Title
A unified approach to inner magnetospheric state prediction

Journal
Journal of Geophysical Research Space Physics, 121(3)

ISSN
2169-9380

Authors
Bortnik, J
Li, W
Thorne, RM
et al.

Publication Date
2016-03-01

DOI
10.1002/2015ja021733

Peer reviewed

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Space Physics

TECHNICAL REPORTS: METHODS

10.1002/2015JA021733

Special Section:
Variability of the Sun and Its Terrestrial Impact VarSITI

A unified approach to inner magnetospheric state prediction

J. Bortnik[1], W. Li[1], R. M. Thorne[1], and V. Angelopoulos[2]

[1]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, California, USA, [2]Earth, Planetary and Space Sciences/Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, USA

Abstract This brief technique paper presents a method of reconstructing the global, time-varying distribution of some physical quantity *Q* that has been sparsely sampled at various locations within the magnetosphere and at different times. The quantity *Q* can be essentially any measurement taken on the satellite including a variety of waves (chorus, hiss, magnetosonic, and ion cyclotron), electrons of various energies ranging from cold to relativistic, and ions of various species and energies. As an illustrative example, we chose *Q* to be the electron number density (inferred from spacecraft potential) measured by three Time History of Events and Macroscale Interactions during Substorms (THEMIS) probes between 2008 and 2014 and use the *SYM-H* index, taken at a 5 min cadence for the 5 h preceding each observed data point as the main regressor, although the predictor can also be any suitable geomagnetic index or solar wind parameter. Results show that the equatorial electron number density can be accurately reconstructed throughout the whole of the inner magnetosphere as a function of space and time, even capturing the dynamics of elementary plasmaspheric plume formation and corotation, suggesting that the dynamics of various other physical quantities could be similarly captured. For our main model, we use a simple, fully connected feedforward neural network with two hidden layers having sigmoidal activation functions and an output layer with a linear activation function to perform the reconstruction. The training is performed using the Levenberg-Marquardt algorithm and gives typical RMS errors of ~1.7 and regression of >0.93, which is considered excellent. We also present a discussion on the different applications and future extensions of the present model, for modeling various physical quantities.

Key Points:
- We present a general method that can be used to model any quantity in the inner magnetosphere
- We demonstrate our method on the spatiotemporal distribution of electron density
- This model could be immensely useful for space weather and other applications

Correspondence to:
J. Bortnik,
jbortnik@gmail.com

Citation:
Bortnik, J., W. Li, R. M. Thorne, and V. Angelopoulos (2016), A unified approach to inner magnetospheric state prediction, *J. Geophys. Res. Space Physics*, *121*, 2423–2430, doi:10.1002/2015JA021733.

Received 27 JUL 2015
Accepted 3 MAR 2016
Accepted article online 9 MAR 2016
Published online 30 MAR 2016

1. Introduction

The Earth's inner magnetosphere is a complex, interconnected system that is driven by energy inputs from the Sun and solar wind (e.g., see introductory discussions in *Kivelson and Russell* [1995] and *Khazanov* [2011]). There are a variety of components that comprise the magnetospheric environment, such as the ionosphere, background electric and magnetic fields, plasma waves, electrons ranging from cool (<1 eV) to ultrarelativistic (>5 MeV) energies, and different species and energies of ions. These components can be externally or internally driven and typically couple to one another in subtle and often unexpected ways. Understanding the dynamical behavior of these various magnetospheric components is interesting and important scientifically and also has important societal implications in that the near Earth space environment can impact human technological systems in dramatic ways (see, for instance, the National Research Council's Decadal Survey report "Solar and Space Physics: A Science for a Technological Society").

A number of different approaches to specify and predict the state of the inner magnetosphere have been employed in previous studies. The first and most familiar involves the construction of first principles physics-based models, and a number of efforts are currently underway in pursuit of this goal [e.g., *Wolf et al.*, 1991; *Toffoletto et al.*, 2003, 2004; *Jordanova et al.*, 2008, 2010; *Wang et al.*, 2004; *Wiltberger et al.*, 2004; *Goodrich et al.*, 2004; *Pembroke et al.*, 2012]. The major advantage of physics-based models is that they employ fundamental physical principles and thus can be useful in extrapolating to situations that have not been previously observed. The disadvantage is that these models are typically large, complex, and computationally intensive. The coupling between various component systems is not straightforward, and if there are missing physical processes from the models (which there inevitably always are), the models can produce erroneous results when operated in unusual regimes.

On the other end of the spectrum, the state of the inner magnetosphere can be specified simply by plotting a particular quantity in a plane or a volume, broadly parameterized by some controlling parameter. For example, *Meredith et al.* [2012] plot the intensity of whistler mode chorus waves in the inner magnetosphere, parameterized by three ranges of the *AE* index representing quiet, moderate, and active conditions, based on data collected by the DE1, Combined Release and Radiation Effects Satellite, Cluster 1, Double Star TC1, and Time

History of Events and Macroscale Interactions during Substorms (THEMIS). Such models give realistic distributions but are essentially static and do not reflect the spatiotemporally changing conditions in the magnetosphere. In fact, this problem was described by *Fung* [1996] in relation to radiation belt fluxes, who also outlined the need for a new generation of models that would be data driven and spatiotemporal in nature.

Recent models have attempted to make such data-based models by performing function fitting to the data under various geomagnetic activity levels. For instance, *Orlova et al*. [2014] derive a model of plasmaspheric hiss intensity in two local time sectors and calculate weighting coefficients using linear stepwise regression for a quadratic polynomial with variables in *L* shell, latitude, and *Kp*. An extension of this approach was employed by *J.-H. Kim et al*. [2015], who fit up to seventh-order polynomials at every 2 h magnetic local time (MLT) interval to chorus wave intensity, in two latitude ranges and driven by either solar wind conditions or geomagnetic indices. While these functional fits make the models more flexible and dynamic, the functional forms used for fitting are only applicable to the parameter at hand and a new approach will inevitably need to be taken if a different parameter is to be predicted. One way to automatically perform nonlinear function fitting is to use the sophisticated machinery of the NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous inputs) technique, as has been recently reported by *Boynton et al*. [2013] in studying energetic particle fluxes at geosynchronous orbit. An alternative approach is using artificial neural network (ANN) to perform the nonlinear function fitting. The use of ANN and other techniques in "artificial intelligence" to problems in solar-terrestrial physics has been discussed for quite some time [*Jocelyn et al*., 1993], but application to magnetospheric problems has been relatively spare to date. A recent example was reported by *K.-C. Kim et al*. [2013, 2015] for the modeling of whistler mode chorus and plasmaspheric hiss, respectively. In these studies the authors perform extensive cross-correlation analyses to determine the parameters and time lags that produce the highest correlation with the quantity to be predicted and then use only those values to train a single-layer ANN to perform the function fitting. While this work represents the most sophisticated approach in this category of models, it nevertheless requires extensive hand selection of the model inputs, which often may not be obvious, and having a shallow ANN precludes the model from constructing its own optimal feature set. In fact, the problem of selecting the correct input parameters for the model was the main barrier in the development of the first ANN (then known as "perceptrons" [e.g., *Minsky and Papert*, 1969]).

In order to overcome some of the difficulties discussed above and come up with a universal model that can specify any given quantity in the inner magnetosphere (based on satellite data) without necessarily knowing the most important driving parameters a priori, we propose the use of deep neural networks, regressed upon a suitable driver, such as the time history of a relevant geomagnetic index. Section 2 describes the methodology and shows an application of this approach to creating a dynamic plasma density model. A summary and conclusions are given in section 3.

2. Methodology and Application

The general problem we aim to address can be stated as follows: given a sequence of measurements of some quantity Q_i, at time t_i and location $\mathbf{r}_i$, where $i = 1, \ldots, N$ and N is a "large" number, specify the distribution of Q at every point $\mathbf{r}$ in the model domain at time t and let that spatial distribution evolve as a function of time. In order to illustrate our technique, we present the application where the quantity Q_i is chosen to be the plasma density N_e, inferred from the spacecraft potential of three THEMIS probes [*Angelopoulos*, 2008], A, D, and E, using the method described in *Li et al*. [2010]. That data set covers the range of altitudes extending from well below $L = 2$ to approximately $L = 12$, over all local times, and extends from 1 June 2008 to 31 October 2014 and is reduced to 5 min averages, which results in ~10^6 data points (942,842 to be precise). We also performed some minimal data cleaning by removing all point with density values below 0.01 el/cm^3, observations at $L > 10$, as well as data points within 2 h of midnight at low L shells having values <10 el/cm^3.

In order to organize the data into a dynamic model, we choose a time series to regress the data against that will act as a suitable predictor and contain enough information embedded within it so as to explain the various behaviors of the data. For our application of density prediction, we use the preceding 5 h of the *SYM-H* index (also at 5 min cadence, giving 60 points) as the only predictor, since this index is simple, readily available through the OMNI database (ftp://spdf.gsfc.nasa.gov/pub/data/omni/high_res_omni/), and has been used by previous researchers in driving more complex plasmasphere-ionosphere models [e.g., *Huba and Sazykin*, 2014], i.e., it should contain sufficient information to act as a predictor for the cold plasma density. Our selection of *SYM-H* is not unique, and we could have just as easily used other geomagnetic indices such as *AE* or *Kp* (suitably subsampled to a 5 min cadence since plasmaspheric dynamics typically do not occur on

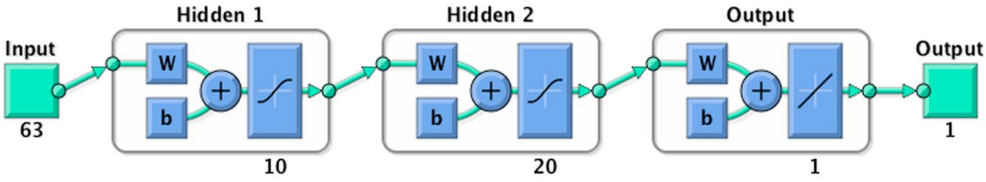

Figure 1. Schematic illustration of the neural network architecture used in our study. There are two hidden layers with 10 and 20 neurons, respectively, having sigmoid activation functions. There is one output neuron with a linear activation function.

timescales faster than 5 min), a combination of indices, or indeed solar wind parameters. However, for the purposes of demonstrating the technique (and to be consistent with *Huba and Sazykin* [2014]), it was not considered necessary to explore alternative inputs such as other geomagnetic indices or other time lags. However, when network performance is inadequate, varying the number and types of input parameters is an obvious thing to try. In addition, solar wind data tend to contain many data gaps which would result in losing many of our data points, whereas *SYM-H* is available for long periods of time with no data gaps.

The design matrix X, and target matrix, that results is

$$X = \begin{bmatrix} symH_{i-60} & \cdots & symH_{i-1} & L_i & \cos\phi_i & \sin\phi_i \\ \vdots & & & & & \\ symH_{N-60} & \cdots & symH_{N-1} & L_N & \cos\phi_N & \sin\phi_N \end{bmatrix}, \quad T = \begin{bmatrix} Q_i \\ \vdots \\ Q_N \end{bmatrix},$$

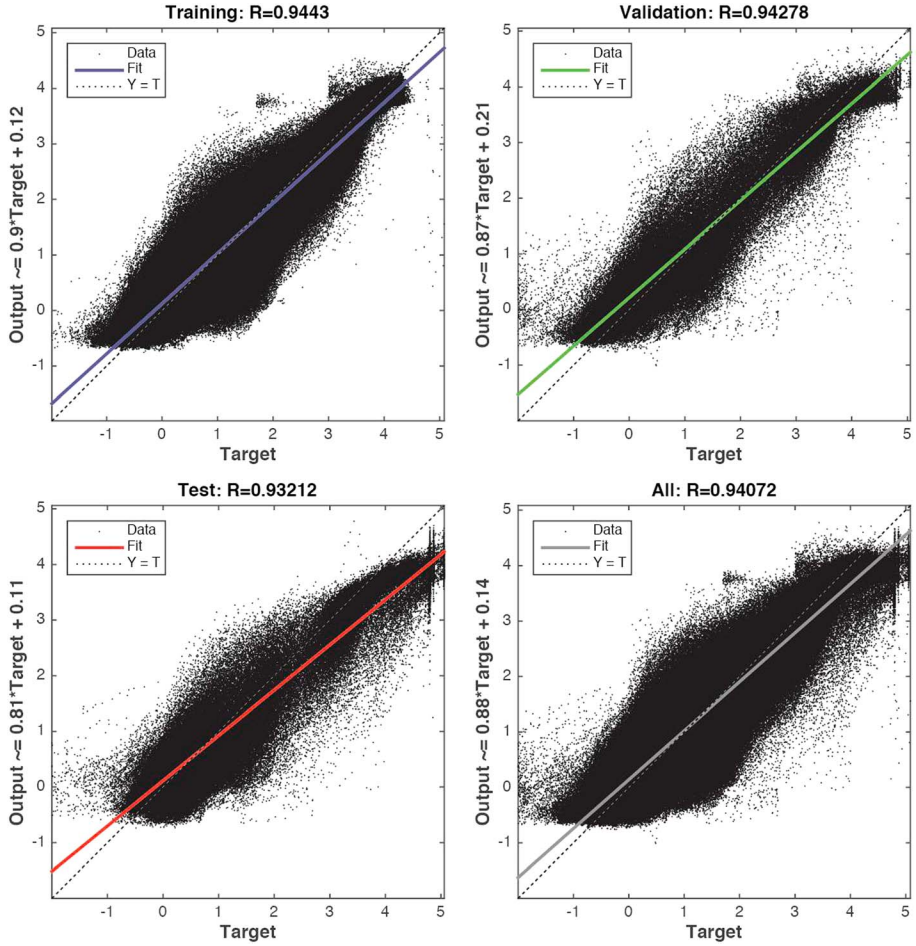

Figure 2. Regression of the various classes of data points as indicated in the panel headers (training, validation, test, and all samples), showing the target data ($\log_{10}N_e$) to be modeled on the abscissa, and the ANN output on the ordinate. The dashed straight line represents a perfect model fit. The colored lines in each plot represent the best fit lines to the data points and corresponding equation on the ordinate label. Regression values are shown in the headers and are everywhere >0.93.

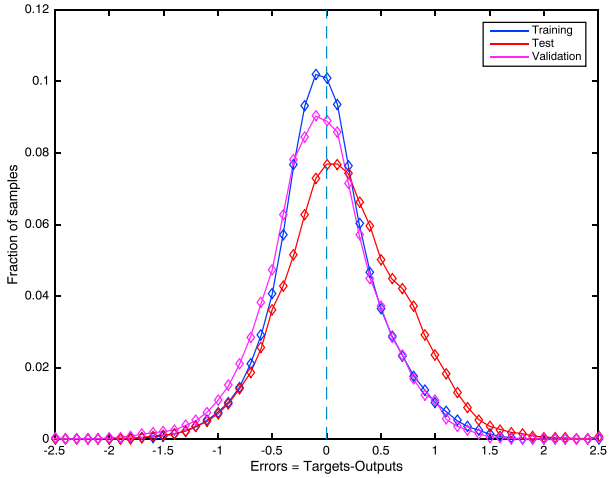

Figure 3. The distribution of errors (in $\log_{10}$ space) for the various classes of data points (training, validation, and test samples). Zero error, i.e., "perfect" fit is shown as the orange vertical line.

where the angle ϕ_i represents the MLT angle (MLT/24 × 2π) which is used with cosine and sine to guarantee continuity across midnight. L_i is the L shell of the satellite at time t_i, together with the corresponding $\log_{10}$ of the electron density Q_i observed by the spacecraft at that time, and $i = 1, \ldots, N$ such that X and T both have ~10^6 rows, and 63 columns. For the purposes of our illustration, we only consider the electron density in the magnetic equatorial plane, but the model could be readily extended to three dimensions by including in the design matrix X a column that contains the latitude λ_i of the satellite at time index i (this is readily accomplished with the present data set but only for near-equatorial regions. For regions further off the equator, additional data sets such as DE1 and POLAR need to be included in the training set, being ever mindful of intercalibration issues between different satellites).

The problem now becomes to create a model that predicts T given the design matrix X. This problem could be solved with a variety of techniques, and essentially, all machine learning techniques that deal with regression aim to solve a problem of this type [e.g., *Marsland*, 2009]. The simplest approach is to create a linear model by creating a set of weight $\mathbf{A} = [a_1, \ldots, a_m]^T$ such that $\mathbf{XA} = \mathbf{T}$ and invert to obtain the ordinary least squares solution, giving: $\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}$, but there is no guarantee that the response of our target quantity (electron density) is necessarily linear and in fact will most likely be a nonlinear function of the inputs.

The technique we present in this study is based on the artificial neural network (ANN), which is a very powerful and general approach to modeling complex, nonlinear functions and in fact has been shown to be a universal approximator to any smooth function, given that there are a sufficient number of neurons [*Hornik et al.*, 1989; *Cybenko*, 1989]. The architecture of our ANN is shown schematically in Figure 1 and contains an input layer with 63 features, two hidden layers with 10 and 20 neurons in the first and second layers, respectively, each having an identical sigmoid activation function (i.e., $S(z) = 1/(1 + \exp(-z))$) and output layer. We choose a two-hidden layer ANN because there is now good theoretical reason to believe that "deep" neural nets are more computationally efficient than "shallow" nets [*Bengio*, 2009] and (more importantly) that deep nets constitute better predictors because they are able to learn a more effective set of input parameters (called a "feature set" in machine learning parlance) than the inputs that are initially presented to the ANN [*Hinton et al.*, 2006]. In our application, we use a two-hidden layer ANN since this is the shallowest deep network, that is, deep in the sense that it has more than one hidden layer. The number of neurons in each hidden layer was chosen somewhat arbitrarily, and no attempt has been made at either optimizing performance or producing the most parsimonious network possible.

In general, the first layer will reconstruct the input parameters into its own optimal set of features, and by choosing 10 neurons in this layer we force a data compression, that is, a reduction in the dimensionality of the system from 63 to 10. This is intuitively plausible since the 63 input parameters are not all independent, and in particular, the 60 point, 5 h time history of *SYM-H* can be reduced significantly by extracting only the relevant information, for example, one might look for features such as the extent and intensity of dips in *SYM-H* and their duration. For our example of density prediction, one might guess at the most important features and present those directly to the ANN inputs, but in the general case, we would not necessarily know which aspects of the input time series contain the most relevant information, and it is best to let the ANN construct its own optimal set of features in the first few layers. The second layer in our ANN contains 20 neurons, and it is in this layer that the optimal features are recombined to capture the physical behavior of our system. In principle, the deep ANN should not normally need more than two hidden layers, since this architecture should be able to approximate any smooth functional mapping (see illustration in *Marsland* [2009, section 3.3.3]).

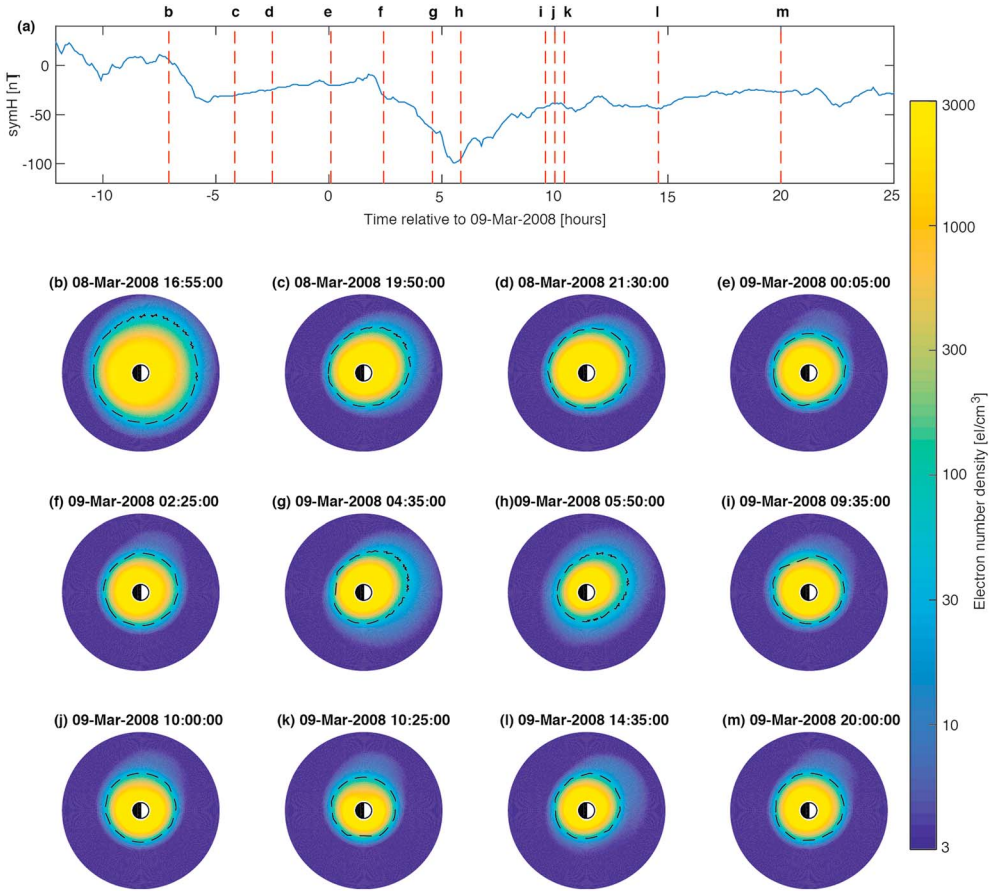

Figure 4. Neural network reconstruction of the global plasma density as a function of L and MLT. (a) The *SYM-H* index relative to 9 March 2008 00:00 UT. The red vertical dashed lines correspond to the times of the snapshots shown in the 12 panels beneath. The common color scale indicating electron density in electrons per cubic centimeter is shown on the right, and the dashed line in Figures 4b–4m represents the 50 el/cm^3 contour. The entire simulation period is outside the training interval (i.e., out of sample).

In order to train the neural network, we divide our data into three sequential blocks: a training set containing 70% of the samples, a validation set containing 15% of the samples, and a test set containing the remaining 15% of the samples. The training set of points is used to train the ANN together with 5 h time series of *SYM-H* preceding each data point, and here we used the Levenberg-Marquardt algorithm to perform the training [*Marsland*, 2009]. The training of the ANN continues, and the RMS error between predicted and observed values is calculated at each time step for all three sets of points. There are a variety of ways to judge neural net "goodness," but the approach we take here is to focus on the generalizability of the network, so we continue training until the RMS error of the validation set of points stops improving for several consecutive time steps (here we arbitrarily choose 10 time steps). Note that the validation set of points is not involved in the training per se and is thus independent (except that it is involved in stopping of the training step). The test set of points is truly an independent set of samples and acts as a check on the generalizability of the ANN and its performance.

Figure 2 shows the regression of ANN predictions versus the THEMIS observations for all three classes of data points and shows that the ANN has a correlation of 0.93 with the observations for the training, validation, and test data points, and further, that all three classes of points have roughly similar correlations so that the ~0.93 level can be considered the general power of the ANN in making predictions. Another way to state this goodness of fit is with r^2 measure, and here we see that the ANN can "explain" (0.93^2=) >86% of the variability of the observed data, which is considered to be excellent by the authors, especially given that this is an illustrative study and no attempt has been made at optimizing the predictive model.

Figure 3 shows the distribution of the errors, that is, the differences between the ANN-predicted values compared to observed values and having an overall RMS value of ~0.24. This translates to roughly a factor of

$10^{0.24} \sim 1.7$, which is well within the factor of ~2 error inherent in the method of obtaining density values from spacecraft potential [*Li et al.*, 2010].

Perhaps the most important aspect of the ANN is its ability to perform global, time-dependent specification and prediction. To demonstrate this capability, we choose a time period which contains a moderate storm and is completely out of sample from our training period, i.e., during March 2008. A few snapshots of the global evolution are shown in Figure 4 below, where the plasmasphere [e.g., *Nishida*, 1966; *Lemaire and Gringuaz*, 1998] is seen to be initially large and extended (particularly on the dayside) at the start of the event. As geomagnetic activity intensifies in 9 March 2008, the plasmasphere becomes smaller due to the erosion of its outer layers and extended on the dayside. In fact, the ANN is able to infer the existence of an elementary "plume" structure on the dayside (observed in the 09:35, 10:00, and 10:25) even though this structure is probably very poorly sampled in the data, and (by design) our model will only capture histories of 5 h or less, thereby neglecting long-lived plumes [*Borovsky et al.*, 2014].

Even a relatively simple model such as the one shown in Figure 4 is already of value and can be immediately deployed in conjunction with physical models or used in its own right to compare against physical models of the plasmasphere.

3. Summary and Discussion

The basic problem we set out to address in this paper was to find a method to reconstruct the global, time-varying distribution of some physical quantity Q that has been sparsely sampled at various locations within the magnetosphere and at different times. As an illustration, we chose the electron number density (inferred from spacecraft potential) measured by three THEMIS probes between 2008 and 2014 and used a simple, fully connected feedforward neural network with two hidden layers having sigmoidal activation functions and an output layer with a linear activation function to perform the reconstruction. The training was performed using the Levenberg-Marquardt algorithm and gave typical RMS errors of ~1.7 and regression of >0.93, which is considered excellent. The main regressor was a time series of the *SYM-H* index, taken at a 5 min cadence for the 5 h preceding each observed data point, giving a time series containing 60 points.

A few important points should be made regarding our choice of model. While the ANN model described in section 2 is one possible archetype of a machine learning approach that could be used for the purposes of specifying and predicting the quantities of interest to us, it should be noted that there are many alternatives to "learning" subtle features in data (e.g., multiple linear regression, linear and nonlinear autoregressive models with exogenous inputs, support vector machines, regression trees, and hidden Markov Models). Although we believe that ANN and particularly ANN with deep architectures are particularly well suited to the problem at hand [*Bengio*, 2012], there is nevertheless good evidence [*Banko and Brill*, 2001] showing that the quantity of data available for training, in fact, trumps the choice of model and that the worst performing model with a given quantity of data typically outperforms the best performing model, when the amount of training data is increased tenfold (this finding might be considered foundational for the recent popularity of "big data" or "data science"). Exploring different models and the effects of varying the quantity of training data should be an important continuation of the present work. Focusing on density alone, with the data we currently have in hand, it is entirely feasible to scale the data set used in our illustrative study by several times and extend the model to off-equatorial latitudes (currently, this simple model is only equatorial, though it could be immediately extended with off-equatorial scaling functions such as presented by *Denton et al.* [2006]).

We have chosen electron number density for our pilot study because this is a simple scalar quantity (as opposed to a vector magnetic field, for instance), and the physical processes controlling its behavior are more easily predictable than the subtle interplay that governs something like radiation belt dynamics, which is a delicate balance between source and loss processes [e.g., *Reeves et al.*, 2003]. However, we emphasize that the electron density in our pilot study can be viewed as a placeholder for a number of other quantities, such as electron (and ion) fluxes at a range of energies of interest (e.g., from thermal to ultra-relativistic), a variety of waves (e.g., chorus, plasmaspheric hiss, electromagnetic ion cyclotron, magnetosonic, and possibly even ultralow frequency waves), and even background quantities such as the magnetic and electric fields.

Another point to note is that we have used a very simple predictor in our pilot study, namely, the *SYM-H* index (chosen here to follow the work of *Huba and Sazykin* [2014]), which has proven to be remarkably powerful in being able to predict the number density distribution and evolution. It should, in principle, reflect all the

geoeffective features in the solar wind that impinge upon the magnetosphere, has the added advantage that it is readily available, and does not contain data gaps, as compared to solar wind data, for example. Nevertheless, it would be important to explore the time series of various solar wind parameters as the regressor in future work, which could shed important insights into the physics governing the state of the inner magnetosphere. Since different physical quantities evolve on different timescales, different cadences and lengths of time history should also be explored (being ever mindful of the "curse of dimensionality"). For very long input time series, convolutional neural networks might prove advantageous, due to the inherent translational invariance properties they possess [*Bengio*, 2009].

Finally, we make the point that statistical predictive techniques are not and could never be a substitute for physical modeling/understanding as suggested in recent reports [*Anderson*, 2008]. Models such as ANN perform well in regions that are densely sampled but generally do not perform well when making extrapolations. Physical models, on the other hand, should perform well in extrapolating from their typical operating domains, assuming that all the physical processes have been correctly represented. In the authors' opinion, the optimal use of machine learning techniques is in the so-called "insight discovery,"i.e., to discover any potentially important physical processes that may not have been included in the physics-based models, so that the physics-based modeling could be used to predict and specify in situations that have not been previously encountered.

Acknowledgments

The authors would like to gratefully acknowledge NASA grants NNX14AN85G, NNX11AR64G, NNX13AI61G, NNX14AI18G, NNX15AF61G, and NNX15AI96G. The authors acknowledge J.W. Bonnell and F.S. Mozer for use of EFI data and C.W. Carlson and J.P. McFadden for use of ESA data, which are used to infer plasma density. We would like to acknowledge THEMIS wave data obtained from http://themis.ssl.berkeley.edu/themisdata/ and thank the NSSDC Omniweb for the provision of the geomagnetic activity indices (ftp://spdf.gsfc.nasa.gov/pub/data/omni/omni_cdaweb/) used in this report.

References

Anderson, C. (2008), The end of theory: The data deluge makes the scientific method obsolete, Wired, June 23, 2008. [Available at http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.]

Angelopoulos, V. (2008), The THEMIS mission, *Space Sci. Rev.*, *141*(1–4), 5–34, doi:10.1007/s11214-008-9336-1.

Banko, M. and E. Brill (2001), Scaling to very very large corpora for natural language disambiguation, In ACL '01 proceedings of the 39th Annual Meeting on Association for Computational Linguistics.

Bengio, Y. (2009), Learning deep architectures for AI, *Found. Trends Mach. Learn.*, *2*(1), 1–127.

Bengio, Y. (2012), Learning deep architectures for AI, *Found. Trends Mach. Learn.*, *2*(1), 1–127, doi:10.1561/2200000006.

Borovsky, J. E., D. T. Welling, M. F. Thomsen, and M. H. Denton (2014), Long-lived plasmaspheric drainage plumes: Where does the plasma come from?, *J. Geophys. Res. Space Physics*, *119*, 6496–6520, doi:10.1002/2014JA020228.

Boynton, R. J., M. A. Balikhin, S. A. Billings, G. D. Reeves, N. Ganushkina, M. Gedalin, O. A. Amariutei, J. E. Borovsky, and S. N. Walker (2013), The analysis of electron fluxes at geosynchronous orbit employing a NARMAX approach, *J. Geophys. Res. Space Physics*, *118*, 1500–1513, doi:10.1002/jgra.50192.

Cybenko, G. (1989), Approximation by superposition of a sigmoidal function, *Math. Control Signals Syst.*, *2*, 303–314.

Denton, R. E., K. Takahashi, I. A. Galkin, P. A. Nsumei, X. Huang, B. W. Reinisch, R. R. Anderson, M. K. Sleeper, and W. J. Hughes (2006), Distribution of density along magnetospheric field lines, *J. Geophys. Res.*, *111*, A04213, doi:10.1029/2005JA011414.

Fung, S. F. (1996), Recent development in the NASA trapped radiation models, in *Radiation Belts: Models and Standards, Geophys. Monogr. Ser.*, vol. 97, edited by J. F. Lemaire, D. Heynerick, and D. N. Baker, pp. 79–91, AGU, Washington, D. C.

Goodrich, C. C., A. L. Sussman, J. G. Lyon, M. Shay, and P. Cassak (2004), The CISM code coupling strategy, *J. Atmos. Sol. Terr. Phys.*, *66*(15-16), 1469–1479.

Hinton, G. E., S. Osindero, and Y. W. Teh (2006), A fast learning algorithm for deep belief nets, *Neural Comput.*, *18*(7), 1527–1554.

Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer feedforward networks are universal approximator, *Neural Netw.*, *2*, 359–3661.

Huba, J. D., and S. Sazykin (2014), Storm time ionosphere and plasmasphere structuring: SAMI3-RCM simulation of the 31 March 2001 geomagnetic storm, *Geophys. Res. Lett.*, *41*, 8208–8214, doi:10.1002/2014GL062110.

Jocelyn, J. A., H. Lundstedt, and J. Trolinger (Eds.) (1993), *Proceedings of International Workshop on Artificial Intelligence Applications in Solar-Terrestrial Physics, Lund*, NOAA, Boulder, Colo.

Jordanova, V. K., J. Albert, and Y. Miyoshi (2008), Relativistic electron precipitation by EMIC waves from self-consistent global simulations, *J. Geophys. Res.*, *113*, A00A10, doi:10.1029/2008JA013239.

Jordanova, V. K., S. Zaharia, and D. T. Welling (2010), Comparative study of ring current development using empirical, dipolar, and self-consistent magnetic field simulations, *J. Geophys. Res.*, *115*, A00J11, doi:10.1029/2010JA015671.

Khazanov, G. V. (2011), *Kinetic Theory of the Inner Magnetospheric Plasma*, Springer, New York, doi:10.1007/978-1-4419-6797-8.

Kim, J.-H., D.-Y. Lee, J.-H. Cho, D.-K. Shin, K.-C. Kim, W. Li, and T. K. Kim (2015), A prediction model for the global distribution of whistler chorus wave amplitude developed separately for two latitudinal zones, *J. Geophys. Res. Space Physics*, *120*, 2819–2837, doi:10.1002/2014JA020900.

Kim, K.-C., Y. Shprits, J. Lee, and J. Hwang (2013), Empirically modeled global distribution of magnetospheric chorus amplitude using an artificial neural network, *J. Geophys. Res. Space Physics*, *118*, 6243–6253, doi:10.1002/jgra.50595.

Kim, K.-C., D.-Y. Lee, and Y. Shprits (2015), Dependence of plasmaspheric hiss on solar wind parameters and geomagnetic activity and modeling of its global distribution, *J. Geophys. Res. Space Physics*, *120*, 1153–1167, doi:10.1002/2014JA020687.

Kivelson, M. G., and C. T. Russell (1995), *Introduction to Space Physics*, pp. 586 , University Press, Cambridge, UK, Cambridge, isbn:0521451043.

Lemaire, J. F., and K. I. Gringuaz (1998), *The Earth's Plasmasphere*, Cambridge Univ. Press, Cambridge, U. K.

Li, W., R. M. Thorne, J. Bortnik, Y. Nishimura, V. Angelopoulos, L. Chen, J. P. McFadden, and J. W. Bonnell (2010), Global distributions of suprathermal electrons observed on THEMIS and potential mechanisms for access into the plasmasphere, *J. Geophys. Res.*, *115*, A00J10, doi:10.1029/2010JA015687.

Marsland, S. (2009), *Machine Learning, an Algorithmic Perspective, Chapman & Hall/CRC Machine Learning & Pattern Recognition*, CRC, Boca Raton, Fla.

Meredith, N. P., R. B. Horne, A. Sicard-Piet, D. Boscher, K. H. Yearby, W. Li, and R. M. Thorne (2012), Global model of lower band and upper band chorus from multiple satellite observations, *J. Geophys. Res.*, *117*, A10225, doi:10.1029/2012JA017978.

Minsky, M. L., and S. A. Papert (1969), *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass.
Nishida, A. (1966), Formation of plasmapause, or magnetospheric plasma knee, by the combined action of magnetospheric convection and plasma escape from the tail, *J. Geophys. Res.*, *71*, 5669–5679, doi:10.1029/JZ071i023p05669.
Orlova, K., M. Spasojevic, and Y. Shprits (2014), Activity-dependent global model of electron loss inside the plasmasphere, *Geophys. Res. Lett.*, *41*, 3744–3751, doi:10.1002/2014GL060100.
Pembroke, A., F. Toffoletto, S. Sazykin, M. Wiltberger, J. Lyon, V. Merkin, and P. Schmitt (2012) Initial results from a dynamic coupled magnetosphere-ionosphere-ring current model, *J. Geophys. Res.*, *117*, A02211, doi:10.1029/2011JA016979.
Reeves, G. D., K. L. McAdams, R. H. W. Friedel, and T. P. O'Brien (2003), Acceleration and loss of relativistic electrons during geomagnetic storms, *Geophys. Res. Lett.*, *30*(10), 1529, doi:10.1029/2002GL016513.
Toffoletto, F. R., S. Sazykin, R. W. Spiro, and R. A. Wolf (2003), Modeling the inner magnetosphere using the rice convection model (review), *Space Sci. Rev.*, WISER special issue, *108*, 175–196.
Toffoletto, F. R., S. Sazykin, R. W. Spiro, R. A. Wolf, and J. G. Lyon (2004), RCM meets LFM: Initial results of one-way coupling, *J. Atmos. Sol. Terr. Phys.*, *66*(15-16), 1361.
Wang, W., M. Wiltberger, A. G. Burns, S. C. Solomon, T. L. Killeen, N. Maruyama, and J. G. Lyon (2004), Initial results from the coupled magnetosphere-ionosphere-thermosphere model: Thermosphere-ionosphere responses, *J. Atmos. Sol. Terr. Phys.*, *66*(15-16), 1425.
Wiltberger, M., W. Wang, A. G. Burns, S. C. Solomon, J. G. Lyon, and C. C. Goodrich (2004), Initial results from the coupled magnetosphere ionosphere thermosphere model: Magnetospheric and ionospheric responses, *J. Atmos. Sol. Terr. Phys.*, *66*(15-16), 1411–1424.
Wolf, R. A., R. W. Spiro, and F. J. Rich (1991), Extension of the rice convection model into the high-latitude ionosphere, *J. Atmos. Terr. Phys.*, *53*, 817–829.