Zhelavskaya, I. S., Shprits, Y., Spasojević, M. (2017): Empirical Modeling of the Plasmasphere Dynamics Using Neural Networks. - *Journal of Geophysical Research*, *122*, 11, pp. 11,227—11,244.

DOI: http://doi.org/10.1002/2017JA024406

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Space Physics

RESEARCH ARTICLE
10.1002/2017JA024406

Empirical Modeling of the Plasmasphere Dynamics Using Neural Networks

Irina S. Zhelavskaya[1,2], **Yuri Y. Shprits**[1,2,3], **and Maria Spasojević**[4]

[1]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, [2]Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany, [3]Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA, USA, [4]Hansen Experimental Physics Laboratory, Stanford University, CA, USA

Key Points:
- We developed a dynamic plasmasphere density model by applying neural networks to in situ density measurements and verifying with global images
- The optimal model takes as input the 96 h time history of geomagnetic indices and accurately captures plume formation and evolution
- Using both solar wind data and geomagnetic indices as inputs slightly overfits the training data, and gaps in solar wind data are problematic

Correspondence to:
I. S. Zhelavskaya,
irina.zhelavskaya@gmail.com

Citation:
Zhelavskaya, I. S., Shprits, Y. Y., & Spasojevic, M. (2017). Empirical modeling of the plasmasphere dynamics using neural networks. *Journal of Geophysical Research: Space Physics*, *122*, 11,227–11,244. https://doi.org/10.1002/2017JA024406

Received 28 MAY 2017
Accepted 7 OCT 2017
Accepted article online 24 OCT 2017
Published online 14 NOV 2017

Abstract We present the PINE (Plasma density in the Inner magnetosphere Neural network-based Empirical) model - a new empirical model for reconstructing the global dynamics of the cold plasma density distribution based only on solar wind data and geomagnetic indices. Utilizing the density database obtained using the NURD (Neural-network-based Upper hybrid Resonance Determination) algorithm for the period of 1 October 2012 to 1 July 2016, in conjunction with solar wind data and geomagnetic indices, we develop a neural network model that is capable of globally reconstructing the dynamics of the cold plasma density distribution for $2 \leq L \leq 6$ and all local times. We validate and test the model by measuring its performance on independent data sets withheld from the training set and by comparing the model-predicted global evolution with global images of He^+ distribution in the Earth's plasmasphere from the IMAGE Extreme UltraViolet (EUV) instrument. We identify the parameters that best quantify the plasmasphere dynamics by training and comparing multiple neural networks with different combinations of input parameters (geomagnetic indices, solar wind data, and different durations of their time history). The optimal model is based on the 96 h time history of *Kp*, *AE*, *SYM-H*, and $F_{10.7}$ indices. The model successfully reproduces erosion of the plasmasphere on the nightside and plume formation and evolution. We demonstrate results of both local and global plasma density reconstruction. This study illustrates how global dynamics can be reconstructed from local in situ observations by using machine learning techniques.

1. Introduction

The plasmasphere is a toroidal region of cold (~1 eV) and relatively dense (~$10-10^4$ cm^{-3}) plasma surrounding the Earth (Lemaire & Gringauz, 1998). It is located in the inner magnetosphere and extends from the ionosphere out to a boundary known as the plasmapause, where the plasma density drops abruptly by several orders of magnitude. The plasmasphere is very dynamic: its shape and size are highly susceptible to the time history of solar and geomagnetic conditions (Chappell et al., 1970a; O'Brien & Moldwin, 2003). Two flow regimes, sunward convection and corotation with the Earth, determine the shape and size of the plasmasphere (Darrouzet et al., 2009; Singh et al., 2011). During quiet geomagnetic time, the plasmasphere is refilled with the ions from the topside ionosphere and expands up to $\sim 4-7\,R_E$ (Goldstein, Sandel, Hairston, et al., 2003; Krall et al., 2008; Singh & Horwitz, 1992); its shape is then roughly circular with a bulge on the duskside (Nishida, 1966). In contrast, during periods of high geomagnetic activity, the plasmasphere is eroded by the dominating sunward magnetospheric convection and, therefore, contracts (Carpenter, 1970; ,Chappell et al., 1970b; Goldstein, Sandel, Hairston, et al., 2003). The stronger the disturbance, the more the plasmasphere contracts (down to $2\,R_E$ during severe geomagnetic storms). Features of various scales (plumes, channels, crenulations, shoulders, fingers, etc.) (Carpenter & Stone, 1967; Foster & Burke, 2002; Garcia et al., 2003; Goldstein et al., 2004; Grebowsky, 1970; LeDocq et al., 1994; Sandel et al., 2001; Spasojević et al., 2003) can be formed during that time and are transformed during the storm recovery phase by the interplay between convection and corotating regimes (Dungey, 1961; Goldstein, Kanekal, et al., 2005; Goldstein et al., 2002). Their formation depends on the previous configuration of the plasmasphere/plasmapause, the time history of convective erosion from electric fields and of refilling from ionospheric outflow (Darrouzet et al., 2009; Gallagher et al., 2005).

The large-scale redistribution of magnetospheric cold plasma during storm intervals has wide-ranging implications for a variety of other magnetospheric and ionospheric processes. The evolving size and shape

of the plasmasphere controls the growth and propagation of plasma waves and directly affects resonant wave-particle interactions, thus profoundly influencing energetic ion and electron distributions over a wide range of energies (e.g., Orlova et al., 2016, Shprits et al., 2016; Spasojević et al., 2004). Eroded plasmaspheric material is transported sunward and is regularly observed near the dayside magnetopause boundary (e.g., Chen & Moore, 2006; Lee et al., 2016). Here the enhanced dayside plasma density may limit the rate of reconnection and thus influence the global convection pattern (e.g., André et al., 2016; Borovsky & Denton, 2006). Storm time plasmaspheric structures have been found to be strongly associated with ionospheric density features including ionospheric storm enhanced density and polar cap patches (Su et al., 2001; Zhang et al., 2013).

One of the parameters that is used to quantify the plasmasphere dynamics is plasma density (Kotova, 2007; Lemaire & Gringauz, 1998, and references therein). Plasma density can be measured on satellites in several ways. These methods include measuring the density directly with active sounders (e.g., Trotignon et al., 2003) or particle counters (e.g., Geiger & Müller, 1928), or it can be derived from the spacecraft potential (e.g., Escoubet et al., 1997). Another method of determining plasma density, one of the most accurate and reliable, is by observing intense upper hybrid resonance (UHR) frequency bands in dynamic spectrograms (Mosier et al., 1973). Although the process of deriving UHR bands can be challenging, significant success has been achieved recently in developing methods for deriving UHR bands from the spectrograms in an automated fashion (Kurth et al., 2015; Zhelavskaya et al., 2016). In both of these works, authors developed automated algorithms for deriving UHR frequency bands from electric and magnetic field measurements made with the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) instrument (Kletzing et al., 2013) on board the Van Allen Probes (Mauk et al., 2013). Moreover, large density databases have been produced in these studies. Although the value of the point density measurements obtained from satellites is considered reliable, measurements from one satellite naturally only cover one point in space at a time and do not provide a global view on the plasmasphere dynamics. Therefore, developing models of plasma density capable of reproducing the global dynamics of the plasmasphere, particularly dynamic plume development, is important.

Numerous empirical and physics-based density models have been developed. The most widely used empirical models in recent years are those developed in the studies of Carpenter and Anderson (1992), Gallagher et al. (2000), and Sheeley et al. (2001). The model Carpenter and Anderson (1992) is based on electron density measurements deduced from sweep frequency receiver (SFR) radio measurements on the International Sun-Earth Explorer (ISEE-1) spacecraft and ground-based whistler measurements. It is a saturated density model and, as such, represents the density distribution after several days of refilling. The model covers the range of $2.25 < L < 8$ and the local time interval of 0 to 15 MLT. It provides the mean density value for different L shells. The plasmasphere and plasma trough models of Sheeley et al. (2001) present statistical averages based on densities derived from the Combined Release and Radiation Effects Satellite (CRRES) swept frequency receiver by identifying the upper hybrid resonance frequency. The models cover the range of $3 \leq L \leq 7$ and all local times. Sheeley et al. (2001) provide the mean and the standard deviation of measurements for the plasmasphere and trough models to represent depleted or saturated density levels for different L and MLT for the trough. The Global Core Plasma Model (GCPM) by Gallagher et al. (2000) combines several previously developed models (including Carpenter and Anderson (1992) and Gallagher et al. (1998)) using transition equations in order to obtain a more complete description of the plasma in the inner magnetosphere.

Despite the extensive use of these empirical density models in space physics simulations, they cannot provide reliable density estimates during extreme events, such as geomagnetic storms, since they are parameterized only by static geomagnetic parameters. The described models do not include the dynamic dependence of plasma density on solar wind conditions, and density is known to be highly variable during storm times (e.g., Moldwin et al., 1995; Park, 1974; Park & Carpenter, 1970). Accurately calculating the evolving density distribution from first principles has also proven elusive due to the sheer number of physical processes involved (Huba & Krall, 2013; Marchaudon & Blelly, 2015).

The IMAGE mission (Imager for Magnetopause-to-Aurora Global Exploration) (Burch, 2000), operating in 2000–2005, provided another important source of information about the plasmasphere. The IMAGE satellite was launched into an orbit with a very high-latitude apogee, which made it capable of remotely observing the azimuthal distribution of plasmaspheric plasma. The Extreme UltraViolet (EUV) instrument (Sandel et al., 2000) on board the IMAGE satellite produced the first global images of the plasmasphere. Analysis of the

images revealed the complex and dynamic evolution of the plasmasphere (Goldstein, Burch, et al., 2005; Spasojević et al., 2003) and highlighted the importance of the cold plasma distribution in controlling other magnetospheric processes (Baker et al., 2004; Spasojević et al., 2004). Various plasmaspheric structures, such as plumes, notches, shoulders, fingers, channels, and crenulations, have been discovered or better understood (Darrouzet et al., 2009, and references therein). The global EUV images can also be used to infer the plasmapause position by looking at the outermost sharp He^+ edge (Goldstein, Spasojević, et al., 2003). However, despite work in this area (Gurgiolo et al., 2005; Nakano et al., 2014), routine conversion of the images to plasma density has not been available, in part due to difficulty inverting the line-of-sight measurements as well as the unknown H^+-to- He^+ density ratio.

In this work we use a different approach to model plasma density and better understand its dynamical dependence on solar wind and geomagnetic conditions. The methodology employed is neural network-based empirical modeling. Neural networks are a powerful tool for deriving a highly multivariate nonlinear functional relation between input and output data, if such a relation exists (Anderson, 1995; Bishop, 1995; Haykin, 2009). An inherent part of neural networks includes finding the optimal inputs that best fit the output parameter. Thus, neural networks can serve as a good tool to identify parameters that are optimal for quantification of the plasmasphere dynamics.

In this work we investigate two important aspects pertaining to the plasmasphere dynamics: (1) the critical combination of geomagnetic and solar wind parameters that determine the evolution of the plasmasphere and (2) the time history duration that is critical for quantification of the plasmasphere dynamics.

The inputs to our neural network models include not only location (L and MLT) like in the empirical models described above but also the time history of solar wind and geomagnetic parameters. Our neural network models have a single output: plasma density. To identify the optimal combination of input parameters and time history that quantifies the plasmasphere dynamic evolution, we train and test multiple neural networks with various combinations of input and internal network parameters. The models undergo an extensive process of validation using in situ density measurements withheld from the training set. We also compare the model-predicted global evolution with global images of helium distribution (from IMAGE EUV). The model that performs well both quantitatively (on the point satellite measurements) and qualitatively (reproduces the global dynamics of the plasmasphere qualitatively well) is selected as the final model.

The rest of the paper is organized as follows. In section 2, we describe the methodology employed. We provide the detailed description of the data used to train neural networks, the procedure of optimal input selection, and the quantitative and qualitative validation methods. In section 3, we explain how the final model was selected and show examples of density determination for different disturbance intervals for both Van Allen Probes and IMAGE EUV data. In sections 4 and 5, we discuss the physical implications of the selected inputs, model performance, its advantages and limitations, and possible improvements.

2. Methodology

The methodology employed in this work is neural network-based empirical modeling. Neural networks are one of the best techniques for finding multivariate nonlinear mapping between input variables (in this case, radial distance, MLT, geomagnetic and solar wind parameters, and their time histories) and output variables (in this case, a single output, cold plasma density). Constructing a neural network requires a set of data, referred to as a training set, for which both the inputs and outputs are known. The input to the model is the time history of geomagnetic and solar wind parameters. The output of the model is electron number density. We construct a large number of neural networks and determine which configuration of input variables (which solar wind parameters, length of time history, etc.) and internal network parameters (number of neurons in the hidden layer) produces a model that (1) is quantitatively accurate and (2) generalizes well and avoids overfitting the training data. To do so, the neural networks undergo an extensive process of validation and testing using data that was withheld from the original training set. Moreover, we validate the model-predicted global evolution by comparing it to global images of the helium distribution from NASA's IMAGE mission.

In the rest of the section, we describe training, validation, and input selection procedures in detail.

2.1. Details of the Neural Network

In this work, we use feedforward neural networks, a type of neural network commonly used for solving regression, approximation, and function-fitting problems (e.g., Hassoun, 1995; Haykin, 1994). Here we describe specific aspects of neural networks concerning the current work. A detailed description of feedforward neural

networks is provided in section 2 of Zhelavskaya et al. (2016). Zhelavskaya et al. (2016) used a feedforward neural network with one hidden layer to derive electron number density from electric and magnetic field measurements made by Van Allen Probes.

Following Zhelavskaya et al. (2016), we also use feedforward neural networks with one hidden layer. Although it is possible to design a neural network with multiple hidden layers, it has been proven both theoretically and empirically that it is sufficient to have one hidden layer of finite size in a network to approximate continuous functions (Cybenko, 1989). A second layer is unlikely to improve results and introduces the additional problem in that the commonly used gradient descent training algorithms are unlikely to find an optimal solution when applied to all layers at once (Hochreiter et al., 2001). Moreover, adding more layers to the neural network increases the number of parameters to be optimized (i.e., it is necessary to search for an optimal number of neurons in every additional hidden layer) and can significantly increase the computational demand for training. Certainly, for more complex tasks, such as object recognition in images or other tasks related to computer vision, deep neural networks (neural networks with many hidden layers) are generally used and have achieved significant success (e.g., Cireşan et al., 2012; Le, 2013; Krizhevsky et al., 2012). However, different approaches to training must be employed for the case of deep neural networks in order to train them efficiently (e.g., Vincent et al., 2010).

The complexity of a neural network is characterized by the number of neurons in the hidden layer. More neurons in the hidden layer means a more complex network. In building a neural network model, it is important to find the model of optimal complexity that will minimize the error and at the same time will not overfit the training data, that is, will generalize well when applied to data that was not used to train the network. Here we determine the optimal number of neurons using fivefold cross validation combined with a learning curve approach. These techniques are described in section 2.4.

In this work, the input to the neural networks is the time history of geomagnetic and solar wind parameters and the location given by L and MLT. The output of the neural networks is plasma density. The process of input selection is described in detail in section 2.3. The plasma density data set used to train the neural networks is described in section 2.2.

2.2. Training Data

All solar wind, magnetic field data, and geomagnetic parameters were obtained from NASA's OMNIWeb data service. The density data to train the neural networks came from the density database obtained using the Neural-network-based Upper hybrid Resonance Determination (NURD) algorithm (Zhelavskaya et al., 2016) for the period 1 October 2012 to 1 July 2016. Zhelavskaya et al. (2016) used neural networks to infer upper hybrid resonance frequency from electric and magnetic field measurements made with the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) instrumentation suite (Kletzing et al., 2013) on the Van Allen Probes satellites. Deriving the plasma density from the upper hybrid frequency is straightforward. This electron density data set is publicly available at ftp://rbm.epss.ucla.edu/ftpdisk1/NURD. Van Allen Probes provide density measurements for $L \sim 2 - 6$ at all local time sectors.

2.3. Inputs to the Neural Network

A wide variety of parameters might potentially be used to quantify the plasmasphere dynamics. They include (1) geomagnetic parameters (*Kp*, *AE*, *AU*, *AL*, *Dst*, *SYM-H*, and *ASYM-H*), (2) solar wind data (n_p, v, P_{dyn}, B_z, and B_y), (3) solar wind coupling functions (vBs, vB_T, $vB_T \sin\theta_c$, and $d\Phi_{MP}/dt$) (e.g., Newell et al., 2007), (4) solar cycle indicators ($F_{10.7}$, sunspot number, and ionospheric IG index), (5) time history of activity inputs (total duration (e.g., 24, 36, and 48 h) and resolution (linear, log)), and (6) averaging technique (weighted average and average from $t = 0$).

The goal of this work is to identify the critical combination of geomagnetic and solar wind parameters that determine the evolution of the cold plasma. However, the number of possible combinations of these parameters is extremely large, and sweeping through the input parameter space for optimal input parameters may not be feasible due to the cost in time and computational resources. In this work, as a first step toward building an accurate plasmasphere model, we consider a limited number of input parameters. We focus on two important scientific questions:

1. How well do models driven solely (a) by geomagnetic parameters, (b) by solar wind, or (c) by a combination of both predict cold plasma dynamics? (i.e., which combination of activity parameters determines the evolution of the cold plasma best?)

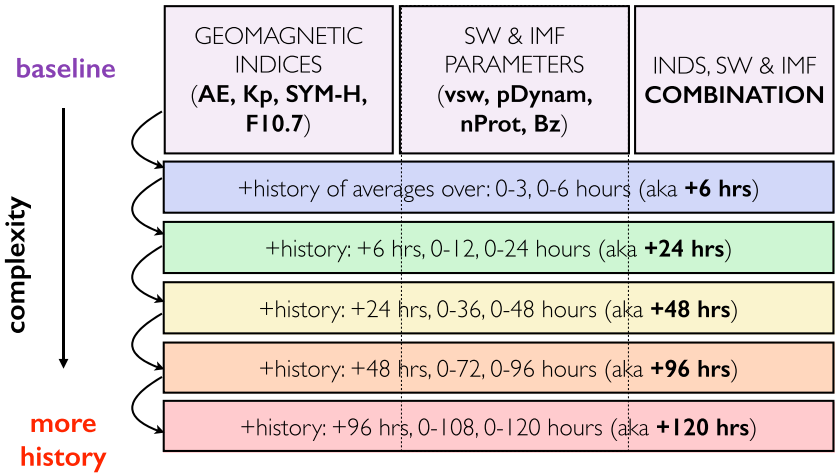

Figure 1. Eighteen combinations of input parameters to the neural network explored in this work. We start from simple models containing only instantaneous values of (1) geomagnetic indices, (2) solar wind and IMF parameters, and (3) their combinations and then subsequently add a time history of these parameters to the models. Each subsequent model includes inputs of the previous model and several more inputs corresponding to more time history added.

2. What is the memory of the magnetosphere? That is, what is the total time history duration of the solar wind and geomagnetic state that is critical in quantifying the distribution of cold plasma within the magnetosphere?

To address these questions, we consider the combinations of input parameters shown in Figure 1. We consider models based solely on geomagnetic indices, solely on solar wind data, and on a combination of both. Since the size and shape of the plasmasphere is known to be a result of the integrated time history of magnetospheric convection, it is important to provide as inputs not only the current value of activity parameters but also their time history. We start with simple models containing only instantaneous values of activity parameters (first row of Figure 1). Every neural network also includes a location input as L and MLT. Then, we subsequently add more time history of the corresponding parameters to the networks, up to 120 h of time history. We use averages of the time histories of activity parameters integrated from hour 0 (e.g., 0–3 h, 0–6 h, and 0–12 h). Initial analysis showed that such averaging technique produces a model that transitions more stably from one state to the next than one that uses successive time histories (0–3 h, 3–6 h, 6–12 h, etc.). Overall, the lower the row in Figure 1, the larger the number of input parameters included to train the neural network.

We use the procedure described in the next section to determine the combination of parameters that best quantifies the cold plasma dynamics.

2.4. Model Selection and Validation

A model is a configuration of input variables, internal neural network parameters, and a learning algorithm. We consider 18 combinations of input variables shown in Figure 1. The internal network parameters include the number of hidden layers in the network, the number of neurons in the hidden layer, and the transfer function of neurons. We use networks with one hidden layer in this work. The transfer function of neurons in the hidden layer is the hyperbolic tangent and is linear in the output layer. The training algorithm is the Levenberg-Marquardt algorithm (Hagan & Menhaj, 1994; Hagan et al., 1996; Levenberg, 1944; Marquardt, 1963) for all neural networks. The only parameter that varies and is being optimized is the number of neurons in the hidden layer. As mentioned before, the number of neurons in the hidden layer characterizes the complexity of the neural network.

To select the optimal model, that is, a neural network model with input variables and internal parameters that (1) is quantitatively accurate and (2) generalizes well and avoids overfitting the training data, we consider two aspects: (1) model performance on the point density measurements made on Van Allen Probes (local validation) and (2) model's ability to reproduce the global dynamic evolution of the plasmasphere (global validation).

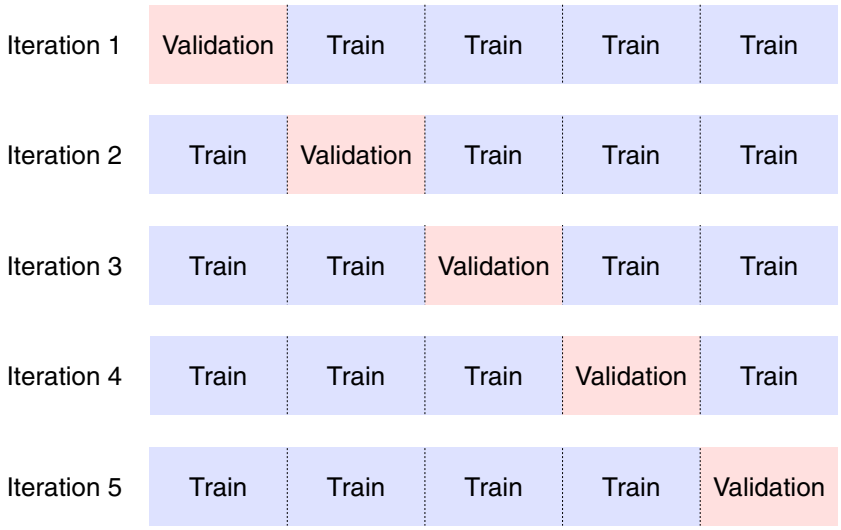

Figure 2. Schematic representation of K-fold cross validation (when $K = 5$). The data set is split into five subsets of approximately equal size. At every iteration, a different subset is left out, and the rest of the data are used for training. Thus, we can estimate the error of the model on five different subsets unseen during the training and obtain the mean and standard deviation of error of the validation error.

We select the model that is optimal in terms of both local and global validation. If two models have a similar performance, we select the less complex model (this is a common practice in machine learning; e.g., Blumer et al., 1987).

Models are validated locally by using K-fold cross validation combined with the learning curve analysis. For the global validation, we compare the neural network output on an L and MLT grid (global view of the plasmasphere in the equatorial plane) to images of He^+ distribution taken with the EUV instrument. Specifically, we compare the plasmapause locations identified manually in the EUV images to the approximation of plasmapause derived from the neural network global output. We discuss aspects of local and global validation in the two following subsections.

2.4.1. Local Validation: Cross Validation and Learning Curves

K-fold cross validation (CV) is a method for estimating the generalization ability of the model by measuring its error on validation data (not shown during the training) (e.g., Devijver & Kittler, 1982, Kohavi, 1995). The optimal model is the model with the lowest validation error among the considered models. It is worth noting here that a model is a configuration of input variables, internal network parameters, and a learning algorithm, as opposed to a particular model instance trained in the cross validation procedure. By using cross validation, we select a model, but not an individual learned instance of a model. Combining CV with the learning curve analysis significantly speeds up the model selection procedure (e.g., Cortes et al., 1994; Meek et al., 2002). Learning curves are also used to determine whether a model overfits or underfits (Perlich et al., 2003). Both K-fold cross validation and learning curves are described below in detail.

The idea of K-fold cross validation is shown in Figure 2. The data set is divided randomly into K subsets of approximately equal size ($K = 5$ in Figure 2). In each iteration, one subset is left aside and not used for training, while the rest of the $K - 1$ subsets are used to train a neural network. The subset left aside is called a validation set. The performance of the model on the validation and training sets is then calculated. One can use an error measure suitable for a specific application at hand (e.g., RMSE, MAPE, or sMAPE) to calculate the performance of the model. This procedure is repeated K times with the validation subset being different in each iteration. This way we can assess how well the model performs on the unseen data and obtain an estimate of generalization error mean and standard deviation. If the procedure is repeated a certain number of times (every time the data set is split differently), the result becomes more statistically significant.

One drawback of cross validation is that it can be computationally expensive if the training procedure takes significant time. One factor that can influence the time required for training is the size of the training set. The larger the training set, the more demanding the training procedure. However, if the training set is extremely large, it is possible that some part of the data does not add any new information to the model.

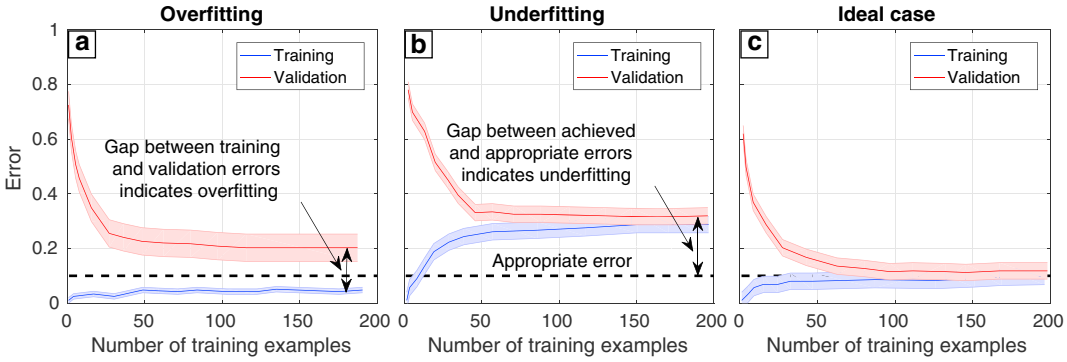

Figure 3. Toy illustration demonstrating learning curves in three different scenarios: (a) overfitting, (b) underfitting, and (c) ideal case. All three panels show plots of error versus number of examples in the training set. In every panel, the blue solid curve denotes cross validation (CV) error on the training set, and the blue shaded area shows the standard deviation of CV training error. The red solid curve represents CV error on the validation set, and the red shaded area shows standard deviation of CV validation error. The black dashed line stands for the appropriate error level in the task at hand.

Thus, for the sake of model selection, it might be sufficient to train models on a smaller portion of the data. Learning curve analysis allows assessment of how much data are needed to build a model and how much a model benefits from adding more data.

A learning curve is a plot of validation and training errors (obtained during cross validation) as a function of the training set size (number of observations in the training data). The learning curve is used to assess improvements in the model performance as the number of observations in the training sample increases. To illustrate how the method works, we describe three possible scenarios that can be observed in practice below: overfitting, underfitting, and desired model performance.

1. Overfitting occurs when a model is too complex for the underlying data structure. Subsequently, it can fit every data point in the training data set perfectly but performs very poorly on the unseen data (validation data). One example of overfitting would be fitting a quadratic function with a high-order polynomial. The fitted polynomial might go through every training data point, but it will not fit well the general shape of a quadratic function. Thus, the generalization ability of such a model is very low, and it cannot be used in practice. The performance of a model that overfits might be improved if more data are used to train it or if the complexity of the model is reduced.
2. Underfitting, on the other hand, occurs when a model is too simple for the underlying data structure; that is, it cannot fit the data because it does not have a necessary level of complexity. An example of underfitting would be fitting a quadratic polynomial with a straight line (a one-order polynomial). It is obvious that no matter what coefficients we choose for the straight line or how much data we use, it is not possible to fit a straight line to a quadratic function. A model that underfits has high error, both on the training and validation data sets. Such a model is also not suitable for use in practice. Its performance can be improved only by increasing its complexity.
3. In the case when the errors on both the training and validation sets are close to each other and at the same time are low enough, a model can be considered well trained and can be used in practice.

Figure 3 shows a toy illustration of these scenarios. In all three plots, the error obtained during cross validation is plotted on the *y* axis against the number of training examples on the *x* axis. The blue curve is the mean error on the training set; the blue shaded region is the standard deviation of the training error. The red curve and shaded region show the mean and standard deviation of the model error on the validation set, respectively. Dashed horizontal line shows the appropriate or desired error chosen for a specific application. All three scenarios start similarly. When the number of samples in the training set is small, the training error is close to 0 because the network can fit all the points in the training set perfectly, and the validation error is very high. Then, for the case of overfitting (Figure 3a), the error on the training set remains small as more observations are added to the training set. At the same time, the error on the validation set decreases but is still too far from the training error or the appropriate error level. The gap between the training and validation errors indicates overfitting. For the underfitting scenario (Figure 3b), training and validation errors move closer to

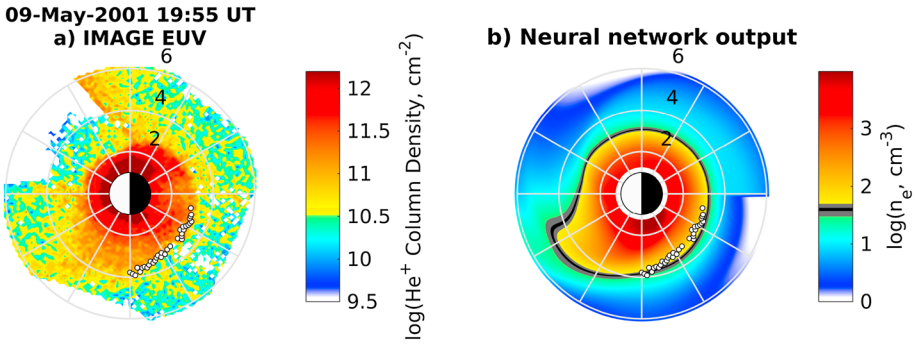

Figure 4. Example of comparison of the neural network output with EUV image. (a) EUV image of the He^+ column density mapped to the equatorial plane with the manually selected plasmapause indicated (white circles). (b) The output of a neural network model trained on 3.6 years of RBSP-A and RBSP-B measurements driven by the time history of the solar wind for the same interval. The neural network captures the erosion on the nightside and the formation of a plume in the afternoon sector. Minimum *Dst* is −68 nT for this event, and maximum *Kp* is 5.3. The Sun is to the left.

each other as the training set size increases but are both high and do not reach the appropriate error level. This generally means that the model is too simple and is not able to fit the data. Ideally (Figure 3c), as the number of training examples increases, the training and validation errors are close to each other and are close to the appropriate error level indicated by the dashed line. Moreover, after some point, increasing the training sample size does not further reduce the error (in Figure 3c, the error is not reduced by increasing the training set from 100 to 200 data points).

Learning curves and cross validation can also be used in identifying whether a model is stable, that is, performs similarly on different sets of data. This aspect is very important, since, if a model is not stable, that is, the spread of its error is large, we cannot trust predictions made by this model in practice. This can be interpreted from the standard deviation of error shown as the blue and red shaded areas in Figure 3 on the training and validation data, respectively. These areas show the spread of the model error on different training and validation subsets. A large spread means that the model is not consistent across different data sets, which we aim to avoid. It is therefore desirable that this spread is small. This aspect is also taken into consideration in the model selection procedure.

2.4.2. Global Validation: Comparison With IMAGE EUV Data

The validation techniques described above allow us to determine whether a model generalizes well on the unseen data and whether its output is quantitatively accurate. However, we also need to evaluate how well the model reproduces the global evolution of the plasmasphere, for example, the development and evolution of the plasmaspheric plume. For this, we compare the output of the neural network-based empirical models to EUV images mapped to the equatorial plane. Specifically, we compare the shape of the plasmasphere, the plasmapause locations derived manually from IMAGE EUV images, to the approximation of the plasmapause predicted by the neural network models. The density threshold of 40 ± 10 el/cm^3 is used as an approximation of the plasmapause in the global density reconstructions predicted by neural networks, that is, equivalent to the lower sensitivity threshold of the EUV instrument (Goldstein, Sandel, Forrester, et al., 2003). Although IMAGE EUV produced images of He^+ distribution and not of H^+ dominant in the plasmasphere, it was shown in the study of Goldstein, Sandel, Forrester, et al. (2003) that the sharp He^+ edge in the EUV images coincides with the actual plasmapause locations. Authors showed this by comparing the *L* shells of steep electron number density gradients derived from dynamic spectrograms produced by the IMAGE Radio Plasma Imager (RPI) in the passive mode with the *L* shells of He^+ edges in the EUV images obtained when the satellite was outside the plasmasphere near the apogee. The IMAGE EUV instrument provided the first global images of the plasmasphere. The IMAGE mission also operated during a different solar cycle than the one used in training. Therefore, we consider the IMAGE EUV images to be the best source of data available for validating the global evolution of the plasmasphere shape predicted by our neural network-based empirical models.

Figure 4a shows an example EUV image from 9 May 2001, a moderate geomagnetic disturbance interval (min *Dst* = −68 nT) examined in Spasojević et al. (2005). The white dots are the manually estimated plasmapause location. Figure 4b shows the output of a neural network empirical model. This particular model was trained, validated, and tested using over 1,000,000 points of density data from RBSP-A and RBSP-B. The

inputs to the network include *L*, MLT (rather than LT or UT since electric field is organized in terms of magnetic local time), and the averages of *AE*, *Kp*, *SYM-H*, and $F_{10.7}$ taken over 0–6, 0–12, 0–24, 0–36, 0–48, and 0–96 previous hours. The optimal network was found to have 45 neurons in the hidden layer with a root-mean-square error (RMSE) of 0.295 and 0.3 on the training data and in cross validation, respectively. The optimal number of neurons in the hidden layer was determined in the procedure described in section 3.2. The neural network was then applied to a grid of *L* and MLT using the solar wind and geomagnetic conditions from 9 May 2001. Goldstein, Sandel, Forrester, et al. (2003) estimated that the lower sensitivity threshold of the EUV instrument was equivalent to $40 \pm 10\ \mathrm{cm}^{-3}$, which is indicated by the black and gray section of the color bar in Figure 4b. However, this threshold depends on many factors including the unknown H^+-to-He^+ density ratio.

We perform such comparisons for multiple events starting from June 2000 to November 2005, when the IMAGE mission was operating. The considered events cover various geomagnetic conditions including both quiet and active times. Manually identified plasmapause locations for this period were taken from http://enarc.space.swri.edu/EUV. We visually compare the manually extracted plasmapause locations to the locations of the density threshold of $40 \pm 10\ \mathrm{cm}^{-3}$ in the global output of the neural network models. For each model, we identify how well it reproduces the size and shape of the plasmasphere in general for many different events. We track the ability of each model to reproduce the plume formation, its size and location, and the plasmasphere erosion on the nightside.

2.4.3. Model Selection

We determine the optimal model (configuration of input variables and internal network parameters) using results of both local and global validation. Local validation aids in identifying which models do not overfit or underfit and are stable (i.e., their performance is consistent over different data sets). By validating these models globally, we determine the optimal model that best reproduces the global dynamic evolution on average.

Finally, after determining the optimal configuration of input variables and internal neural network parameters, a model that can be used in practice should be built. To do that, all available data (from 1 October 2012, to 1 July 2016) are divided into two subsets that are called design and test sets. This division must be done before cross validation, so that the test set is completely independent and is not used in model selection or training. The division is in 9:1 ratio (design:test). The selected model is then trained on the whole design set. After the model is trained, its performance is evaluated on the test set and can be regarded as the final model error.

3. Implementation

3.1. Details of Neural Networks Implementation

All operations on neural networks are performed using the Matlab Neural Network Toolbox. To train the neural networks, we use the Levenberg-Marquardt algorithm (Hagan & Menhaj, 1994; Hagan et al., 1996; Levenberg, 1944; Marquardt, 1963). The training set is composed of density measurements made on both Van Allen probes A and B (Zhelavskaya et al., 2016). Density measurements coming from Van Allen Probes initially have a 6 s cadence. The minimum resolution of the solar wind measurements from OMNIWeb is 1 min. Therefore, we use 1 min averages of density measurements to train the neural networks. That way, we can preserve the variability in the plasma density measurements and at the same time have the maximum possible measurement resolution for the OMNIWeb data. After averaging, a total of 3,274,908 density measurements are available for training.

The data set is randomly divided into two parts in a 9:1 ratio, as discussed above. The larger part, the design set, is used to select the optimal model by using fivefold cross validation and learning curve analysis. Eighteen neural networks models corresponding to 18 input combinations from Figure 1 have been considered in the model selection process. The smaller part, the test set, is used to estimate the accuracy of the final model. The test set is used neither in training nor in validation.

For each neural network model with different inputs, it is also necessary to select the optimal number of neurons in the hidden layer (internal parameter of the neural network). The optimal number of neurons might depend on the number of inputs to the model, which contributes to the complexity of the model. The number of inputs to the models ranges from 6 in the simplest models based solely on instantaneous values of solar wind data or geomagnetic parameters to 90 in the most complex model, which includes 120 h of time history of both solar wind and geomagnetic parameters. Depending on the number of inputs, different numbers of neurons in the hidden layer are considered when selecting the optimal number of neurons for each model.

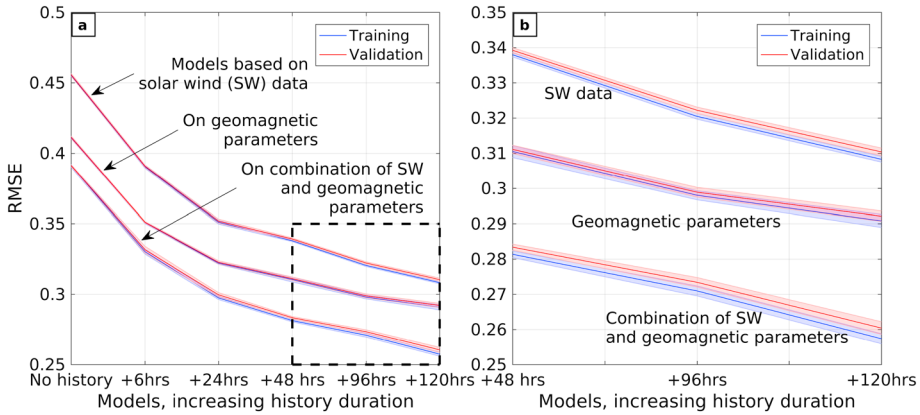

Figure 5. Root-mean-square error (RMSE) on the *y* axis versus models having different time history included on the *x* axis for (a) all models and (b) models having time histories of 48, 96, and 120 h. The farther on the *x* axis, the more time history is included into a model. In both panels, the blue solid curve denotes the cross validation (CV) error on the training set, and the blue shaded area shows standard deviation of the CV training error. The red solid curve represents the CV error on the validation set; the red shaded area shows standard deviation of the CV validation error. Models based on combination of solar wind and geomagnetic parameters perform best, while models based only on solar wind data have the worst performance.

The following empirical formula was used to calculate the range of neurons to consider for each number of inputs: round$((x - \mod(x,5)) \times \text{coef}_i)$, where x is the number of inputs to the model and coef_i is each number from the range $\{0.75, 1.0, 1.25, 1.5, 1.75\}$. Using this formula, we obtain five different numbers of neurons to consider. For example, if a number of inputs is equal to 34, this formula will yield 23, 30, 38, 45, 53. In total, we consider 90 neural networks in our analysis, which arise from 18 possible model configurations (input combinations), with each of them having five possible numbers of neurons in the hidden layer. The following procedure is conducted for each neural network model.

The design data set is randomly divided into five subsets of approximately equal size for K-fold cross validation (the division is the same for all models). Here we use $K = 5$, and thus a total of five cross-validation iterations are performed for each model. The size of one subset is approximately 590,000 samples. Thus, the total number of measurements that can be used for training in one iteration of cross validation is approximately 2,360,000 samples. In each iteration, a different subset was left aside to estimate the accuracy of the model trained on the remaining four subsets. The learning curve plot is produced as follows. In each iteration, the validation set remains the same, while the training set is increased from 100 to 1,000,000 training samples. Training samples are drawn from the four subsets used for training in each CV iteration. Depending on the model complexity, a different number of training samples might be sufficient for the training and validation errors to become close to each other. Specifically, it is sufficient to train simple models (such as ones from the first to the third rows in Figure 1) on the training set containing 500,000 samples, but for complex models sometimes more than 1,000,000 samples might be required. In that case, we increase the size of the training set and train the network on this set. In all our tests, the training set size does not reach 2,360,000 samples (size of the total training set in one iteration). On average, training and validation errors get close to each other when the training set has 1,000,000 samples.

3.2. Optimal Model Selection

3.2.1. Local Validation

After the training procedure described above is conducted for all neural network models, the model with the optimal number of neurons in the hidden layer is selected for each of the 18 considered neural networks with different inputs (see Figure 1). We select models that minimize the error on the validation set in the same way done by Zhelavskaya et al. (2016). Below we plot the errors of the 18 resulting models with the optimal number of neurons selected.

Figure 5a shows the root-mean-square error (RMSE) of the resulting models against the time history included in the model inputs. The three groups of curves correspond to models with different input combinations as indicated by arrows in the plot: solely based on the time history of solar wind data, of geomagnetic indices, and on their combination. Blue curves show errors on the training set, and the blue shaded areas

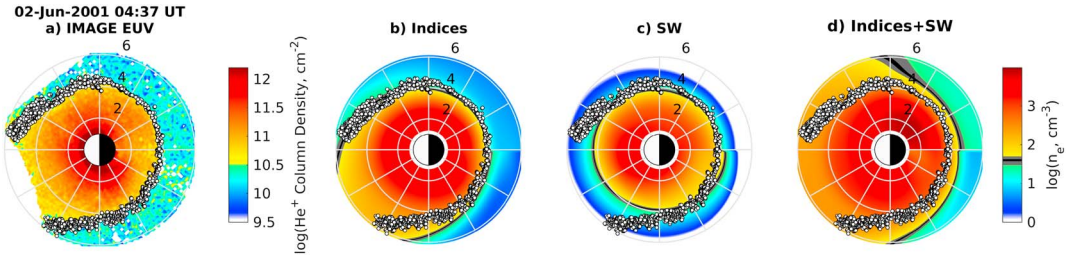

Figure 6. Example comparison of the EUV image of the He^+ column density mapped to the equatorial plane (a) with the output of three neural network models: (b) based solely on geomagnetic indices, (c) based on solar wind parameters, and (d) based on both geomagnetic and solar wind parameters. White circles indicate the manually selected plasmapause. Minimum *Dst* for this event is −27 nT, maximum *Kp* is 5.3. The Sun is to the left.

show the standard deviation of error. Red curves and shaded areas show the same but for the validation set correspondingly. Figure 5b is a zoomed-in fragment of Figure 5a indicated with the dashed rectangle. All three groups of curves decrease quite steeply in the beginning when not much history is included in the models. As the duration of the time history included in the models grows, the curves become less steep. This suggests that there is a certain saturation point in the time history, after which the performance of models stops improving. Although in the frame of considered durations of time history, the error is still reduced to 120 h.

Figure 5 shows that models based solely on solar wind data perform worse than models based on geomagnetic parameters or a combination of both. Models based on the combination of the solar wind and geomagnetic parameters have the lowest error. Thus, although solar wind alone might not be the best driver of the plasmasphere dynamics, its contribution is significant when combined with geomagnetic parameters. The improved performance of the combination model implies a nonlinear relationship between solar wind and geomagnetic parameters and plasmasphere dynamics.

Finally, the error spread (indicated by the shaded areas in Figure 5) of the models based solely on geomagnetic parameters is smaller than that of the models based on the combination of parameters. The difference between the mean error on the training and validation sets is also smaller for the models based solely on geomagnetic parameters. This can indicate slight overfitting of the models based on the combination of solar wind and geomagnetic parameters.

3.2.2. Global Validation

Output of all models is also compared to the images of the global He^+ distribution made with the IMAGE EUV instrument. Particularly, plasmapause locations manually identified in IMAGE EUV images were compared to the approximation of the plasmapause locations in the neural network output, which was determined by a density threshold of $40 \pm 10\,cm^{-3}$ (Goldstein, Sandel, Forrester, et al., 2003).

Comparison with IMAGE EUV data showed that models trained solely on solar wind parameters perform worse than the models based on the geomagnetic parameters or their combination, which confirms the quantitative results in Figure 5. Models based on the geomagnetic parameters and the combination of the solar wind and geomagnetic parameters perform significantly better in reproducing the global dynamics of the plasmasphere and capturing plume formation and evolution. However, it was observed that despite being quantitatively more accurate, models based on the combination of parameters perform slightly worse during increased geomagnetic activity when compared with IMAGE data. This might be due to the slight overfitting indicated above (seen in Figure 5).

An example of such a comparison is shown in Figure 6. Figure 6a shows an example EUV image from 2 June 2001, a mild geomagnetic disturbance interval (min *Dst* = −27 nT) examined in Goldstein, Sandel, Hairston, et al. (2003). The white dots are the manually estimated plasmapause location. Figures 6b–6d show the output of three neural network models correspondingly: (b) based solely on geomagnetic indices, (c) on solar wind parameters, and (d) both geomagnetic and solar wind parameters. The black and gray section of the color bar indicates a $40 \pm 10\,cm^{-3}$ density threshold. Inputs to all three networks contain 96 h of time history. Each model contains its optimal number of neurons selected in the local validation procedure described above. The number of neurons in the hidden layer of these models are 45 (Figure 6b), 45 (Figure 6c), and 81 (Figure 6d). It can be seen from Figure 6 that the model based solely on geomagnetic parameters captures the erosion on the nightside and the formation of a plume in the afternoon sector quite well, whereas

the model based solely on solar wind does not reproduce the plume formation. The model based on the combination of parameters seems to overestimate the size of the plume and does not reproduce the plasmasphere shape accurately. Such a performance of the models based solely on solar wind (SW) data was quite often observed when compared to IMAGE EUV data. Frequently, independent of how much time history is included into the SW-based model, it does not do a very good job at reproducing plume formation and evolution. Additionally, the models based on the combination of parameters do not always accurately reconstruct plasmasphere evolution. Another significant disadvantage of the models that include solar wind parameters is the presence of gaps in the solar wind data, which cause gaps in the density reconstruction. In general, the models based solely on geomagnetic parameters produce the most reliable and stable global density reconstruction.

3.2.3. Resulting Model

As a result of local and global validation procedures, for further use in practice we selected the model based solely on the geomagnetic parameters. In the future, gaps in the solar wind data might be filled (Kondrashov et al., 2010), and as our quantitative analysis shows, models based on the combination of parameters should perform better. The total duration of the geomagnetic parameters time history that is critical for quantifying plasmasphere dynamics was found to be 96 h. A visual comparison with IMAGE EUV plasmapause data shows that there is no significant improvement in performance of the model that includes 120 h of history. Therefore, the less complex model is selected. The optimal number of neurons in the hidden layer, that is, at which the model has the minimum error on the validation set, was found to be 45.

Altogether, the resulting final neural network model is based on the 96 h time history of the following indices: *AE*, *Kp*, *SYM-H*, and $F_{10.7}$ and has 45 neurons in the hidden layer. After selecting the optimal combination of input parameters and number of neurons in the hidden layer, the final neural network model is trained on the whole design data set and evaluated on the test set. The root-mean-square error (RMSE) of the resulting model is 0.3015 on the test set and 0.2950 on the design set. The linear correlation coefficient between the density measurements made on the Van Allen Probes and the density determined with the developed model is 0.9458 on the test set and 0.9462 on the design set. The final model is further referred to as the PINE model — the Plasma density in the Inner magnetosphere Neural network-based Empirical model.

4. Results

4.1. Examples of the Resulting Model Output

Figures 7a and 7b show plots of the electron density during the March storm in 2015. The blue curve shows the density derived with the NURD algorithm that is referred to as the true density on Van Allen Probes. The red curve shows density determined by the PINE model, and the black curve shows the density calculated using the plasmasphere and trough density model by Sheeley et al. (2001). Figure 7c shows the *Kp* index during this event. It can be seen from the plot that the PINE model is capable of capturing the plasmasphere dynamics during varying geomagnetic activity. The developed model does a significantly better job following the changing density levels than the model by Sheeley et al. (2001). The resulting model was tested for multiple events, and its performance is consistently better than Sheeley et al. (2001).

The top row of Figure 8 shows an example of global density reconstruction for a series of time frames during 26–27 June 2001, a minor geomagnetic disturbance interval (min *Dst*=−21 nT) examined in detail in Spasojević et al. (2003). The black and gray section of the color bar is 40 ± 10 el cm^{-3} and is equivalent to the estimated lower sensitivity threshold of the EUV instrument (Goldstein, Sandel, Forrester, et al., 2003). The *Kp* index for the duration of the event is shown in the bottom row of Figure 8, and red vertical lines indicate the particular times considered in the top row. To reconstruct the density in the equatorial plane, the developed model was applied to a grid of *L* and MLT using geomagnetic conditions from 26 to 27 June 2001. Several stages of this minor geomagnetic disturbance interval are shown. The first snapshot (Figure 8a) shows the onset of the disturbance. The neural network captures very well the nightside edge of the plasmasphere. Figures 8b–8d show the density reconstructed during the disturbance. The model does a good job of reconstructing the size of the plume, its rotation, and the plasmapause location. Figure 8e shows the recovery period after the disturbance when the plume begins to corotate with the main plasmasphere from the afternoon sector across the nightside. As discussed in Spasojević et al. (2003), this is a particularly interesting event because of its isolated nature. Due to its isolated type, this event is particularly difficult to model. Prior to the disturbance, *Kp* remained below 2 for over 40 h, and then it dropped to 1- for over 30 h immediately after the disturbance. The neural network model experiences difficulties in accurately reconstructing the

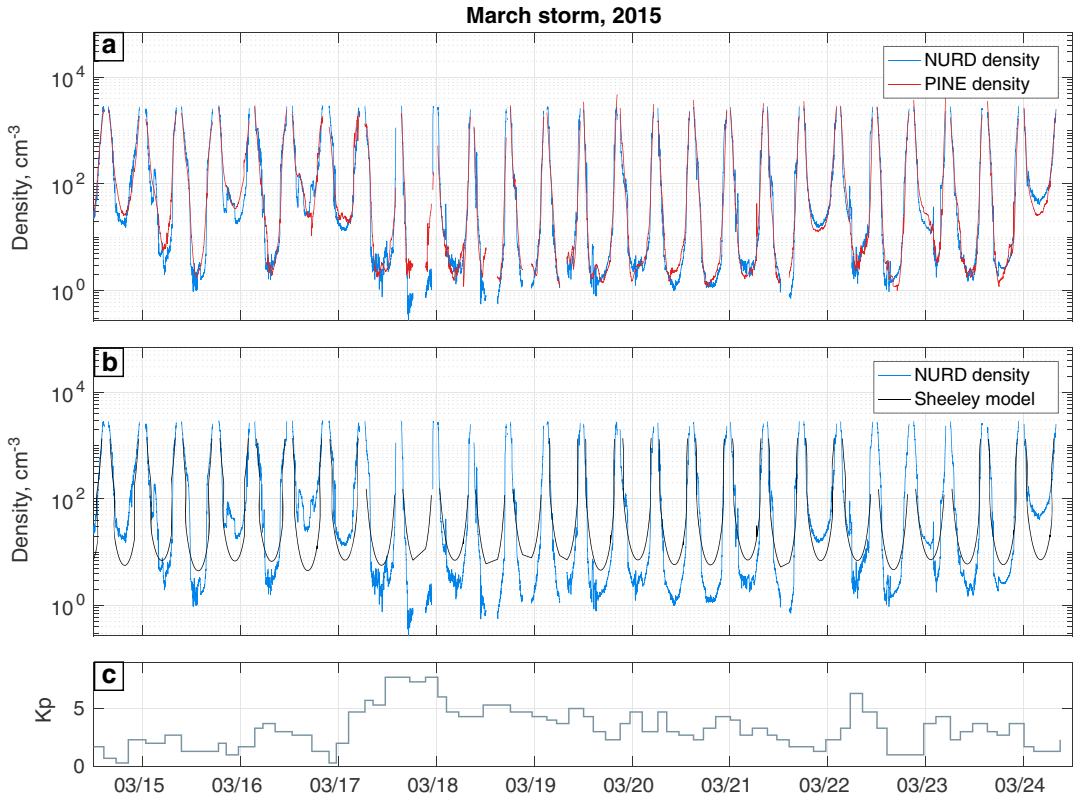

Figure 7. Example of density determination for the March storm in 2015. (a and b) Plots of density versus time. Blue curves are the density derived from Van Allen Probes using NURD algorithm (Zhelavskaya et al., 2016) for Probe A. Red curve is the density determined by the PINE model. Black curve is the density calculated using the trough and plasmasphere density model by Sheeley et al. (2001). (c) *Kp* geomagnetic activity index during this event.

plasmapause location in the recovery phase of the storm shown in Figure 8e. Looking carefully at this panel, we can see that the shape of the plasmasphere is slightly "flattened out," and the plasmasphere is deformed from having a spherical shape with an extended plume. Such unusual shape of the plasmasphere can pose difficulties for neural network since these types of events are underrepresented in data, which might explain why the neural network is not able to accurately capture the plasmapause location. However, despite the

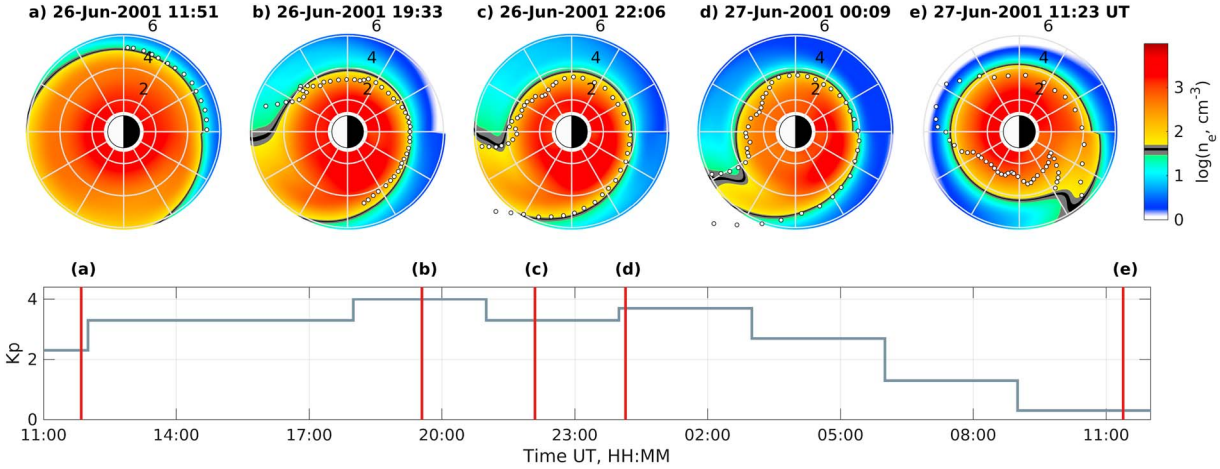

Figure 8. (top row) Example of global density reconstruction by the resulting neural network model during 26–27 June 2001, a minor geomagnetic disturbance interval (min *Dst* = −21 nT) examined in detail in Spasojević et al. (2003). The white dots denote the manually estimated plasmapause location. (bottom row) *Kp* geomagnetic activity index in gray; red vertical lines correspond to the density snapshots in the top row as indicated by letters. The Sun is to the left.

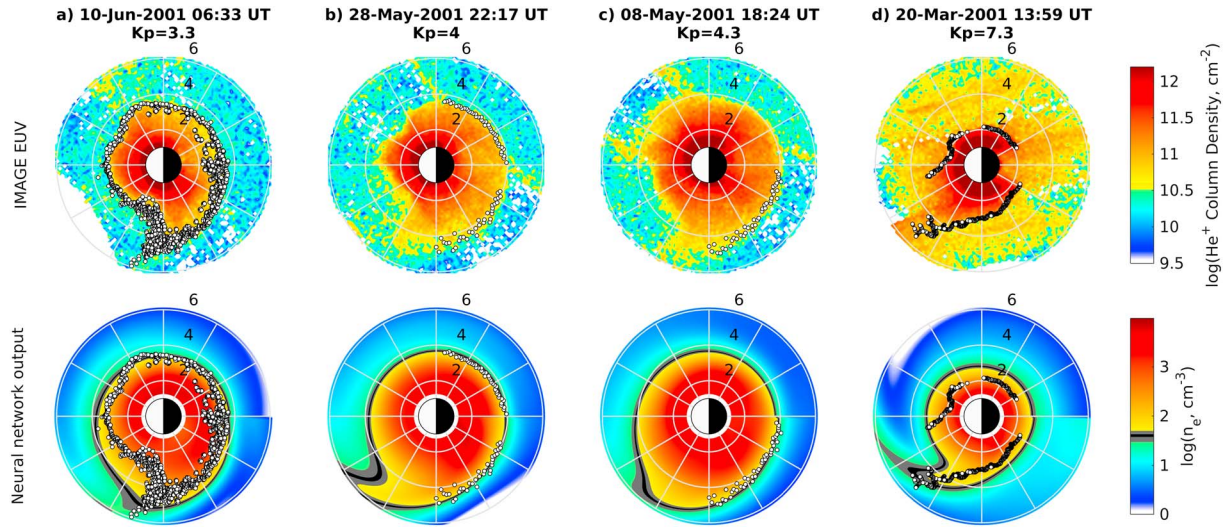

Figure 9. Examples of global density reconstruction by the resulting neural network model for four different events during the main phase plume formation. (top row) The EUV images for the times indicated in the titles and (bottom row) the final model output for those times. Events are ordered from left to right according to *Kp* (from low to high). *Kp* is shown in the titles as well. The Sun is to the left.

complexity of this case, the neural network was able to reproduce the corotation of the plume and accurately determine the plasmapause locations during the disturbance and its onset shown in Figures 8a–8d.

Further examples of global density reconstruction are shown in Figure 9 for four different events during their respective main phase plume formations: (a) 10 June 2001 at 06:33 UT, (b) 28 May 2001 at 22:17 UT, (c) 8 May 2001 at 18:24 UT, and (d) 20 March 2001 at 13:59 UT (considered in Spasojević et al. (2003, 2005); Spasojevic and Sandel (2010), and Goldstein et al. (2007), respectively). The top row shows the EUV images, and the bottom row shows the output of the resulting model for these times. Events are sorted from left to right in accordance with *Kp*: from low to high *Kp*. Again, the white dots indicate the manually selected plasmapause. It can be seen from the figures that the resulting model reproduces the plume formation quite well. Erosion on the nightside is very well captured for all four events. However, the model slightly overestimates the plasmapause boundary on the nightside for the time frame shown in Figure 9d, in the high geomagnetic disturbance interval where $Kp = 7.3$. This is caused by the fact that the training density data obtained from the Van Allen Probes is underrepresented for highly disturbed times. Nevertheless, the model does a fairly good job of reproducing the global shape of the plume for this event.

Overall, we find that the developed model does a very good job of capturing the erosion of the plasmapause on the nightside and the formation of the plume in the afternoon sector.

4.2. Model Use

The developed model was applied to multiple events from IMAGE and Van Allen Probes eras. The obtained density reconstructions can be found at ftp://rbm.epss.ucla.edu/ftpdisk1/PINE.

5. Discussion

Our results have shown that solar wind and geomagnetic parameters can be used to accurately quantify the plasmasphere dynamics. The relationship between solar wind and geomagnetic parameters and plasma density is not linear, and both types of parameters have their own effect on the cold plasma dynamics. Both quantitative (comparison with the satellite point measurements) and qualitative (comparison with the global EUV images) validations showed that neural network models based solely on the time history of solar wind data perform worse than models based solely on the time history of geomagnetic parameters. This might be caused by the fact that geomagnetic indices reflect the state of the geomagnetic system that is affected by the incoming solar wind and thus the indices already contain the information from the solar wind in a processed form. Also, geomagnetic indices are produced from the measurements made at ground observatories around the world, which makes them "global" in the sense of local time unlike the solar wind parameters measured at the Lagrangian L_1 point. Therefore, only solar wind parameters might not be fully representative of

all processes occurring in the geomagnetic system and hence might be a worse proxy for the plasmasphere dynamics than geomagnetic indices. Models based on the combination of both types of parameters are the most accurate quantitatively; however, they tend to slightly overfit the training data. Overfitting might occur since the models based on the combination of parameters have more inputs and thus are more complex. More complex models (with more inputs) are more prone to overfitting (e.g., Linoff & Berry, 2011). However, the fact that the models based on the combination of parameters are quantitatively more accurate than the models based solely on either the solar wind or geomagnetic parameters suggests that there is a nonlinear relationship between solar wind and geomagnetic parameters, and the plasmasphere dynamics.

In terms of reproducing the global plasmasphere dynamics, models based on the solar wind data have a hard time reproducing plumes, regardless of how much time history is included. This again might be potentially caused by the inability of solar wind parameters to fully describe the state of geomagnetic system as opposed to the geomagnetic indices. The two other types of models perform better in this regard. However, models based on the combination of parameters have a significant disadvantage: there are gaps present in the solar wind data, and hence there are gaps in density determination (the gaps in solar wind data for years 2012 to 2016 range from 22.96 to 28.89% per year for solar wind data with 1 min cadence). These results answer the first question stated in section 1: the critical combination of solar wind and geomagnetic parameters that best quantify the plasmasphere dynamics includes the *Kp*, *AE*, *SYM-H*, and $F_{10.7}$ indices.

The time history of the solar wind and geomagnetic parameters plays a very important role in quantifying the global dynamic evolution of the plasmasphere. Models that contain little to no time history are not able to reproduce plume formations and are not very reliable during high geomagnetic activity. Starting from 48 h, time history models tend to perform better during active times and in capturing the formation of the plume. The optimal duration of time history of the solar wind and geomagnetic state that quantifies the distribution of cold plasma within the magnetosphere was found to be 96 h. By employing such a time history, it becomes possible to reconstruct density for interesting and complex events as the one shown in Figure 8. In this event, models including less time history were not able to reproduce the plume formation and rotation after the disturbance. This result answers the second question stated in this work: the critical duration of time history of parameters necessary for quantification of the plasmasphere dynamics is 96 h.

Neural networks proved to be effective in finding the multivariable nonlinear mapping from the time history of solar wind and geomagnetic parameters to plasma density. It is worth noting, however, that neural networks learn better when data are abundant. For our case, a large volume of data is available for the times of quiet geomagnetic activity. However, the available data are not sufficient for the times of high geomagnetic activity. Data used for training the PINE model covers the period from October 2012 to July 2016, which is a relatively quiet period with few storms. The PINE model still produces satisfactory results during periods of strong geomagnetic activity ($Kp > 7$), but at times it has difficulties in accurately reproducing the plasmasphere shape. It is also important to note that the IMAGE EUV images were obtained between 2000 and 2005, whereas density data from Van Allen Probes is measured between 2012 and 2016. These are different periods in the solar cycle, which might also affect neural network training. This may be improved by using a larger training set covering a full solar cycle and with more disturbance intervals on which to train the network behavior, such as adding the density databases from the CRRES and THEMIS (Time History of Events and Macroscale Interactions during Substorms) missions.

The PINE model performs remarkably well in reproducing the global plasmasphere dynamics and large-scale irregularities, such as plumes. However, the model's current limitation is capturing small- and medium-scale irregularities (on the order of a few tenths of an R_E). This might be due to several reasons. First, the influence of many activity parameters on the plasmasphere has not been explored in this work and therefore is not included in the model (such as *AU*, *AL*, B_x, B_y, solar wind coupling functions (Newell et al., 2007), etc.). Second, a rather coarse averaging of time (from hour 0) was included in the models. However, it is difficult to know which parameters would help model these fine-scale features, since they are not very well understood (Darrouzet et al., 2009) and probably involve very complex coupling that is unlikely to come up out of a set of parameters.

Despite the limitations, the PINE model does a very good job of reconstructing the global evolution of cold plasma dynamics. It successfully reproduces the erosion of plasma on the nightside, sunward surge of plasma on the dayside, and rotation of the plume toward midnight during disturbed times (e.g., Darrouzet et al., 2009; Goldstein et al., 2004; Sandel et al., 2003; Spasojević et al., 2003). During quiet times, the PINE model can also reproduce the size of the plasmasphere remarkably well. Output of the PINE model agrees significantly

better with both Van Allen Probes and IMAGE EUV observations than output of the existing empirical model by Sheeley et al. (2001), both on a case-by-case comparison and on average.

6. Conclusions

We have built an empirical model of cold plasma density that is capable of globally reconstructing the plasmasphere dynamics. The model is based on a 96 h time history of geomagnetic activity parameters. To find the nonlinear mapping between activity parameters and plasma density, neural networks were used.

To determine the critical combination of activity parameters, we explored the influence of geomagnetic and solar wind parameters and their time history on the plasmasphere dynamics separately and in combination by means of neural network-based empirical modeling. Neural networks with different combinations of input parameters have been developed and compared among each other. To train the networks, a density data set for Van Allen Probes developed by Zhelavskaya et al. (2016) was used. The performance of the networks was estimated quantitatively by means of cross validation and learning curve analysis on the data from Van Allen Probes. Qualitatively, the ability of the networks to reconstruct the global plasmasphere dynamics was tested by comparing the approximate plasmapause locations in the output of the neural networks with the plasmapause data obtained by the IMAGE Extreme UltraViolet (EUV) instrument.

The comparison showed that models improve their performance as more time history is included. The critical duration of the time history of parameters, after which no significant improvement is observed, was found to be 96 h. Models based on the combination of solar wind and geomagnetic parameters achieved the best quantitative performance. However, due to gaps in the solar wind data and slight overfitting of the models based on the combination of solar wind and geomagnetic parameters, it was decided to use the model based solely on geomagnetic parameters at the moment. If the gaps in the solar wind data are filled in the future and the same analysis is repeated, models based on the combination of solar wind and geomagnetic parameters show potential to perform better. Nevertheless, the improved quantitative performance of the combination model implies that a nonlinear relationship exists between solar wind and geomagnetic parameters, and the plasmasphere dynamics.

In this work, we have demonstrated that neural networks are an efficient technique for building complex models based on space weather data.

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 by the NURD algorithm and is available at ftp://rbm.epss.ucla.edu/ftpdisk1/NURD. This work was supported by Helmholtz-Gemeinschaft (HGF) [10.13039/501100001656], NSF GEM AGS-1203747, NASA grant NNX12AE34G, NASA grant NNX16AF91G, and project PROGRESS funded by EC | Horizon 2020 Framework Programme (H2020) [10.13039/100010661] (637302)). The research has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through grant SFB 1294. The work at Stanford was supported by NASA award NNX15AI94G. I. S. Z. is grateful to Alexander Drozdov and Nikita Aseev for helpful discussions and support of this project. The global electron density reconstructions obtained in this work are available from ftp://rbm.epss.ucla.edu/ftpdisk1/PINE.

References

Anderson, J. A. (1995). *An introduction to neural networks*. London, UK: MIT Press.
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*, 6705–6712. https://doi.org/10.1002/2016GL069665
Baker, D., Kanekal, S., Li, X., Monk, S., Goldstein, J., & Burch, J. (2004). An extreme distortion of the Van Allen belt arising from the 'Hallowe'en' solar storm in 2003. *Nature*, *432*(7019), 878–881.
Bishop, C. M. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press.
Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's razor. *Information Processing Letters*, *24*(6), 377–380.
Borovsky, J. E., & Denton, M. H. (2006). Effect of plasmaspheric drainage plumes on solar-wind/magnetosphere coupling. *Geophysical Research Letters*, *33*, L20101. https://doi.org/10.1029/2006GL026519
Burch, J. (2000). IMAGE mission overview. *Space Science Reviews*, *91*(1-2), 1–14.
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*, *97*(A2), 1097–1108. https://doi.org/10.1029/91JA01548
Carpenter, D. L., & Stone, K. (1967). Direct detection by a whistler method of the magnetospheric electric field associated with a polar substorm. *Planetary and Space Science*, *15*(2), 395–397.
Chappell, C., Harris, K., & Sharp, G. (1970a). 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.
Chappell, C., Harris, K., & Sharp, G. (1970b). The morphology of the bulge region of the plasmasphere. *Journal of Geophysical Research*, *75*(19), 3848–3861.
Chen, S.-H., & Moore, T. E. (2006). Magnetospheric convection and thermal ions in the dayside outer magnetosphere. *Journal of Geophysical Research*, *111*, A03215. https://doi.org/10.1029/2005JA011084
Cireśan, D., Meier, U., Masci, J., & Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. *Neural Networks*, *32*, 333–338.
Cortes, C., Jackel, L. D., Solla, S. A., Vapnik, V., & Denker, J. S. (1994). Learning curves: Asymptotic values and rate of convergence. *Advances in Neural Information Processing Systems*, *6*, 327–327.
Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303–314.
Darrouzet, F., Gallagher, D. L., André, N., Carpenter, D. L., Dandouras, I., Décréau, P. M., … Tu, J. (2009). Plasmaspheric density structures and dynamics: Properties observed by the CLUSTER and IMAGE missions. *Space Science Reviews*, *145*(1-2), 55–106.

Devijver, P. A., & Kittler, J. (1982). *Pattern recognition: A statistical approach*. London: Prentice Hall.
Dungey, J. W. (1961). Interplanetary magnetic field and the auroral zones. *Physical Review Letters*, *6*(2), 47–48.
Escoubet, C., Pedersen, A., Schmidt, R., & Lindqvist, P.-A. (1997). Density in the magnetosphere inferred from ISEE 1 spacecraft potential. *Journal of Geophysical Research*, *102*(A8), 17,595–17,609.
Foster, J., & Burke, W. (2002). SAPS: A new categorization for sub-auroral electric fields. *Eos, Transactions American Geophysical Union*, *83*(36), 393–394.
Gallagher, D., Adrian, M., & Liemohn, M. (2005). Origin and evolution of deep plasmaspheric notches. *Journal of Geophysical Research*, *110*, A09201. https://doi.org/10.1029/2004JA010906
Gallagher, D., Craven, P., & Comfort, R. (1998). A simple model of magnetospheric trough total density. *Journal of Geophysical Research*, *103*(A5), 9293–9297.
Gallagher, D. L., Craven, P. D., & Comfort, R. H. (2000). Global core plasma model. *Journal of Geophysical Research*, *105*(A8), 18,819–18,833. https://doi.org/10.1029/1999JA000241
Garcia, L. N., Fung, S. F., Green, J. L., Boardsen, S. A., Sandel, B. R., & Reinisch, B. W. (2003). Observations of the latitudinal structure of plasmaspheric convection plumes by IMAGE-RPI and EUV. *Journal of Geophysical Research*, *108*(A8), 1321. https://doi.org/10.1029/2002JA009496
Geiger, H., & Müller, W. (1928). Elektronenzählrohr zur messung schwächster aktivitäten. *Naturwissenschaften*, *16*(31), 617–618.
Goldstein, J., Burch, J., Sandel, B., Mende, S., C:son Brandt, P., & Hairston, M. (2005). Coupled response of the inner magnetosphere and ionosphere on 17 April 2002. *Journal of Geophysical Research*, *110*, A03205. https://doi.org/10.1029/2004JA010712
Goldstein, J., Kanekal, S., Baker, D., & Sandel, B. (2005). Dynamic relationship between the outer radiation belt and the plasmapause during March–May 2001. *Geophysical Research Letters*, *32*, L15104. https://doi.org/10.1029/2005GL023431
Goldstein, J., Sandel, B., Forrester, W., & Reiff, P. (2003). IMF-driven plasmasphere erosion of 10 July 2000. *Geophysical Research Letters*, *30*(3), 1146. https://doi.org/10.1029/2002GL016478
Goldstein, J., Sandel, B., Frey, H., & Mende, S. (2007). Multiple plasmapause undulations observed by the image satellite on 20 March 2001. *Journal of Atmospheric and Solar-Terrestrial Physics*, *69*(3), 322–333.
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), 2243. https://doi.org/10.1029/2003GL018390
Goldstein, J., Sandel, B., Thomsen, M., Spasojević, M., & Reiff, P. (2004). Simultaneous remote sensing and in situ observations of plasmaspheric drainage plumes. *Journal of Geophysical Research*, *109*, A03202. https://doi.org/10.1029/2003JA010281
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*, *108*(A4), 1147. https://doi.org/10.1029/2002JA009475
Goldstein, J., Spiro, R., Reiff, P., Wolf, R., Sandel, B., Freeman, J., & Lambour, R. (2002). IMF-driven overshielding electric field and the origin of the plasmaspheric shoulder of May 24, 2000. *Geophysical Research Letters*, *29*(16), 66–1–66-4. https://doi.org/10.1029/2001GL014534
Grebowsky, J. (1970). Model study of plasmapause motion. *Journal of Geophysical Research*, *75*(22), 4329–4333.
Gurgiolo, C., Sandel, B., Perez, J., Mitchell, D., Pollock, C., & Larsen, B. (2005). Overlap of the plasmasphere and ring current: Relation to subauroral ionospheric heating. *Journal of Geophysical Research*, *110*, A12217. https://doi.org/10.1029/2004JA010986
Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design* (3632 pp.). Boston: PWS Company.
Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*, *5*(6), 989–993.
Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. Cambridge, MA: MIT Press.
Haykin, S. (1994). *Neural networks: A comprehensive foundation*. New York: MacMillan.
Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ: Pearson.
Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). *Gradient flow in recurrent nets: The difficulty of learning long-term dependencies*. Los Alamitos: IEEE Press.
Huba, J., & Krall, J. (2013). Modeling the plasmasphere with SAMI3. *Geophysical Research Letters*, *40*, 6–10. https://doi.org/10.1029/2012GL054300
Kletzing, C., Kurth, W., Acuna, M., MacDowall, R., Torbert, R., Averkamp, T., … Tyler, J. (2013). The Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) on RBSP. *Space Science Reviews*, *179*(1-4), 127–181.
Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Paper presented at 14th International Joint conference on Artificial Intelligence (Vol. 14, pp. 1137–1145). Stanford, CA.
Kondrashov, D., Shprits, Y., & Ghil, M. (2010). Gap filling of solar wind data by singular spectrum analysis. *Geophysical Research Letters*, *37*, L15101. https://doi.org/10.1029/2010GL044138
Kotova, G. (2007). The Earth's plasmasphere: State of studies (a review). *Geomagnetism and Aeronomy*, *47*(4), 409–422.
Krall, J., Huba, J., & Fedder, J. (2008). Simulation of field-aligned H^+ and He^+ dynamics during late-stage plasmasphere refilling. *Annales Geophysicae: Atmospheres, Hydrospheres and Space Sciences*, *26*, 1507–1516.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097–1105). Lake Tahoe, Nevada.
Kurth, W., De Pascuale, S., Faden, J., Kletzing, C., Hospodarsky, G., Thaller, S., & Wygant, J. (2015). Electron densities inferred from plasma wave spectra obtained by the Waves instrument on Van Allen Probes. *Journal of Geophysical Research: Space Physics*, *120*, 904–914. https://doi.org/10.1002/2014JA020857
Le, Q. V. (2013). *Building high-level features using large scale unsupervised learning, Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 8595–8598). Vancouver, BC, Canada: Vancouver Convention & Exhibition Center, IEEE.
LeDocq, M., Gurnett, D., & Anderson, R. (1994). Electron number density fluctuations near the plasmapause observed by the CRRES spacecraft. *Journal OF Geophysical Research*, *99*, 23–661.
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*, 492–506. https://doi.org/10.1002/2014JA020943
Lemaire, J. F., & Gringauz, K. I. (1998). *The Earth's plasmasphere*. Cambridge, MA: Cambridge University Press.
Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, *2*(2), 164–168.
Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: For marketing, sales, and customer relationship management*. Crosspoint Boulevard Indianapolis, IN: Wiley.
Marchaudon, A., & Blelly, P.-L. (2015). A new interhemispheric 16-moment model of the plasmasphere-ionosphere system: IPIM. *Journal of Geophysical Research: Space Physics*, *120*, 5728–5745. https://doi.org/10.1002/2015JA021193

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, *11*(2), 431–441.
Mauk, B., Fox, N. J., Kanekal, S., Kessel, R., Sibeck, D., & Ukhorskiy, A. (2013). Science objectives and rationale for the Radiation Belt Storm Probes mission. *Space Science Reviews*, *179*(1-4), 3–27.
Meek, C., Thiesson, B., & Heckerman, D. (2002). The learning-curve sampling method applied to model-based clustering. *Journal of Machine Learning Research*, *2*, 397–418.
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*, *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.
Nakano, S., Fok, M.-C., Brandt, P. C., & Higuchi, T. (2014). Estimation of the helium ion density distribution in the plasmasphere based on a single IMAGE/EUV image. *Journal of Geophysical Research: Space Physics*, *119*, 3724–3740. https://doi.org/10.1002/2013JA019733
Newell, P., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research*, *112*, A01206. https://doi.org/10.1029/2006JA012015
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*, A01206. https://doi.org/10.1029/2006JA012015
Orlova, K., Shprits, Y., & Spasojevic, M. (2016). New global loss model of energetic and relativistic electrons based on Van Allen Probes measurements. *Journal of Geophysical Research: Space Physics*, *121*, 1308–1314. https://doi.org/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.
Perlich, C., Provost, F., & Simonoff, J. S. (2003). Tree induction vs. logistic regression: A learning-curve analysis. *Journal of Machine Learning Research*, *4*, 211–255.
Sandel, B. R., Broadfoot, A. L., Curtis, C., King, R., Stone, T., Hill, R., … Gallagher, D. L. (2000). The Extreme Ultraviolet Imager investigation for the IMAGE mission. *Space Science Reviews*, *91*, 197–242.
Sandel, B., Goldstein, J., Gallagher, D., & Spasojevic, M. (2003). Extreme Ultraviolet Imager observations of the structure and dynamics of the plasmasphere. *Space Science Reviews*, *109*(1), 25–46.
Sandel, B. R., King, R. A., Forrester, W., Gallagher, D. L., Broadfoot, A. L., & Curtis, C. (2001). Initial results from the IMAGE Extreme Ultraviolet Imager. *Geophysical Research Letters*, *28*(8), 1439–1442.
Sheeley, B., Moldwin, M., Rassoul, H., & Anderson, R. (2001). An empirical plasmasphere and trough density model: CRRES observations. *Journal of Geophysical Research*, *106*(A11), 25,631—25,641.
Shprits, Y. Y., Drozdov, A. Y., Spasojevic, M., Kellerman, A. C., Usanova, M. E., Engebretson, M. J., … Aseev, N. A. (2016). Wave-induced loss of ultra-relativistic electrons in the Van Allen radiation belts. *Nature Communications*, *7*, 12883.
Singh, N., & Horwitz, J. (1992). Plasmasphere refilling: Recent observations and modeling. *Journal of Geophysical Research*, *97*(A2), 1049–1079.
Singh, A., Singh, R., & Siingh, D. (2011). State studies of Earth's plasmasphere: A review. *Planetary and Space Science*, *59*(9), 810–834.
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*, L04803. https://doi.org/10.1029/2003GL018389
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*, *108*, 1340. https://doi.org/10.1029/2003JA009987
Spasojevic, M., & Sandel, B. (2010). Global estimates of plasmaspheric losses during moderate disturbance intervals. *Annales Geophysicae*, *28*, 27–36. Copernicus GmbH.
Spasojević, M., Thomsen, M., Chi, P., & Sandel, B. (2005). Afternoon subauroral proton precipitation resulting from ring current-plasmasphere interaction. In J. Burch, M. Schulz & H. Spence (Eds.), *Inner magnetosphere interactions: New perspectives from imaging* (pp. 85–99). Washington, DC: American Geophysical Union.
Su, Y.-J., Thomsen, M. F., Borovsky, J. E., & Foster, J. C. (2001). Linkage between polar patches and plasmaspheric drainage plumes. *Geophysical Research Letters*, *28*(1), 111–113.
Trotignon, J., Rauch, J., Décréu, P., Canu, P., & Lemaire, J. (2003). Active and passive plasma wave investigations in the Earth's environment: The Cluster/Whisper experiment. *Advances in Space Research*, *31*(5), 1449–1454.
Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, *11*, 3371–3408.
Zhang, Q.-H., Zhang, B.-C., Lockwood, M., Hu, H.-Q., Moen, J., Ruohoniemi, J. M., … Baker, J. B. (2013). Direct observations of the evolution of polar cap ionization patches. *Science*, *339*(6127), 1597–1600.
Zhelavskaya, I., Spasojevic, M., Shprits, Y., & Kurth, W. (2016). Automated determination of electron density from electric field measurements on the Van Allen Probes spacecraft. *Journal of Geophysical Research: Space Physics*, *121*, 4611–4625. https://doi.org/10.1002/2015JA022132