J. S., Kuznetsova, M., Rasitter, L., Bititza, D., Butala, M., Codrescu, M., et al. (2014). *Systematic evaluation of ionosphere/thermosphere (IT) models, Modeling the ionosphere-thermosphere system* (pp. 145–160). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118704417.ch13

Sutton, E. K. (2018). A new method of physics-based data assimilation for the quiet and disturbed thermosphere. *Space Weather*, 16, 736–753. https://doi.org/10.1029/2017SW001785