A combined neural network- and physics-based approach for modeling plasmasphere dynamics

Irina S. Zhelavskaya[1], Nikita A Aseev[2], and Yuri Shprits[3]

[1]Helmholtz Centre Potsdam, GFZ German Research Centre For Geosciences, Potsdam, Germany
[2]GFZ German Research Centre for Geosciences
[3]GFZ Helmholtz Center Potsdam

November 24, 2022

Abstract

In recent years, feedforward neural networks (NNs) have been successfully applied to reconstruct global plasmasphere dynamics in the equatorial plane. These neural network-based models capture the large-scale dynamics of the plasmasphere, such as plume formation and erosion of the plasmasphere on the nightside. However, their performance depends strongly on the availability of training data. When the data coverage is limited or non-existent, as occurs during geomagnetic storms, the performance of NNs significantly decreases, as networks inherently cannot learn from the limited number of examples. This limitation can be overcome by employing physics-based modeling during strong geomagnetic storms. Physics-based models show a stable performance during periods of disturbed geomagnetic activity, if they are correctly initialized and configured. In this study, we illustrate how to combine the neural network- and physics-based models of the plasmasphere in an optimal way by using the data assimilation Kalman filtering. The proposed approach utilizes advantages of both neural network- and physics-based modeling and produces global plasma density reconstructions for both quiet and disturbed geomagnetic activity, including extreme geomagnetic storms. We validate the models quantitatively by comparing their output to the in-situ density measurements from RBSP-A for an 18-month out-of-sample period from 30 June 2016 to 01 January 2018, and computing performance metrics. To validate the global density reconstructions qualitatively, we compare them to the IMAGE EUV images of the He+ particle distribution in the Earth's plasmasphere for a number of events in the past, including the Halloween storm in 2003.

A combined neural network- and physics-based approach for modeling plasmasphere dynamics

I. S. Zhelavskaya[1,2], N. A. Aseev[1,2], and Y. Y. Shprits[1,2,3]

[1]GFZ Potsdam, Potsdam, Germany
[2]Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany
[3]Earth, Planetary and Space Sciences, University of California, Los Angeles, Los Angeles, CA, USA

Key Points:

- We develop an approach to combine a neural network with a physics-based model of the plasmasphere using Kalman filtering.
- The approach is extensively validated using in-situ density measurements and observed plasmapause position derived from the IMAGE EUV.
- The developed model reproduces the plasmasphere dynamics during quiet, moderate, disturbed, and extreme geomagnetic events.

Corresponding author: Irina Zhelavskaya, `irina.zhelavskaya@gmail.com`

Abstract

In recent years, feedforward neural networks (NNs) have been successfully applied to reconstruct global plasmasphere dynamics in the equatorial plane. These neural network-based models capture the large-scale dynamics of the plasmasphere, such as plume formation and erosion of the plasmasphere on the nightside. However, their performance depends strongly on the availability of training data. When the data coverage is limited or non-existent, as occurs during geomagnetic storms, the performance of NNs significantly decreases, as networks inherently cannot learn from the limited number of examples. This limitation can be overcome by employing physics-based modeling during strong geomagnetic storms. Physics-based models show a stable performance during periods of disturbed geomagnetic activity, if they are correctly initialized and configured. In this study, we illustrate how to combine the neural network- and physics-based models of the plasmasphere in an optimal way by using the data assimilation Kalman filtering. The proposed approach utilizes advantages of both neural network- and physics-based modeling and produces global plasma density reconstructions for both quiet and disturbed geomagnetic activity, including extreme geomagnetic storms. We validate the models quantitatively by comparing their output to the in-situ density measurements from RBSP-A for an 18-month out-of-sample period from 30 June 2016 to 01 January 2018, and computing performance metrics. To validate the global density reconstructions qualitatively, we compare them to the IMAGE EUV images of the He^{+} particle distribution in the Earth's plasmasphere for a number of events in the past, including the Halloween storm in 2003.

1 Introduction

The plasmasphere is a region of cold (< 10 eV) and dense ($10-10^{4}\text{cm}^{-3}$) plasma encircling the Earth and corotating with it (Lemaire & Gringauz, 1998). It is located in the inner magnetosphere and extends several Earth radii (R_{E}) into space out to a boundary known as the plasmapause (Gringauz, 1963; Carpenter, 1963). At this boundary, the plasma density decreases drastically by several orders of magnitude. The plasmasphere is a very dynamic region, and its shape and size are strongly dependent on solar and geomagnetic conditions (O'Brien & Moldwin, 2003; Chappell et al., 1970b). The size and shape of the plasmasphere are controlled by two regimes: sunward convection and corotation with the Earth (Darrouzet et al., 2009; A. Singh et al., 2011). The corotation regime dominates during quiet geomagnetic times, and the plasma trapped inside the closed magnetic field lines corotates with the Earth (Carpenter, 1966). At the same time, the plasmasphere is refilled from the dayside ionosphere (N. Singh & Horwitz, 1992; Goldstein, Sandel, Hairston, & Reiff, 2003; Krall et al., 2008). It has a nearly circular shape with a dusk side bulge (Nishida, 1966). Contrastingly, the sunward magnetospheric convection begins to dominate during intervals of high geomagnetic activity (Carpenter, 1970; Chappell et al., 1970a; Goldstein, Sandel, Hairston, & Reiff, 2003) and erodes the plasmasphere. The stronger the geomagnetic disturbance, the more severely the plasmasphere is eroded (as far as 2 R_{E} during severe disturbances). The combination of convection and corotation electric fields causes the development of a plasmaspheric plume in the dusk local time sector (e.g., Spasojević et al., 2003; Grebowsky, 1970).

The plasmasphere is important for a number of physical processes. Its size and shape control the propagation and growth of plasma waves, and affect wave-particle interactions, thus greatly influencing distributions of energetic ions and electrons across a broad range of energies (e.g., Spasojević et al., 2004; Horne et al., 2005; Y. Y. Shprits et al., 2016; Orlova et al., 2016). The plasmaspheric material eroded during periods of strong convection is transported sunward and is observed near the dayside magnetopause regularly (e.g., Chen & Moore, 2006; Lee et al., 2016). The enhanced plasma density at the dayside magnetopause can limit the rate of reconnection, thus affecting the global convection pattern (e.g., André et al., 2016; Borovsky & Denton, 2006). The plasma density is also a crucial parameter in a variety of applications in the field of space weather, such as spacecraft anomaly analysis due to spacecraft charging (e.g., Reeves et al., 2013) and GPS navigation (e.g., Mazzella, 2009; Xiong et al., 2016). It is therefore important to model the dynamics of the plasmasphere accurately in order to be able to reliably predict the aforementioned processes.

A number of physics-based and empirical models have been developed in recent decades. The most commonly used empirical models are the Carpenter and Anderson (1992), D. L. Gallagher et al. (2000), and Sheeley et al. (2001) models. The Carpenter and Anderson (1992) model is based on measurements of electron density derived from radio measurements made with the sweep frequency receiver (SFR) on board the International Sun-Earth Explorer (ISEE-1) spacecraft and ground-based whistler measurements. It is a model of saturated density and, thus, represents the distribution of density after several days of refilling. The model covers the range from 2.25 to 8 in L-shell, and the interval of 0-15 MLT (magnetic local time). The model provides the mean value of density at different L-shells.

D. L. Gallagher et al. (2000) developed the Global Core Plasma Model (GCPM), which combined several previously developed models (such as Carpenter and Anderson (1992) and D. Gallagher et al. (1998)) by means of transition equations, in order to provide a more comprehensive description of the inner-magnetospheric plasma. The models of plasmasphere and plasma trough developed by Sheeley et al. (2001) provide statistical averages of density based on density measurements obtained from the swept frequency receiver onboard the Combined Release and Radiation Effects Satellite (CRRES) by identifying the upper hybrid resonance frequency. The models cover all local times and $3 \leq L \leq 7$. Moreover, Sheeley et al. (2001) also provide the standard deviation of density for both the plasmasphere and trough models to describe depleted and saturated density levels for various L-shells and MLT sectors for the trough.

Despite the extensive use of these empirical density models for the inner-magnetospheric simulations, they provide statistically averaged values and do not account for the changing geomagnetic conditions, and therefore, cannot produce reliable estimates of density during extreme geomagnetic events. The models described above do not include the dynamic dependence of plasma density on geomagnetic and solar wind conditions, and density is known to vary substantially during periods of strong geomagnetic activity (e.g., Park & Carpenter, 1970; Park, 1974; Moldwin et al., 1995).

This fact motivated the development of time-dependent models of plasma density. In recent years, a number of models utilizing neural networks and taking into account solar or geomagnetic conditions have been developed (Bortnik et al., 2016; I. Zhelavskaya et al., 2017; Chu, Bortnik, Li, Ma, Angelopoulos, & Thorne, 2017; Chu, Bortnik, Li, Ma, Denton, et al., 2017). In all these studies, the authors used feedforward neural networks with different architectures to model the plasma density in the equatorial plane or in 3D (in Chu, Bortnik, Li, Ma, Denton, et al. (2017)). Feedforward neural networks are a powerful mathematical tool for finding nonlinear multivariate mappings from input to output variables, if such a mapping exists (J. A. Anderson, 1995; C. M. Bishop, 1995; Haykin et al., 2009). Bortnik et al. (2016) used density measurements inferred from the spacecraft potential (Li et al., 2010) on board the THEMIS (Time History of Events and Macroscale Interactions during Substorms) mission (Angelopoulos, 2009) to train their neural network model. They used a 5-hour time history of Sym-H index and location as an input to the model. The Chu, Bortnik, Li, Ma, Angelopoulos, and Thorne (2017) model is based on the same density measurements. The inputs to the model were location and the time histories of Sym-H for the preceding 3 days, AL for 2 hours, and F10.7 for 3 days. Chu, Bortnik, Li, Ma, Denton, et al. (2017) built up on those two studies and developed a three-dimensional electron density (DEN3D) model based on density measurements from the plasma wave experiment on board ISEE (D. A. Gurnett et al., 1978), the plasma wave experiment on board the CRRES (R. R. Anderson et al., 1992), the plasma wave instrument on board Polar (D. Gurnett et al., 1995), and the radio plasma imager (RPI) on board the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE). They used location and the time histories of Sym-H for the preceding 3 days, AL for 5 hours, and F10.7 for 3 days as inputs to their model. The model of I. Zhelavskaya et al. (2017), Plasma density in the Inner magnetosphere Neural network-based Empirical (PINE) model, is based on the density measurements obtained from the upper-hybrid resonance frequency measured with the EMFISIS instrument on board the Van Allen Probes. This technique is known to be one of the most reliable methods for obtaining plasma density (Mosier et al., 1973). The inputs to the model were the 96-hour time history of Kp, AE, Sym-H, and F10.7 indices and the location given by L and MLT. They showed that neural networks-based models can accurately reproduce the dynamics of the plasmasphere (with correlation coefficient ≈ 0.95), and can successfully reproduce the asymmetric shape of the plasmasphere, including plume formation and erosion on the nightside.

Neural networks learn from data and are very powerful when data are abundant. However, when the data are limited or lacking, their performance may significantly decrease (Priddy & Keller, 2005). This implies that neural networks can be difficult to apply to highly unbalanced regression problems and to predict rare events. Extreme geomagnetic storms are an example of such events. Figure 1 shows the distribution of the Kp index over the training period of the PINE model (Oct 2012-Jul 2016, I. Zhelavskaya et al. (2017)). As can be seen from the figure, its distribution is highly skewed. Observations for Kp > 7 are limited. In fact, there is not a single example of Kp $= 9$ during this period and, hence, in the training dataset. That implies that NNs may not be reliable during periods of high geomagnetic activity, which are the most interesting events.

One possible way to overcome this limitation is to employ a different approach to model the plasmasphere dynamics during disturbed geomagnetic conditions. In particular, physics-based modeling is a more stable approach than neural network-based modeling for high Kp, since it does not depend on data availability. A number of physics-based models have been developed in recent years. Pierrard and Stegen (2008) used the

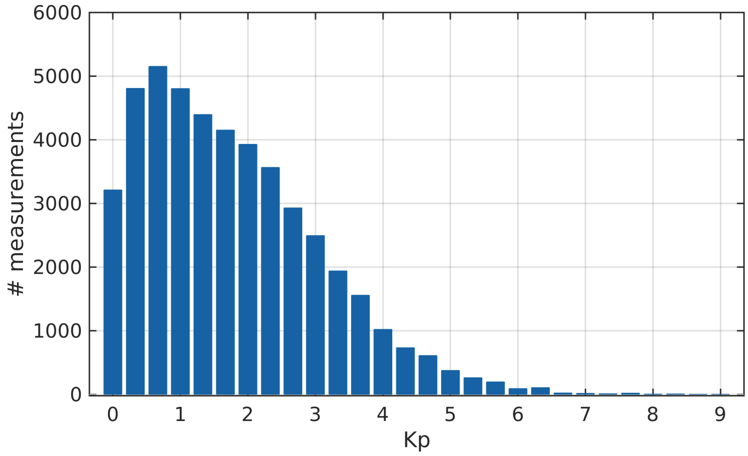

Figure 1. Distribution of Kp over October 2012 – June 2016 (the training time interval for the PINE model). The cadence of bins is 1/3, i.e., the same as the cadence of the Kp index.

kinetic exospheric approach to model the dynamics of the plasmasphere. V. Jordanova et al. (2006) coupled their ring current-atmosphere interactions model (RAM) with a 3-D equilibrium code (SCB) and a cold plasma model (CPL). The RAM-SCB-CPL model calculates the cold electron density in the equatorial plane by following the motion of individual flux tubes, using a model of electric field which includes a corotation potential and a convection potential that is chosen from either semi-empirical models (Volland, 1973; Stern, 1975; Weimer, 2005), or self-consistently calculated electric potential (Yu et al., 2015), mapped to the equatorial plane along the SCB field lines. Krall et al. (2016) coupled this model with SAMI3 (Sami3 is Also a Model of the Ionosphere) to model the plasmasphere dynamics during two events in 2001. De Pascuale et al. (2018) used RAM-CPL to simulate equatorial plasmaspheric electron densities during two storm events in 2013, and compared them to in-situ measurements from the Van Allen Probes (Radiation Belt Storm Probes). Huba and Krall (2013) used the first-principles physics-based model SAMI3 to model the plasmasphere in 3D. They incorporated the neutral wind dynamo potential, the corotation potential, and a time-dependent potential from Volland (1973) and Maynard and Chen (1975) to model the convection potential for an idealized magnetic storm. An overview of various other physics-based models of the plasmasphere based on the fluid and the kinetic approaches is given in Pierrard et al. (2009).

The physics-based models rely on a number of physical processes, which are usually parameterized empirically in the model (e.g., refilling, electric and magnetic fields, etc.). Such parameterizations tend to be simplified as they are based on statistical averages over certain parameters (such as L-shell, MLT, or others). This can lead to inaccuracies in the physics-based model associated with such simplified assumptions. Therefore, it would be ideal to develop an optimal approach combining the advantages of both neural network- and physics-based models, namely the stability of physics-based models during geomagnetic storms, and the ability of neural networks to reproduce realistic density distributions for various events as they are independent of other parameterizations, such as refilling, magnetic and electric field models, etc..

One possible way to combine the models is to employ data assimilation. Data assimilation is a mathematical tool designed for combining a model with typically sparse observations in an optimal way (Kalman, 1960). In data assimilation, the information provided by both the physical model and the available observations is used to find the most likely estimate of the unknown true state of a dynamic system, while taking into account their uncertainties. The sequential Kalman filter (Kalman, 1960) is one popular algorithm of data assimilation. It uses predictions and observations in a recursive manner to improve the system measurements. It has numerous applications in technology, including the navigational system on Global Positioning System devices and the Apollo mission (Grewal & Andrews, 2010), image processing (Salti et al., 2014; Bresson et al., 2015), ocean modeling, operational weather forecasting (Kalnay, 2003; Lahoz et al., 2010; Sorenson, 1985), and reconstruction of the global state of the radiation belts (e.g., Y. Shprits et al., 2007, 2013).

In this study, we employ the Kalman filter technique to combine a neural network- and physics-based models in an optimal way. We use a version of the four-dimensional physics-based Versatile Electron Radiation Belt code (Y. Y. Shprits et al., 2015; Aseev et al., 2016), VERB-CS code (CS stands for "Convection Simplified"), to model the plasmasphere dynamics in the equatorial plane. The physics-based VERB-CS code (Aseev and Shprits, 2019) was initially developed to model the dynamics of the ring current, but can be adjusted to model the plasmasphere dynamics as well, which is done in this study. We treat the output of the neural network model PINE (I. Zhelavskaya et al., 2017) as "observations" of plasma density in the data assimilation setup. PINE is purely data-driven and produces realistic density reconstructions that have a remarkably similar distribution to actual density measurements and reproduces the shape of the plasmasphere bulge and plumes. To ensure that the models perform well quantitatively and reproduce point satellite measurements accurately, we compare their output to in-situ electron density measurements obtained from the Van Allen Probes for an 18-month out-of-sample interval from 30 June 2016 to 01 January 2018. Additionally, we compare the global evolution of plasma density predicted by the models with the global He^+ images obtained from the IMAGE EUV to validate the models qualitatively and ensure that they reproduce the global dynamics of the plasmasphere correctly.

The paper is structured as follows: In section 2, we describe the data used for training and validation of the models, i.e., in situ density measurements from the Van Allen Probes and the plasmapause position derived from the IMAGE EUV. We describe the neural network, the physics-based VERB-CS code, and the Kalman filter and how it is used to develop the assimilative model in section 3. In section 4, we present the results obtained with the models for the Halloween storm in 2003 and several events from 2001, and also for a long-term density reconstruction. Finally, in sections 5 and 6, we discuss implications and possible improvements to the models developed in this study.

2 Data

All magnetic field, solar wind data, and geomagnetic indices have been downloaded from the OMNIWeb data service. We have used the density dataset obtained with the NURD (Neural-network-based Upper hybrid Resonance Determination) algorithm (I. Zhelavskaya et al., 2016) for the period from 01 October 2012 to 01 July 2016, to train the neural networks. I. Zhelavskaya et al. (2016) employed feedforward neural networks to identify the upper hybrid resonance bands in the dynamic spectrograms made with the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) suite (Kletzing et al., 2013) onboard the Van Allen Probes satellites and calculated the plasma density from the upper-hybrid resonance frequency. The electron density data set is publicly available at the GFZ Data Services (I. Zhelavskaya et al., 2020). The Van Allen Probes provide electron density measurements for all local time sectors and $L \sim 2 - 6\ R_{\mathrm{E}}$. We use density measurements for a period of 30 June 2016 to 1 January 2018 (obtained with the same method) to quantify the performance of all the models developed in this study in section 4.3.

To validate the global output of our models, we use the plasmapause locations derived from the EUV instrument on board the IMAGE satellite (Sandel et al., 2000). The

IMAGE EUV instrument provided the global images of the plasmasphere for the first time. These images can be used to derive the location of the plasmapause by considering the outermost sharp edge of He^+ (Goldstein, Spasojević, et al., 2003). Goldstein, Sandel, Forrester, and Reiff (2003) showed that the sharp edge of He^+ in the EUV images corresponds to the actual location of plasmapause. We use the density threshold of 40 ± 10 cm^{-3} as an approximation of the plasmapause position in the global reconstructions of density produced by the models, which corresponds to the lower sensitivity threshold of the IMAGE EUV instrument (Goldstein, Sandel, Forrester, & Reiff, 2003). It is worth noting that the IMAGE mission operated in 2000 – 2005, which was a different solar cycle compared to the one we used in the training of the neural network. Therefore, the IMAGE EUV images are the best available data source for validating the global evolution of the shape of the plasmasphere produced by the models developed in this study. The plasmapause database derived from the IMAGE EUV instrument was obtained from `http://enarc.space.swri.edu/EUV/`.

3 Methodology

3.1 The neural network-based model of plasma density PINE

We utilize the output of the PINE model (I. Zhelavskaya et al., 2017) as "observations", which we combine with the physics-based VERB-CS code modeling the evolution of plasma density in the data assimilation setup. I. Zhelavskaya et al. (2017) used feedforward neural networks to model the global plasmasphere dynamics in the equatorial plane. They used geomagnetic parameters, their time histories, and the location given by L and MLT as input variables to the model. The plasma density in the equatorial plane of the Earth was the only output. The neural networks were trained on a 4-year plasma density dataset obtained from the Van Allen Probes plasma wave measurements. These density measurements were derived using the Neural-network-based Upper hybrid Resonance Determination (NURD) algorithm for automatic inference of the electron number density from plasma wave measurements made by the Van Allen Probes (I. Zhelavskaya et al., 2016). The model was extensively validated by means of K-fold cross validation to ensure that it does not overfit and generalizes well on unseen data. Furthermore, its global output was compared to the collection of global images of the He^+ distribution in the plasmasphere obtained with the EUV instrument of NASA's IMAGE mission to ensure that the model produces reasonable global density reconstructions (e.g., Figures 4, 8, and 9 in I. Zhelavskaya et al. (2017)). The model works well for quiet and moderately disturbed events (Kp $<$ 7), but its performance is limited during strong geomagnetic storms due to the lack of such examples in the training data set.

In the original study, the authors used K-fold cross validation with K=5 to train and validate the model. They used this procedure also to find the optimal inputs to the model (for more details, please see Appendix A). The training and validation datasets were constructed by randomly dividing the whole dataset into K=5 subsets, where in each iteration, one subset was left aside and used to validate the model (not used for training), while the rest of the K $-$ 1 subsets were used to train a neural network. It should be noted that while the division of data into training, validation, and test sets is carried out in a random fashion in that study, the more secure way to perform splitting for the time series is to divide data sequentially. The sequential division guarantees independence of all three subsets, while random division may produce optimistic evaluations on the validation and test sets for the events outside of the time period of the dataset. Nonetheless, the network resulting from training conducted using the random division would still have a good performance when reconstructing the past events.

In this study, we expand the analysis performed in I. Zhelavskaya et al. (2017) by conducting the K-fold cross validation procedure using sequential division of data into training and validation sets. We use an approach similar to the one implemented in I. S. Zhelavskaya et al. (2019). In that study, the authors implemented an approach incorporating both sequentiality and randomness in splitting the data into training and validation sets. The motivation behind that is that, as discussed above, random division into folds may lead to optimistic evaluations on the validation set, since such splitting causes a correlation between the training and validation sets. The sequential splitting, in turn, may lead to a significantly different distribution of the target variable in the training and validation sets. For example, it may occur that the validation or training set does not contain periods of high geomagnetic activity due to the way the data were split. Therefore, I. S. Zhelavskaya et al. (2019) implemented an intermediate solution. They first split the data into 35-day blocks sequential in time, and then assigned these 35-day blocks randomly to the CV folds for either training or validation. The reason for using blocks of a 35-day length is to avoid the possible effect of the 27-day recurrence caused by the solar rotation.

We apply the K-fold CV procedure described above to the density measurements from the Van Allen Probes. The obtained results are described in detail in Appendix A. In summary, we confirm the findings of the original study of I. Zhelavskaya et al. (2017). The models based on the geomagnetic indices yield the best performance, compared to the models based only on solar wind or on both solar wind and geomagnetic indices. As discussed in the original study, the models based on the solar wind inputs are less accurate than models based on geomagnetic indices. At the same time, the models based on both data sources tend to overfit the training data. We find that the optimal model is based on the 48-hour time history of geomagnetic indices AE, Kp, Sym-H, and also F10.7. The model also includes the location input given by L and MLT. This updated version of the PINE model is used in this study.

3.2 The physics-based model of plasma density VERB-CS

The evolution of the plasmasphere density in the equatorial plane can be described by the following equation:

$$\frac{\partial n}{\partial t} + v_\phi \frac{\partial n}{\partial \phi} + v_R \frac{\partial n}{\partial R} = S - L \tag{1}$$

where n denotes the plasma density; ϕ is the MLT; R is the radial distance in the equatorial plane; v_ϕ and v_R are drift velocities in MLT and radial distance, respectively; S is the source of charged particles; and L includes loss processes. The second and third terms describe the transport of the plasmaspheric particles due to the $E\times B$ drift. Refilling is taken into account by the source term S, and the loss term L accounts for the loss of the particles into the interplanetary medium.

We calculate $E\times B$ drift velocities using the dipole magnetic field approximation and assuming that the electric field is a superposition of co-rotation, convection, and sub-auroral polarization stream (SAPS)-driven electric fields. The co-rotation electric field is calculated from the electrostatic potential:

$$\phi_{\mathrm{CR}} = -\frac{A_{\mathrm{CR}}}{r},\ A_{\mathrm{CR}} \approx 92\ \mathrm{kV/R_E} \tag{2}$$

To calculate the convection electric field, we use the Kp-dependent Volland-Stern electric field model (Maynard & Chen, 1975; Stern, 1975; Volland, 1973):

$$\phi_{\mathrm{VS}} = -A_{\mathrm{MC}} r^2 \sin(\phi),\ A_{\mathrm{MC}} = \frac{0.045}{1 - 0.159\mathrm{Kp} + 0.0093\mathrm{Kp}^2} \tag{3}$$

We use a shielding parameter $\gamma = 1.8$ instead of standard $\gamma = 2$, as our experiments show that using $\gamma = 1.8$ agrees better with observations (more details are provided in the Results and Discussion sections). We include the effect of SAPS in the model by using the Kp-dependent model of the SAPS electric field developed by Goldstein et al. (2005). This model quantitatively includes the average properties of SAPS reported in Foster and Vo (2002). The SAPS has an effect on the location of the dusk side plasmapause and influences the shape and location of plasmaspheric plumes.

To account for refilling, we use refilling rates of equatorial electron density from Denton et al. (2012). These rates were inferred from passive radio emissions measured by the IMAGE RPI instrument during quiet geomagnetic times and are valid for the range $L = 2$ to 9. The model provides median, mean, 1st and 3rd quartiles of the refilling rates. In our study, we have used the median refilling rate: $\log_{10}(\frac{dn_{\mathrm{e,eq}}}{dt}) = 2.22 - 0.006L - 0.0347L^2$ (in cm^{-3}/day).

The escape of particles from the plasmasphere into the interplanetary medium through the magnetopause can be described by the loss term L of the form

$$L = \frac{n}{\tau} \tag{4}$$

where τ is a lifetime parameter. To model magnetopause loss, we set τ close to 0 outside of the magnetosphere, and to a very large number inside the magnetosphere. The boundary of the magnetosphere, the magnetopause, is calculated using the Shue et al. (1998) model.

To solve equation (1) numerically, we employ the VERB-CS code (Aseev & Shprits, 2019). The VERB-CS code models electron transport in ambient electric and magnetic fields and loss due to interaction with plasma waves. The VERB-CS code solves the two-dimensional advection equation that describes the particle drift, and we have extended the code to solve equation (1) by introducing losses to the magnetopause and the source term S.

Equation (1) must be complemented by initial and boundary conditions. To specify the initial conditions, we use the empirical density model of Sheeley et al. (2001) and the model of plasmapause by Carpenter and Anderson (1992). The models by Sheeley et al. (2001) provide the mean and the standard deviation of measurements for the plasmasphere and trough, and are valid for $3 \leq L \leq 7$ and all local times. To extend the density to lower L-shells, we use the density at $L = 3$ for $L < 3$. We start our simulations using the VERB-CS code during geomagnetically quiet intervals (Kp ≤ 2) and therefore assume that the plasmasphere is symmetric in MLT at the beginning of each simulation.

The boundary conditions are periodic in MLT and constant in R. They are set at $R_0 = 1.75$ and 10 R_{E} with 0.2-R_{E} and 0.5-hour grid steps in radial distance and MLT, respectively. We use the Sheeley et al. (2001) model to set the inner boundary conditions at 1.75 R_{E}. We use a statistical model of electron plasma sheet density developed by Dubyagin et al. (2016) to set the outer boundary conditions at 10 R_{E}. The model is valid for the nightside MLT sectors and distances between 6 and 11 R_{E} and is based on ~ 400 h of particle measurements from the THEMIS mission. The model is parameterized by the average of the solar wind proton density over 4 h and the average of the southward component of interplanetary magnetic field (IMF B_{S}) over 6 h. We assume that the electrons at 10 R_{E} reside on the open drift paths at 10 R_{E} on the dayside and set the outer boundary conditions to 0 from 6 to 18 MLT.

The plasmasphere is known to reach saturation after prolonged periods of quiet geomagnetic conditions (Park, 1974; Xiao-Ting et al., 1988; Lawrence et al., 1999; Su et al., 2001). To account for this effect, we have imposed a saturation upper limit of density on the code output. We have used the saturated density model of Carpenter and Anderson (1992). It is worth noting that this model provides an average of plasmasphere density observed after periods of relatively quiet geomagnetic conditions for at least 62 hours, rather than a theoretical upper limit. However, the ease of use of this model makes it a good choice for the purposes of this study, namely to illustrate the application of data assimilation to combining neural network and physics-based models together in an optimal way.

3.3 The assimilative model

In this section, we outline the Kalman filter technique and describe its application to the fusion of the physics-based and empirical models of the plasmasphere.

The Kalman filter is a popular technique for data assimilation. It is commonly used to adjust model predictions in accordance with available, typically sparse, observations, while taking into account uncertainties of both the model and observations (Kalman, 1960).

In this study, we adapt the Kalman filter technique to combine the predictions of two models, namely the physics-based VERB-CS code and the neural network-based PINE model. For this purpose, we consider the VERB-CS code as a model that propagates a state of the system in time. The output of the data-driven PINE model, in turn, is used as observations.

3.3.1 The Kalman Filter

The Kalman filter consists of two steps: the forecast step and the analysis step. These steps are repeated in cycles. In the forecast step, the model is used to issue predictions at the current time step t_k, using the previous state of the system, if available. The output of this step is called the forecast of the system. In the analysis step, this forecast is updated in an optimal way, given the observations at time t_k. The output of this step is called the analysis. At this point, the cycle of the Kalman filter is complete and the next iteration can start at time t_{k+1}. Before describing each of the steps in more detail, several notations need to be introduced.

Let us denote the model forecast of the density at time t_k by $\mathbf{n}_k^f$. Hereinafter, subscript k is an index of time t_k, bold lowercase letters denote vectors that are obtained after discretization of a physical quantity, and bold upper-case letters denote matrices. Please note that all discretized variables are assumed to be vectors.

If equation (1) is linear, its discretized version can generally be written as

$$\mathbf{n}_k^f = \mathbf{M}_{k-1}\mathbf{n}_{k-1}^a, \tag{5}$$

where superscripts f and a denote forecast and analysis, respectively, and $\mathbf{M}_{k-1}$ is a matrix, also referred to as the model matrix or the model operator. This matrix can be obtained, for example, by applying a finite difference method to (1). At a given time t_{k-1}, the model matrix $\mathbf{M}_{k-1}$ propagates the current state of the system $\mathbf{n}_{k-1}^a$ to the next state in time $\mathbf{n}_k^f$. The analysis $\mathbf{n}_{k-1}^a$ is the best estimate of the state vector at time t_{k-1}, based on the model and the available observations. The evolution of plasma density can be modeled by sequentially solving equation (5) for $k = 1, 2, \ldots$.

If applied to a real ("true") state $\mathbf{n}_{k-1}^t$ of the system, the model matrix propagates $\mathbf{n}_{k-1}^t$ with some error ϵ_k^M. This error can originate from the uncertainties of the model, such as errors due to missing physical processes in the model or numerical errors due to discretization of the continuous equation (1). It is also referred to as the model error:

$$\mathbf{n}_k^t = \mathbf{M}_{k-1}\mathbf{n}_{k-1}^t + \epsilon_k^M. \tag{6}$$

The vector ϵ_k^m is usually assumed to be a Gaussian white-noise random variable with zero mean and covariance matrix $\mathbf{Q}_k$, which is referred to as the model error covariance matrix (i.e., $\mathbb{E}(\epsilon_k^M) = 0$ and $\mathbb{E}(\epsilon_k^M \epsilon_k^{M\top}) = \mathbf{Q}_k$, where $\mathbb{E}$ is the expectation operator). To correct the model error ϵ_k^M, we can exploit the information that observations provide. Given a true state of the system $\mathbf{n}_k^t$, that is defined on the same grid as the forecast $\mathbf{n}_k^f$, the measurements $\mathbf{n}_k^{obs}$ can be represented as follows:

$$\mathbf{n}_k^{obs} = \mathbf{H}_k\mathbf{n}_k^t + \epsilon_k^{obs}, \tag{7}$$

where $\mathbf{H}_k$ is referred to as the observation operator and ϵ_k^{obs} is the observation error. The role of the observation operator is to convert the true state from the model grid onto the grid of observations (these two grids are generally different). The observation error ϵ_k^{obs} can be associated with the measurement technique. Note that when we treat the output of the data-driven PINE model as observations, the error ϵ_k^{obs} includes errors of the PINE model predictions. The typical assumption is that vector ϵ_k^{obs} is a Gaussian white-noise random variable with zero mean and covariance matrix $\mathbf{R}_k$, also referred to as the observation error covariance matrix (i.e., $\mathbb{E}(\epsilon_k^{obs}) = 0$ and $\mathbb{E}(\epsilon_k^{obs}\epsilon_k^{obs\top}) = \mathbf{R}_k$).

The Kalman filter then combines the model forecast $\mathbf{n}_k^f$ with observations $\mathbf{n}_k^{obs}$ to obtain a prediction that is closest to the truth in the least squares sense, given the information about the model and observation error covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$. The optimal combination of the forecast and observations is referred to as analysis, $\mathbf{n}_k^a$, as mentioned above. The analysis $\mathbf{n}_k^a$ at time t_k can be obtained from the analysis $\mathbf{n}_{k-1}^a$ at the previous time step by sequentially solving the equations that constitute the Kalman filter described below.

Forecast step

The forecast step advances the forecast and the forecast error covariance. First, the analysis $\mathbf{n}_{k-1}^a$ obtained at time t_{k-1} is propagated to the next time t_k using the model matrix $\mathbf{M}_{k-1}$:

$$\mathbf{n}_k^f = \mathbf{M}_{k-1}\mathbf{n}_{k-1}^a. \tag{8}$$

Then, the forecast error covariance matrix $\mathbf{P}_k^f$ is updated according to:

$$\mathbf{P}_k^f = \mathbf{M}_{k-1}\mathbf{P}_{k-1}^a\mathbf{M}_{k-1}^T + \mathbf{Q}_{k-1}, \tag{9}$$

where $\mathbf{P}_k^a$ is the analysis error covariance matrix. The matrices $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$ are estimates of forecast and analysis errors, respectively. The forecast error covariance matrix $\mathbf{P}_k^f$ is used later in the analysis step.

Analysis step

In the analysis step, the forecast obtained in the previous step is updated according to observations:

$$\mathbf{n}_k^a = \mathbf{n}_k^f + \mathbf{K}_k\left(\mathbf{n}_k^{obs} - \mathbf{H}_k\mathbf{n}_k^f\right), \tag{10}$$

where $\mathbf{K}_k$ is referred to as the Kalman gain. The Kalman gain is a matrix of optimal weights that is used to correct the forecast based on available observations. The last term in the equation represents the correction to the forecast given the observations, weighted by the Kalman gain. The Kalman gain $\mathbf{K}_k$ is updated at time t_k as follows:

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}. \tag{11}$$

Finally, the analysis error covariance matrix $\mathbf{P}_k^a$ is updated as follows:

$$\mathbf{P}_k^a = \mathbf{P}_k^f - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^f. \tag{12}$$

This finishes the iteration k of the Kalman filter.

3.3.2 Details of implementation

There are several details of the Kalman filter implementation that should be taken into account, which we describe below.

The nonlinear term S in equation (1) does not allow us to write the discretization of the equation in the form (8). The non-linearity of the equation requires the extension of the Kalman filter equations (8) and (9) by linearizing the model operator. In order to simplify the implementation of the Kalman filter, we avoid the linearization of the model operator by running one step of the VERB-CS code instead of solving equation (8) to obtain the plasma density forecast $\mathbf{n}_k^f$ from a previous (optimal) state $\mathbf{n}_{k-1}^a$. The VERB-CS code solves the partial differential equation (1) numerically by discretizing density n, drift velocities v_φ and v_R, sources S, losses L, and spatial and temporal derivatives

$\frac{\partial}{\partial v_\varphi}$, $\frac{\partial}{\partial v_R}$, and $\frac{\partial}{\partial t}$. Discretization allows us to consider plasma density and other parameters at discrete times t_k, where k is an integer, on the grid consisting of discrete values of MLT and R.

To update the forecast error covariance matrix $\mathbf{P}_k^f$ in equation (9), we obtain the model matrix $\mathbf{M}_{k-1}$ by discretizing equation (1) without the source term S. For this, we use the first-order explicit upwind finite difference scheme with a time step that automatically adapts to changing drift velocities to satisfy the Courant stability condition. Such an approach allows us to take into account the refilling only in the forward model. Neglecting the refilling rates does not significantly affect the optimality of the Kalman filter, if the step of data assimilation is chosen to be much smaller than the characteristic time of the refilling (that is on the order of days, Denton et al. (2012) and references therein). In this study, the data assimilation is performed every 4 hours.

As mentioned in the previous paragraph, assimilation of the VERB-CS and the PINE model output is performed every 4 hours. This time allows the physics-based code to evolve the state of the system starting from the initial "blended" state. We note that this time was chosen empirically. Comparison with other times (3 and 5 hours, not shown here) showed that using 4 hours provides a slightly better performance. The assimilation is not performed when Kp $>$ 6 and for one day after the storm, i.e., only the output of the VERB-CS code is taken into account during the storm times and shortly after them, and the PINE output is not considered. This is done in order to avoid possible errors that can be propagated from the neural network model, as it is not reliable for Kp > 6.

Another aspect that should be noted is the implementation of the observation operator $\mathbf{H}_k$. This operator transforms the forecast of the model n_k^f from the model grid onto the grid of observations (see eq. (7)). In our case, the model grid is that of VERB-CS, and the observations grid is that of the PINE model. As discussed in the previous section, the spatial grid of VERB-CS ranges from 0 to 24 hours with 0.5-hour grid step in MLT and from 1.75 to 10 R_E with 0.2 R_E grid step in radial distance. In order to obtain the global output using the PINE model (i.e., on the whole equatorial plane and not just at specific L and MLT), we need to assume a spatial grid, on which the output is produced. The PINE model is valid for all MLT sectors, and from 1.75 to 6.15 R_E in radial distance due to the use of density measurements from Van Allen Probes for training. Therefore, the lower and upper boundaries of the PINE grid are set at 1.75 and 6.15

R_E in radial distance, respectively. In order to simplify the implementation of the Kalman filter, we use a spatial grid of the same resolution as the VERB-CS, namely with 0.5-hour and 0.2-R_E grid steps in MLT and radial distance, respectively. Thus, the spatial grid of PINE is a subset of the VERB-CS grid, which makes it easier to assimilate the PINE output. $\mathbf{H}_k$ is then defined as a matrix consisting of zeros and ones, where 1 corresponds to an element of this matrix when the model's (VERB-CS') grid point coincides with the observation (PINE's) grid point, and 0 otherwise. The number of rows in $\mathbf{H}_k$ is the number of grid points of PINE, and the number of columns is the number of grid points of VERB-CS.

In the standard formulation of the Kalman filter, the model and observation error covariance matrices are assumed to be known., That is rarely the case in practice, and simple approximations are typically made. One approach is to set up the model and observation covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ as diagonal matrices with elements $\alpha^m(n^f)^2$ and $\alpha^{obs}(n^{obs})^2$, respectively (Kondrashov et al., 2011). α^m and α^{obs} are referred to as model and observation errors, respectively, and are usually empirically chosen constants. If they are chosen to be equal to each other, both model and data contribute equally to the result of data assimilation, otherwise the result is dominated by either data or model. This approach is used in a number of studies in space physics, in particular for the radiation belt reanalysis (Daae et al., 2011; Y. Shprits et al., 2013; Kellerman et al., 2014; Cervantes et al., 2020). In this study, we employ an approach that builds on and extends this methodology. We use the same form of the model and observation covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, namely, diagonal matrices with elements $\alpha^m(n^f)^2$ and $\alpha^{obs}(n^{obs})^2$, respectively. However, we choose the model and observation errors, α^m and α^{obs}, to depend on the plasmapause position (instead of just being constant). Specifically, we assign different values to m and obs depending on whether n^f and n^{obs} are located inside or outside the plasmapause. The plasmapause is calculated according to a fixed density threshold of 40 cm^{-3} (the densities larger than the threshold are assumed to be inside the plasmasphere, otherwise – outside). We assign the model error inside the plasmapause $\alpha^m_\mathrm{inside} = 0.407$, the model error outside the plasmapause $\alpha^m_\mathrm{outside} = 0.507$, the observation error inside the plasmapause $\alpha^{obs}_\mathrm{inside} = 0.335$, and the observation error outside the plasmapause $\alpha^{obs}_\mathrm{outside} = 0.333$. The description of how these values were obtained is presented in Appendix B. Using such an approach, we obtain a better agreement between the assimilative model and observations, compared to using single constants for $\alpha^m(n^f)^2$ and $\alpha^{obs}(n^{obs})^2$, respectively. However, we choose the model and observation errors, α^m and α^{obs}.

4 Results

In this section, we perform several tests to compare the performance of the PINE, VERB-CS, and assimilative model, which is henceforth referred to as the Assimilative Magnetospheric Plasma density (AMP) model. We compare the outputs of the models during the 2003 Halloween storm and a number of events during March-June 2001. We validate the models by comparing the modeled and observed shape of the plasmasphere by using the plasmapause location obtained from the IMAGE EUV instrument. We also perform a long-term density reconstruction for the period of 30 June 2016 to 1 January 2018, using all the models. For the long-term run, we validate the models by comparing their output to the in-situ density measurements from RBSP-A. This period was not used in the training of the PINE model. The setup of all the models used in these tests is described in section 3.

4.1 Test 1: Halloween storm 2003

The first test we perform is to compare the performance of the models for the 2003 Halloween storm. The Halloween storm occurred from late October to early November 2003 and was one of the strongest solar storms observed during the satellite era. During this period, a series of energetic eruptions occurred, including two CMEs (coronal mass ejections), which struck the Earth, one shortly after another, with an extremely short (less than a day) Sun-Earth shock transit time (e.g., Gopalswamy, 2006). At the Earth, Kp reached 9 and Dst nearly -400 nT. Fortunately, the plasmapause locations derived from IMAGE EUV are available during some parts of the storm, which makes it an ideal event for testing the models for extreme geomagnetic conditions.

Figure 2 shows the global electron density reconstruction during the Halloween storm 2003 using the PINE (left column), VERB-CS (middle column), and assimilative (right column) models. The first four rows show the global snapshots of density, and the bottom row shows the Kp index during the event. The first four rows correspond to the specific times during the event when the plasmapause from the IMAGE EUV instrument

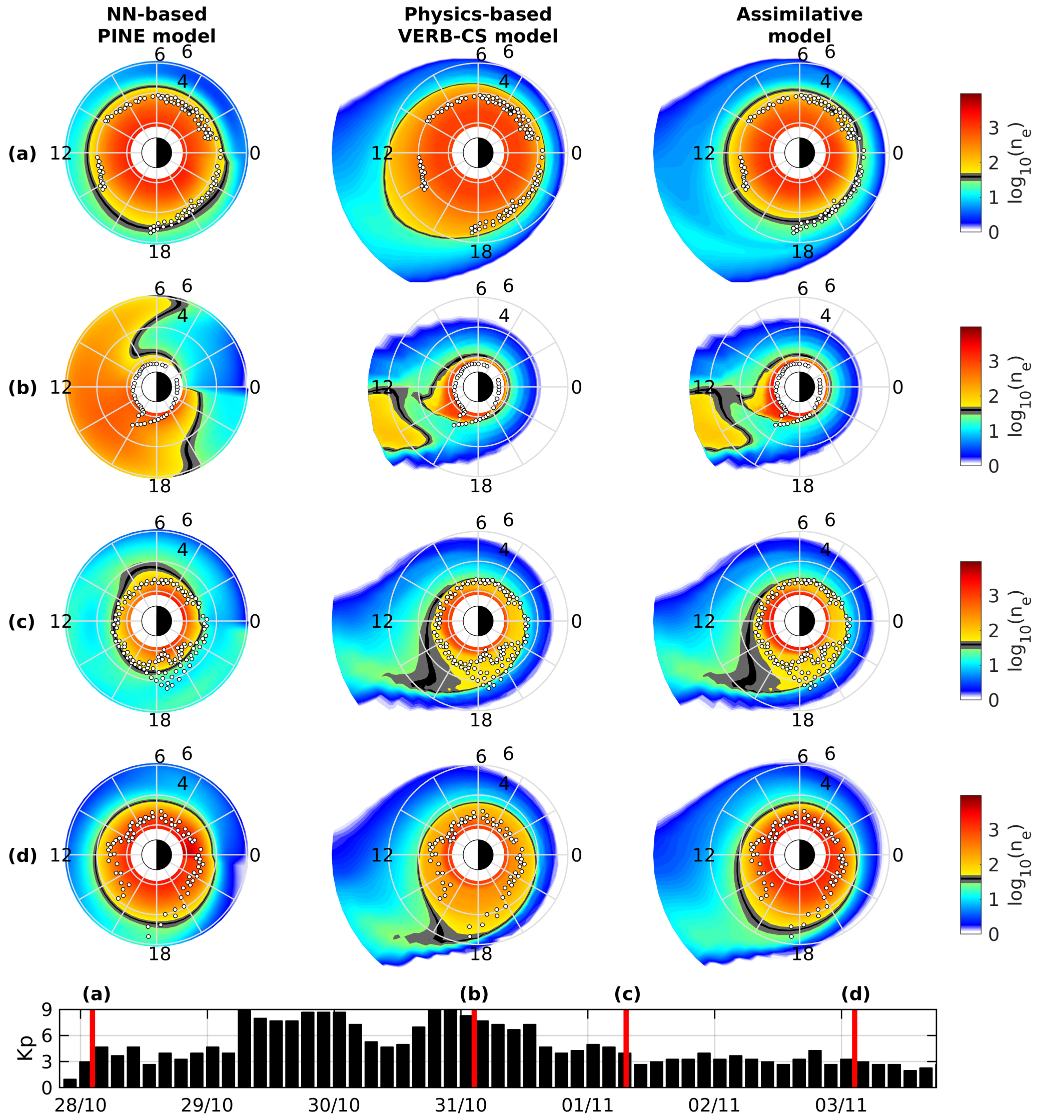

Figure 2. Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the 2003 Halloween storm. The first four rows show the outputs of the models corresponding to the times marked with the red lines in the bottom panel showing the Kp index during the 2003 Halloween storm. The black-and-white dots show the location of the plasmapause derived from the IMAGE EUV images. The color in the first four rows indicates the logarithm of density (the scale of the colorbar is the same for all models and all times). The gray and black section of the colorbar indicates a density threshold of 40 ± 10 cm^{-3} and can be considered a rough approximation of the plasmapause location for the sake of comparison to the observed plasmapause position obtained from IMAGE EUV (more details on that are given in section 2). The Sun is to the left. Row (a) corresponds to the time before the storm, (b) to the period during the storm (second CME), (c) and (d) to the recovery phase of the storm.

was available. These times are marked with the red lines and labels (a-d) in the bottom panel.

In order to obtain a global output using the PINE model, it was applied to each point on its spatial grid independently (described in section 3.3.2), and the smoothed global output shown in Figure 2 was obtained by interpolating between the points. The output of the VERB-CS code was obtained by running the model starting from quiet geomagnetic conditions (27-Oct-2003 20:00 UT, Kp = 1.7) with the setup described in section 3.2. The output of the assimilative model was obtained by running the model from the same time. Its setup is described in section 3.3. The time step of the simulations is 15 minutes for all the models. The time of the IMAGE plasmapause location is chosen to be as close as possible to the temporal grid of the models, but not further than 14 minutes away. Therefore, the time of the IMAGE plasmapause may not exactly coincide with the time of the simulation and may deviate from it by 14 minutes maximum. Although it is possible to set the exact time of the simulation for the PINE model, we choose to select the same temporal grid as in the VERB-CS and assimilative AMP models to ensure an equal comparison between all the models.

As seen in Figure 2, the PINE model agrees well with the plasmapause derived from the IMAGE EUV before the storm, but produces unrealistic global density reconstruction during the main phase of the storm. As discussed in the introduction, the reason for that is the absence of training examples during extreme geomagnetic events (there is no single Kp = 9 in the training dataset of PINE). After the storm (row (c)), the size of the plasmasphere reproduced with PINE is in good agreement with the IMAGE observations. On the contrary, the VERB-CS model produces an overly extended plasmasphere during the quiet time before the storm, but successfully reproduces the massive erosion of the plasmasphere (row (b)) observed in the IMAGE EUV observations as well. Several days after the storm (row (d)), VERB-CS produces lower densities inside the plasmasphere than those produced by PINE (this can be seen from the color in the density snapshots: yellow color in VERB-CS, compared to the red color in the PINE output).

The assimilative AMP model is in good agreement with IMAGE observations for all phases of the disturbance. The size of the plasmasphere before the storm is in better agreement with IMAGE plasmapause observations, compared to the VERB-CS output, and is closer to the size of the plasmasphere modelled with PINE. During the storm, the PINE output is not taken into account and, therefore, the assimilative model produces results similar to the physics-based VERB-CS model, successfully reproducing the massive erosion of the plasmasphere. After the storm, it produces densities closer to those obtained with PINE inside the plasmasphere (as can be seen from the color in the density snapshots), while the shape of the plasmapause is retained from both PINE and VERB-CS models. This example illustrates how a neural network-based and physics-based models can be combined in an optimal way to produce a more accurate global density reconstruction than each of them separately. Such a result is a good indication that the assimilative methodology is useful to model the plasmasphere dynamics during extreme geomagnetic events.

4.2 Test 2: Multiple events (March-June 2001)

In the previous section, we showed that the assimilation of the neural network- and physics-based models demonstrated good agreement with the plasmapause observations during the 2003 Halloween storm, and performed better than either of the models separately. In this section, we test the models further by comparing their output for a number of events in March-June 2001. We have selected 5 events corresponding to different Kp levels, starting from quiet geomagnetic conditions (Kp = 2.7) and reaching disturbed ones (Kp = 8). The motivation behind this selection was to test how the models perform separately and when combined by means of data assimilation for different levels of geomagnetic disturbance.

Figure 3 shows snapshots of global density reconstructions using the PINE, VERB-CS, and assimilative models for 5 different events in 2001. The format is similar to Figure 2. The columns correspond to models, as labeled in the top row. The rows correspond to events. The times of the density snapshots and the corresponding Kp values are labeled in each row on the left. The events are ordered by increasing Kp index, rather than by time. The format of the density snapshots is the same as in the top four rows of Figure 2.

The global density reconstructions are obtained in the same fashion, as described in the previous section. Spatial and temporal grids of the models and their setup are also the same as used there. We note again that the time grid step is 15 minutes, and therefore may not exactly coincide with the timing of the plasmapause observations derived

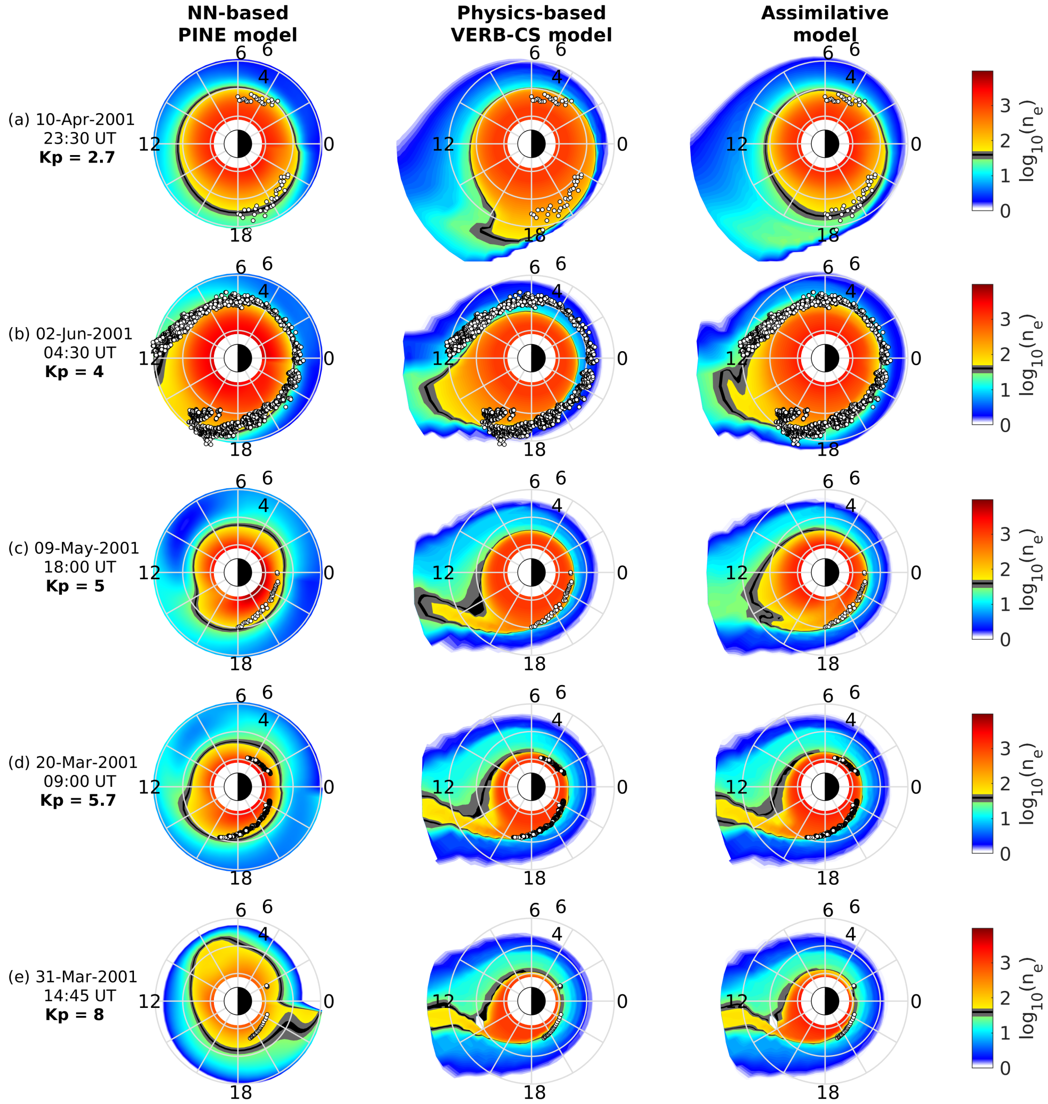

Figure 3. Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during a series of events in 2001, as indicated in the labels on the left in each row. The format of the density snapshots is the same as in Figure 2.

from IMAGE EUV (but is not farther than 14 minutes away). The simulations were run separately for 5 different events, each starting from quiet geomagnetic conditions. The starting time and Kp at the beginning of the events were: (a) 10 April 2001 03:00 UT (Kp = 1.7), (b) 01 June 2001 00:00 UT (Kp = 0.3), (c) 08 May 2001 03:00 UT (Kp = 0.7), (d) 18 March 2001 12:00 UT (Kp = 1), and (e) 30 March 2001 11:00 UT (Kp = 2).

It can be seen that PINE performs well during low and moderate geomagnetic activity (rows a-c), i.e., the modelled plasmapause agrees well with the one observed with IMAGE, similar to the results of the previous section. However, for a more disturbed event, such as in row (e), when Kp = 8, it produces an abnormal artifact on the night side. On the contrary, the physics-based VERB-CS model performs very well for the disturbed times (rows d and e): the modelled plasmapause matches exactly the one observed with IMAGE. However, for the quiet event shown in row (a), it produces an overly expanded plasmasphere, compared to the observed plasmapause. For the event in row (b), when Kp = 4, the plasmasphere produced by VERB-CS is more eroded than was observed.

The assimilative model blends the outputs of both models in an optimal way for all the tested events. Its output is closer to the output of the VERB-CS code during the disturbed intervals (rows d-e) and to the output of the PINE model for the quiet times (rows a-b). For the event (c), the output of the assimilation appears to be somewhat in between the outputs of the PINE and VERB-CS models. This test illustrates that the output of the combined model agrees better with the plasmapause observations from IMAGE than the output of each of the models used separately, not only for extreme geomagnetic storm, but also for quiet and moderately disturbed events.

4.3 Test 3: Long-term reconstruction of density

In the two previous sections, we have illustrated the performance of the assimilative, PINE, and VERB-CS models for different geomagnetic conditions, including quiet, moderate, and extreme conditions. The assimilative model demonstrated a better performance compared to PINE and VERB-CS used separately for all considered events. In this section, we test the performance of all models further by performing a long-term reconstruction of plasma density using all the models. We compare the modeled density

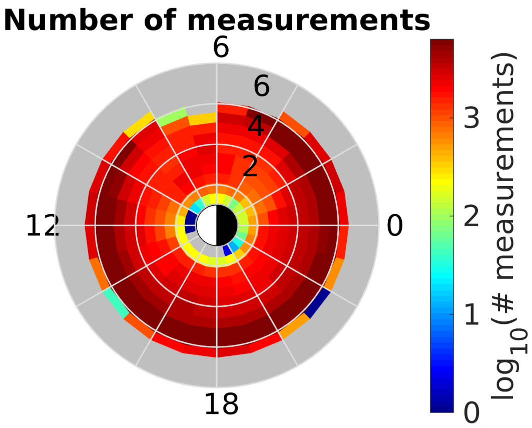

Figure 4. Coverage of RBSP-A during 30 June 2016 – 01 January 2018.

with the in-situ density measurements from RBSP-A during the period between 30 June 2016 and 01 January 2018. We choose this particular interval for testing, as the density measurements during this time were mostly not used in the training of the PINE model. Therefore, this choice ensures a fair comparison between the performance of PINE, VERB-CS, and the assimilative model. Furthermore, RBSP-A crosses all MLT sectors during this interval, as shown in Figure 4, which allows us to evaluate the performance of the models in different MLT sectors. The setup of all three models is the same as in the two previous sections. Summary plots demonstrating performance metrics calculated during this period for all three models are shown at the end of this section. It is worth noting that the PINE model was trained on the interval 01 October 2012 – 01 July 2016 (I. Zhelavskaya et al., 2017), and therefore, we exclude the period 30 June – 01 July 2016 when calculating the performance metrics here. We choose 30 June 2016 as the start time of the simulation as the Kp index was smaller than on 02 July 2016 (0.3 vs. 0.7), and also since there was a minor disturbance (Kp = 3.3) between 30 June and 02 July, which could negatively influence the initial conditions for VERB-CS.

Figure 5 presents a comparison of the output of the neural network density model in-situ density measurements from RBSP-A from 30 June 2016 to 01 January 2018. Panel (a) shows the in-situ density observations from RBSP-A. Panel (b) shows the output of the PINE model. These two panels have the same format: the x-axis corresponds to time, the y-axis to the L-shell, and the color indicates the logarithm of electron density. The next two panels (c) and (d) show the difference between the observations and the out-

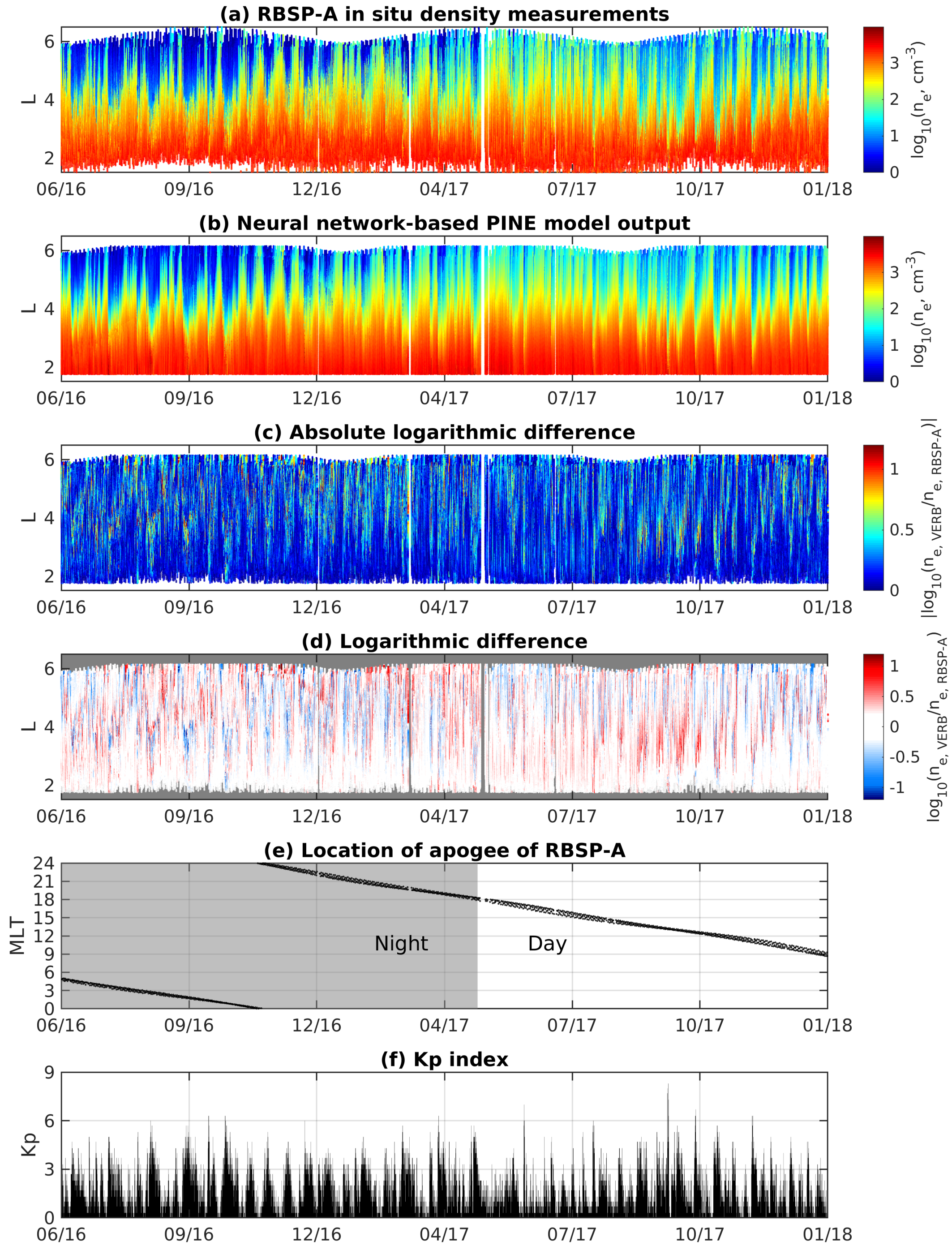

Figure 5. Long-term comparison of the PINE model and the RBSP-A density measurements during July 2016 – January 2018. Panels (a) and (b) show the RBSP-A density measurements and the output of the PINE model, respectively, where the L-shell is on the y-axis, time is on the x-axis, and the color indicates the log of density. Panels (c), (d), and (e) show the absolute difference, the sign of the difference, and the difference between log of the model and data, respectively.

put of the model. Panel (c) shows the absolute difference between logarithms of modelled and observed density. Panel (d) shows the difference itself. The blue color in this panel implies underestimation of density by the model (the modelled density is smaller than observed), the red color overestimation. Panel (e) shows the location of RBSP-A's apogee during this interval. The gray shaded area implies that the apogee of RBSP-A was at the nightside (i.e., from 18 to 6 MLT). It is worth noting that the apogee of RBSP-A is located in the night sector during approximately the first half of the interval, and therefore, the densities on the farther L-shells are smaller (dark blue color), compared to the second half of the interval (where the color is light green on the farther L-shells). During the second half of the interval, RBSP-A's apogee was located on the dayside, and therefore, the density is higher there due to plasmaspheric bulge and plume. The bottom panel shows the Kp index during this period.

To obtain the model output at the L- and MLT-coordinates of RBSP-A, the model was first applied to the full spatial grid of L and MLT. Then, a virtual satellite was flown through the model output at the coordinates closest to the L- and MLT-coordinates of RBSP-A, and after that, the output was interpolated to these coordinates. Although the PINE model can be directly applied to specific L and MLT coordinates without the need to make a virtual flyby, such a procedure was nonetheless employed in order to obtain a consistent comparison with VERB-CS and the assimilative model.

It can be seen that the PINE model output is very similar to the observations. The model captures the expansion of the plasmasphere that occurs during periods of quiet geomagnetic conditions and erosion associated with geomagnetic disturbances. For example, the massive erosion of the plasmasphere during the September 2017 storm is captured by the model. Moreover, the positive and negative differences between the model output and observations (shown in panel (d)) are spread randomly over the duration of the simulation, which indicates that there is no systematic bias in the model. Overall, these results show that the PINE model performs well on the out-of-sample period (i.e., the period not used in the training).

Figure 6 shows the comparison between in-situ density from RBSP-A and the output of the physics-based VERB-CS code. The format of the figure is the same as in Figure 5, where panel (b) presents the output of VERB-CS, and panels (c) and (d) show the difference between the modelled and observed density in different formats. The model

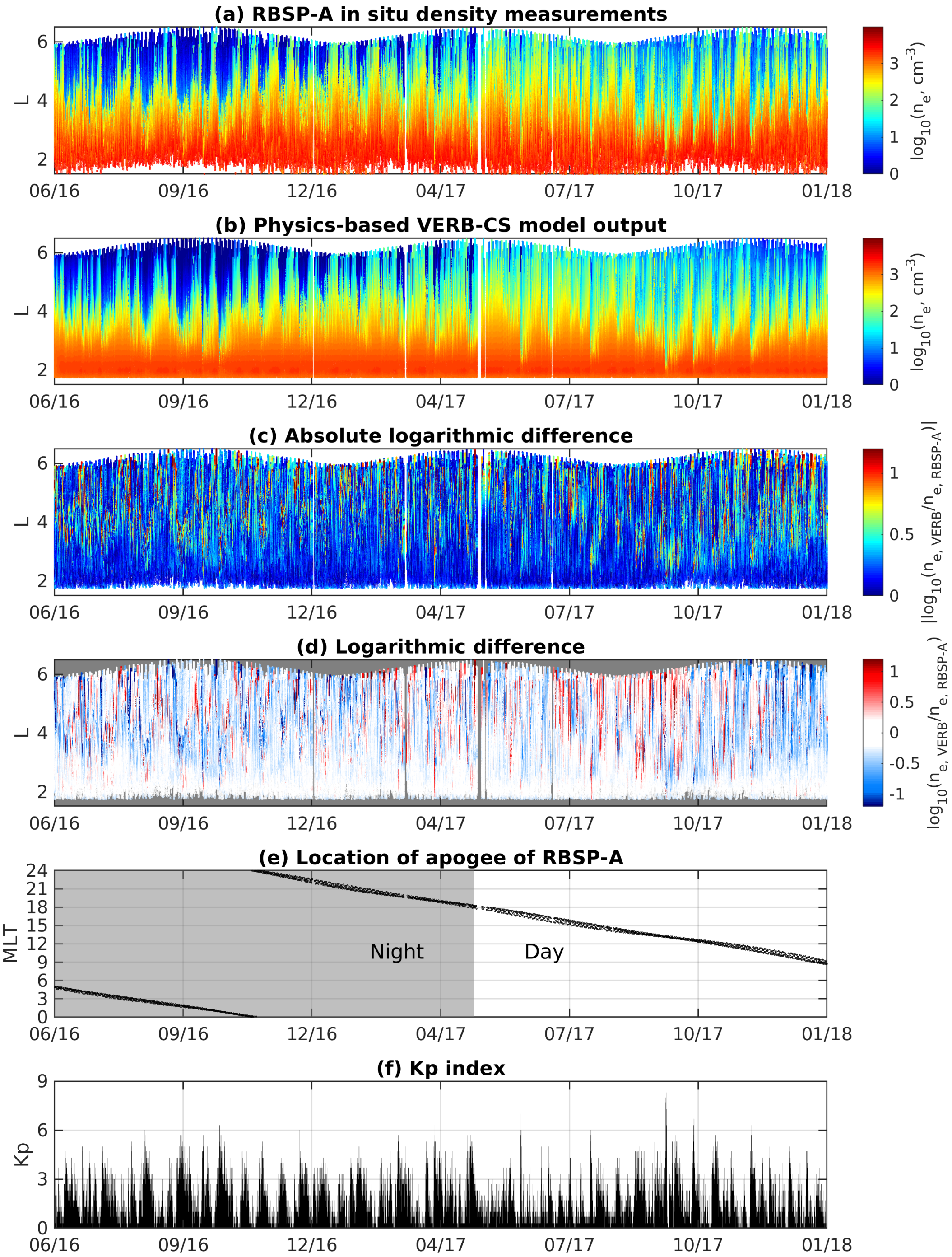

Figure 6. Long-term comparison of the physics-based model and the RBSP-A density measurements during July 2016 – January 2018. Panels (a) and (b) show the RBSP-A density measurements and the output of the physics-based model, respectively, where the L-shell is on the y-axis, time is on the x-axis, and the color shows the log of density. Panels (c), (d), and (e) show the absolute difference, the sign of the difference, and the difference between log of the model and data, respectively.

output at the coordinates of RBSP-A was obtained in the same fashion as in Figure 5: the model was first run on the full spatial grid, and then the virtual satellite was flown through the output at the L- and MLT-coordinates of RBSP-A.

It can be seen that the VERB-CS model captures the general dynamics of the plasmasphere, i.e., its erosion and expansion, well. Again, the model reproduces a massive erosion of the plasmasphere during the September 2017 storm. It can be seen, however, that the differences between observations and the output of VERB-CS shown in panel (c) are larger than those of the PINE model (shown in Figure 5). As can be seen from panel (d), the VERB-CS model tends to systematically underestimate observations (visually, the blue color is predominant). In particular, an underestimation can be seen in the first half of the time interval, when RBSP-A's apogee was located at the nightside, and in the 9-12 MLT sector.

Finally, Figure 7 presents the comparison between in-situ density measurements from RBSP-A and the output of the assimilative model. The format is the same as in Figures 5 and 6, where panel (b) shows the output of the assimilative model and panels (c) and (d) show the difference between the model output and observations in different formats (as described below in Figure 5). The output of the model was obtained in the same manner as for the other models. The assimilative model was first run on the full spatial grid, and then a virtual satellite was flown through the global output of the assimilative model along the RBSP-A coordinates.

It can be seen from the figure that the assimilative model successfully captures the general dynamics of the plasmasphere, i.e., erosion and expansion associated with corresponding geomagnetic conditions. Although, similarly to VERB-CS, it produces lower densities on the nightside (first half of the interval), its output is in better agreement during the rest of the interval, compared to the VERB-CS model: the underestimation that was observed in the VERB-CS output is reduced. Consequently, the errors of the assimilative model are larger than those of PINE on the nightside but are comparable or even lower than those of PINE on the dayside. The densities inside the plasmasphere (at low L-shells) are lower compared to the observations from RBSP-A, which is caused by the use of the saturation density model (Carpenter & Anderson, 1992) in the assimilative model setup as well. Overall, the performance of the assimilative model improves on the dayside compared to VERB-CS and is similar to that of PINE. On the nightside, the as-

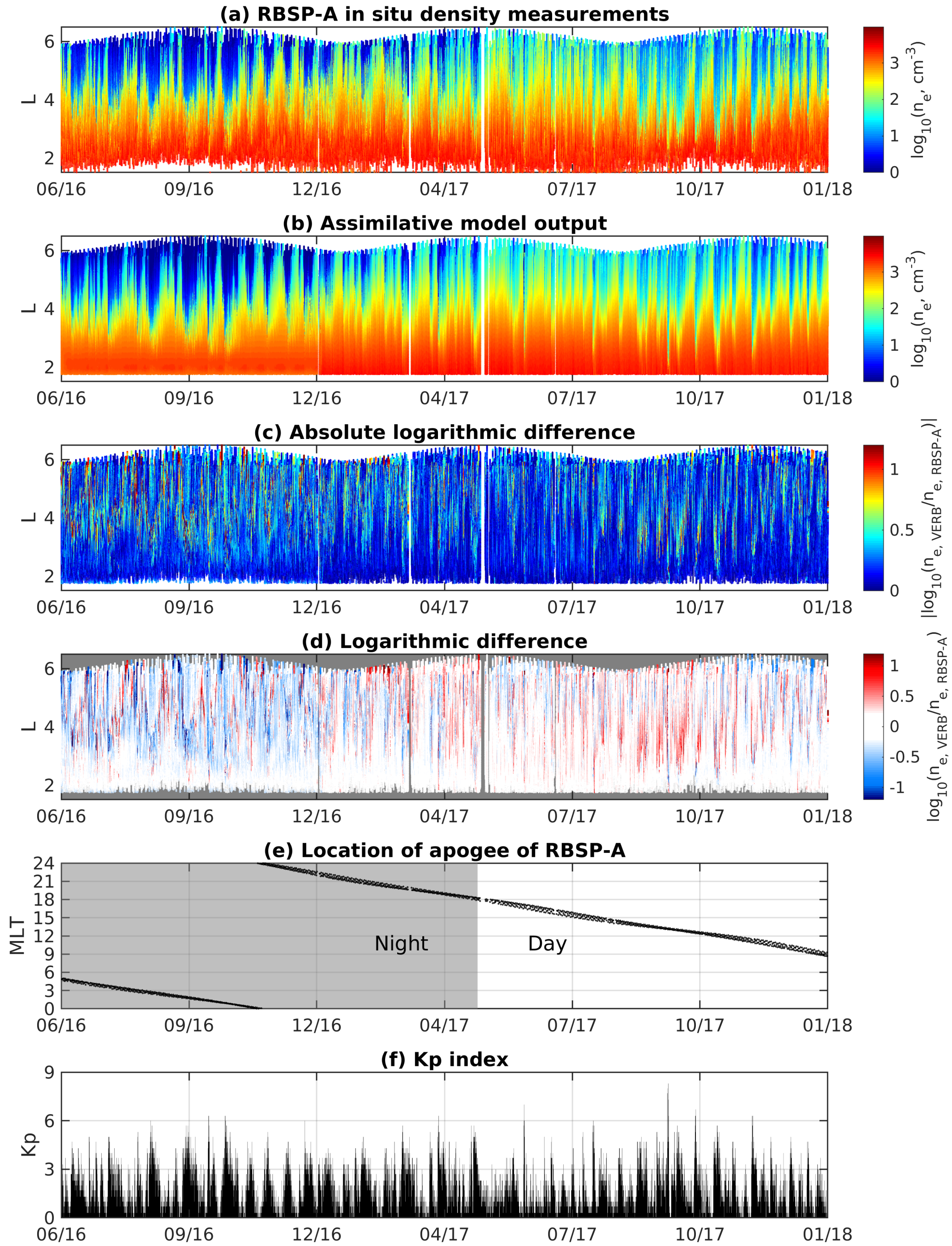

Figure 7. Long-term comparison of the assimilative model and the RBSP-A density measurements during July 2016 – January 2018. Panels (a) and (b) show the RBSP-A density measurements and the output of the assimilative model, respectively, where the L shell is on the y-axis, time is on the x-axis, and the color shows the log of density. Panels (c), (d), and (e) show the absolute difference, the sign of the difference, and the difference between log of the model and data, respectively.

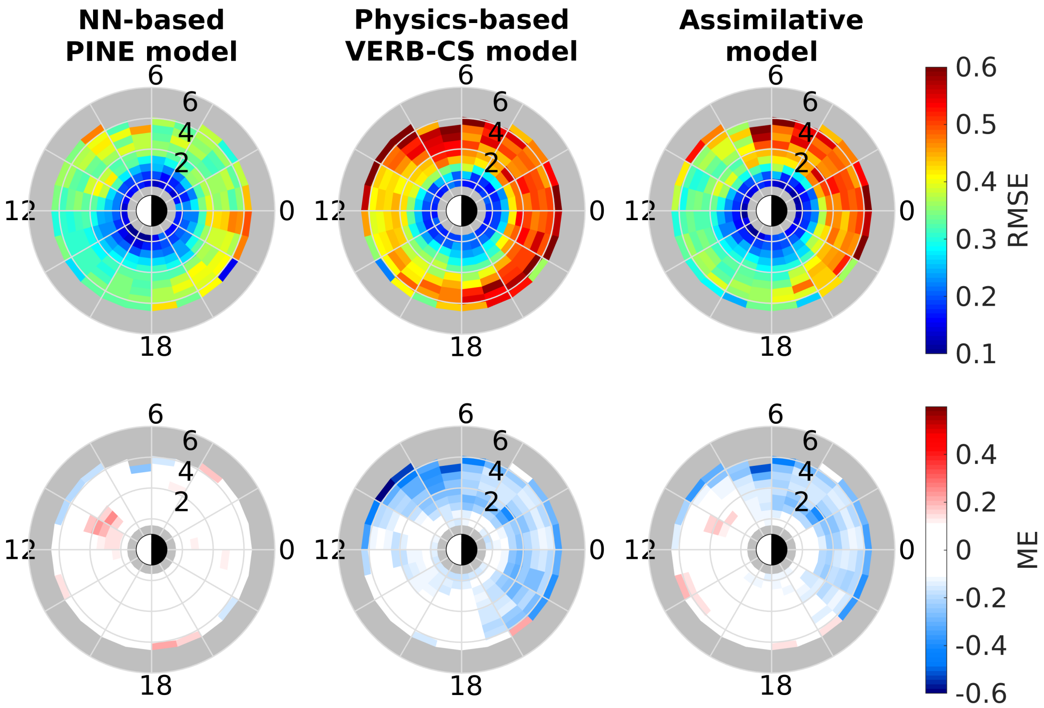

Figure 8. The root-mean-square error (top row) and the bias or mean error (bottom row) of the PINE, physics-based, and assimilative models for the 02 July 2016 – 01 January 2018 period. The Sun is to the left. The colorbar of each row shows the value of the corresponding metric (RMSE or ME). The colorbar limits are the same for all models in each row.

similative model produces results closer to the VERB-CS and tends to underestimate the density.

4.3.1 Performance metrics

To obtain a general overview of the performance of all the models, it is helpful to examine the performance metrics calculated for each model over the whole time period under consideration. We use the root-mean-square error (RMSE) and mean error (ME) or bias to analyse the performance of all the models in different L and MLT sectors.

Figure 8 shows the RMSE (top row) and ME/bias (bottom row) of the PINE (left column), assimilative (middle), and VERB-CS (right) models calculated over the period from 02 July 2016 to 01 January 2018, used in the long-term simulations shown in Figures 5-7. It is worth mentioning again that the PINE model was trained on the interval 01 October 2012 – 01 July 2016 (I. Zhelavskaya et al., 2017), and therefore we exclude the period 30 June – 01 July 2016, when calculating the performance metrics here. In order to calculate the metrics, the data are binned in L and MLT, and the performance metrics are computed separately for each bin. The bins in L range from 1.5 to 6.5 with 0.5 bin size, and in MLT from 0 to 24 with bin size 1.

It can be seen that all the models have lower errors closer to the Earth (inside the plasmasphere), and that the errors increase with L. The errors of the PINE model are the lowest out of all models in terms of both RMSE and bias in all bins. The errors of the physics-based model are larger on the nightside and smaller for 10-18 MLT. This result is similar to the one shown in Figure 6, and implies that the VERB-CS model systematically produces a more eroded plasmasphere on the nightside than is observed. After performing sensitivity tests to all the input parameters of the VERB-CS code (magnetic field, electric field, initial conditions, boundary conditions, etc.), we found that changes in the electric field have the most impact on this behavior (not shown here). Modifying the shielding parameter γ changes the extent of erosion significantly. From sensitivity tests (not shown), we found that using $\gamma = 1.8$ provides better agreement with observations than using the standard $\gamma = 2$ (Maynard & Chen, 1975). Therefore, we use $\gamma = 1.8$ in these simulations. This aspect of the VERB-CS code requires further investigation and testing, which we discuss in more detail in section 5.

It can be seen that the errors of the assimilative model are significantly reduced in the day and dusk sectors, compared to the physics-based model, but are still large on the nightside (21-7 MLT). This implies that the assimilative technique works well for blending the models on the dayside: the error of the assimilative model is smaller than that of VERB-CS and is closer to the PINE error. However, on the nightside, the assimilative model performance is similar to that of VERB-CS rather than PINE. The reason for that could lie in the performance of VERB-CS and in the choice of model and observation errors α^m and α^{obs} in the Kalman filter. VERB-CS has considerably larger errors on the nightside than PINE does, and it is probable that α^m and α^{obs} used here do not account for such a difference in errors between VERB-CS (model) and PINE (used as observations). If the VERB-CS model is improved, the results of data assimilation will consequently be improved as well. We discuss this in more detail in the Discussion section.

This test still illustrates that the assimilative methodology provides quantitative improvement in performance compared to the performance of the VERB-CS model (in particular, on the dayside). The performance of the assimilative model is comparable to the PINE model performance on the dayside but is closer to the performance of VERB-CS on the night side. It is clear that further improvement in the physics-based model (VERB-CS), or using a more advanced model, will improve the performance of the assimilative model.

As discussed above and can be seen from the results obtained in all three tests, neural networks have a good performance in general (in terms of performance metrics), but their performance decreases during extreme geomagnetic storms. At the same time, the physics-based VERB-CS code reproduces the plasmasphere dynamics during geomagnetic storms well but tends to be less accurate quantitatively, compared to the neural network PINE model. The assimilative methodology employed here performs well at combining both models during different levels of geomagnetic disturbance and shows the best agreement with the plasmapause derived from the IMAGE EUV instrument out of all models. The comparison with the in-situ density from RBSP-A over a long-term reconstruction of plasma density shows that the assimilative model can reach the performance of PINE on the dayside, but at the moment, its errors are closer to the VERB-CS model errors on the nightside, and consequently are larger than PINE's. This aspect can be improved in the future by either improving the VERB-CS model and/or by adjusting model and observation errors in the assimilative model. Overall, the assimilative model developed in this study demonstrates a potential to combine the advantages of both neural network and physics-based models, namely to have a good quantitative performance on average, and produce realistic global density reconstructions during the extreme geomagnetic events.

5 Discussion

Our results show that the assimilative methodology employed in this study for combining the neural network PINE model and the physics-based VERB-CS code demonstrates great potential for combining advantages of both models. Namely, the assimilative model demonstrated good performance on a series of test events from the IMAGE era for a variety of geomagnetic conditions: quiet, moderate, disturbed, and extreme geomagnetic storms. The output of the model showed better agreement with the plasmapause locations derived from IMAGE EUV than PINE or VERB-CS, when used separately (Figures 2 and 3). As discussed above, the performance of neural networks is limited by the training data. As there are no examples of extreme geomagnetic storms in the training dataset of PINE, its performance is reduced during such events. The assimilative methodology helps eliminate this limitation. The assimilative model also demonstrated good capabilities in combining quantitative advantages of models in terms of performance metrics. Long-term simulations of plasma density using all the models show that the current setup of the assimilative model allows for improving the performance of VERB-CS considerably in the dayside local time sector.

An advantageous feature of the assimilative model is that it can reconstruct the dynamics of the plasma density beyond the domain of the neural network-based model. The domain of the data assimilative model extends to 10 R_{E} in radial distance (as in the physics-based VERB-CS model), compared to 6.15 R_{E} of the PINE model. The PINE model is valid from ~ 1.75 to 6.15 R_{E} due to the use of density from the Van Allen Probes for training, and the domain of the assimilative model is the same as that of the physics-based model, VERB-CS. Therefore, the predictions of the neural network PINE model can be extended further to the plasma sheet by using the physics-based VERB-CS model as a “smart” extrapolator. It is worth noting that we have used in-situ density measurements from RBSP-A to validate the models. Therefore, all the models in this study were quantitatively validated up to 6.15 R_{E}. The quantitative validation beyond this radial distance is out of the scope of this study, but including density measurements from other missions, such as THEMIS, will aid in the quantitative validation of both the VERB-CS and the assimilative models beyond 6.15 R_{E}. Moreover, including such density measurements into the training dataset of the neural network will also allow for extending it to larger radial distances.

An important aspect of the assimilative approach employed here is the choice of model and observation errors α^{m} and α^{obs}. In this study, we employed an approach similar to Kondrashov et al. (2011), which was adjusted to use different constant values for model and observation errors α^{m} and α^{obs} inside and outside of the plasmasphere. We have compared the results obtained using such an approach to using constant values of errors throughout all radial extent of models (not shown here). We found that using different values of errors for inside and outside the plasmapause works better in our case and provides better agreement with observations. It is worth noting that selecting the model and observation errors is one of the most difficult topics in data assimilation and is still an active area of research (e.g., Berry & Harlim, 2017; C. H. Bishop, 2019; Hamilton et al., 2019); the existing approaches to select them are mostly empirical. Investigating the selection of these errors in a systematic way and experimenting with the dependence of the errors on other parameters such as L, MLT, and/or geomagnetic activity should be the subject of future research.

As this is the first study in which a neural network model was combined with a physics-based model of the plasmasphere, it is focused on demonstrating the assimilative methodology and its potential rather than reaching the best possible accuracy for either of the models in this study. Consequently, we made a number of assumptions and simplifications, in particular regarding the electric field, refilling, and saturation density models used in the physics-based VERB-CS code. We discuss them below in the context of the results obtained in this study.

In this study, we have used the electric field model of Volland (1973) and Stern (1975) parameterized by Kp (Maynard & Chen, 1975). Since Kp is a 3-hour index, the model inherently does not take into account the electric field variations on timescales less than 3 hours, which may not be sufficient time to account for changes in the plasmasphere dynamics on shorter timescales (Goldstein et al., 2005). Using a realistic electric field from global magnetospheric models or different parameterization accounting for shorter timescales, such as the parameterization of Goldstein et al. (2005) based on solar wind and IMF parameters, can potentially improve the model and needs to be investigated further. It is also worth noting that our tests showed that using a smaller shielding parameter $\gamma = 1.8$ instead of standard $\gamma = 2$ with the Maynard and Chen (1975) parameterization provides better agreement with both the IMAGE plasmapause and in-situ density observations from RBSP-A. Changes in this parameter significantly influence the extent of the erosion of the plasmasphere on the nightside.

It can also be seen from the results that, in some cases, the plasmasphere produced by the VERB-CS is more extended than was observed, in particular during geomagnetically quiet times (e.g., first row of Figure 3). This could be attributed to the refilling rates used. We have used median refilling rates from Denton et al. (2012) (without accounting for solar-cycle dependence). The model assumes that there is no significant dependence of the refilling rate on MLT. The same refilling rates are assumed for all geomagnetic conditions. At the moment, there still remain unsolved problems in the processes of plasmasphere ion refilling from the ionosphere (D. L. Gallagher & Comfort, 2016). Therefore, this topic should be explored further in regard to the physics-based modeling of the plasmasphere, and other approaches to parameterizing the refilling should be tested. For example, De Pascuale et al. (2018) used the approach of Rasmussen et al. (1993) to model the refilling using the Cold PLasma physics-based model CPL (V. K. Jordanova & Miyoshi, 2005; V. Jordanova, Zaharia, & Welling, 2010; V. Jordanova, Thorne, et al., 2010; V. K. Jordanova et al., 2014), where the approach of equatorial plasma densities toward equilibrium depends on the variation from the saturation level and a timescale. The timescale of refilling depends on the local time in addition to L, and was calculated from the differences in outgoing ion flux into the plasmasphere at hemispheric boundaries using empirical atmosphere models, including the MSIS-86 thermosphere model (Hedin, 1987), and IRI ionosphere model (Bilitza, 1986), in their study.

From long-term simulations (Figure 6), it can be seen that density inside the plasmasphere (on low L-shells) is slightly smaller on average than that of RBSP-A. This difference can be explained by the fact that the saturation model (Carpenter & Anderson, 1992) incorporated into the VERB-CS model provides lower saturation density on average than observed, using density measurements from the Van Allen Probes. Further investigation of other saturation density models or constructing a new saturation model that includes density measurements from the Van Allen Probes is required to improve the VERB-CS model performance.

The results obtained in this study illustrate that the assimilative methodology can be applied to combine both the qualitative and quantitative advantages of the VERB-CS and PINE models. It is clear that further improvement of the mentioned models or use of more sophisticated models in the physics-based VERB-CS code will improve the performance of the assimilative model. The methodology developed in this study will be especially useful for modeling the plasmasphere dynamics during geomagnetic storms and extreme events, such as the Halloween storms, while also providing realistic density values during quiet and moderate geomagnetic conditions. The combined data assimilative model is not computationally expensive and can be used as a part of global models of the magnetosphere or coupled with ring current and radiation belt codes.

6 Summary and Conclusions

In this study, we demonstrate for the first time how a neural network and physics-based models of the plasmasphere electron density can be combined in an optimal way by using data assimilation. We use the Kalman filter technique to optimally blend the neural network PINE model developed by I. Zhelavskaya et al. (2017) and the physics-based VERB-CS code (Aseev & Shprits, 2019) adjusted to model the plasmasphere dynamics.

We conduct three tests to evaluate the performance of the PINE, VERB-CS, and the assimilative model developed in this study. In the first two tests, we compare the model-predicted global evolution of plasma density to the global images of the He^+ distribution from IMAGE EUV; namely, we compare the modelled shape of the plasmasphere to the observed one using the plasmapause locations derived from the IMAGE EUV for the 2003 Halloween storm and for five events during March-June 2001. In the third test, we conduct a long-term reconstruction of electron density using all three models for an out-of-sample interval from 30 June 2016 to 01 January 2018. We compare the output of the models to the in-situ density obtained from RBSP-A and compute performance metrics.

The tests conducted in this study show that the neural network model PINE has a good quantitative performance on average and reproduces the general dynamics of the plasmasphere well, such as erosion on the nightside and plume formation. Its performance is limited, however, for $Kp > 7$ due to the lack of training data. The physics-based VERB-CS code also reproduces the dynamics of the plasmasphere well, and is especially effective during high geomagnetic activity and extreme geomagnetic events. However, its quantitative performance is lower than PINE's. Using the Kalman filter technique of data assimilation, we were able to combine the advantageous features of both models. The assimilative model is capable of reproducing the dynamics of the plasmasphere well during both quiet and disturbed geomagnetic activity, including extreme geomagnetic events. Its quantitative performance is better than that of VERB-CS and is comparable to PINE's for the dayside local time sector.

Future work includes considering different and more realistic electric field, refilling, and saturation density models. More work should be done regarding the selection of model and observation error in the Kalman filter setup. The assimilative model can be extended by assimilating in-situ density measurements in the model, as well (e.g., from Van Allen Probes, IMAGE RPI, or other sources, depending on the time period), in addition to the output of the neural network model PINE.

Appendix A The updated version of the PINE model

As discussed in section 3.1, I. Zhelavskaya et al. (2017) used K-fold cross validation with random splitting of data into training and validation folds to validate the models. They also used this procedure to select optimal input variables to the model. They considered several different combinations of solar wind parameters and geomagnetic indices as potential inputs to the neural network. In particular, they considered models based solely on geomagnetic indices (Kp, AE, Sym-H, and also F10.7), solely on solar wind data (solar wind speed, dynamic pressure, proton density, and the interplanetary magnetic field (IMF) B_{z}), and on a combination of both. As the time history of previous conditions is important for the plasmasphere dynamics, they also considered different durations of time history of these parameters as inputs, starting with simple models based only on instantaneous values of activity parameters and subsequently adding more time history of the corresponding parameters to the networks, up to 120 h of time history. The time history was represented as averages of the time histories of activity parameters integrated from hour 0 (e.g., 0-3, 0-6, 0-12 h, etc.). Every neural network also included a location input, as given by L and MLT.

In this study, we extend this analysis using the K-fold cross validation procedure described in section 3.1. We consider the same combinations of input parameters to the neural networks. The neural networks are trained on the density measurements from both RBSP-A and RBSP-b during 01 October 2012 – 01 July 2016. We use cross validation to obtain validation and training errors, and the standard deviations of errors. As described in section 3.1, all available data for this time interval are split into 35-day blocks. At first, 10% of the data are left aside as a testing dataset. Then the remaining 35-day blocks are randomly assigned to the training or validation sets. This type of data split allows the sequentiality of data to be preserved, and also introduces randomness and representation of different geomagnetic conditions in both validation and training sets. The rest of the methodology is identical to that of I. Zhelavskaya et al. (2017).

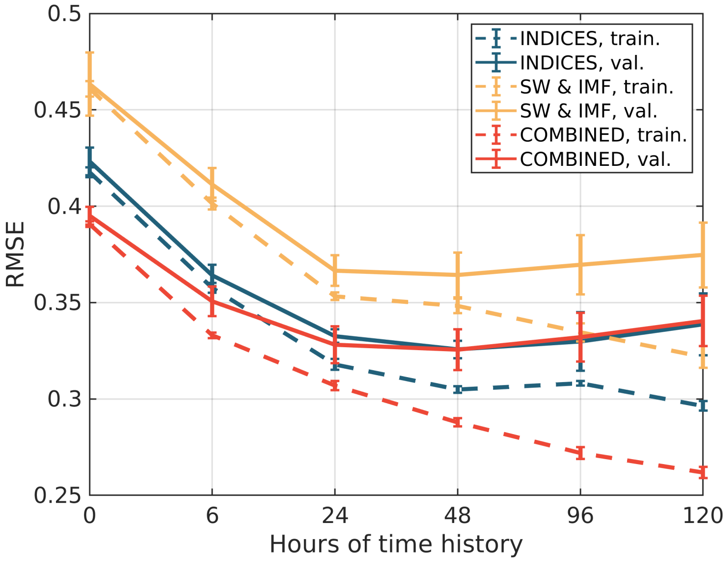

Figure A1. Root-mean-square error (RMSE) on the y-axis versus the hours of time history included in the models. The yellow color shows the errors of models based on solar wind, the blue color is for the models based on geomagnetic indices, and red is for the models based on both of them combined. Solid lines show validation errors and dashed lines show training errors. The error bars show the standard deviation of error on the validation set obtained during the CV procedure.

Figure A1 shows the root-mean-square error (RMSE) plotted against the number of hours of time history included into the model. The yellow color indicates the errors of models based on solar wind, the blue color is for the models based on geomagnetic indices, and red is for the models based on both of them combined. Solid lines show the validation errors and dashed lines the training errors. The error bars show the standard deviation of error obtained during the CV procedure. The validation error represents how well a model performs on the unseen data and is the error we aim to minimize. The difference between the training and validation errors indicates if a model overfits the data or not. When the difference is too large, this means that a model "learned" the training data too well – it memorized it essentially, and due to that performs poorly on the unseen data. As a consequence, it does not have good generalization capabilities.

It can be seen that the models based only on solar wind have the largest errors. The errors of the models based only on geomagnetic indices are significantly lower. The validation errors of all models are large when no time history is included into them and decrease as more time history is included. After a certain point (around 48-hour time history), however, the validation errors start to slightly increase again. At the same time, the training errors always decrease as more time history is included. The moment when the validation error starts increasing indicates that a model starts to overfit. That is not desirable in the models and needs to be avoided. In this case, the overfitting starts approximately after a 48-hour time history (for all models). The inclusion of longer time history does not bring additional improvement. The models based on the combination of solar wind and geomagnetic indices have similar errors to the models based only on indices, but overfit much more. This implies that the model based only on geomagnetic indices contains a sufficient amount of information to model the plasmasphere dynamics accurately. In this case, the optimal model is based on the 48-hour time history of geomagnetic indices, since the validation error is the smallest for that particular combination, and the model does not overfit significantly. The inputs to the model are L, MLT, and averages of Kp, AE, Sym-H, and F10.7 over previous 3, 6, 12, 24, 36, and 48 hours.

Appendix B The model and observation error of the Kalman filter

The model and observation errors α^m and α^{obs} were obtained as outlined below. We use the results of the long-term density reconstruction obtained in section 4.3. There, PINE and VERB-CS were run for the period of 30 June 2016 – 01 January 2018, and their output was compared to the in-situ density measurements from RBSP-A during that period. The performance of both models was analysed using the RMSE performance metric. Here, we have also computed the RMSE of both models for locations inside and outside the plasmapause of the respective model separately for the period of 02 July 2016 – 01 January 2018. The plasmapause was calculated using the density threshold of 40 cm^{-3}, as described in section 2. Figure B1 (panels a-d) shows the histograms of error distributions during this period for both models at different locations. The RMSEs are labeled inside the respective panels. We have employed the RMSE values obtained in this analysis as model and observation errors α^m and α^{obs} inside and outside the plasmapause of each model.

The same analysis was performed for the output of the assimilative model. Its RMSEs inside and outside the plasmapause are shown in panels (e-f) of Figure B1. It can be seen that the RMSEs of the assimilative model are equal to approximately an average of those of PINE and VERB-CS RMSEs (and also MEs). After conducting a series of experiments with different values of m and obs including just constant values, i.e., without dependence on the plasmapause location (not shown), we found that these values provide the best agreement between the assimilative model and in-situ density observations.

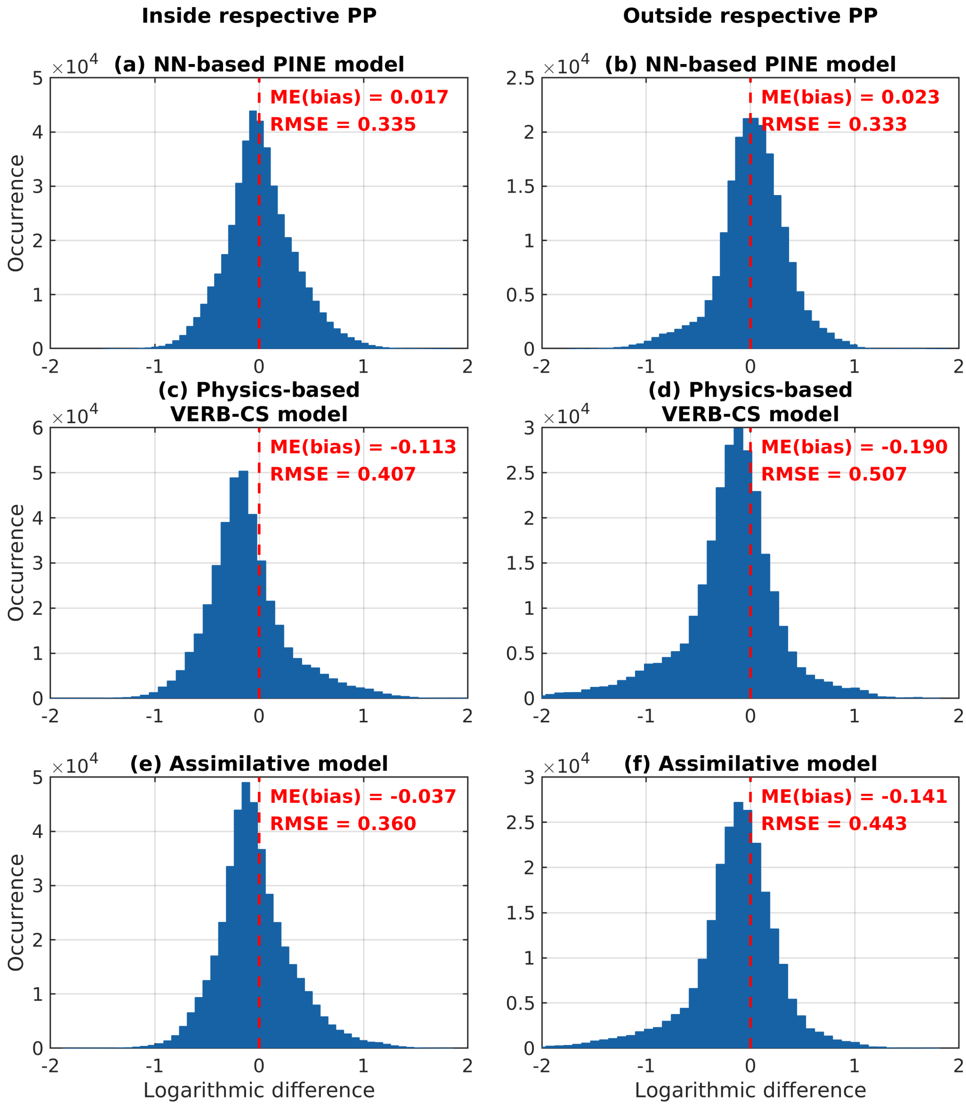

Figure B1. Distribution of errors of PINE, VERB-CS, and the assimilative model inside and outside the plasmapause during the out-of-sample period of 02 July 2016 – 01 January 2018 (compared to density measurements from RBSP-A). The respective RMSE and ME are given inside each panel.

Acknowledgments

The authors used geomagnetic indices and solar wind data provided by OMNIWeb (`http://omniweb.gsfc.nasa.gov/form/dx1.html`). IMAGE EUV plasmapause database was obtained from `http://enarc.space.swri.edu/EUV/`. This work was funded by the NASA Heliophysics Guest Investigator Program under NASA grant NNX07AG48G, P.I. Jerry Goldstein. The electron density data set was derived using the NURD algorithm and is available at the GFZ Data Services (I. Zhelavskaya et al., 2020). I. Z. was supported by Geo.X, the Research Network for Geosciences in Berlin and Potsdam, under Grant No SO_087_GeoX. This work was supported by the pilot project "MAP" which is a part of the 'Helmholtz Pilot projects Information & Data Science II' funded by the Initiative and Networking Fund of the Helmholtz Association, and by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870452 (PAGER). The research has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through grant SFB 1294. We would like to thank Maria Spasojević for her contribution to the development of the PINE model and the NURD algorithm. I. Z. is grateful to Jerry Goldstein, T. Paul O'Brien, Alexander Drozdov, and Hayley Allison for helpful discussions.

References

Anderson, J. A. (1995). *An introduction to neural networks*. MIT press.

Anderson, R. R., Gurnett, D. A., & Odem, D. L. (1992). Crres plasma wave experiment. *Journal of Spacecraft and Rockets*, *29*(4), 570-573. Retrieved from `https://doi.org/10.2514/3.25501` doi: 10.2514/3.25501

André, M., N., Li, W., Toledo-Redondo, S., Khotyaintsev, Y. V., Vaivads, A., Graham, D. B., ... Saito, Y. (2016). Magnetic reconnection and modification of the Hall physics due to cold ions at the magnetopause. *Geophysical Research Letters*, *43*(13), 6705–6712. Retrieved from `http://dx.doi.org/10.1002/2016GL069665` (2016GL069665) doi: 10.1002/2016GL069665

Angelopoulos, V. (2009). The themis mission. In *The themis mission* (pp. 5–34). Springer.

Aseev, N., & Shprits, Y. (2019). Reanalysis of ring current electron phase space densities using van allen probe observations, convection model, and log-normal kalman filter. *Space Weather*, *17*(4), 619–638.

Aseev, N., Shprits, Y., Drozdov, A., & Kellerman, A. (2016). Numerical applications of the advective-diffusive codes for the inner magnetosphere. *Space Weather*, *14*(11), 993–1010.

Berry, T., & Harlim, J. (2017). Correcting biased observation model error in data assimilation. *Monthly Weather Review*, *145*(7), 2833–2853.

Bilitza, D. (1986). International reference ionosphere: Recent developments. *Radio Science*, *21*(3), 343–346.

Bishop, C. H. (2019). Data assimilation strategies for state-dependent observation error variances. *Quarterly Journal of the Royal Meteorological Society*, *145*(718), 217–227.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford university press.

Borovsky, J. E., & Denton, M. H. (2006). Effect of plasmaspheric drainage plumes on solar-wind/magnetosphere coupling. *Geophysical Research Letters*, *33*(20), n/a–n/a. Retrieved from http://dx.doi.org/10.1029/2006GL026519 (L20101) doi: 10.1029/2006GL026519

Bortnik, J., Li, W., Thorne, R. M., & Angelopoulos, V. (2016). A unified approach to inner magnetospheric state prediction. *Journal of Geophysical Research: Space Physics*, *121*(3), 2423–2430. Retrieved from http://dx.doi.org/10.1002/2015JA021733 (2015JA021733) doi: 10.1002/2015JA021733

Bresson, G., Féraud, T., Aufrère, R., Checchin, P., & Chapuis, R. (2015). Real-time monocular slam with low memory requirements. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 1827–1839.

Carpenter, D. L. (1963). Whistler evidence of a kneein the magnetospheric ionization density profile. *Journal of Geophysical Research*, *68*(6), 1675–1682.

Carpenter, D. L. (1966). Whistler studies of the plasmapause in the magnetosphere: 1. temporal variations in the position of the knee and some evidence on plasma motions near the knee. *Journal of Geophysical Research*, *71*(3), 693–709.

Carpenter, D. L. (1970). Whistler evidence of the dynamic behavior of the duskside bulge in the plasmasphere. *Journal of Geophysical Research*, *75*(19), 3837–3847.

Carpenter, D. L., & Anderson, R. R. (1992). An ISEE/whistler model of equatorial electron density in the magnetosphere. *Journal of Geophysical Research: Space Physics*, *97*(A2), 1097–1108. Retrieved from http://dx.doi.org/10.1029/

91JA01548 doi: 10.1029/91JA01548

Cervantes, S., Shprits, Y., Aseev, N., Drozdov, A., Castillo, A., & Stolle, C. (2020). Identifying radiation belt electron source and loss processes by assimilating spacecraft data in a three-dimensional diffusion model. *Journal of Geophysical Research: Space Physics*, *125*(1), e2019JA027514.

Chappell, C., Harris, K., & Sharp, G. (1970a). The morphology of the bulge region of the plasmasphere. *Journal of Geophysical Research*, *75*(19), 3848–3861.

Chappell, C., Harris, K., & Sharp, G. (1970b). A study of the influence of magnetic activity on the location of the plasmapause as measured by OGO 5. *Journal of Geophysical Research*, *75*(1), 50–56.

Chen, S.-H., & Moore, T. E. (2006). Magnetospheric convection and thermal ions in the dayside outer magnetosphere. *Journal of Geophysical Research: Space Physics*, *111*(A3), n/a–n/a. Retrieved from http://dx.doi.org/10.1029/2005JA011084 (A03215) doi: 10.1029/2005JA011084

Chu, X., Bortnik, J., Li, W., Ma, Q., Angelopoulos, V., & Thorne, R. (2017). Erosion and refilling of the plasmasphere during a geomagnetic storm modeled by a neural network. *Journal of Geophysical Research: Space Physics*, *122*(7), 7118–7129.

Chu, X., Bortnik, J., Li, W., Ma, Q., Denton, R., Yue, C., ... others (2017). A neural network model of three-dimensional dynamic electron density in the inner magnetosphere. *Journal of Geophysical Research: Space Physics*, *122*(9), 9183–9197.

Daae, M., Shprits, Y., Ni, B., Koller, J., Kondrashov, D., & Chen, Y. (2011). Reanalysis of radiation belt electron phase space density using various boundary conditions and loss models. *Advances in space research*, *48*(8), 1327–1334.

Darrouzet, F., Gallagher, D. L., André, N., Carpenter, D. L., Dandouras, I., Décréau, P. M., ... others (2009). Plasmaspheric density structures and dynamics: Properties observed by the CLUSTER and IMAGE missions. *Space science reviews*, *145*(1-2), 55–106.

Denton, R., Wang, Y., Webb, P., Tengdin, P., Goldstein, J., Redfern, J., & Reinisch, B. (2012). Magnetospheric electron density long-term (¿ 1 day) refilling rates inferred from passive radio emissions measured by image rpi during geomagnetically quiet times. *Journal of Geophysical Research: Space Physics*, *117*(A3).

De Pascuale, S., Jordanova, V., Goldstein, J., Kletzing, C., Kurth, W., Thaller, S., & Wygant, J. R. (2018). Simulations of van allen probes plasmaspheric electron density observations. *Journal of Geophysical Research: Space Physics*, *123*(11), 9453–9475.

Dubyagin, S., Ganushkina, N. Y., Sillanpää, I., & Runov, A. (2016). Solar wind-driven variations of electron plasma sheet densities and temperatures beyond geostationary orbit during storm times. *Journal of Geophysical Research: Space Physics*, *121*(9), 8343–8360.

Gallagher, D., Craven, P., & Comfort, R. (1998). A simple model of magnetospheric trough total density. *Journal of Geophysical Research: Space Physics*, *103*(A5), 9293–9297.

Gallagher, D. L., & Comfort, R. H. (2016). Unsolved problems in plasmasphere refilling. *Journal of Geophysical Research: Space Physics*, *121*(2), 1447-1451. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JA022279` doi: 10.1002/2015JA022279

Gallagher, D. L., Craven, P. D., & Comfort, R. H. (2000). Global core plasma model. *Journal of Geophysical Research: Space Physics*, *105*(A8), 18819–18833. Retrieved from `http://dx.doi.org/10.1029/1999JA000241` doi: 10.1029/1999JA000241

Goldstein, J., Burch, J., & Sandel, B. (2005). Magnetospheric model of subauroral polarization stream. *Journal of Geophysical Research: Space Physics*, *110*(A9).

Goldstein, J., Sandel, B., Forrester, W., & Reiff, P. (2003). IMF-driven plasmasphere erosion of 10 July 2000. *Geophysical research letters*, *30*(3).

Goldstein, J., Sandel, B., Hairston, M., & Reiff, P. (2003). Control of plasmaspheric dynamics by both convection and sub-auroral polarization stream. *Geophysical research letters*, *30*(24).

Goldstein, J., Spasojević, M., Reiff, P., Sandel, B., Forrester, W., Gallagher, D., & Reinisch, B. (2003). Identifying the plasmapause in IMAGE EUV data using IMAGE RPI in situ steep density gradients. *Journal of Geophysical Research: Space Physics*, *108*(A4).

Gopalswamy, N. (2006). Highlights of the october-november 2003 extreme events.

Solar Extreme Events: Fundamental Science and Applied Aspects. Yerevan: Alikhanyan Physics Institute, *24*.

Grebowsky, J. (1970). Model study of plasmapause motion. *Journal of Geophysical Research*, *75*(22), 4329–4333.

Grewal, M., & Andrews, A. (2010, Feb). How good is your gyro [ask the experts]. *IEEE Control Systems Magazine*, *30*(1), 12-86. doi: 10.1109/MCS.2009.935122

Gringauz, K. (1963). The structure of the ionized gas envelope of earth from direct measurements in the ussr of local charged particle concentrations. *Planetary and Space Science*, *11*(3), 281–296.

Gurnett, D., Persoon, A., Randall, R., Odem, D., Remington, S., Averkamp, T., ... others (1995). The polar plasma wave instrument. *Space Science Reviews*, *71*(1-4), 597–622.

Gurnett, D. A., Scarf, F., Fredricks, R., & Smith, E. (1978). The isee-1 and isee-2 plasma wave investigation. *IEEE Transactions on Geoscience Electronics*, *16*(3), 225–230.

Hamilton, F., Berry, T., & Sauer, T. (2019). Correcting observation model error in data assimilation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(5), 053102.

Haykin, S. S., Haykin, S. S., Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Pearson Upper Saddle River, NJ, USA:.

Hedin, A. E. (1987). Msis-86 thermospheric model. *Journal of Geophysical Research: Space Physics*, *92*(A5), 4649–4662.

Horne, R. B., Thorne, R. M., Shprits, Y. Y., Meredith, N. P., Glauert, S. A., Smith, A. J., ... others (2005). Wave acceleration of electrons in the van allen radiation belts. *Nature*, *437*(7056), 227–230.

Huba, J., & Krall, J. (2013). Modeling the plasmasphere with sami3. *Geophysical research letters*, *40*(1), 6–10.

Jordanova, V., Miyoshi, Y., Zaharia, S., Thomsen, M., Reeves, G., Evans, D., ... Fennell, J. (2006). Kinetic simulations of ring current evolution during the geospace environment modeling challenge events. *Journal of Geophysical Research: Space Physics*, *111*(A11).

Jordanova, V., Thorne, R., Li, W., & Miyoshi, Y. (2010). Excitation of whistler mode chorus from global ring current simulations. *Journal of Geophysical Research: Space Physics*, *115*(A5).

Jordanova, V., Zaharia, S., & Welling, D. (2010). Comparative study of ring current development using empirical, dipolar, and self-consistent magnetic field simulations. *Journal of Geophysical Research: Space Physics*, *115*(A12).

Jordanova, V. K., & Miyoshi, Y. (2005). Relativistic model of ring current and radiation belt ions and electrons: Initial results. *Geophysical Research Letters*, *32*(14). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL023020` doi: 10.1029/2005GL023020

Jordanova, V. K., Yu, Y., Niehof, J., Skoug, R., Reeves, G., Kletzing, C., ... Spence, H. E. (2014). Simulations of inner magnetosphere dynamics with an expanded ram-scb model and comparisons with van allen probes observations. *Geophysical Research Letters*, *41*(8), 2687–2694.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability.* Cambridge university press.

Kellerman, A., Shprits, Y., Kondrashov, D., Subbotin, D., Makarevich, R., Donovan, E., & Nagai, T. (2014). Three-dimensional data assimilation and reanalysis of radiation belt electrons: Observations of a four-zone structure using five spacecraft and the verb code. *Journal of Geophysical Research: Space Physics*, *119*(11), 8764–8783.

Kletzing, C., Kurth, W., Acuna, M., MacDowall, R., Torbert, R., Averkamp, T., ... others (2013). The electric and magnetic field instrument suite and integrated science (EMFISIS) on RBSP. *Space Science Reviews*, *179*(1-4), 127–181.

Kondrashov, D., Ghil, M., & Shprits, Y. (2011). Lognormal kalman filter for assimilating phase space density data in the radiation belts. *Space Weather*, *9*(11). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011SW000726` doi: 10.1029/2011SW000726

Krall, J., Huba, J., & Fedder, J. (2008). Simulation of field-aligned H+ and He+ dynamics during late-stage plasmasphere refilling. In *Annales geophysicae: atmospheres, hydrospheres and space sciences* (Vol. 26, p. 1507).

Krall, J., Huba, J., Jordanova, V., Denton, R., Carranza, T., & Moldwin, M. (2016). Measurement and modeling of the refilling plasmasphere during 2001. *Journal* *of Geophysical Research: Space Physics*, *121*(3), 2226–2248.

Lahoz, W., Khattatov, B., & Ménard, R. (2010). Data assimilation and information. In *Data assimilation* (pp. 3–12). Springer.

Lawrence, D., Thomsen, M., Borovsky, J., & McComas, D. (1999). Measurements of early and late time plasmasphere refilling as observed from geosynchronous orbit. *Journal of Geophysical Research: Space Physics*, *104*(A7), 14691–14704.

Lee, S.-H., Zhang, H., Zong, Q.-G., Otto, A., Rème, H., & Liebert, E. (2016). A statistical study of plasmaspheric plumes and ionospheric outflows observed at the dayside magnetopause. *Journal of Geophysical Research: Space Physics*, *121*(1), 492–506.

Lemaire, J. F., & Gringauz, K. I. (1998). *The Earth's plasmasphere.* Cambridge Univ. Press (Cambridge, MA).

Li, W., Thorne, R., Bortnik, J., Nishimura, Y., Angelopoulos, V., Chen, L., … Bonnell, J. (2010). Global distributions of suprathermal electrons observed on THEMIS and potential mechanisms for access into the plasmasphere. *Journal of Geophysical Research: Space Physics*, *115*(A12).

Maynard, N., & Chen, A. (1975). Isolated cold plasma regions: Observations and their relation to possible production mechanisms. *Journal of Geophysical Research*, *80*(7), 1009–1013.

Mazzella, A. J. (2009). Plasmasphere effects for gps tec measurements in north america. *Radio Science*, *44*(5).

Moldwin, M. B., Thomsen, M. F., Bame, S. J., McComas, D., & Reeves, G. D. (1995). The fine-scale structure of the outer plasmasphere. *Journal of Geophysical Research: Space Physics*, *100*(A5), 8021–8029.

Mosier, S. R., Kaiser, M. L., & Brown, L. W. (1973). Observations of noise bands associated with the upper hybrid resonance by the IMP 6 radio astronomy experiment. *Journal of Geophysical Research*, *78*(10), 1673–1679.

Nishida, A. (1966). Formation of plasmapause, or magnetospheric plasma knee, by the combined action of magnetospheric convection and plasma escape from the tail. *Journal of Geophysical Research*, *71*(23), 5669–5679.

O'Brien, T., & Moldwin, M. (2003). Empirical plasmapause models from magnetic indices. *Geophysical research letters*, *30*(4).

Orlova, K., Shprits, Y., & Spasojević, M. (2016). New global loss model of energetic and relativistic electrons based on van allen probes measurements. *Journal of Geophysical Research: Space Physics*, *121*(2), 1308-1314. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JA021878 doi: 10.1002/2015JA021878

Park, C. (1974). Some features of plasma distribution in the plasmasphere deduced from Antarctic whistlers. *Journal of geophysical research*, *79*(1), 169–173.

Park, C., & Carpenter, D. L. (1970). Whistler evidence of large-scale electron-density irregularities in the plasmasphere. *Journal of Geophysical Research*, *75*(19), 3825–3836.

Pierrard, V., Goldstein, J., André, N., Jordanova, V. K., Kotova, G. A., Lemaire, J. F., ... Matsui, H. (2009). Recent progress in physics-based models of the plasmasphere. In *The earths plasmasphere* (pp. 193–229). Springer.

Pierrard, V., & Stegen, K. (2008). A three-dimensional dynamic kinetic model of the plasmasphere. *Journal of Geophysical Research: Space Physics*, *113*(A10).

Priddy, K. L., & Keller, P. E. (2005). *Artificial neural networks: an introduction* (Vol. 68). SPIE press.

Rasmussen, C. E., Guiter, S. M., & Thomas, S. G. (1993). A two-dimensional model of the plasmasphere: refilling time constants.

Reeves, G., Spence, H. E., Henderson, M., Morley, S., Friedel, R., Funsten, H., ... others (2013). Electron acceleration in the heart of the van allen radiation belts. *Science*, *341*(6149), 991–994.

Salti, S., Lanza, A., & Di Stefano, L. (2014). Synergistic change detection and tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, *25*(4), 609–622.

Sandel, B. R., Broadfoot, A. L., Curtis, C., King, R., Stone, T., Hill, R., ... others (2000). The extreme ultraviolet imager investigation for the IMAGE mission. In *The image mission* (pp. 197–242). Springer.

Sheeley, B., Moldwin, M., Rassoul, H., & Anderson, R. (2001). An empirical plasmasphere and trough density model: CRRES observations. *Journal of Geophysical Research: Space Physics*, *106*(A11), 25631–25641.

Shprits, Y., Kellerman, A., Kondrashov, D., & Subbotin, D. (2013). Application of a new data operator-splitting data assimilation technique to the 3-d verb diffusion code and crres measurements. *Geophysical Research Letters*, *40*(19), 4998–5002.

Shprits, Y., Kondrashov, D., Chen, Y., Thorne, R., Ghil, M., Friedel, R., & Reeves, G. (2007). Reanalysis of relativistic radiation belt electron fluxes using crres satellite data, a radial diffusion model, and a kalman filter. *Journal of Geophysical Research: Space Physics*, *112*(A12).

Shprits, Y. Y., Drozdov, A. Y., Spasojević, M., Kellerman, A. C., Usanova, M. E., Engebretson, M. J., ... others (2016). Wave-induced loss of ultra-relativistic electrons in the Van Allen radiation belts. *Nature Communications*, *7*, 12883.

Shprits, Y. Y., Kellerman, A. C., Drozdov, A. Y., Spence, H. E., Reeves, G. D., & Baker, D. N. (2015). Combined convective and diffusive simulations: Verb-4d comparison with 17 march 2013 van allen probes observations. *Geophysical Research Letters*, *42*(22), 9600–9608.

Shue, J.-H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., ... Kawano, H. (1998). Magnetopause location under extreme solar wind conditions. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17691-17700. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98JA01103` doi: 10.1029/98JA01103

Singh, A., Singh, R., & Siingh, D. (2011). State studies of Earth's plasmasphere: A review. *Planetary and Space Science*, *59*(9), 810–834.

Singh, N., & Horwitz, J. (1992). Plasmasphere refilling: Recent observations and modeling. *Journal of Geophysical Research: Space Physics*, *97*(A2), 1049–1079.

Sorenson, H. W. (1985). *Kalman filtering: theory and application.* IEEE.

Spasojević, M., Frey, H., Thomsen, M., Fuselier, S., Gary, S., Sandel, B., & Inan, U. (2004). The link between a detached subauroral proton arc and a plasmaspheric plume. *Geophysical research letters*, *31*(4).

Spasojević, M., Goldstein, J., Carpenter, D., Inan, U., Sandel, B., Moldwin, M., & Reinisch, B. (2003). Global response of the plasmasphere to a geomagnetic disturbance. *Journal of Geophysical Research: Space Physics*, *108*(A9).

Stern, D. P. (1975). The motion of a proton in the equatorial magnetosphere. *Journal of Geophysical Research*, *80*(4), 595–599.

Su, Y.-J., Thomsen, M. F., Borovsky, J. E., & Lawrence, D. J. (2001). A comprehensive survey of plasmasphere refilling at geosynchronous orbit. *Journal of*

Geophysical Research: Space Physics, *106*(A11), 25615–25629.

Volland, H. (1973). A semiempirical model of large-scale magnetospheric electric fields. *Journal of Geophysical Research*, *78*(1), 171–180.

Weimer, D. (2005). Predicting surface geomagnetic variations using ionospheric electrodynamic models. *Journal of Geophysical Research: Space Physics*, *110*(A12).

Xiao-Ting, S., Gendrin, R., & Caudal, G. (1988). Refilling process in the plasmasphere and its relation to magnetic activity. *Journal of Atmospheric and Terrestrial Physics*, *50*(3), 185 - 195. Retrieved from `http://www.sciencedirect.com/science/article/pii/0021916988900670` doi: https://doi.org/10.1016/0021-9169(88)90067-0

Xiong, C., Stolle, C., & Lühr, H. (2016). The swarm satellite loss of gps signal and its relation to ionospheric plasma irregularities. *Space Weather*, *14*(8), 563–577.

Yu, Y., Jordanova, V., Zou, S., Heelis, R., Ruohoniemi, M., & Wygant, J. (2015). Modeling subauroral polarization streams during the 17 march 2013 storm. *Journal of Geophysical Research: Space Physics*, *120*(3), 1738–1750.

Zhelavskaya, I., Shprits, Y., Spasojevic, M., & Kurth, W. (2020). *Electron density derived with the Neural-network-based Upper-hybrid Resonance Determination algorithm from the Van Allen Probes EMFISIS measurements. V. 1.1.* GFZ Data Services. doi: 10.5880/GFZ.2.8.2020.002

Zhelavskaya, I., Shprits, Y. Y., & Spasojević, M. (2017). Empirical modeling of the plasmasphere dynamics using neural networks. *Journal of Geophysical Research: Space Physics*, *122*(11), 11,227-11,244. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JA024406` doi: 10.1002/2017JA024406

Zhelavskaya, I., Spasojević, M., Shprits, Y., & Kurth, W. (2016). Automated determination of electron density from electric field measurements on the Van Allen Probes spacecraft. *J. Geophys. Res. Space Physics*, *121*, 4611-4625. doi: 10.1002/2015JA022132

Zhelavskaya, I. S., Vasile, R., Shprits, Y. Y., Stolle, C., & Matzka, J. (2019). Systematic analysis of machine learning and feature selection techniques for prediction of the kp index. *Space Weather*, *17*(10), 1461-1486. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019SW002271 doi: 10.1029/2019SW002271