Development of Data Assimilation Systems for the Ionosphere, Thermosphere, and Mesosphere

Primary Author:
Nicholas M. Pedatella
High Altitude Observatory, National Center for Atmospheric Research
303-497-2619
nickp@ucar.edu

Contributing Authors: Jeffrey L. Anderson (NCAR), Chih-Ting Hsu (NCAR), Kayo Ide (University of Maryland), Timothy Kodikara (German Aerospace Center), Fazlul Laskar (University of Colorado), Tomoko Matsuo (University of Colorado), Eric K. Sutton (University of Colorado), Valery Yudin (The Catholic University of America), Xinan Yue (Chinese Academy of Sciences)

Synopsis: The past decade saw the development of several data assimilation systems for the ionosphere, thermosphere, and mesosphere (ITM). To fully realize the capabilities of ITM data assimilation systems for both scientific investigations and operations, several critical advances are needed. This white paper outlines some of the outstanding challenges facing ITM data assimilation that need to be addressed in the coming decade in order to achieve robust, high-quality, ITM data assimilation systems. Benefits to both the scientific and operational communities of advancing ITM data assimilation capabilities are also provided. These include, but are not limited to, providing the framework for investigating ITM predictability, scientific investigations into day-to-day ITM variability driven by the lower atmosphere and geomagnetic storms, as well as advancing space weather forecasting capabilities.

Background and Motivation

Data assimilation refers to the process of optimally combining background model estimates with observations to produce the best possible estimate of the state. It provides a number of advantages compared to numerical models or observations alone and has proven to be fundamental for improving troposphere numerical weather prediction (NWP) and for producing atmospheric reanalyses (e.g., NASA MERRA) in the troposphere and stratosphere. Data assimilation provides an observational constraint to numerical models, removing biases and providing a better representation of the atmospheric state. The improved state estimate can serve as initial conditions for short- and medium-range forecasting using physics-based, first-principle, models. Data assimilation can also extend the utility of observations by providing information on quantities that are not directly observed. For example, ionosphere electron density observations can be used to better constrain the thermosphere neutral mass density, leading to improved estimates of satellite drag (Matsuo et al., 2013; Dietrich et al, 2022).

There have been significant developments in the ionosphere, thermosphere, and mesosphere (ITM) data assimilation research in the past decade. This is especially true with regards to ITM data assimilation systems that use physics-based background models. In this context, data assimilation system refers to the combination of the background model and assimilation algorithm. Note that we focus on the use of physics-based, first-principle, models as they enable fundamental understanding of the coupled ITM system as well as initialized numerical forecasting of the ITM. Significant recent developments have occurred in ionosphere-thermosphere (~90-500 km altitude domain) data assimilation systems that use both state and driver estimation (Lee et al., 2012; Morozov et al., 2013; Chartier et al., 2016; Chen et al., 2016a; Codrescu et al., 2018; Sutton, 2018), and whole atmosphere (surface to ~500 km altitude domain) data assimilation systems (Pedatella et al., 2018; 2020) that represent the extension of NWP schemes into the ITM. The data assimilation systems developed for the ITM have been applied to study variability of the coupled atmosphere-ionosphere system forced by geomagnetic storms and lower atmosphere perturbations (e.g., Chen et al., 2016a; Rajesh et al., 2017; Pedatella et al., 2018). They have also been used for initialized forecast experiments, providing important understanding of the ITM predictability (Chartier et al., 2016; Chen et al., 2016b; Pedatella et al., 2018; 2019). In addition, ITM data assimilation systems have operational applications for nowcasting and forecasting the ITM variability. Continuing to advance ITM data assimilation capabilities would thus benefit both the scientific and operational communities.

Despite the advantages afforded by data assimilation, the techniques and applications of data assimilation systems focused on the ITM remain in their infancy. This white paper outlines some of the outstanding challenges in ITM data assimilation, the scientific and operational advances that these would enable, and concludes with recommendations for advancing ITM data assimilation techniques and applications.

Outstanding Challenges in ITM Data Assimilation

Although there is a long history of developing data assimilation techniques for the lower atmosphere (troposphere-stratosphere), these techniques cannot be directly applied to the ITM due to important differences in chemistry, dynamics, and physics of these domains. For example, the ITM is a partly driven system due to forcing from solar and geomagnetic activity. The time scales of variability, especially in the ionosphere, also tend to be shorter than those in the lower atmosphere. It is thus necessary to develop data assimilation techniques that are tailored to the unique demands of the ITM. Below we outline some of the outstanding challenges in ITM data assimilation.

There are two primary approaches currently taken for ITM data assimilation. The first approach aims to estimate the driver inputs (e.g., solar flux) to the background model (Morozov et al., 2013; Codrescu et al., 2018; Sutton, 2018). This leverages the spatial correlations built into the physical model, providing global-scale corrections to the background state (i.e., not only at locations where observations are assimilated). The second approach adjusts the model state (e.g., temperature, electron density) directly from the observations (Lee et al., 2012; Hsu et al., 2014; Pedatella et al., 2020) without any adjustment to the global model drivers. This approach, as an extension of lower atmosphere NWP systems, is much more sensitive to the availability and location of the assimilated observations. It is, however, less reliant on the physical model being capable of fully capturing the dynamics of the ITM compared to only adjusting the global model drivers. At present, it remains unclear which of the two approaches is best suited for the ITM, and it may be that each approach is best suited for different applications. Combining the two approaches of driver and state estimation may also prove useful, though this is something that has yet to be fully explored. It is thus critical to understand the capabilities, including their advantages and disadvantages, of these two different approaches in order to understand how they can be used to advance ITM data assimilation capabilities in the future.

A major challenge in atmospheric data assimilation is the treatment of small-scale waves, which can be artificially introduced by unbalanced data assimilation adjustments. Small-scale waves are often removed in lower atmosphere data assimilation through various filtering techniques (e.g., Lynch and Huang, 1992; Bloom et al., 1996). However, wave-driven variability is an important source of ITM variability, and the filtering method can significantly impact the dynamics of the mesosphere and thermosphere (Sankey et al., 2007). It is thus critical to develop techniques that minimize the introduction of artificial small-scale waves due to the data assimilation adjustments while preserving the physical waves that are present and cause ITM variability. Minimizing the introduction of artificial small-scale waves is particularly crucial for whole atmosphere data assimilation systems due to the fact that certain small waves introduced in the troposphere-stratosphere can grow exponentially with altitude and obtain large amplitudes by the time they reach the ITM.

Despite increases in the number of observations, the ITM system remains relatively poorly observed. This is especially the case in the mesosphere and thermosphere. Developing techniques to best use the sparse mesosphere and thermosphere observations is therefore critical. For example, assimilation of radiance observations greatly enhanced the impact of satellite observations in NWP compared to retrieved parameters (Eyre et al., 2020). Similarly, assimilating radiance observations directly in the ITM may lead to greater impact of the sparse available observations (Cantrall et al., 2019). Furthermore, the observations that do exist are available from different sources, in different file formats, and often do not include realistic specification of uncertainties. This presents a challenge for using all of the available observations within ITM data assimilation systems and the appropriate data quality control measures of diverse measurements needs to be rigorously developed. The development of centralized databases containing all observations with standard file-formats and uncertainty information would improve usage of all observations in ITM data assimilation systems, as well as be beneficial to the broader community. It should also be noted that although the ionosphere is generally well observed, it is important to continue to develop techniques to maximize the impact of ionosphere observations to constrain the state of the thermosphere. As the ionosphere and thermosphere are a highly coupled system, ionosphere observations can effectively constrain the thermosphere if the appropriate correlations are established between the ionosphere and thermosphere (Matsuo et al., 2013; Hsu et al., 2014). Using ionosphere observations to constrain the thermosphere is important not only due to the deficiency of thermospheric observations, but also since the thermosphere is critical for driving ionosphere variability and improving forecast skill.

Currently, the ensemble Kalman filter is arguably the most used technique for performing ITM data assimilation using physics-based background models. Ensuring that the ensemble spread is representative of the model uncertainty is fundamental for the ensemble Kalman filter technique to perform correctly. Generating ensemble spread in ITM models is, however, complicated by the fact that the ITM is a partly driven system, and the ensemble spread will tend to be too small when all ensemble members are subject to identical external forcing. To overcome this obstacle, ITM ensemble data assimilation models have typically relied on perturbing the external forcing parameters, such as the solar flux at 10.7 cm (F10.7), to generate ensemble spread (e.g., Lee et al., 2012). The perturbations are typically done in an ad-hoc manner, and are not representative of the forcing uncertainty. Development of new methodologies for generating ensemble spread such as perturbations to model physics parameterizations that are currently used in modern NWP systems (Buizza et al., 1999; Berner et al., 2009) are an alternative approach for generating ensemble spread that is more reflective of the model uncertainty. Accurate representation of the ensemble spread is also critical for space weather forecasts and enables reliable probabilistic forecasts if the uncertainty in the ensemble predictions is reflective of the actual uncertainty (Lewis, 2005).

In addition to the ensemble Kalman filter, there are several alternative techniques for data assimilation, such as variational (3D-Var and 4D-Var) as well as hybrid approaches (Rabier et al., 2000; Hamill and Snyder, 2000; Clayton et al., 2013). Despite their potential advantages, the current state-of-the-art NWP data assimilation approaches have yet to be adopted for use in the ITM. This is, at least in part, due to the relative complexity of implementing these techniques compared to the ensemble Kalman filter. It is critical that the community pursue the development and implementation of advanced data assimilation techniques for the ITM. This will ensure the continued advancement of ITM data assimilation systems. Implementation of alternative methods will further enable data assimilation using high-resolution background models, something that is currently inhibited by the computational cost of the ensemble Kalman filter.

Scientific and Operational Advances Enabled by Enhanced Data Assimilation

Advances in data assimilation techniques, such as those outlined in the previous section, would lead to significant advances in both ITM science as well as for space weather operational forecasts and applications. Below we outline some of the key areas that would benefit from advances in ITM data assimilation.

Data assimilation systems provide the best estimate of the current state of the ITM, which can be used as initial conditions for physics-based model forecasts. Data assimilation algorithm improvements have a significant role in enhancement of NWP forecast skill (Bauer et al., 2015). Advances in the data assimilation techniques for the ITM, including making optimal use of observations, are therefore considered as highly likely to lead to improved space weather forecasts.

The continued development of ITM data assimilation systems also provides the opportunity to perform initialized forecast/hindcast experiments to investigate the predictability of the ITM. There is currently only a rudimentary understanding of the ITM predictability, with little understanding of how it may vary with, for example, local time, season, and solar activity. Initialized forecast/hindcast experiments using ITM data assimilation systems can be used to improve knowledge of the ITM predictability. Additionally, they can be applied to develop insight into how ITM predictability can be improved. For example, Observing System Simulation Experiments can be used to investigate future observing systems and how they may impact the ITM forecast skill.

Atmospheric reanalyses (e.g., NASA MERRA2, ECMWF ERA5) are widely used in the troposphere-stratosphere communities due to the fact that they provide the best available estimate of the atmospheric state globally. No such product currently exists for the ITM. Investment in ITM data assimilation systems in the next decade would enable them to reach the level of maturity where a reliable ITM reanalysis product could be developed and made available to the scientific community. By providing the best estimate of the ITM state, such a reanalysis product would enable significant advances in a broad range of ITM science, such as day-to-day variability driven by the lower atmosphere and storm time variations in the ionosphere-thermosphere.

Data assimilation can also be a powerful tool for model development. The data assimilation increments provide information on the locations of large model-observation discrepancies, guiding where to focus efforts during the model development process. Data assimilation techniques can also be used for parameter estimation, potentially accelerating the model tuning and development process. Model parameterizations, such as, for example, gravity wave drag and eddy diffusion, remain significant sources of uncertainty in ITM models (Qian et al, 2009; Pedatella et al., 2014; Siskind et al, 2014; Pilinski and Crowley, 2015). Data assimilation would provide an observationally informed estimate of these parameters, which are currently often adjusted as model tuning parameters.

Conclusions and Recommendations

The past decade has seen the advancement of data assimilation systems for the ITM. However, as outlined in this white paper, there remain several challenges when it comes to ITM data assimilation systems. It is recommended that ITM data assimilation systems be considered as a critical component for performing ITM research in the coming decade. To realize the full potential of ITM data assimilation systems for ITM research, we recommend the following:

- **The 2023 Solar and Space Physics Decadal Survey should recognize the important role that ITM data assimilation systems will have in the coming decade.** As ITM data assimilation systems continue to develop, they will be fundamental for advancing both scientific understanding of the ITM as well as space weather forecasts.

- **Prioritize the development of fundamental techniques for performing data assimilation in the ITM.** Current efforts are typically single PI led, limiting the rate of ITM data assimilation development. Moving beyond the single PI led approach requires investment in the development of community ITM data assimilation systems and would serve to accelerate their development. Interdisciplinary proposal opportunities to support a broader collaboration between the ITM and NWP data assimilation communities would also enhance ITM data assimilation capabilities.

- **Develop more transparent and robust O2R/R2O research pipelines by creating open solicitations to participate in operational data assimilation efforts.** There is currently a gap between the ITM data assimilation systems developed in the research community and those developed and used operationally. This limits the impact of the efforts in the research community on operational systems, while also inhibiting the research community from focusing efforts where they may be most useful operationally.

- **Establish quantifiable metrics to track improvements in ITM data assimilation system nowcasts and forecasts.** The ITM community currently lacks accepted metrics for quantifying the performance of ITM data assimilation systems. This makes comparing results from different data assimilation systems difficult, and also inhibits quantification of nowcasting and forecasting improvements over time.

- **Incorporate research into ITM predictability into existing research programs.** Current knowledge of the ITM predictability is extremely poor, with little knowledge of how it varies on different spatial scales, seasons, and levels of solar and geomagnetic activity. The sources of predictability and how it can be improved are also unknown. Targeted research programs into ITM predictability will eliminate the gaps in our current understanding.

References

Bauer, P., Thorpe, A., and Brunet, G. (2015), The quiet revolution of numerical weather prediction, *Nature*, 525, 47–55, https://doi.org/10.1038/nature14956

Berner J., Shutts, G. J., Leutbecher, M., and Palmer, T. N. (2009), A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system, *J. Atmos. Sci.*, 66 , 603-626, doi:10.1175/2008JAS2677.1.

Bloom, S. C., Takacs, L. L., DaSilva, A. M., and Ledvina, D. (1996), Data assimilation using incremental analysis updates, *Mon. Weather Rev.*, 124, 1256-1271. https://doi.org/10.1175/1520-0493(1996)124%3C1256:DAUIAU%3E2.0.CO;2

Buizza R, Miller M., and Palmer T. N. (1999), Stochastic representation of model uncertainties in the ECWMF ensemble prediction system, *Q. J. R. Meteorol. Soc.*, 125, 2887-2908, doi:10.1002/qj.49712556006.

Cantrall, C. E., Matsuo, T., and Solomon, S. C. (2019), Upper atmosphere radiance data assimilation: A feasibility study for GOLD far ultraviolet observations, *J. Geophy. Res.*, 124, 8154-8164, https://doi.org/10.1029/2019JA026910

Chartier, A. T., Matsuo, T., Anderson, J. L., Collins, N., Hoar, T. J., Lu, G., Mitchell, C. N., Coster, A. J., Paxton, L. J., and Bust, G. S. (2016), Ionospheric data assimilation and forecasting during storms, *J. Geophys. Res.*, 121, 764-778, doi:10.1002/2014JA020799.

Clayton, A.M., Lorenc, A.C. and Barker, D.M. (2013), Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, *Q.J.R. Meteorol. Soc.,* 139, 1445-1461, https://doi.org/10.1002/qj.2054

Codrescu, S. M., Codrescu, M. V., and Fedrizzi, M. (2018), An Ensemble Kalman Filter for the thermosphere-ionosphere, *Space Weather*, 16, 57-68, https://doi.org/10.1002/2017SW001752

Chen, C. H., Lin, C. H., Matsuo, T., Chen, W. H., Lee, I. T., Liu, J. Y., Lin, J. T., and Hsu, C. T. (2016a), Ionospheric data assimilation with thermosphere-ionosphere-electrodynamics general circulation model and GPS-TEC during geomagnetic storm conditions, *J. Geophys. Res.*, 121, 5708-5722, doi:10.1002/2015JA021787.

Chen, C. H., Lin, C. H., Matsuo, T., and Chen, W. H. (2016b), Ionosphere data assimilation modeling of 2015 St. Patrick's Day geomagnetic storm, *J. Geophys. Res.*, 121, 11,549-11,559, doi:10.1002/2016JA023346.

Dietrich, N., Matsuo, T., and Hsu, C.-T. (2022), Specifying satellite drag through coupled thermosphere-ionosphere data assimilation of radio occultation electron density profiles, *Space Weather*, 20, e2022SW003147, https://doi.org/10.1029/2022SW003147

Eyre, J. R., English, S. J., and Forsythe, M. (2020), Assimilation of satellite data in numerical weather prediction. Part I: The early years. *Q J R Meteorol Soc.,* 146, 49-68, https://doi.org/10.1002/qj.3654

Hamill, T. M., and Snyder, C., (2000), A hybrid ensemble Kalman filter-3D variational analysis scheme, *Mon. Wea. Rev.*, 128, 2905-2919, https://doi.org/10.1175/1520-0493(2000)128%3C2905:AHEKFV%3E2.0.CO;2

Hsu, C.-T., Matsuo, T., Wang, W., and Liu, J.-Y. (2014), Effects of inferring unobserved thermospheric and ionospheric state variables by using an Ensemble Kalman Filter on global ionospheric specification and forecasting, *J. Geophys. Res.*, 119, 9256-9267, doi:10.1002/2014JA020390.

Lee, I. T., Matsuo, T., Richmond, A. D., Liu, J. Y., Wang, W., Lin, C. H., Anderson, J. L., and Chen, M. Q. (2012), Assimilation of FORMOSAT-3/COSMIC electron density profiles into a coupled thermosphere/ionosphere model using ensemble Kalman filtering, *J. Geophys. Res.*, 117, A10318, doi:10.1029/2012JA017700.

Lewis, J. M. (2005), Roots of ensemble forecasting, *Mon. Weather Rev,* 133 1865-1885, doi:10.1175/MWR2949.1

Lynch, P., and Huang, X.-Y. (1992), Initialization of the HIRLAM model using a digital filter, *Mon. Weather Rev*, 120, 1019-1034. https://doi.org/10.1175/1520-0493(1992)120%3C1019:IOTHMU%3E2.0.CO;2

Matsuo, T., Lee, I.-T., and Anderson, J. L. (2013), Thermospheric mass density specification using an ensemble Kalman filter, *J. Geophys. Res.*, 118, 1339-1350, doi:10.1002/jgra.50162.

Morozov, A. V., Ridley, A. J., Bernstein, D. S., Collins, N., Hoar, T. J., and Anderson, J. L. (2013), Data assimilation and driver estimation for the Global Ionosphere-Thermosphere Model using the Ensemble Adjustment Kalman Filter, *J. Atmos. Sol. Terr. Phys.*, 104, 126-136, https://doi.org/10.1016/j.jastp.2013.08.016.

Pedatella, N. M., et al. (2014), The neutral dynamics during the 2009 sudden stratosphere warming simulated by different whole atmosphere models, *J. Geophys. Res.*, 119, 1306-1324, doi:10.1002/2013JA019421.

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, *J. Geophys. Res.*, 123, 3131-3153, https://doi.org/10.1002/2017JA025107

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

Pilinski, M. D., and Crowley, G. (2015), Seasonal variability in global eddy diffusion and the effect on neutral density, *J. Geophys. Res.*, 120, 3097-3117, doi:10.1002/2015JA021084.

Qian, L., Solomon, S. C., and Kane, T. J. (2009), Seasonal variation of thermospheric density and composition, *J. Geophys. Res.*, 114, A01312, doi:10.1029/2008JA013643.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. (2000), The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Q.J.R. Meteorol. Soc.*, 126, 1143-1170, https://doi.org/10.1002/qj.49712656415

Rajesh, P. K., Lin, C. H., Chen, C. H., Lin, J. T., Matsuo, T., Chou, M. Y., Chen, W. H., Chang, M. T., and You, C. F. (2017), Equatorial plasma bubble generation/inhibition during 2015 St. Patrick's Day storm, *Space Weather*, 15, 1141-1150, doi: doi:10.1002/2017SW001641.

Sankey, D., Ren, S., Polavarapu, S., Rochon, Y. J., Nezlin, Y., and Beagley, S. (2007), Impact of data assimilation filtering methods on the mesosphere, *J. Geophys. Res.*, 112, D24104, doi:10.1029/2007JD008885.

Siskind, D. E., Drob, D. P., Dymond, K. F., and McCormack, J. P. (2014), Simulations of the effects of vertical transport on the thermosphere and ionosphere using two coupled models, *J. Geophys. Res.*, 119, 1172-1185, doi:10.1002/2013JA019116.

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/10.1002/2017SW001785.