MODELING OF THE PLASMASPHERE DYNAMICS

Irina Zhelavskaya

Kumulative Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Computational Physics"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Physik und Astronomie
der Universität Potsdam
und
Deutsches GeoForschungsZentrum (GFZ)

22. Oktober 2020

BETREUER*INNEN:
Hauptbetreuer: Prof. Dr. Yuri Shprits
Zweitbetreuerin: Dr. Maria Spasojević

GUTACHTER*INNEN:
Prof. Dr. Jerry Goldstein
Dr. Viacheslav Merkin
Prof. Dr. Arkady Pikovsky
Prof. Dr. Claudia Stolle
Prof. Dr. Fabrice Cotton
Prof. Dr. Matthias Holschneider
Prof. Dr. Jens Wickert

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-48243
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-482433

Declaration

I, Irina Sergeevna Zhelavskaya, hereby declare that this thesis entitled "Modeling of the Plasmasphere Dynamics" has been composed entirely by myself. The work contained herein is my own, except where explicitly stated otherwise. I confirm that:

- I have fully acknowledged and referenced the ideas and work of others, whether published or unpublished.
- This dissertation has not been submitted for any other degree or professional qualification.

Irina Sergeevna Zhelavskaya

Contents

Acknowledgments

First, I would like to express my deepest gratitude to my principal thesis advisor, Prof. Yuri Shprits, for his constant support and guidance during my time at GFZ Potsdam and University of Potsdam. I am immensely grateful for his openness and always being willing to help and share his broad knowledge of the field, and for giving me freedom to explore my own ideas, while seamlessly guiding me in the right direction. He taught me to focus on the big picture and to see the forest for the trees, for which I will be forever grateful.

I would also like to extend my sincerest gratitude to my second thesis advisor, Dr. Maria Spasojević, for her support, sharing her extensive experience and knowledge, and always having been very generous with her time whenever I needed advice. Maria has set an example of a careful and excellent researcher for me, which I will always strive to be, and I am extremely grateful for the opportunity to work and learn from her.

I would like to sincerely thank the reviewers of my dissertation, Prof. Jerry Goldstein and Dr. Viacheslav Merkin, for their valuable comments and feedback that helped me improve this thesis further. I would also like to extend my gratitude to other members of my Examination Committee: Prof. Arkady Pikovsky, Prof. Claudia Stolle, Prof. Fabrice Cotton, Prof. Jens Wickert, and Prof. Matthias Holschneider.

I wish to pay special thanks to Dr. Alexander Drozdov for his help during my days at UCLA, supporting my work, and always being ready to help. I would also like to extend my deepest thanks to Prof. Jerry Goldstein and Dr. T. Paul O'Brien for their always insightful and valuable comments and suggestions on my research. I am very grateful for having the opportunity to discuss my work with them and get their feedback.

I am very grateful to my colleagues at GFZ for their help with the work that has become part of this dissertation: Dr. Nikita Aseev, Dr. Ruggero Vasile, Prof.

Claudia Stolle, and Dr. Jürgen Matzka. I would also like to sincerely thank Dr. Dominika Sörgel, Sharon Uy and Melanie Burns for their careful proofreading of all my papers and valuable feedback, and Dr. Nikita Aseev, Sandra Paulin, Dr. Michael Wutzig, and Dr. Dominika Sörgel for proofreading parts of this thesis and providing very helpful feedback. I would like to further express my gratitude to scientists, with whom I had an opportunity and pleasure to collaborate with during my PhD: Dr. William Kurth, Dr. Dedong Wang, Dr. Guram Kervalishvili, Dr. Chao Xiong, Dr. Maria Usanova, Dr. Michael Hartinger, Prof. Hermann Lühr, Dr. Hayley Allison, Dr. Rakesh Sarma, Dr. Adam Kellerman, Dr. Dieter Bilitza, Dr. James Weygand, Dr. Ryan McGranaghan, Dr. Steve Morley, Dr. Joe Borovsky, Dr. Vania Jordanova, Prof. Michael Liemohn, Dr. Janet Green, Dr. Anthony Saikin, Prof. János Lichtenberger, Dr. Elena Kronberg, Dr. Hui Zhu, Prof. David Jackson, Dr. Mikhail Sitnov, Dr. Xing Cao, Dr. Leonie Pick, Ingo Michaelis, Daji Landis, Artem Smirnov, Dr. Milad Asgarimehr, Dr. Bernhard Fluche, Dr. Enrico Camporeale, and I am very grateful to many other scientists, with whom I had a pleasure to interact and learn from during these four years.

I wish to also thank all my colleagues from sections 2.8 and 2.3 at GFZ, in particular, Dr. Sanja Panovska, Dr. Juan Rodriguez-Zuluaga, Dr. Leonie Pick, Dr. Robin Senftleben, Dr. Dedong Wang, Angelica Castillo, Alexander Jordan, Sandra Paulin, Melanie Burns, Dr. Frederic Effenberger, Sebastian Cervantes-Villa, Dr. Hayley Allison, Dr. Stefano Bianco, Matyas Szabo-Roberts, Dr. Foteini Vervelidou, Dr. Tarique Siddiqui, Dr. Stefanie Weege, Dr. Dominika Sörgel, Dr. Achim Morschhauser, Dr. Martin Rother, Martina Krüger, and Antje Timmermann, for making this PhD such an enjoyable experience.

My sincerest thanks go to my former advisor at MEPhI and CONT SRISA RAS, Dr. Vladimir Krizhanovsky, for igniting my interest in scientific research and for his always valuable advice and support. I am also very privileged to be a part of the first Geo.X cohort of young scientists and wish to thank Dr. Hildegard Gödde, Dr. Annett Hüttges, Lisa Rheinheimer, Dr. Manja Luzi-Helbing for their support and providing with numerous opportunities, and also all Geo.X fellows for productive and fun times that we spent together. I am also very grateful for the opportunity to be a part of the Viadrina mentoring program, and would like to thank Dr. Luana Martin-Russu, Sara Kauer-Bonin, Bettina Gebhardt, Susann Senkpiel, Prof. Christian Wenger, and Dr. Neela Enke for their support.

Finally, I would like to thank my parents for their love, support, always being there for me and believing in me. Mom, dad, you rock!

Irina Zhelavskaya
Potsdam, Germany
November 11, 2020

The results of this dissertation were funded by Geo.X, the Research Network for Geosciences in Berlin and Potsdam, under Grant No SO_087_GeoX, and by GFZ Potsdam. I would also like to acknowledge NSF (GEM), COSPAR, IUGG, and Potsdam Graduate School for the travel grants that allowed to partially cover my participation in a number of international conferences.

Peer-reviewed publications

First author publications

1. **Zhelavskaya I. S.**, N. Aseev, Y. Y. Shprits, M. Spasojević (2020). A combined neural network- and physics-based approach for modeling the plasmasphere dynamics. (*submitted to Journal of Geophysical Research: Space Physics*).

2. **Zhelavskaya I. S.**, R. Vasile, Y. Y. Shprits, C. Stolle, J. Matzka (2019). Systematic analysis of machine learning techniques for Kp prediction, *Space Weather*, 17. `https://doi.org/10.1029/2019SW002271`.

3. **Zhelavskaya I. S.**, Y. Y. Shprits and M. Spasojević (2018). Chapter 12 - Reconstruction of Plasma Electron Density From Satellite Measurements Via Artificial Neural Networks, in *Machine Learning Techniques for Space Weather*, edited by Enrico Camporeale, Simon Wing and Jay R. Johnson, Elsevier, pp. 301-327, ISBN 9780128117880, `https://doi.org/10.1016/B978-0-12-811788-0.00012-3` *(6 citations).*

4. **Zhelavskaya I. S.**, Y. Y. Shprits, M. Spasojević (2017). Empirical modeling of the plasmasphere dynamics using neural networks. *Journal of Geophysical Research: Space Physics*, 122, 11,22711,244. `https://doi.org/10.1002/2017JA024406` *(12 citations).*

5. **Zhelavskaya, I. S.**, M. Spasojevic, Y. Shprits, and W. Kurth (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, `doi:10.1002/2015JA022132` *(26 citations).*

Co-authored publications

1. Smirnov, A. G., Berrendorf, M., Shprits, Y. Y., Kronberg, E. A., Allison, H. J., Aseev, N. A., **Zhelavskaya, I.**, et al. (2020). Medium Energy Electron Flux in Earth's Outer Radiation Belt (MERLIN): A Machine Learning Model. *Space Weather*, 18, e2020SW002532. https://doi.org/10.1029/2020SW002532.

2. Jackson, D. R., Bruinsma, S., Negrin, S., Stolle, C., Budd, C. J., Gonzalez, R. D., Down, E., Griffin, D. J., Griffith, M. J., Kervalishvili, G., ... & **Zhelavskaya, I.** (2020). The Space Weather Atmosphere Models and Indices (SWAMI) Project: Overview and first results. *Journal of Space Weather and Space Climate, 10, 18.*

3. Wang, D., Shprits, Y. Y., **Zhelavskaya, I. S.**, Effenberger, F., Castillo, A., Drozdov, A. Y., ... & Cervantes, S. (2020). The Effect of Plasma Boundaries on the Dynamic Evolution of Relativistic Radiation Belt Electrons. *Journal of Geophysical Research: Space Physics, e2019JA027422.*

4. Sarma, R., Chandorkar, M., **Zhelavskaya I. S.**, Shprits, Y., Drozdov, A., & Camporeale, E. (2020). Bayesian inference of quasilinear radial diffusion parameters using Van Allen Probes, *Journal of Geophysical Research: Space Physics*, 125, e2019JA027618. `https://doi.org/10.1029/2019JA027618`.

5. Shprits Y. Y., R. Vasile, **Zhelavskaya I. S.** (2019). Now-casting and Predicting the Kp index Using Historical Values and Real-time Observations, *Space Weather*, 17. `https://doi.org/10.1029/2018SW002141`.

6. Pick, L., Effenberger, F., **Zhelavskaya, I.**, Korte, M. (2019). A statistical classifier for historical geomagnetic storm drivers derived solely from ground-based magnetic field measurements. *Earth and Space Science*, 6, 2000 2015, `https://doi.org/10.1029/2019EA000726`.

7. Asgarimehr M., **Zhelavskaya I. S.**, Foti G., Reich S., and Wickert, J. (2019). A GNSS-R Geophysical Model Function: Machine Learning for Wind Speed Retrievals, *IEEE Geoscience and Remote Sensing Letters.* `doi:10.1109/LGRS.2019.2948566`.

8. Wang D., Shprits Y. Y., **Zhelavskaya I. S.**, Agapitov O., Drozdov A., Aseev N. A. (2019). Analytical Chorus Wave Model Derived from Van Allen Probe

Observations, *Journal of Geophysical Research: Space Physics*, 124. https://doi.org/10.1029/2018JA026183.

9. Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., Tobiska W. K., Wintoft, P., Ganushkina, N. Y., Welling, D. T., Bingham, S., Balikhin, M. A., Opgenoorth H. J., Engel, M. A., Weigel, R. S., Singer, H. J., Buresova, D., Bruinsma, S., **Zhelavskaya, I. S.**, Shprits. Y. Y., Vasile, R. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, 16. https://doi.org/10.1029/2018SW002067.

10. Shprits, Y. Y., **Zhelavskaya, I. S.**, Green, J. C., Pulkkinen, A. A., Horne, R. B., Pitchford, D., O'Brien, T. P., Glover, A. (2018). Discussions on stakeholder requirements for space weather-related models. *Space Weather*, 16, 341342, https://doi.org/10.1002/2018SW001864.

11. Green, J. C., Shprits, Y. Y., Haggarty, E., Onsager, T., **Zhelavskaya, I.** (2018). Summary of stakeholder inputs on development of a satellite anomaly database from the European Space Weather Week. *Space Weather*, 16, 343344, https://doi.org/10.1029/2018SW001823.

12. A. A. Saikin, V. K. Jordanova, J. C. Zhang, C. W. Smith, H. E. Spence, B. A. Larsen, G. D. Reeves, R. B. Torbert, C. A. Kletzing, **I. S. Zhelavskaya**, Y. Y. Shprits (2018). Comparing simulated and observed EMIC wave amplitudes using in situ Van Allen Probes measurements. *Journal of Atmospheric and Solar-Terrestrial Physics*, ISSN 1364-6826, https://doi.org/10.1016/j.jastp.2018.01.024.

13. Cao X., Y. Y. Shprits, B. Ni, **I. S. Zhelavskaya** (2017). Calculation of the minimum resonant energy for interactions between radiation belt electrons and EMIC waves accounting for hot plasma effects. *Scientific reports*, 7(1), 17719.

14. Aseev, N. A., Shprits, Y. Y., Drozdov, A. Y., Kellerman, A. C., Usanova, M. E., Wang, D., **I. S. Zhelavskaya** (2017). Signatures of Ultrarelativistic Electron Loss in the Heart of the Outer Radiation Belt Measured by Van Allen Probes. *Journal of Geophysical Research: Space Physics*, *122*, 10,10210,111. https://doi.org/10.1002/2017JA024485.

15. Shprits, Y. Y., A. Drozdov, M. Spasojevic, A. Kellerman, M. E. Usanova, M. J. Engebretson, O. V. Agapitov, **I. S. Zhelavskaya**, T. J. Raita, H. E. Spence, D. N. Baker, H. Zhu, N. A. Aseev (2016). Wave-induced loss of ultra-relativistic electrons in the Van Allen radiation belts. *Nat. Commun.*, 7, 12883, doi:10.1038/ncomms12883.

Data publications

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

2. Wang, D., Shprits, Y., & **Zhelavskaya, I.** (2020). MLT-averaged Plasmapause Position calculated from the PINE Plasmasphere Model for the GEM challenge events. *GFZ Data Services.* https://doi.org/10.5880/GFZ.2.8.2020.001.

Presentations at international meetings

Solicited presentations at international meetings

1. "A combined neural network- and physics-based approach for modeling plasmasphere dynamics", **Workshop on machine Learning, data Mining and data Assimilation in Geospace (LMAG2020)**, September 21–24, 2020, online because of the pandemic.

2. "Application of Geomagnetic Indices to the Plasmasphere Modeling", **27th IUGG General Assembly**, July 8–18, 2019, Montréal, Québec, Canada.

3. "A combined neural network- and physics-based approach for modeling the plasmasphere dynamics", **The Plasmasphere and the Warm Plasma Cloak workshop**, September 18–20, 2018, Los Alamos, NM, USA.

4. "Prediction of the dynamics of the plasmasphere", **EGU 2018**, April 8–13, 2018, Vienna, Austria.

5. "Empirical modeling of the plasmasphere dynamics using neural networks", **Space Weather: A Multi-Disciplinary Approach Workshop**, Lorentz Center, September 25–29, 2017, Leiden, the Netherlands.

6. "Deriving electron density from electric field measurements on the Van Allen Probes spacecraft and building a global dynamic model of plasma density using neural networks", **IAPSO-IAMAS-IAGA 2017**, August 27 – September 1, 2017, Cape Town, South Africa.

Invited seminars and schools

1. "Study of the plasmasphere using machine learning techniques", **invited lecturer at the SPP "Dynamic Earth" Summer School on "Machine Learning in Geosciences"**, September 8–10, 2020, Neustadt/Weinstr., Germany.

2. "Modeling the plasmasphere electron density: from local Van Allen Probe observations to the global plasmasphere dynamics", **Los Alamos National Laboratory**, September 17, 2018, Los Alamos, NM, USA.

3. "Reconstruction of electron density from satellite measurements and Empirical modelling of the plasmasphere dynamics using neural networks", **Van Allen Probes SWG Telecon**, 5 January, 2018.

4. "Empirical modeling of the plasmasphere with neural networks", **POF PT I Workshop: From Atmosphere to Space Weather**, May 30, 2017, Potsdam, Germany.

Oral presentations at international meetings

1. "Reconstructing the Plasmasphere Dynamics from Geomagnetic Indices Using Neural Networks", **AGU 2019**, December 9 – 13, 2019, San Francisco, California, USA.

2. "A combined neural network- and physics-based approach for modelling the plasmasphere dynamics during extreme geomagnetic storms", **Machine Learning in Heliophysics 2019**, September 16 – 20, 2019, Amsterdam, the Netherlands.

3. "A combined neural network- and physics-based approach for modeling the plasmasphere dynamics during extreme geomagnetic events", **AGU 2018**, December 9 – 14, 2018, Washington D.C, USA. *AGU Outstanding Student Presentation Award (OSPA) for best student presentation.*

4. "Modelling the global plasmasphere dynamics using neural network- and physics-based approaches", **COSPAR 2018**, July 14–22, 2018, Pasadena, CA, USA.

5. "Systematic analysis of machine learning techniques for Kp prediction in the framework of the H2020 project SWAMI", **COSPAR 2018**, July 14–22, 2018, Pasadena, CA, USA.

6. "International standard for validating the Kp index predictive models", **COSPAR 2018**, July 14–22, 2018, Pasadena, CA, USA.

7. "Empirical modeling of the plasmasphere dynamics using neural networks", **European Space Weather Week 2017**, November 27 – December 1, 2017, Ostend, Belgium.

8. "Global dynamic evolution of the cold plasma inferred with neural networks", **DGG 2017**, March 27–30, 2017, Potsdam, Germany. *Award for best student oral presentation.*

Poster presentations at international meetings

1. "A combined neural network- and physics-based approach for modeling the plasmasphere dynamics", **EGU 2019**, April 8 – 13, 2019, Vienna, Austria.

2. "Systematic analysis of machine learning techniques for Kp prediction", **EGU 2018**, April 8–13, 2018, Vienna, Austria. *EGU Outstanding Student Poster and PICO (OSPP) award for best student poster presentation.*

3. "Empirical modeling of the plasmasphere dynamics using neural networks", **AGU 2017**, December 11–15, 2017, New Orleans, LA, USA.

4. "Empirical modeling of the plasmasphere dynamics using neural networks", **GEM 2017 Summer Workshop**, June 18–23, 2017, Portsmouth, VA, USA. *Award for best student poster presentation.*

5. "Global dynamic evolution of the cold plasma inferred with neural networks", **EGU 2017**, April 23–28, 2017, Vienna, Austria.

6. "Global dynamic evolution of the bulk plasma inferred with neural networks", **AGU 2016**, December 12–16, 2016, San Francisco, CA, USA.

7. "Automated determination of electron density from electric field measurements on the Van Allen Probes spacecraft", **AGU 2015**, December 14–18, 2015, San Francisco, CA, USA.

List of figures

List of tables

Abstract

The plasmasphere is a dynamic region of cold, dense plasma surrounding the Earth. Its shape and size are highly susceptible to variations in solar and geomagnetic conditions. Having an accurate model of plasma density in the plasmasphere is important for GNSS navigation and for predicting hazardous effects of radiation in space on spacecraft. The distribution of cold plasma and its dynamic dependence on solar wind and geomagnetic conditions remain, however, poorly quantified. Existing empirical models of plasma density tend to be oversimplified as they are based on statistical averages over static parameters. Understanding the global dynamics of the plasmasphere using observations from space remains a challenge, as existing density measurements are sparse and limited to locations where satellites can provide in-situ observations. In this dissertation, we demonstrate how such sparse electron density measurements can be used to reconstruct the global electron density distribution in the plasmasphere and capture its dynamic dependence on solar wind and geomagnetic conditions.

First, we develop an automated algorithm to determine the electron density from in-situ measurements of the electric field on the Van Allen Probes spacecraft. In particular, we design a neural network to infer the upper hybrid resonance frequency from the dynamic spectrograms obtained with the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) instrumentation suite, which is then used to calculate the electron number density. The developed Neural-network-based Upper hybrid Resonance Determination (NURD) algorithm is applied to more than four years of EMFISIS measurements to produce the publicly available electron density data set.

We utilize the obtained electron density data set to develop a new global model of plasma density by employing a neural network-based modeling approach. In addition to the location, the model takes the time history of geomagnetic indices and location

as inputs, and produces electron density in the equatorial plane as an output. It is extensively validated using in-situ density measurements from the Van Allen Probes mission, and also by comparing the predicted global evolution of the plasmasphere with the global IMAGE EUV images of He^+ distribution. The model successfully reproduces erosion of the plasmasphere on the night side as well as plume formation and evolution, and agrees well with data.

The performance of neural networks strongly depends on the availability of training data, which is limited during intervals of high geomagnetic activity. In order to provide reliable density predictions during such intervals, we can employ physics-based modeling. We develop a new approach for optimally combining the neural network- and physics-based models of the plasmasphere by means of data assimilation. The developed approach utilizes advantages of both neural network- and physics-based modeling and produces reliable global plasma density reconstructions for quiet, disturbed, and extreme geomagnetic conditions.

Finally, we extend the developed machine learning-based tools and apply them to another important problem in the field of space weather, the prediction of the geomagnetic index Kp. The Kp index is one of the most widely used indicators for space weather alerts and serves as input to various models, such as for the thermosphere, the radiation belts and the plasmasphere. It is therefore crucial to predict the Kp index accurately. Previous work in this area has mostly employed artificial neural networks to nowcast and make short-term predictions of Kp, basing their inferences on the recent history of Kp and solar wind measurements at L1. We analyze how the performance of neural networks compares to other machine learning algorithms for nowcasting and forecasting Kp for up to 12 hours ahead. Additionally, we investigate several machine learning and information theory methods for selecting the optimal inputs to a predictive model of Kp. The developed tools for feature selection can also be applied to other problems in space physics in order to reduce the input dimensionality and identify the most important drivers.

Research outlined in this dissertation clearly demonstrates that machine learning tools can be used to develop empirical models from sparse data and also can be used to understand the underlying physical processes. Combining machine learning, physics-based modeling and data assimilation allows us to develop novel methods benefiting from these different approaches.

Zusammenfassung

Die Plasmasphäre ist eine die Erde umgebende dynamische Region aus kaltem, dichtem Plasma. Ihre Form und Größe sind sehr anfällig für Schwankungen der solaren und geomagnetischen Bedingungen. Ein przises Modell der Plasmadichte in der Plasmasphäre ist wichtig für die GNSS-Navigation und für die Vorhersage gefährlicher Auswirkungen der kosmischen Strahlung auf Raumfahrzeuge. Die Verteilung des kalten Plasmas und seine dynamische Abhängigkeit vom Sonnenwind und den geomagnetischen Bedingungen sind jedoch nach wie vor nur unzureichend quantifiziert. Bestehende empirische Modelle der Plasmadichte sind in der Regel zu stark vereinfacht, da sie auf statistischen Durchschnittswerten statischer Parameter basieren. Das Verständnis der globalen Dynamik der Plasmasphäre anhand von Beobachtungen aus dem Weltraum bleibt eine Herausforderung, da vorhandene Dichtemessungen spärlich sind und sich auf Orte beschränken, an denen Satelliten In-situ-Beobachtungen liefern können. In dieser Dissertation zeigen wir, wie solche spärlichen Elektronendichtemessungen verwendet werden können, um die globale Elektronendichteverteilung in der Plasmasphäre zu rekonstruieren und ihre dynamische Abhängigkeit vom Sonnenwind und den geomagnetischen Bedingungen zu erfassen.

Zunächst entwickeln wir einen automatisierten Algorithmus zur Bestimmung der Elektronendichte aus In-situ-Messungen des elektrischen Feldes der Van Allen Probes Raumsonden. Insbesondere entwerfen wir ein neuronales Netzwerk, um die obere Hybridresonanzfrequenz aus den dynamischen Spektrogrammen abzuleiten, die wir durch die Instrumentensuite Electric and Magnetic Field Instrument Suite (EMFISIS) erhielten, welche dann zur Berechnung der Elektronenzahldichte verwendet wird. Der entwickelte Neural-network-based Upper Hybrid Resonance Determination (NURD)-Algorithmus wird auf mehr als vier Jahre der EMFISIS-Messungen angewendet, um den öffentlich verfügbaren Elektronendichte-Datensatz zu erstellen.

Wir verwenden den erhaltenen Elektronendichte-Datensatz, um ein neues globales

Modell der Plasmadichte zu entwickeln, indem wir einen auf einem neuronalen Netzwerk basierenden Modellierungsansatz verwenden. Zusätzlich zum Ort nimmt das Modell den zeitlichen Verlauf der geomagnetischen Indizes und des Ortes als Eingabe und erzeugt als Ausgabe die Elektronendichte in der äquatorialebene. Dies wird ausführlich anhand von In-situ-Dichtemessungen der Van Allen Probes-Mission und durch den Vergleich der vom Modell vorhergesagten globalen Entwicklung der Plasmasphäre mit den globalen IMAGE EUV-Bildern der He^+ -Verteilung validiert. Das Modell reproduziert erfolgreich die Erosion der Plasmasphäre auf der Nachtseite sowie die Bildung und Entwicklung von Fahnen und stimmt gut mit den Daten überein.

Die Leistung neuronaler Netze hängt stark von der Verfügbarkeit von Trainingsdaten ab, die für Intervalle hoher geomagnetischer Aktivität nur spärlich vorhanden sind. Um zuverlässige Dichtevorhersagen während solcher Intervalle zu liefern, können wir eine physikalische Modellierung verwenden. Wir entwickeln einen neuen Ansatz zur optimalen Kombination der neuronalen Netzwerk- und physikbasierenden Modelle der Plasmasphäre mittels Datenassimilation. Der entwickelte Ansatz nutzt sowohl die Vorteile neuronaler Netze als auch die physikalischen Modellierung und liefert zuverlässige Rekonstruktionen der globalen Plasmadichte für ruhige, gestörte und extreme geomagnetische Bedingungen.

Schließlich erweitern wir die entwickelten auf maschinellem Lernen basierten Werkzeuge und wenden sie auf ein weiteres wichtiges Problem im Bereich des Weltraumwetters an, die Vorhersage des geomagnetischen Index Kp. Der Kp-Index ist einer der am häufigsten verwendeten Indikatoren für Weltraumwetterwarnungen und dient als Eingabe für verschiedene Modelle, z.B. für die Thermosphäre, die Strahlungsgürtel und die Plasmasphäre. Es ist daher wichtig, den Kp-Index genau vorherzusagen. Frühere Arbeiten in diesem Bereich verwendeten hauptsächlich künstliche neuronale Netze, um Kurzzeit-Kp-Vorhersagen zu treffen, wobei deren Schlussfolgerungen auf der jüngsten Vergangenheit von Kp- und Sonnenwindmessungen am L1-Punkt beruhten. Wir analysieren, wie sich die Leistung neuronaler Netze im Vergleich zu anderen Algorithmen für maschinelles Lernen verhält, um kurz- und längerfristige Kp-Voraussagen von bis zu 12 Stunden treffen zu können. Zusätzlich untersuchen wir verschiedene Methoden des maschinellen Lernens und der Informationstheorie zur Auswahl der optimalen Eingaben für ein Vorhersagemodell von Kp. Die entwickelten Werkzeuge zur Merkmalsauswahl können auch auf andere Probleme in der Weltraumphysik angewendet werden, um die Eingabedimensionalität zu reduzieren und die wichtigsten Treiber

zu identifizieren.

Die in dieser Dissertation skizzierten Untersuchungen zeigen deutlich, dass Werkzeuge für maschinelles Lernen sowohl zur Entwicklung empirischer Modelle aus spärlichen Daten als auch zum Verstehen zugrunde liegender physikalischer Prozesse genutzt werden können. Die Kombination von maschinellem Lernen, physikbasierter Modellierung und Datenassimilation ermöglicht es uns, kombinierte Methoden zu entwickeln, die von unterschiedlichen Ansätzen profitieren.

Chapter 1

Introduction

The plasmasphere is a toroidal region of cold, dense plasma surrounding the Earth and approximately corotating with it. It is located in the inner magnetosphere and is an extension of the ionosphere covering the altitudes from ∼1000 km to ∼10,000 – 40,000 km in the equatorial plane. The dynamics and configuration of the plasmasphere are highly susceptible to variations in solar conditions and geomagnetic activity in the near-Earth space.

The plasmasphere is host to many complex interactions between charged particles and electromagnetic fields in the near-Earth space, and therefore knowledge of its global-scale dynamics is crucial for understanding the flow of mass and energy within the solar-terrestrial environment. It also plays an important role for a number of space weather related effects.

Space weather is a collective term used to describe hazardous events in the near-Earth space environment that can have an effect on humans and technology in space and can have adverse effects on the ground. In particular, having an accurate model of plasma density in the plasmasphere is crucial for the navigation of satellites and for predicting spacecraft charging.

Despite the importance of plasmaspheric modeling, previous models of plasma density tend to be simplified. They are mostly based on statistical averages, in the case of empirical models, or use simplified treatments of physical processes, in the case of physics-based models. Understanding the global dynamics of the plasmasphere using observations from space remains a challenge, as existing density measurements are sparse and limited to locations where satellites can provide in-situ observations. Nonetheless, recent advances in machine learning allow us to create more advanced

models of plasma density in the plasmasphere based on such sparse observations, that are capable of capturing the global dynamic dependence of the plasma density on geomagnetic and solar wind conditions.

In this chapter, we start by introducing the notion of space weather and providing a general overview of the Earth's magnetosphere, a cavity in the solar wind where the plasmasphere is located. We then describe basic physical processes governing the dynamics of the plasmasphere, observational techniques, and previous models of the plasmaspheric dynamics. Finally, we outline the employed methodology, concluding with an overview and major contributions of this dissertation.

1.1 Space weather background

The term "space weather" generally refers to conditions on the Sun, in the solar wind, and within the Earth's magnetosphere, ionosphere and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health (definition used by the U.S. National Space Weather Plan). Adverse conditions in the near-Earth space environment can cause the disruption of electric power distribution grids, navigation, communications, and satellite operations, leading to potentially huge socio-economic losses.

The Sun is the source of most of space weather hazards. It constantly emits large amounts of plasma, called the solar wind, buffeting the Earth's magnetosphere. The Earth's magnetosphere, a region around the Earth dominated by its magnetic field, partially shields us from this constant stream of particles, as charged particles cannot easily penetrate the lines of a magnetic field. However, some amount of energy, mass, and momentum are still transferred from the solar wind to the Earth's magnetosphere. The interaction of the solar wind with the Earth's magnetic field and the influence of the underlying atmosphere and ionosphere produce various regions inside the magnetosphere populated by different types of plasma, such as the ring current, the radiation belts, and the plasmasphere. The next section describes the magnetosphere system and its interaction with the solar wind in more detail.

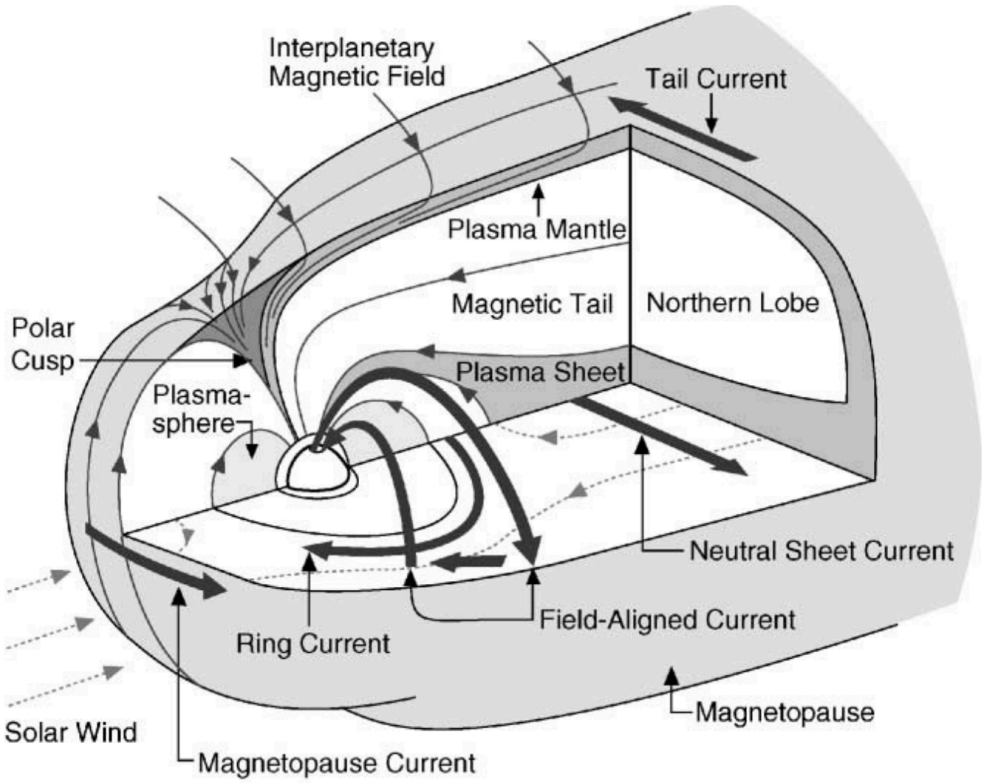

Figure 1.1: A schematic 3D view of the Earth's magnetosphere with the different plasma regions and current systems indicated (after Kivelson and Russell (1995)).

1.1.1 The magnetosphere and the Sun-Earth system

The magnetosphere is a complex system encompassing a variety of plasma populations with different temperatures and energies, a zoo of electromagnetic waves, and a number of large-scale current systems. Its configuration is shown schematically in Figure 1.1 illustrating some of these phenomena. Such a configuration of the magnetosphere is a result of the interaction between the Earth's magnetosphere and the solar wind blowing from the Sun (Cowley, 1996).

The Sun emits large volumes of plasma into interplanetary space at supersonic speeds (300 to 800 km/s), and this radially outflowing plasma is referred to as the solar wind (Parker, 1963; Neugebauer and Snyder, 1966). It is comprised mainly of protons and electrons with a small percentage ($< 5\%$) of alpha particles and heavier ions. The average electron number density of the solar wind is $\sim$5 cm^{-3} and the average energy is $\sim$10 electronvolt (eV). Due to the high conductivity of the plasma, the solar magnetic field is embedded or "frozen in" the solar wind flow and is carried

outwards from the Sun with it. This magnetic field is known as the interplanetary field (IMF) and has an average strength of ∼5 nT.

The magnetic field of the Earth (Chapman and Bartels, 1940) is believed to be generated by electric currents in the Earth's outer core caused by the convective motion of charged, molten iron below the surface in this region. The resulting field can be approximated, to first order, as a dipole with the magnetic axis tilted ∼11° from the Earth's rotation axis. The magnetic field is directed down towards the surface of the Earth in the northern hemisphere and away from the Earth in the southern hemisphere. The average strength of the magnetic field on the surface of the Earth is approximately 50 μT.

The incoming solar wind and IMF distort the approximately dipolar shape of the geomagnetic field. When reaching the magnetosphere, the solar wind is not able to easily penetrate the dipole-like geomagnetic field but is instead slowed down and deflected around it (Chapman and Ferraro, 1931). The kinetic pressure of the solar wind deforms the outer part of the Earth's magnetic field, compressing the field on the dayside and stretching it on the nightside, which results in the characteristic bullet shape of the magnetosphere. The boundary between the region where the Earth's magnetic field is dominant and the IMF is dominant is called the magnetopause. The dayside magnetopause boundary is typically located at a distance of 8 – 10 Earth radii at the geomagnetic equator (1 $R_{\mathrm{E}} = 6378$ km), whereas downstream from the Earth, the magnetotail formed by the solar wind typically extends for several hundred R_{E} on the nightside. Near the Earth (at radial distances of less than ∼3 – 4 R_{E}), however, the dipolar shape of the geomagnetic field remains relatively unaffected by the solar wind flow.

The Dungey cycle

The dipolar configuration of the terrestrial magnetic field can be further distorted by a process known as magnetic reconnection or magnetic merging (Dungey, 1961). The process of magnetic field line reconnection occurs when the interplanetary and terrestrial magnetic field lines have opposing directions, i.e., since the terrestrial field line has a northward direction on the dayside magnetopause, the IMF line should be directed southwards. This process is schematically shown in Figure 1.2, where the IMF and terrestrial field lines are labeled 1' and 1, respectively. When the two lines meet at the dayside magnetopause, they merge or reconnect in a small diffusion

region. After merging, the two field lines which previously were closed, i.e., with both ends connected to the Earth, split into two open field lines (labeled 2 and 2'), having one footprint on the Earth and the other footprint stretching out into the solar wind and ultimately reaching back to the Sun. The solar wind drags the merged field lines across the polar cap (the region in each hemisphere containing the footprints of the open field lines) down the magnetotail (labeled 3 – 5). Continual loading of magnetic flux into the magnetotail increases the magnetic pressure, forcing the open field lines attached to different hemispheres to reconnect again in a diffusion region located at a distance of ∼100 to 200 R_E downtail, reforming closed field lines (labeled 6). The stretched tail field line begins to move Earthward (labeled 7 – 8) due to magnetic tension and pressure gradients, and is eventually brought back to the dayside magnetosphere (labeled 9), where the whole process repeats again. This large scale motion of the magnetic field and plasma is known as the Dungey cycle and is most important for various processes and plasma populations in the magnetosphere (Dungey, 1961).

The basic configuration of the magnetic field resulting from the magnetic field line reconnection remains the same when the IMF is oriented in an arbitrary direction, except for when it is directed purely northward, in which case no open field lines exist (Dungey, 1963; Stern, 1973). The magnetospheric sunward flow of plasma (also called return flow) is referred to as global magnetospheric convection. This flow generates an electric field known as the convection electric field that will be discussed in more detail in section 1.2.4.

Coordinate systems

When describing various magnetospheric phenomena, it is convenient to employ a coordinate system developed around the geometry of the Earth's magnetic field. The focus of this dissertation is on the plasmasphere, the region of the inner magnetosphere, which can be roughly defined as a region lying inside the geosynchronous orbit, located at a geocentric equatorial distance of 6.6 R_E. We assume, throughout this dissertation, that the terrestrial magnetic field in this region can be approximated by a centered tilted dipole, with the dipole axis tilted by 11° with respect to the geographic rotation axis.

Often, the solar magnetic (SM) coordinate system is used when describing phenomena in the inner magnetosphere (Kivelson and Russell, 1995). The SM coordinate

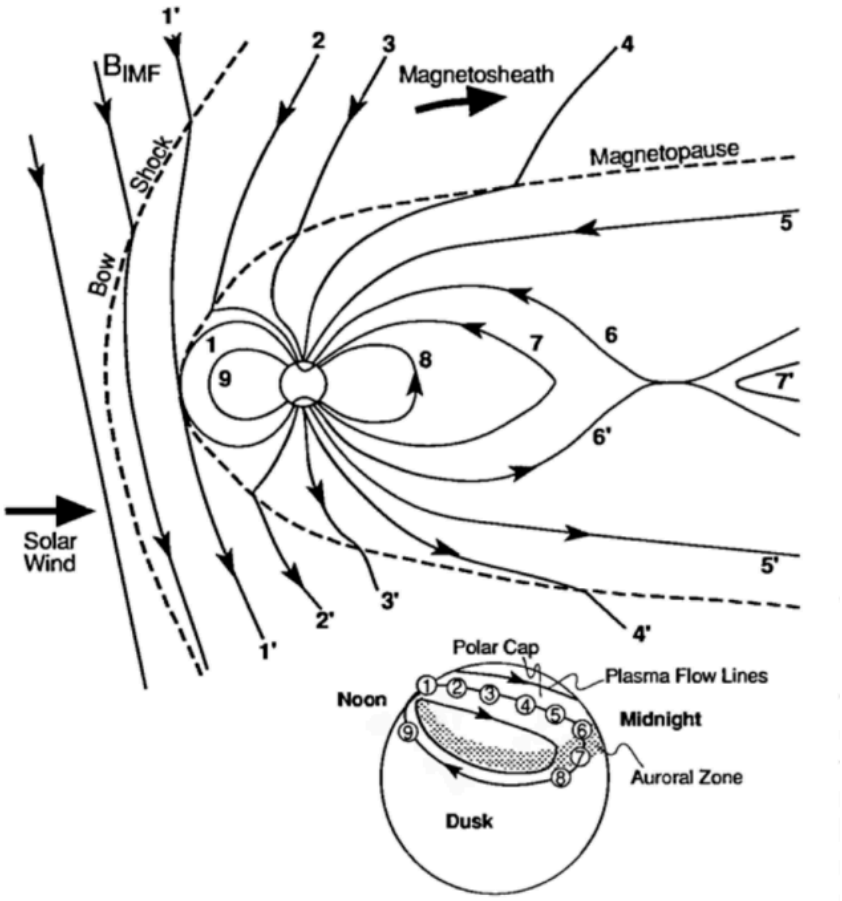

Figure 1.2: Schematic illustration of magnetic field line reconnection and the corresponding flow of plasma during periods of southward interplanetary magnetic field. The inset shows the movement of the feet of the field lines in the northern high-latitude ionosphere and the corresponding high-latitude plasma flows: an anti-sunward flow in the polar cap and a return flow at lower latitudes (after Kivelson and Russell (1995)).

system is a right-handed Cartesian coordinate system, in which the z-axis coincides with the geomagnetic dipole axis, the y-axis is perpendicular to the Earth-Sun line and directed towards dusk, and the x-axis completes the right-handed set. The SM system rotates with both a yearly and daily period with respect to inertial coordinates. The x-axis does not point directly towards the sun, and thus the angle that it makes with respect to the ecliptic plane changes by 22° over a 24-hour period.

Since magnetospheric plasma is strongly ordered by the magnetic field, in practice, it is more convenient to use another coordinate system with the following three coordinates: L shell (or L value, L), magnetic latitude (λ), and magnetic local time (MLT). The L shell parameter was proposed by McIlwain (1986) in order to organize energetic charged particle data in a realistic magnetic field. Its definition is greatly simplified within the centered dipole approximation, and it describes the distance at which a field line crosses the magnetic equatorial plane, in units of Earth radii. For example, all field lines that cross the geomagnetic equator at 4 R_{E} from the center of the Earth are described by $L = 4$. The magnetic latitude λ is measured from the equatorial plane and describes the position along a given L shell ranging from -90 to 90°. Finally, MLT, defines the position of the field line with respect to the direction of the sun and is measured in decimal hours, where an MLT of 00 is in the anti-sunward direction (i.e., magnetic midnight), 06 corresponds to dawn, 12 to noon, and 18 to dusk. The equation of a field line in a dipole magnetic field (Walt, 1994) can be expressed as:

$$r = L\cos(\lambda)^2. \tag{1.1}$$

The magnetospheric plasma regions

Several distinct populations of plasmas exist within the magnetosphere (Figure 1.1). Some of these populations directly affect the plasmasphere, such as the ionosphere that serves as its source of plasma, and the dynamics of other populations in turn depend on the dynamics of the plasmasphere, such as the radiation belt and the ring current. These plasma populations are introduced below.

Ionosphere

The ionosphere is the base of the Earth's plasma environment. It is located at altitudes between 70 to 1000 km above the Earth and consists of molecules from the Earth's neutral atmosphere that become partially ionized by solar ultraviolet radiation (Appleton and Barnett, 1925; Bilitza and Reinisch, 2008; Bauer, 2012). Another

source of ionization, although not as significant as the UV radiation, comes from solar and galactic cosmic rays and also energetic particles from the magnetosphere that precipitate into the neutral upper atmosphere. This secondary ionization is important at high latitudes and at night, when photoionization from the Sun halts. The ionosphere, together with the solar wind, serves as a source for all other plasma populations in the magnetosphere.

The energy of constituent electrons and ions of the ionosphere is generally less than several tenths of an electronvolt. The ionospheric electron density peaks in the F-region of the ionosphere at an altitude of ∼300 km with density of $10^5 - 10^6$ cm^{-3} (depending on the time of the day), which is also the region with the highest density in the whole magnetosphere. The electron density gradually decreases with altitude and blends into the plasmasphere at ∼1000 km. The transition between the topside ionosphere and the plasmasphere is characterized by the transition from oxygen as the dominant ion in the ionosphere to hydrogen as the dominant ion in the plasmasphere.

Unlike other parts of the magnetosphere, the ionosphere is only partially ionized and also contains neutral air molecules. The density of the latter ones is relatively high leading to collisions between neutral and charged particles. In other parts of the magnetosphere, collisions between particles do not occur frequently, and therefore the plasmas there are treated as collisionless.

Plasmasphere

The plasmasphere is a torus of cold plasma surrounding the Earth and corotating with it. The source of the plasmaspheric particles is the ionosphere, and the outflow of ions and electrons from the ionosphere along mid- and low-latitude field lines populates the plasmasphere (e.g., Lemaire, 1989). H^+ is the dominant ion in the plasmasphere (∼80%), followed by He^+ (∼10 – 20%), O^+ (∼5 – 10%) and trace amounts of heavier ions, with an equal number of electrons preserving the charge neutrality (Comfort et al., 1988). Temperatures of ions and electrons in the plasmasphere are on the order of 1 eV, which is on average approximately three times higher than in the ionosphere, likely due to heating by plasma waves or Coulomb collisions with photoelectrons slipping away from the ionosphere (Comfort, 1986; Newberry et al., 1989).

Typical densities in the plasmasphere are on the order of $\sim 10^3$ electrons per cm^3, which are quite high compared with other regions of the magnetosphere. There is a sharp drop in density at radial distances ranging from 2 to 6 Earth radii in the

equatorial plane, depending on the level of geomagnetic activity, which is known as the plasmapause. The plasmapause is an approximately field-aligned surface at the outer boundary of the plasmasphere and the density beyond the plasmapause is typically on the order of $\sim$10 cm^{-3}. This region of low-density cold plasma is usually referred to as the plasmatrough (Kivelson and Russell, 1995).

Bulk motion of plasma in the plasmasphere is mainly driven by two flow regimes: convection and corotation. These flow regimes are caused by the corresponding electric fields, which are discussed in detail in section 1.2.4. The corotation of plasma with the Earth is caused by the Earth's rotation. This regime dominates during quiet geomagnetic conditions. The convection flow is caused by the magnetic reconnection of the IMF with the terrestrial magnetic field and dominates during periods of elevated geomagnetic activity. The plasma is stripped away from the plasmasphere and is transported sunwards by convection, often creating a filament between the plasmasphere and the magnetosphere known as a plasmaspheric plume (section 1.2.4); the plasmasphere shrinks due to the loss of plasma and compression. Once the geomagnetic activity decreases and the convection electric field subsides, the plasmasphere expands as it is being refilled from the ionosphere.

Radiation belts and ring current

The radiation belts (Van Allen and Frank, 1959; Vernov and Chudakov, 1960) are energetic charged particles magnetically trapped in the terrestrial magnetic field. They reside on magnetic field lines in the range of $L = 2$ to 7, and extend in energy from several 100 keV up to 10 MeV for electrons and up to 500 MeV for protons. These energetic particles undergo a helical gyro-motion around field lines, a bounce motion between the hemispheres, and a slow drift motion around the Earth (Walt, 1994). Typically, the radiation belt particles exhibit a two-belt structure comprised of an inner belt and an outer belt, separated by a local minimum of particle flux known as the slot region. The inner belt extends from $\sim$1.2 to 2.5 R_{E} and is mainly composed of energetic protons produced by cosmic ray albedo neutron decay (CRAND) and trapped solar protons (Selesnick et al., 2007, 2014). Electrons are also found in the inner belt region (e.g., Fennell et al., 2015); they originate mostly from injections, as the CRAND process is not considered to be a significant source of trapped low-energy electrons (Selesnick, 2015). The outer radiation belt extends from $\sim$3 to $\sim$8 R_{E} and is mostly composed of energetic electrons. It is a highly dynamic region, compared to the inner belt, and depends strongly on the changing geomagnetic conditions and

associated processes (Shprits et al., 2008a,b).

As opposed to the cold, low-energy ionospheric and plasmaspheric plasmas, which are sensitive mostly to electric field-driven drifts, such as $\mathbf{E} \times \mathbf{B}$ drift (section 1.2.4), the hot, high-energy particles of the radiation belts experience a relatively slow longitudinal gradient and curvature drifts due to the spatial variations of the terrestrial magnetic field closer to the Earth. Electrons drift eastwards around the Earth, while ions drift westwards, which results in a net westward current known as the ring current. This opposing motion of ions and electrons also generates a magnetic field, as prescribed by Ampère's law, which tends to oppose the intrinsic geomagnetic field close to the equator. As the strength of the ring current increases, the terrestrial field is depressed more near the equator (Dessler and Parker, 1959; Sckopke, 1966). The ring current is located in the region where electrons and ions may complete closed magnetic drifts around the Earth and, therefore, extends over geocentric distances between $\sim$2 and 9 R_{E} (Daglis et al., 1999). All trapped particles in the inner magnetosphere contribute to the ring current, however, ions in the medium-energy range of 10 keV to a few hundred of keV contribute most significantly to the total current density (Williams, 1987). The contribution of electrons to the net current is usually small due to their negligible energy density (Baumjohann, 1993; Daglis et al., 1999), however, during storm times it can reach up to 25% (Liu et al., 2005).

Plasma sheet

The plasma sheet is a sheet-like region of plasma located in the magnetotail in the nightside magnetosphere. It occupies the region of closed field lines near the equatorial plane and separates the northern and southern magnetosphere lobes. The plasma sheet consists of both ions and electrons with energies on the order of a few keV and number densities on the order of $\sim 0.1 - 1$ cm^{-3}. The material of the plasma sheet originates from the solar wind, through the magnetic reconnection in the tail, and from the ionosphere, via the outflow in the polar cap regions. Reconnection between the IMF and the terrestrial magnetic field governs the behavior of plasma sheet particles (Knipp et al., 2011). During storms and substorms, disturbances in the plasma sheet cause its particles to travel along the magnetic field lines and collide with the neutral atmospheric particles producing the visible aurora (Chamberlain, 1961).

1.1.2 Geomagnetic disturbances

Geomagnetic disturbances arise from the interaction of the magnetosphere with the incoming solar wind. The solar wind carries a large amount of energy, mass and momentum, some of which is transferred to the magnetosphere during the process of magnetic reconnection (section 1.1.1). The most efficient energy transfer between the solar wind and the magnetosphere happens when the direction of the interplanetary magnetic field embedded in the solar wind is opposite to the Earth's magnetic field (Burton et al., 1975), i.e., directed southward. The southwardly directed vertical component of the IMF ($B_z < 0$) is known to be well correlated with the level of geomagnetic activity (Murayama, 1982).

The duration of geomagnetic disturbances can span from hours, such as in the case of substorms (Akasofu, 1964), to days, as in the case of large geomagnetic storms (Lui et al., 1987). These disturbances generally cause changes in the magnetic field strength measured at the surface of the Earth, resulting from enhancements in magnetospheric and ionospheric current systems. Another way the geomagnetic disturbance manifests itself is through intensification and expansion to lower latitudes of the visible aurora.

Geomagnetic storms typically have three phases that can be observed on the ground, in measurements of the horizontal magnetic field component at the Earth's surface close to the equator. During the first, initial phase, the solar wind compresses the magnetosphere, subsequently strengthening the magnetic field on the dayside and causing the magnetic disturbance on the ground to be positive. Then, in the main phase of the storm, the ring current becomes enhanced as plasma sheet particles are carried Earthward and become trapped in the inner magnetosphere. The ring current produces a magnetic field that is directed opposite to the Earth's magnetic field. Thus, the total field at the Earth's surface decreases, and the resulting magnetic disturbance is negative. The main phase is triggered by a sustained period of southward IMF, allowing continual substorm activity to energize the ring current. The last phase, referred to as the recovery phase, happens when the IMF weakens or turns northward, resulting in ring current particles being lost faster than they are supplied. As the ring current decays, the depression of the geomagnetic field subsides, and geomagnetic conditions gradually return to the pre-storm state (this happens on timescales on the order of several days).

Variations in the horizontal component of the magnetic field are monitored with

magnetometers at ground-based magnetic observatories located at mid- and equatorial latitudes (Dessler and Parker, 1959). These measurements are then used to construct the disturbance time (Dst) index (Sugiura et al., 1964), which provides a measure of the global strength of the ring current. Specifically, Dst provides an estimate of the total energy content of the ring current particles and, therefore, can be used to monitor geomagnetic storms. Other frequently used geomagnetic indices include the AE index (Davis and Sugiura, 1966), which measures changes in the auroral electrojet current system in the ionosphere and thus indicates the substorm activity, and the Kp index (Bartels, 1949), a more global index that includes contributions from several current systems. The Kp index is considered to be a particularly good proxy for the magnetospheric convection (Thomsen, 2004) and is used as an input in a number of models of the global convection electric field (e.g., Grebowsky, 1970; Volland, 1973; McIlwain, 1986; Maynard and Chen, 1975; Chen and Grebowsky, 1974). The Kp index is important for plasmasphere modeling (e.g., Carpenter and Stone, 1967; Gallagher et al., 1998; O'Brien and Moldwin, 2003) and for many other applications in space weather, and therefore we describe it below in more detail.

The Kp index

The Kp index (Bartels, 1949) is a measure of the general level of geomagnetic activity and is widely considered as a proxy for the energy input from the solar wind to Earth and a good measure of the magnetospheric convection (Thomsen (2004) and references therein). It is used as an input to many scientific applications, including the parameterization of ionospheric ion outflow (Yau et al., 2011) and aurora particle precipitation (Emery et al., 2008) in the ionosphere, thermosphere (Bruinsma et al., 2018), hot plasma particle density (Korth et al., 1999; Denton et al., 2016), cold plasma density in the plasmasphere (Maynard and Chen, 1975; Pierrard et al., 2009; Goldstein et al., 2014), plasmapause location (Carpenter and Anderson, 1992; O'Brien and Moldwin, 2003), and radiation belt models and wave parameterizations (Brautigam and Albert, 2000; Shprits et al., 2007b; Orlova et al., 2014; Ozeke et al., 2014; Agapitov et al., 2015) in magnetospheric physics, among others.

Kp is derived from magnetic field measurements at 13 mid-latitude magnetic observatories. To derive Kp, first, the maximum fluctuations of the horizontal component of the geomagnetic field, H, for a given 3-hour interval are determined. After the quiet-day variation pattern is subtracted from them, they are transformed to the

local K index (Bartels et al., 1939), using look-up tables specific to each observatory. They take values from 0 to 9 in discrete intervals (in the form 0o , 0+, 1-, ... , 9-, 9o) according to a quasi-logarithmic scale and describe the local disturbance level. These local K indices are then averaged to obtain the 3-hour global Kp index.

Kp does not attempt to separate the effects of different current systems affecting its measurements and is a global index. It is used by a number of space weather agencies to decide whether geomagnetic alerts and warnings need to be issued for users who may be affected by these disturbances (such as electrical power grid, spacecraft operators, etc.). Geomagnetic storms are indicated by Kp ≥ 5.

Effects of geomagnetic activity on the plasmasphere

Geomagnetic disturbances strongly affect the configuration of the plasmasphere. The interplay between the two large-scale plasma flow regimes in the inner magnetosphere, magnetospheric convection driven by the solar wind and corotation driven by the rotation of the Earth, control the dynamics of the plasmasphere. During the periods of enhanced geomagnetic activity, the plasmasphere shrinks (e.g., Carpenter and Anderson, 1992) and may become eroded down to $L = 2$ during strong storms. During recovery times, it slowly expands as it is being refilled from the ionosphere and can expand globally beyond $L = 6$ after several days of low geomagnetic activity. These processes and the dynamics of the plasmasphere are described in detail in the next section.

1.2 The plasmasphere

1.2.1 Geophysical importance of the plasmasphere

The plasmasphere hosts a variety of complex interactions between charged particles and electromagnetic fields in the magnetosphere, and is important for a number of physical processes. Its size and shape control the generation and propagation of plasma waves and influence the interactions of these waves with energetic ions and electrons, thus greatly affecting distributions of these particles for a wide range of energies (e.g., Spasojević et al., 2004; Horne et al., 2005; Shprits et al., 2016; Orlova et al., 2016). Wave-particle interactions play a significant role in the loss of energetic particles (via precipitation into the upper atmosphere), and can also contribute to

the acceleration of radiation belt electrons during geomagnetic storms (e.g., Horne, 2002; Kozyra et al., 1997).

The current configuration of the plasmasphere provides an indication of the recent time history of magnetospheric convection (although in a complex integral sense), an important global mechanism within the magnetosphere. Furthermore, the plasmasphere contains a relatively large amount of mass, and a substantial amount of that mass is removed during geomagnetic storms (Spasojević and Sandel, 2010; Goldstein et al., 2019). Some of this plasma can be lost to the solar wind or ionosphere, and some of it is energized and redistributed within the magnetosphere. In fact, the plasmaspheric material that is eroded during periods of strong magnetospheric convection is carried sunward and is regularly observed near the dayside magnetopause (e.g., Chen and Moore, 2006; Lee et al., 2016). The enhanced plasma density at the dayside magnetopause can limit the rate of reconnection, thus affecting the global convection pattern (e.g., André et al., 2016; Borovsky and Denton, 2006).

Finally, the plasma density in the plasmasphere is a crucial parameter for a variety of applications in the field of space weather, such as spacecraft anomaly analysis due to spacecraft charging (e.g., Reeves et al., 2013) and GPS navigation (e.g., Mazzella, 2009; Xiong et al., 2016). Therefore, knowledge of the global-scale plasmasphere dynamics and ability to model it accurately is very important in order to better understand and reliably predict the processes mentioned above.

1.2.2 The discovery of the plasmasphere and early studies

The plasmasphere was discovered in the early 1960s independently, but approximately simultaneously, by two scientists: Konstantin Gringauz from the USSR (Gringauz, 1963) and Don Carpenter from the USA (Carpenter, 1963). The discovery was made using different measurement techniques of plasma density.

Gringauz observed an abrupt drop in the magnetospheric plasma density distribution at an altitude of approximately 4 R_{E} in the in-situ measurements of thermal ($\sim$1 eV) ions from particle detectors onboard the LUNIK 1 and 2 spacecraft. Those were two Soviet lunar probes that were launched to the Moon in 1959. These observations led Gringauz to hypothesize the existence of an upward extension of the ionosphere, what we now know as the plasmasphere. It is worth noting that until the late 1950s, the traditional methods of probing the upper atmosphere were limited to an altitude

of ∼1000 km (Lemaire and Gringauz, 1998), and the notion of the magnetosphere did not even exist yet. It was expected that the Earth's atmosphere was in direct equilibrium with the interplanetary medium extending from the Sun. Therefore, at that time, the discovery of Gringauz was difficult to comprehend and accept as a physical reality, and it was met with utter skepticism in the Soviet Union.

At approximately the same time, in 1963, Carpenter observed a sharp drop in the equatorial plasma density distribution in the ground-based observations of plasma waves called "whistlers" at an altitude of $\sim 3 - 5\ R_{\mathrm{E}}$ (Figure 1.3). The whistler-based analysis and measurement technique were pioneered by L. R. O. Storey in his PhD research at Cambridge University (Storey, 1953) and are described in more detail in section 1.2.3. Storey found that the electron concentration was approximately 400 cm^{-3} at a distance of 12,000 km from the Earth. Storey (1958) hypothesized, following the theoretical work of Dungey (1954, 1955), that this plasma was an upward extension of the ionosphere composed of protons. In later years, Gringauz suggested the name "geocorona" for this region (Gringauz et al., 1961).

Carpenter built up on the research of Storey and continued the whistler-based analysis. Carpenter called the sharp drop that he observed in the equatorial electron density profiles a "whistler knee", to which in 1966, he gave the name "plasmapause". The term "geocorona", proposed by Gringauz, was replaced by the term "plasmasphere", by analogy with the term "magnetosphere" (Gold, 1959). Carpenter's original analysis also revealed that the plasmapause tends to shrink during periods of increased geomagnetic activity (Carpenter, 1963). Finally, in 1963, Carpenter and Gringauz met personally for the first time at the XIVth URSI Assembly in Tokyo, where they were able to compare and confirm the results of each other.

By the early 1970s, satellite measurements had once and for all confirmed the existence of the plasmasphere. Using the whistler observations Carpenter (1966) showed that the plasmasphere not only shrank but also became asymmetric with a bulge in the dusk sector during periods of geomagnetic activity, while it expanded and appeared to be more circular during quiet times. Later those findings were confirmed by satellite measurements (Chappell et al., 1971). It was also observed that although the plasmasphere seemed to corotate with the Earth in the local time sector from 00 to 17 MLT (i.e., moved at the same angular velocity as the Earth), its motion was mixed and less clear in the dusk to midnight local time sectors. It was later demonstrated by Grebowsky (1970) in his numerical modeling that there appeared to be a

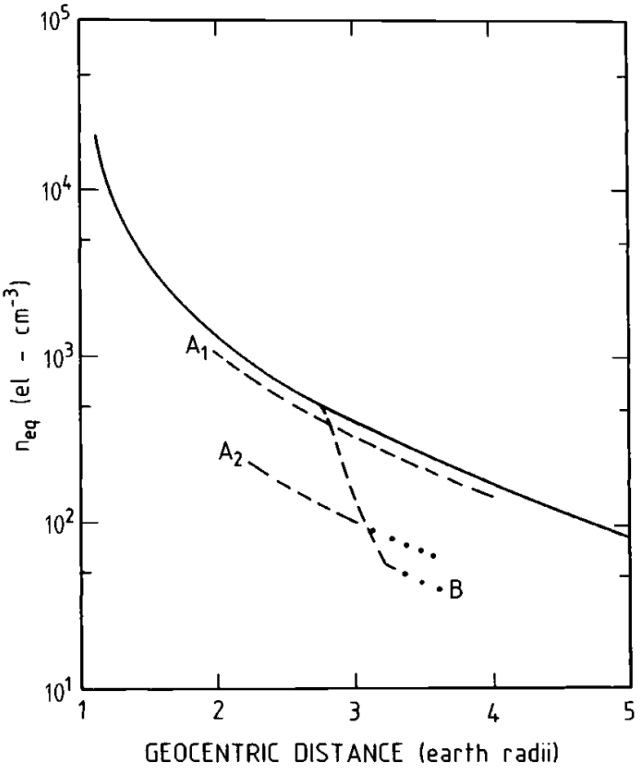

Figure 1.3: Initial diagram by Carpenter showing a quiet day profile of equatorial electron density (solid curve) and various storm-time forms of the density profile observed using whistler measurements (dashed curves) (after Carpenter (1962)).

tail of cold plasma in the dusk local time sector extending all the way to the dayside magnetopause, caused by an enhancement in the magnetospheric convection. In-situ observations of plasma density from different spacecraft revealed the high-density regions of "detached plasma" that seemed to be separated from the main body of the plasmasphere by regions of low-density plasma. These observations appeared to be consistent with plumes connected to the main body of the plasmasphere as shown by the calculations of Grebowsky (1970) and Chen and Wolf (1972). However, at that time there was still no definitive observational evidence to distinguish detached plasma regions from plumes. The existence of plumes connected to the main body of the plasmasphere was directly confirmed in the early 2000s, using the first global images of the plasmasphere made with the Extreme Ultraviolet Imager (EUV) instrument onboard the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE) mission (Burch et al., 2001a; Sandel et al., 2001).

1.2.3 Observational techniques

A number of measurement techniques to observe the plasmasphere dynamics exist. These techniques can be divided into three main categories: in-situ density measurements from satellites, density derived from ground-based observations, and global imaging of the plasmasphere. These techniques have been successfully used for many decades to study the plasmasphere and plasmapause region.

Satellites typically provide high-cadence measurements of the electron number density along the orbit of the satellite. There are numerous ways of how electron density can be measured on satellites, including direct methods, such as particle detectors or Langmuir probes, and indirect measurement techniques. The latter include quasi-thermal noise spectroscopy (Meyer-Vernet and Perche, 1989), employing the spacecraft potential as a proxy for the electron density (Escoubet et al., 1997; Li et al., 2010), or determining the electron density from intense upper-hybrid band emissions (Mosier et al., 1973). In-situ density measurements are useful for studying density structures of various scale sizes (e.g., Moldwin et al., 1995; Carpenter et al., 2000). However, in-situ density measurements are usually quite sparse due to the typically long orbital periods of high altitude satellites (> 9 hours). Therefore, it is difficult to obtain insight into the global dynamics of the plasmasphere using such sparse measurements. The large data sets produced by satellite missions have been

employed to develop empirical models of the plasmasphere density and the plasmapause location. Examples of such models include the widely used Carpenter and Anderson (1992) model based partially on the measurements from the International Sun-Earth Explorer (ISEE-1) mission (in addition to whistler measurements), and models based on measurements from the Combined Release and Radiation Effects Satellite (CRRES) mission (Sheeley et al., 2001; Moldwin et al., 2002).

Electron density can also be derived from ground-based observations of plasma waves called "whistlers". Whistlers are natural radio waves generated by lightning flashes. They propagate along the Earth's magnetic field lines from one hemisphere to the other through the dispersive plasma environment of the magnetosphere (Helliwell, 1965). Whistler waves are observable on the ground and in space and typically occur at a frequency range of 1 to 30 kHz. This range overlaps with audio frequencies, and whistlers can therefore be converted to audio using an appropriate receiver. They usually produce a gliding high-to-low-frequency sound, or whistle, which occurs because the reflected high-frequency waves arrive at the receiver earlier than the lower-pitched signals. Whistlers propagate along the so-called "ducts", the discrete magnetic field-aligned paths, which physically are believed to have the form of field-aligned columns of enhanced ionization (density enhancements) (Smith and Carpenter, 1961). A typical lightning flash generates a number of discrete field-aligned paths that represent a range of L values. The dispersion properties of whistlers depend mainly on electron density and magnetic-field strength along the ducts. Information on the electron density distribution in the inner magnetosphere can therefore be extracted from the analysis of the dispersion characteristics of whistler signals (Carpenter and Smith, 1964; Park, 1972). Many of the whistler components exhibit a frequency of minimum travel time, the so-called "nose" frequency, that is proportional to the minimum value of the magnetic field along their path. Generally, the nose frequency indicates the approximate L value of the path, and travel time at that frequency provides a measure of the integrated electron density along the path (Smith and Carpenter, 1961). The travel time of a whistler at a given frequency is weighted inversely with the magnetic field strength. Thus the dispersion properties of whistlers are particularly sensitive to conditions along the outer, higher-altitude part of their paths (i.e., near the equator), where variations in the plasma parameters per unit distance along the field lines are minimal. As a result, whistlers provide a measure of the electron density near the magnetic equator that is relatively insensitive to the functional form of the plasma

distribution along the field lines used in the calculations (Angerami and Carpenter, 1966). Thus, multi-component whistler measurements can be used to construct instantaneous multipoint equatorial profiles of the electron density. Long observation periods allow monitoring the changes in equatorial density and total flux tube content, and as such, whistler measurements provide capabilities to determine cross-L motions of bulk plasma in a frame of reference rotating with the Earth (Carpenter and Smith, 2001). Modern whistler research includes efforts on the automatization of whistlers detection. In particular, Lichtenberger et al. (2008) developed an automated algorithm for whistler detection: the Automatic Whistler Detector and Analyzer (AWDA) system. This system was installed at multiple ground stations of the AWDANet network for plasma density monitoring (Lichtenberger et al., 2010, 2011).

Both satellite and ground-based observations provide local, typically sparse density measurements within the plasmasphere and cannot provide a global perspective of the plasmasphere. With the launch of the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE) mission in 2000 (Burch, 2000), another source for observing the plasmasphere from a global perspective has become available. The Extreme Ultraviolet Imager (EUV) (Sandel et al., 2000) onboard the IMAGE satellite was designed to study the plasmasphere by imaging the distribution of He^+ in the Earth's plasmasphere. Global images of the He^+ distribution were produced by detecting resonantly scattered solar 30.4-nm radiation. UV radiation from the Sun is resonantly absorbed by ions and atoms in the near-Earth space environment and then re-radiated by them, generating a glow, which can be imaged. He^+ ions, making up a fraction of $10-15\%$ of the plasmasphere, are directly detected by imaging of extreme ultraviolet radiation at 30.4 nm. The rest of the plasmasphere is dominated by H^+, which has no optical emissions. The IMAGE spacecraft operated in a highly elliptical polar orbit with a period of $\sim$14 hours. Its initial apogee was 8.22 R_{E} in geocentric radial distance in the northern hemisphere and its initial perigee was at the altitude of 1000 km in the southern hemisphere. The spacecraft spun with a 2-minute period, and its spin axis was perpendicular to the orbital plane. The EUV instrument consisted of three identical sensor heads with a 30° conical field of view. The sensor heads were tilted by 27° relative to each other to cover a fan-shaped field of view of $84° \times 30°$. As the satellite spun, the EUV instrument captured an $84° \times 360°$ swath across the sky. It integrated the intensity of 30.4-nm emissions over 5 spins of the satellite. To form one full image, the overlapping fields of view from 3 camera heads were merged every

10 minutes. EUV collected images for approximately $7-9$ hours out of each 14-hour orbit. The spatial resolution of EUV was $\sim$0.1 R_{E} or $\sim$650 km in the equatorial plane as seen from apogee ($\sim$8 R_{E}). Such unprecedented measurement capabilities made it possible for the first time to study the plasmasphere system from a global perspective.

1.2.4 Dynamics of the plasmasphere

This section describes the basic dynamics of the plasmasphere, starting with plasma flow regimes governing the bulk plasma motion in the plasmasphere and electric fields associated with them, and concluding with specific processes pertaining to the dynamics of the plasmasphere.

$\mathbf{E} \times \mathbf{B}$ drift

Before describing the dynamics of the plasmasphere in more detail, it is useful to review basic forces that act on charged particles in electric and magnetic fields and the $\mathbf{E} \times \mathbf{B}$ drift that is experienced by charged particles in the presence of electric and magnetic fields. We will consider $\mathbf{E} \times \mathbf{B}$ drift in a non-relativistic case.

The motion of a charged particle in the electric and magnetic field is governed by the action of Coulomb and Lorentz forces. The Coulomb force $\mathbf{F}_{\mathrm{C}}$ is given by:

$$\mathbf{F}_{\mathrm{C}} = q\mathbf{E} \tag{1.2}$$

where q is the charge of the particle, and $\mathbf{E}$ is the electric field. It can be seen from this equation that the charged particle is accelerated along electric fields. A positively charged particle is accelerated in the direction of the electric field, and a negatively charged particle in the opposite direction. Charged particles themselves are the source of an electric field, which points away from protons and points towards electrons. This causes opposite charges to attract, and charges of the same sign to repel each other.

The force experienced by a moving charged particle due to a magnetic field is called the Lorentz force and is given by:

$$\mathbf{F}_{\mathrm{L}} = q(\mathbf{v} \times \mathbf{B}) \tag{1.3}$$

where $\mathbf{v}$ is the velocity of a particle, and $\mathbf{B}$ is the magnetic field. Here, the charged particle is accelerated in the direction perpendicular to both its direction of motion and the magnetic field, and thus, no work is done. This causes a particle to gyrate around a magnetic field line. Furthermore, eq. (1.3) depends on the charge of the particle and therefore results in the clockwise gyration of positive ions, and the anti-clockwise gyration of electrons. The centre of gyration of a particle is called the guiding centre. A particle moving completely parallel to $\mathbf{B}$ will not experience the Lorentz force. If a particle has velocity components both parallel and perpendicular to the magnetic field, it will move along a helical trajectory.

When a particle is placed into both uniform electric and magnetic fields, it experiences both the Coulomb and Lorentz forces. Its equation of motion then becomes a sum of these forces and is called the Lorentz equation:

$$m\frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{1.4}$$

It can be shown from the Lorentz equation that charged particles drift in the presence of uniform electric and magnetic fields (e.g., Northrop, 1963). The drift velocity of the guiding centre of a particle takes the following form:

$$\mathbf{v}_{\mathbf{E}\times\mathbf{B}} = \frac{\mathbf{E} \times \mathbf{B}}{B^2} \tag{1.5}$$

This drift is referred to as $\mathbf{E} \times \mathbf{B}$ drift. $\mathbf{E} \times \mathbf{B}$ drift does not depend on the charge of a particle, and therefore electrons and positive ions drift in the same direction, resulting in no net electric current.

The $\mathbf{E} \times \mathbf{B}$ drift velocity of particles in uniform $\mathbf{E}$ and $\mathbf{B}$ fields can be derived using the Lorentz transformation. For that, we need to move to a reference frame moving with some constant velocity $\mathbf{V}$ relative to the original reference frame and perpendicular to $\mathbf{B}$:

$$\mathbf{E}' = \mathbf{E} + \mathbf{V} \times \mathbf{B} \tag{1.6}$$

where $\mathbf{E}$ is the electric field that particles sense in the non-moving reference frame, $\mathbf{E}'$ is the electric field they sense in the moving reference frame, $\mathbf{B}$ is the magnetic field and $\mathbf{V}$ is the velocity of the moving reference frame.

We can choose $\mathbf{V}$ such that $\mathbf{E}' = 0$, i.e., the electric field sensed by particles in the moving system is zero. In that case, $\mathbf{V}$ will be the drift velocity of particles in $\mathbf{E}$ and

B fields. Without loss of generality, we also assume that $\mathbf{V} \perp \mathbf{B}$, since $\mathbf{V}_{\parallel} \times \mathbf{B} = 0$ in eq. (1.6). By multiplying eq. (1.6) by **B** from the left we obtain:

$$\begin{aligned} \mathbf{B} \times \mathbf{E}' = \mathbf{B} \times (\mathbf{E} + \mathbf{V} \times \mathbf{B}) \implies \\ \implies 0 = \mathbf{B} \times \mathbf{E} + \mathbf{B} \times (\mathbf{V} \times \mathbf{B}) \overset{\text{(using bac-cab rule)}}{=} \mathbf{B} \times \mathbf{E} + \mathbf{V}(\mathbf{B}\mathbf{B}) - \mathbf{B}(\mathbf{B}\mathbf{V}) \overset{\text{(since } \mathbf{B} \perp \mathbf{V})}{=} \\ = \mathbf{B} \times \mathbf{E} + \mathbf{V}B^2 - 0 \implies 0 = \mathbf{B} \times \mathbf{E} + B^2\mathbf{V} \implies \mathbf{V} = \frac{\mathbf{E} \times \mathbf{B}}{B^2}. \end{aligned} \tag{1.7}$$

From eq. (1.6), we can also derive that

$$\mathbf{E} = -\mathbf{V} \times \mathbf{B} \tag{1.8}$$

where again **V** is the velocity of the moving system. This equation will be used in the sections below to describe the bulk motion of particles in the plasmasphere.

Here, we considered the case of uniform electric and magnetic fields. Non-uniform magnetic fields cause other types of drifts of charged particles, such as gradient and curvature drifts (Baumjohann and Treumann, 1997). These drifts depend on the energy of particles as well as their charge. For cold plasma, these drifts are usually neglected, since due to the low energies they are much smaller than the $\mathbf{E} \times \mathbf{B}$ drift.

Bulk plasma motion

Two main external forces drive the bulk motion of cold plasma within the magnetosphere: (1) the daily rotation of the Earth, and (2) the kinetic energy of the solar wind. These forces lead to the creation of the corotation and convection electric fields, respectively.

Corotation

The rigidly rotating magnetic field of the Earth induces an electric field directed perpendicular to the geomagnetic field. This field is usually referred to as the corotation electric field (Baumjohann and Treumann, 1997). In the non-rotating frame of reference, the corotation electric field is given by (see eq. (1.8):

$$\mathbf{E}_{\text{cor}} = -(\Omega_{\text{E}} \times \mathbf{r}) \times \mathbf{B}, \tag{1.9}$$

where **r** is the radial position vector, and $\Omega_{\text{E}} = 7.27 \times 10^{-5}$ rad s^{-1} is the angular velocity of the Earth's rotation. If we assume that **B** is a dipole magnetic field

centered on the rotation axis of the Earth, then $\mathbf{E}_{\text{cor}}$ will be directed radially inwards toward the Earth in the equatorial plane and decrease as $1/r^2$.

We can use eq. (1.8) and (1.9) to determine the velocity of the charged particles in the inner magnetosphere moving in this field:

$$\begin{aligned}\mathbf{v} = \frac{\mathbf{E}_{\text{cor}} \times \mathbf{B}}{B^2} = -\frac{((\Omega_{\text{E}} \times \mathbf{r}) \times \mathbf{B}) \times \mathbf{B}}{B^2} = \frac{\mathbf{B} \times (\Omega_{\text{E}} \times \mathbf{r}) \times \mathbf{B}}{B^2} \overset{\text{(using bac-cab rule)}}{=} \\ = \frac{(\Omega_{\text{E}} \times \mathbf{r})B^2 - \mathbf{B}(\mathbf{B}(\Omega_{\text{E}} \times \mathbf{r}))}{B^2} \overset{\text{(since } \mathbf{B} \perp (\Omega_{\text{E}} \times \mathbf{r}))}{=} \Omega_{\text{E}} \times \mathbf{r}.\end{aligned} \tag{1.10}$$

This equation implies that the velocity of the charged particles is the same as the Earth's rotation, meaning that the cold plasma in the vicinity of the Earth moves together, i.e., corotates with the Earth. As discussed below, the plasmasphere can also lag in corotation or exhibit supercorotatation.

The electric corotation potential in the equatorial plane can be obtained from the following expression:

$$\mathbf{E}_{\text{cor}} = -\nabla\phi_{\text{cor}}. \tag{1.11}$$

In the equatorial plane, it can be expressed as

$$\phi_{\text{cor}}(L) = -\frac{\Omega_E B_E R_E^2}{L}, \tag{1.12}$$

where the term $\Omega_E B_E R_E^2$ equals 92 kV. The corotation potential is radially symmetric in the equatorial plane, and consists of concentric circles. The distance between those circles decreases as L^{-1} (shown in the top right panel in Figure 1.4). These circles are corotating orbits of plasma.

Convection

The interaction of the solar wind and geomagnetic field results in a plasma flow across open polar cap field lines (connected to IMF), as discussed in section 1.1.1. This flow in turn leads to the creation of a large scale electric field $\mathbf{E} = -\mathbf{V}_{\text{sw}} \times \mathbf{B}$ (in the frame of reference of the Earth, see eq. (1.6)), where $\mathbf{V}_{\text{sw}}$ is the solar wind velocity and $\mathbf{B}$ is the magnetic field of the Earth. This field is known as the convection electric field and is directed from dawn to dusk.

In the inner magnetosphere, the electric fields are generally orthogonal to the magnetic fields, as it consists of a highly conducting plasma. If an electrostatic field is parallel to the magnetic field in such a plasma, then the positively charged particles are accelerated in the direction of $\mathbf{E}_{\|}$ and negatively charged particles in the opposite

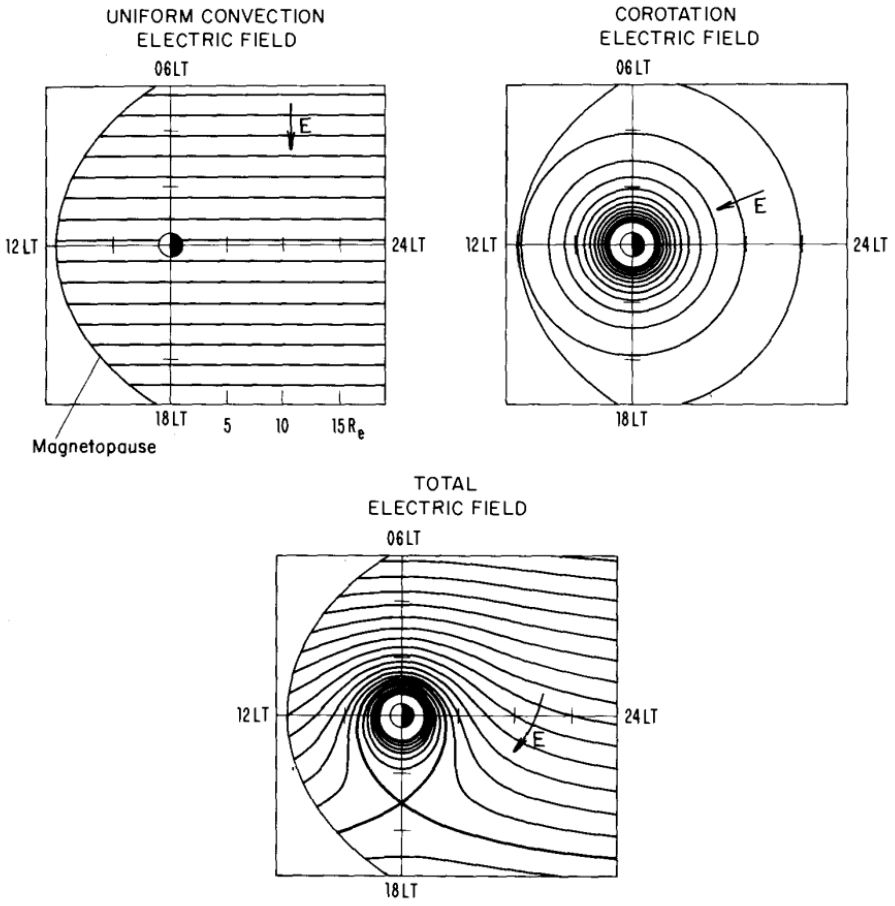

Figure 1.4: Equipotential contours for magnetospheric electric fields in the equatorial plane (Lyons and Williams, 1984). The convection electric field is assumed to have a strength of $2.5 \times 10^{-4} Vm^{-1}$.

direction. This produces a charge separation, which in turn can produce an additional electrostatic field. This field cancels out the original parallel electric field (Kivelson and Russell, 1995). Therefore, parallel electrostatic fields are rarely found in the trapping region of the magnetosphere. Consequently, the electric potential lines are parallel to the magnetic field lines, and the electrostatic potential is assumed to be constant along the field lines. The polar cap potential, therefore, can be mapped to the equator along the Earth's magnetic field lines.

The simplest approximation of the equatorial convection is a spatially uniform dawn-to-dusk electric field (Stern, 1974):

$$\mathbf{E}_{\text{con}} = E_0 \hat{\mathbf{y}} \tag{1.13}$$

where E_0 is the magnitude of the field, which is typically less than 1 mV/m. The equipotentials of such a uniform convection electric field in the equatorial plane are shown in the top left panel of Figure 1.4. In polar coordinates, the convection potential in the equatorial plane can be expressed as

$$\phi_{\text{conv}} = -E_{\text{conv}} L R_{\text{E}} \sin(\psi) \tag{1.14}$$

where E_{conv} is the uniform convection electric field strength in the equatorial plane, LR_{E} is the radial distance, and ψ denotes azimuth (Baumjohann and Treumann, 1997).

The bottom panel of Figure 1.4 shows the combined action of the corotation and convection electric field potentials. In such a configuration of the electric field, the corotation electric field is dominant close to the Earth, and thus the charged particles corotate around the Earth, moving on the closed drift paths. The convection electric field dominates the drift motion at farther distances from the Earth, and the particles move on the open drift paths from the nightside out through the magnetopause.

Shielding

When the magnetospheric convection is enhanced, the charged particles from the plasma sheet move in a sunward direction, from the magnetotail towards the Earth. Energetic electrons drift dawnward and energetic protons drift duskward around the Earth, due to the magnetic gradient and curvature drifts. A charge separation is created due to that, which induces a polarization electric field. This field is directed

from dusk to dawn, and thus acts to shield the inner magnetosphere from the dawn-to-dusk cross-tail convection electric field (e.g., Jaggi and Wolf, 1973). Due to that, the convection electric potential in the inner magnetosphere is slightly weaker than described by eq. (1.14).

It takes a finite time for the shielding to be established, and changes in the convection field or other magnetospheric parameters can cause it to be ineffective for extended periods. If convection grows faster than the shielding timescale (the characteristic timescale is 1 hour (Kelley et al., 1979)), then "undershielding" occurs, and a dawn-to-dusk convection electric field is imposed upon the inner magnetosphere until the shielding adjusts to the new conditions. If, on the other hand, a residual dusk-to-dawn shielding electric field remains after a sudden decrease in the strength of convection (such as after a sudden northward turning of the IMF B_z component), then "overshielding" occurs (Goldstein et al., 2002).

To take the shielding effect into account, the convection potential can be presented in the following form:

$$\phi_{\rm cs} = -A_\gamma (LR_{\rm E})^\gamma \sin(\psi) \tag{1.15}$$

where γ is the shielding factor and A_γ is a constant described by $A_\gamma = 0.5\Delta\phi\Delta y^{-\gamma}$, with $\Delta\phi$ denoting the cross-tail potential difference and Δy half the distance between the dawn and dusk magnetopause along the $\psi = \pm 90^\circ$ axis. The shielding factor γ ranges between 2 and 3 under typical geomagnetic conditions. The electric field amplitude can be calculated from eq. (1.15) and is given by

$$E_{\rm cs} = A_\gamma (LR_{\rm E})^{\gamma-1} [(\gamma^2 - 1)\sin^2\psi + 1]^{1/2} \tag{1.16}$$

The shielding factor $\gamma = 1$ provides a uniform electric field. When using a realistic shielding factor $\gamma \approx 2 - 3$, the electric field amplitude decreases toward the inner magnetosphere and varies with local time.

General plasmaspheric dynamics

The dynamics and configuration of the plasmasphere is governed by the interplay of corotation and convection electric fields (Darrouzet et al., 2009; Singh et al., 2011) described in previous subsections.

During quiet geomagnetic conditions, the corotation regime prevails and the plasma

material trapped inside the closed magnetic field lines corotates with the Earth (Carpenter, 1966). In the meantime, the plasmasphere is being refilled from the topside ionosphere and expands in size up to $\sim 4 - 7\ R_{\rm E}$ (Goldstein et al., 2003b; Singh and Horwitz, 1992; Krall et al., 2008). It has a roughly circular shape with a bulge on the dusk side (Nishida, 1966). However, at farther distances from the Earth where the equipotential contours are open, the flow of plasma is dominated by the solar-wind induced convection, even under quiet geomagnetic conditions (Axford and Hines, 1961). Although ionospheric outflow still occurs at those latitudes, the time during which the open field lines drift in from the magnetotail out to the dayside magnetosphere is not sufficient for them to be filled to the plasmaspheric levels. It is also worth noting that during extended periods of quiet geomagnetic conditions a sharp plasmapause density gradient may not necessarily be present, and instead, the density profile can decrease gradually with distance.

During periods of elevated geomagnetic activity, the magnetospheric convection starts to dominate and the plasmasphere is depleted: the closed magnetic field lines at the dayside magnetopause boundary are torn apart and the plasmaspheric material is carried sunward. As its outer layers are eroded, the plasmasphere shrinks in size (Carpenter, 1970; Chappell et al., 1970a; Goldstein et al., 2003b). The stronger the disturbance, the more the plasmapause contracts (down to 2 $R_{\rm E}$ during severe geomagnetic storms). The sunward transport of the eroded plasmaspheric material forms a plasma tail or plume in the afternoon sector extending toward the dayside magnetopause.

After the disturbance, the corotation regime starts to dominate again and the plasmasphere is refilled. Due to long refilling times and recurring geomagnetic storm and substorm activity, the outer plasmasphere is more likely to be in a stage of partial refilling, while the inner plasmasphere is expected to approach a diffusive equilibrium with the topside ionosphere (Park, 1974). The plumes formed during the geomagnetic disturbance wrap around the Earth, as they corotate with it (Goldstein et al., 2002, 2003b; Goldstein and Sandel, 2005; Sandel et al., 2003; Kim et al., 2007)

Refilling

Refilling of the plasmasphere from the ionosphere is dominant during geomagnetically quiet times and storm recovery phases. It is a complex process and is described here briefly (for more details, see references herein).

The main constituents of the plasmasphere are hydrogen, helium, and oxygen

ions, with hydrogen being the dominant ion. The main source of hydrogen in the plasmasphere is the charge exchange reaction $H + O^+ \rightleftharpoons H^+ + O$, which happens in the upper F region of the ionosphere. The F layer is produced by extreme ultraviolet solar radiation that ionizes atomic oxygen. Protons, formed in that charge exchange reaction, move into the plasmasphere under the influence of diffusion, electric fields and pressure gradients.

Ideally, in equilibrium, the plasma pressure along the flux tube will be such that there would be no flow. During the day, ion production in the topside ionosphere generally causes an excess of plasma pressure in the topside, which results in the upward flow of plasma, from the ionosphere into the plasmasphere (Hanson and Patterson, 1963). At night, the solar EUV production decreases, and the plasma pressure in the equatorial region can exceed that in the topside ionosphere, causing H^+ to flow downward, into the topside ionosphere, where it may exchange charge with O to provide a night-time source of O^+ (Park, 1971). This results in a complex interaction between the ionosphere and the plasmasphere. The total quantity of plasma flowing in during the day will equal that flowing out at night. This is not true during geomagnetic storms, when the electric field convects plasma across field lines (Canuto et al., 1978).

During quiet geomagnetic conditions, refilling proceeds steadily until the plasmasphere reaches a density level where a diffusive equilibrium with the ionosphere is reached. Observations showed that the timescale for the plasmasphere to refill depends on the L value and is $\sim$1 day at $L = 2.5$ and $\sim$8 days at $L = 4$ (Park, 1974).

Duskside bulge and plume

Since the time of the first studies of the plasmasphere, large-scale bulges (1 R_E or greater) have been observed in the equatorial plasmapause, particularly in the dusk local time sector (Carpenter, 1966). In general, the westward edge of the bulge was found to be rather abrupt with an increase in the plasmapause radius of 0.5 to 2.5 R_E within approximately one hour of local time, while the radius on the eastern edge decreased more gradually with increasing local time. During periods of increased geomagnetic disturbance, bulges in the dusk sector were observed at earlier local times, which is assumed to be caused by intense westward flows occurring due to enhanced convection. During quiet periods, bulges tended to be observed at later local times (or not at all), which can be interpreted as the result of the increased influence of the rotation of the Earth during those periods (Carpenter, 1970).

In addition to that, first observations from the Orbiting Geophysical Observatory

(OGO) spacecraft revealed the regions of high-density cold plasma outside the main plasmasphere (Chappell et al., 1970b). These regions were found across the entire dayside magnetosphere during moderate to disturbed magnetic conditions, with a maximum occurrence in the afternoon-dusk sector. These regions were at first interpreted to be detached from the main body of the plasmasphere. They were mostly located close to the duskside plasmapause and continuously farther away at earlier local time sectors. Such regions were thought of as being peeled off and therefore detached from the main plasmasphere near dusk and drifting outward to the dayside magnetopause. Numerical models have been widely used to understand and explain the observations near the dusk-side plasmasphere (see section 1.2.5 for the description of early models). A number of models have been developed that were able to reproduce a "teardrop" plasmasphere with a bulge in the dusk sector and additionally they produced tails of plasma or plumes in the day-to-duskside sector resulting from the enhanced convection (Grebowsky, 1970; Chen and Wolf, 1972; Chen and Grebowsky, 1974; Kurita et al., 1985). The model by Chen and Grebowsky (1974) demonstrated that the observations of detached density structures in the noon to dusk sectors may actually be crossings of satellites through such plumes, and that the "detached" regions may in fact be attached to the main body of the plasmasphere. The existence of plumes was directly confirmed in the 2000s, when the IMAGE mission provided the first global images of the plasmasphere (Burch et al., 2001a; Sandel et al., 2001).

Using the EUV images of the plasmasphere, Goldstein and Sandel (2005) found that plumes are formed and developed in three stages: sunward surge, plume narrowing, and plume rotating. They showed that plumes are developed even with a small increase of geomagnetic activity. With a significant increase of geomagnetic activity, the IMF component B_z turns southward and particles that were previously located on the outermost closed trajectories suddenly become located on open trajectories and drift toward the dayside magnetopause. This results in an inward movement of the nightside plasmapause, towards the Earth, while the dayside plasmapause moves outward, towards the Sun and a broad plume directed sunward is formed. The decrease in geomagnetic activity weakens the plume's sunward orientation and the plume begins to rotate eastward with the main plasmasphere and may wrap around it (Goldstein et al., 2002, 2003b; Goldstein and Sandel, 2005; Sandel et al., 2003; Kim et al., 2007).

More recently, using observations from the MPA instruments onboard the LANL

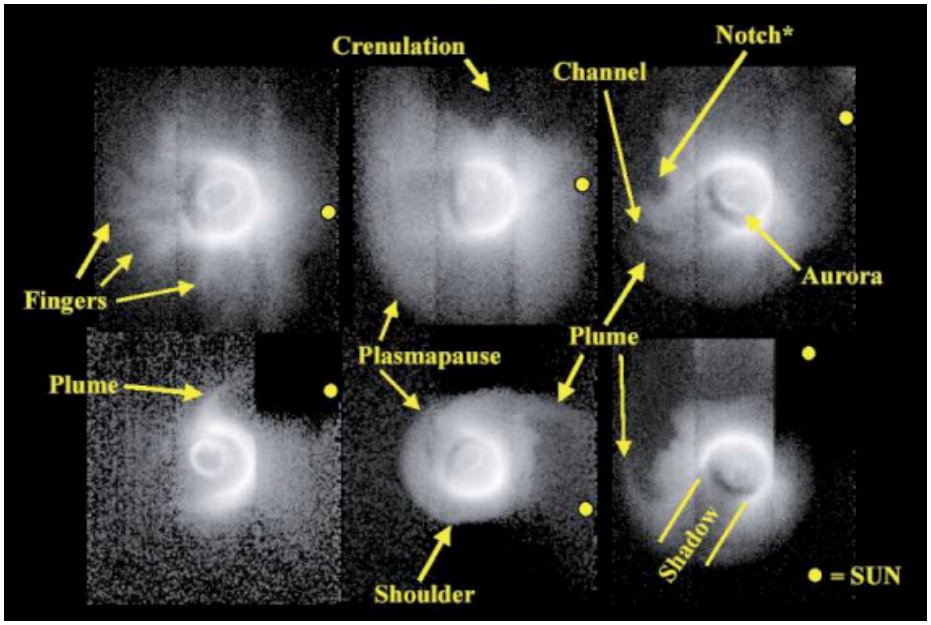

Figure 1.5: Examples of structures at the plasmapause observed by the IMAGE EUV. The direction to the Sun is shown as a yellow dot in each image (figure from (Darrouzet et al., 2009)).

geosynchronous spacecraft ($L = 6.6$), Borovsky et al. (2014) reported long-lived plasmaspheric drainage plumes that sustained for as long as 11 days, which exceeded the lifetimes expected from refilling. They were observed during high-speed-stream-driven storms, which usually last for a few to several days. There have been several theories on where the plasma sustaining these plumes is coming from, but none of them were able to explain the observed plumes in a satisfactory manner, and the origin of this plasma is still unknown (Gallagher and Comfort, 2016).

Structures at the plasmapause

In addition to plumes, the IMAGE EUV images revealed various other structures and irregularities of different scales, such as notches, shoulders, fingers, channels, crenulations, and plume bifurcations. Figure 1.5 shows several examples of these structures. They have been studied extensively by the IMAGE team and provided additional insights into the plasmasphere dynamics.

Notches are typically distinguished by depletion of the radial densities in the outer plasmasphere (Carpenter et al., 2002). Monitoring of notches for a long time (3060 h) provided evidence that the plasmasphere does not strictly corotate with the Earth and departs from corotation. Sandel et al. (2003) found that the rotation rates of notches were between 77% and 93% of strict corotation. A possible explanation for

the departures from corotation lies in the corresponding motions of plasma in the ionosphere, where departures from corotation are observed regularly (Sandel et al., 2003; Burch et al., 2004; Galvan et al., 2008).

Channels of low-density plasma are generally located between the plasmasphere and the plume. Their formation can be explained by differential rotation of the western edge of a plume in L and stagnation of the eastern edge (Spasojević et al., 2003). This may lead to the wrapping of the plume around the plasmasphere.

Shoulders appear in the form of sharp azimuthal gradients and are usually observed during sudden increases in geomagnetic activity (Burch et al., 2001a,b). Goldstein et al. (2002) explained the formation of the shoulder by the residual of overshielding of the convection electric field following the sudden decrease of convection during the northward turning of the IMF.

Crenulations are mesoscale variations on the plasmapause surface that are on the order of a few tenths of an R_{E} and appear to be limited to the local time sector between dawn and the western edge of the plasmaspheric plume, and they grow while rotating eastward towards noon (Spasojević et al., 2003; Goldstein and Sandel, 2005). The latter studies also observed the formation of a double plume (or bifurcation of a single plume) that similarly to crenulations also occurs in the sunward surge phase. The second plume may be formed by the rotation of the shoulder into the dusk sector.

Fingers are irregular features observed at the plasmapause, even during deep quiet geomagnetic conditions. They are considered to arise from resonances of ultra-low frequency waves (Adrian et al., 2004).

Erosion and loss of plasma

Erosion is the dominant driver of the plasmasphere dynamics during the main phase of the geomagnetic disturbance and is driven by enhanced magnetospheric convection, transporting cold plasma toward the dayside magnetopause. During this process, the outer layers of the plasmasphere are stripped away and as a result, the plasmasphere shrinks. Early observational studies found that the reduction in the plasmapause radius during storm times occurs mostly in the post-midnight local time sector. Chappell et al. (1970b) showed, using in-situ ion density measurements, that the plasmapause location in the night-to-dawn local time sector was strongly correlated with the average geomagnetic activity over the previous several hours (2 to 6 hours). The density profile at the plasmapause steepened in that sector and the plasmapause moved inwards to lower L shells, in response to enhanced geomagnetic

activity. On the dayside, however, the plasmapause did not change its position immediately and instead depended on the level of geomagnetic activity during the interval when that region previously corotated through the nightside region (Chappell et al., 1971). Observations obtained from the IMAGE mission showed that the initial erosion of the plasmasphere starts close to the local midnight and then broadens and spreads in the eastward and westward directions covering the entire nightside plasmasphere in several hours (Goldstein and Sandel, 2005; Gallagher and Adrian, 2007). Removal of plasma occurs at different times for different MLTs, therefore the erosion propagates with a finite speed from the initial location (Goldstein and Sandel, 2005).

The subauroral polarization stream (SAPS) may aid the global magnetospheric convection by driving rapid erosion of the plasmasphere in the evening local time sector. The SAPS refers to the broad poleward-directed electric field driving sunward plasma convection at sub-auroral latitudes in the evening local time sector (Foster and Burke, 2002). Enhanced convection injects the ring current energetic particles into the inner magnetosphere, causing pressure gradients driving field-aligned currents (FAC) in the ionosphere that subsequently produce the SAPS (Burke et al., 1998; Anderson et al., 2001; Foster and Burke, 2002). The SAPS region in the ionosphere is located below the auroral oval, extending $\sim$3 – 5 degrees in invariant latitude on average and concentrating mostly in the dusk and pre-midnight local time sectors (Foster and Vo, 2002). The SAPS can considerably increase the rate at which the plasma is transported from the duskside plasmasphere to the dayside magnetopause during intervals of elevated geomagnetic activity (Goldstein et al., 2003b) and may act to remove the duskside bulge (Goldstein et al., 2007).

Large amounts of plasma can be removed from the plasmasphere during enhanced geomagnetic activity, as a result of magnetospheric convection peeling off the outer layers of the plasmasphere. Park and Carpenter (1970) estimated, using whistler data, that $\sim 3 \times 10^{31}$ electrons and ions were removed from the plasmasphere during a magnetic storm on June 15, 1965. This comprised more than half of the plasma population of the plasmasphere during quiet times, as basically all plasma in the range of $L = 3.5$ to 5 was removed from the plasmasphere. IMAGE measurements of in-situ density using the radio plasma imager (RPI) instrument also showed the massive loss of plasma along the magnetic field lines. Reinisch et al. (2004) analyzed the data of the March 31, 2001 storm and showed that outer plasmaspheric layers lost more than 66% of plasma in less than 14 hours and refilling occurred over a period

of 10 days. Spasojević and Sandel (2010) estimated using the IMAGE EUV images that the global loss of He^+ ions during moderate disturbance intervals comprised between ∼0.6 and 2.2 $\times 10^{30}$ He^+ ions, constituting between 20% and 42% of the initial He^+ distribution. When the material eroded from the plasmasphere reaches the dayside magnetopause, it may become trapped in the afternoon-dusk side of the outer magnetosphere, if convection weakens (Carpenter et al., 1993; Chen and Moore, 2006; Lee et al., 2016). The enhanced plasma density at the dayside magnetopause can limit the rate of reconnection (Borovsky and Denton, 2006; Borovsky et al., 2014), thus affecting the global convection pattern (e.g., André et al., 2016; Borovsky and Denton, 2006).

1.2.5 Models of the plasmasphere

Physics-based models

Early modeling of the plasmasphere dynamics focused on explaining the formation of the plasmapause, the sharp drop in density observed by the satellites and on the ground. After first findings on the global structure of the plasmasphere and how it responds to geomagnetic conditions based on the measurements from whistlers, Nishida (1966) and Brice (1967) independently proposed models that explained the formation of the plasmapause as a result of the interplay between the corotation and convection plasma flow regimes. These models defined the steady state plasmapause as the last closed equipotential contour surrounding the Earth (i.e., bottom panel in Figure 1.4). They also modeled the duskside bulge, which resulted from the stagnation of the combined flow of corotation and convection in the dusk sector. Grebowsky (1970) modeled the dynamics of the plasmapause using the ideal magnetohydrodynamics (MHD) equations. He also assumed that the initial position of the plasmapause corresponded to the last closed equipotential in a combined corotation and convection field, resulting in a "teardrop" shape of plasmasphere, with a bulge on the duskside. As the strength of convection increased, the bulge rotated westward and the outer parts of the plasmasphere were transported sunward, resulting in a plasma tail or plume in the afternoon sector (Figure 1.6). Chen and Grebowsky (1974) used a similar modeling approach to demonstrate that the observations of detached density regions in the afternoon local time sector may be crossings of satellites through these plumes, and that the "detached regions" could in fact be attached to

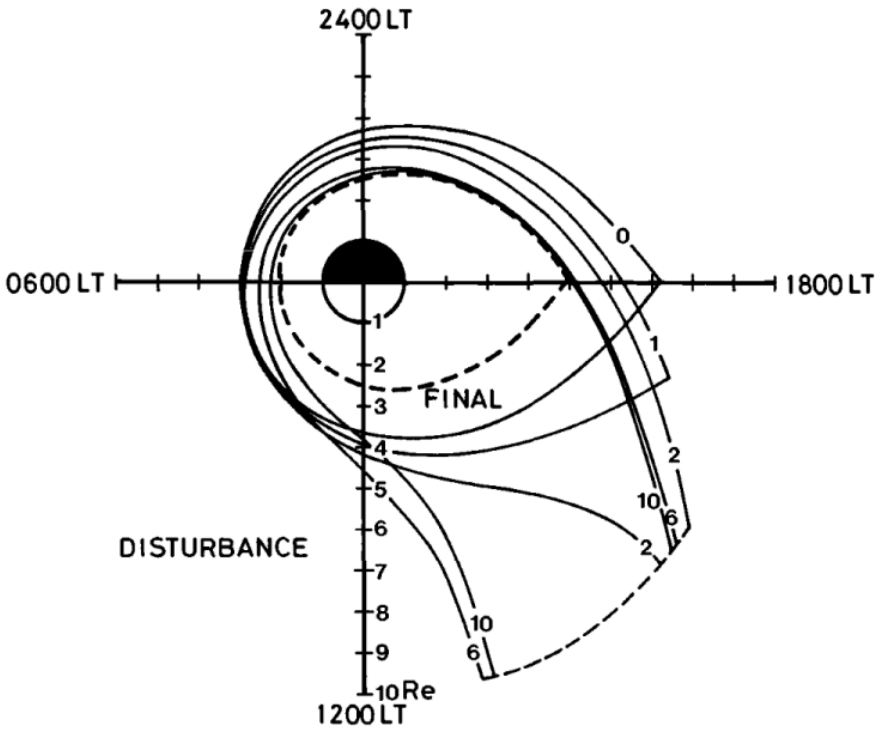

Figure 1.6: Plasmapause location for selected times (0, 1, 2, 6 and 10 hours, labeled) after the increase of the dawn-dusk electric field component from its initial steady state value of 0.28 mV/m to 0.58 mV/m (after Grebowsky (1970)).

the main plasmasphere. As the strength of convection decreased, the modeled plume or bulge rotated eastward and wrapped around the main plasmasphere (e.g., Chen and Wolf, 1972; Kurita et al., 1985). It should be noted that the resulting temporal development of the plume may vary in different models and depends strongly on the assumed initial shape of the plasmasphere and the model of the convection field.

More recently, Goldstein et al. (2003b) developed a model of the plasmapause dynamics using the so-called test particle simulation approach. In this "plasmapause test particle" (PTP) model, the dynamics of the plasmapause are modeled by assuming that the plasmapause boundary is comprised of test particles subject to $\mathbf{E} \times \mathbf{B}$ drift. The evolution of the plasmapause in a time-dependent convection electric field is then modeled by the changing shape of the curve defined by the aggregate of the test particles. In their model, they employed the Volland (1973) and Stern (1975) convection electric field driven by solar wind data and the Kp geomagnetic index, which is representative of the traditional convection-based picture. Using the PTP model and the IMAGE EUV images of He^+ abundance, Goldstein et al. (2003b) demonstrated

that in addition to convection, the subauroral polarization stream played a major role in the plasmasphere dynamics during erosion, in particular, the SAPS flow channel can dominate the duskside when convection was weak. Subsequently, Goldstein et al. (2005a) developed a model of the SAPS electric potential, parameterized by the Kp index and based on the study of Foster and Vo (2002) that provided average characteristics of the SAPS. This model significantly altered the flow paths of plasma in the dusk sector (see Figure 1.7) and provided good agreement with the plasmasphere observations from the IMAGE EUV instrument. Goldstein et al. (2014) updated the PTP model by including a simple treatment of refilling, and compared it to plume encounters derived from the Van Allen Probe in-situ density measurements. The model showed a good agreement with observations which provided yet more evidence that the $\mathbf{E} \times \mathbf{B}$ paradigm produces good quantitative predictions for plasmaspheric dynamics. They found that the greatest discrepancies between the model and observations occurred during the quietest geomagnetic conditions. They concluded that "while $\mathbf{E} \times \mathbf{B}$ drift may dominate during strong convection, there are numerous weaker processes such as plasma interchange (Lemaire and Kowalkowski, 1981; Pierrard and Lemaire, 2004; Pierrard et al., 2009) or neutral wind coupling (Burch et al., 2004) that may contribute during storms and may be dominant during recovery and quiet interval".

In addition to modeling the plasmapause dynamics, a number of physics-based plasma density models have been developed in recent years. Pierrard and Stegen (2008) used the kinetic exospheric approach to model the dynamics of the plasmasphere. Jordanova et al. (2006) coupled their ring current-atmosphere interactions model (RAM) with a 3-D equilibrium code (SCB) and a cold plasma model (CPL). The RAM-SCB-CPL model calculates the cold electron density in the equatorial plane by following the motion of individual flux tubes, using a model of electric field which includes a corotation potential and a convection potential that is chosen from either semi-empirical models (Volland, 1973; Stern, 1975; Weimer, 2005), or a self-consistently calculated electric potential (Yu et al., 2015), mapped to the equatorial plane along the SCB field lines. Krall et al. (2016) coupled this model with SAMI3 (Sami3 is Also a Model of the Ionosphere) Huba et al. (2008) to model the plasmasphere dynamics during two events in 2001. De Pascuale et al. (2018) used RAM-CPL to simulate equatorial plasmaspheric electron densities during two storm events in 2013, and compared them to in-situ measurements from the Van

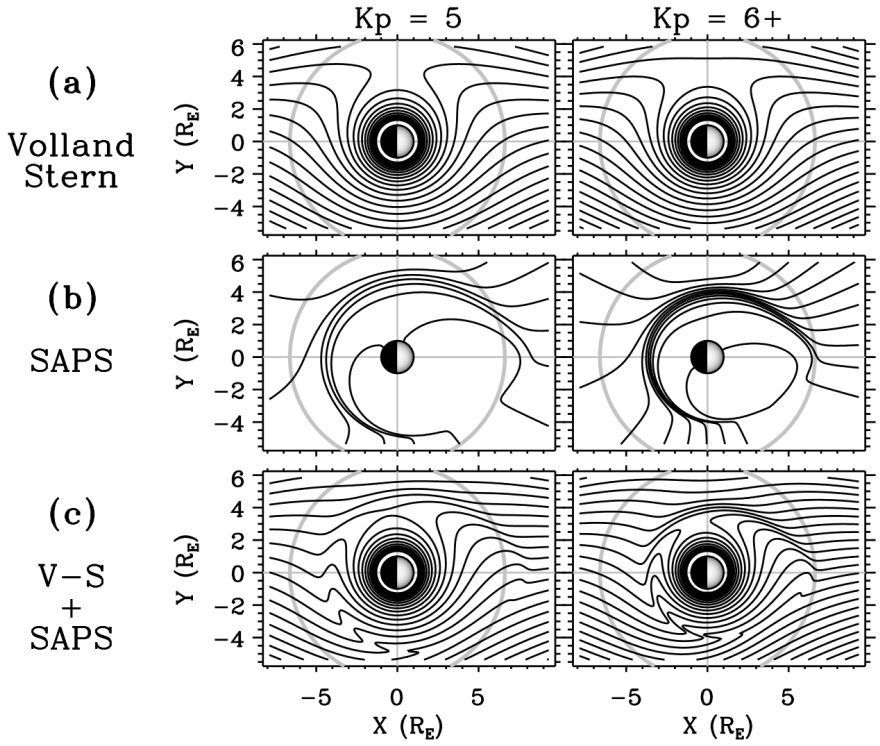

Figure 1.7: Equatorial SAPS potential model for (left) Kp = 5 and (right) Kp = 6+. A gray circle is drawn at geosynchronous orbit (6.62 R_{E}). (a) Volland-Stern model plus corotation (4 kV potential spacing). Note the large flow stagnation region near dusk. (b) SAPS model (2.5 kV spacing), a narrow westward flow channel. (c) Volland-Stern and SAPS combined (4 kV spacing). Duskside sunward flows are significantly enhanced (after Goldstein et al. (2005a)).

Allen Probes (Radiation Belt Storm Probes). Huba and Krall (2013) used the first-principles physics-based model SAMI3 to model the plasmasphere in 3D. The 3D ionospheric SAMI3 model is based on the 2D ionospheric code SAMI2 (Sami2 is Another Model of the Ionosphere) Huba et al. (2000). SAMI3 takes into account wind-driven dynamo electric fields, which is solved as a two-dimensional electrostatic potential equation based on current conservation ($\nabla \cdot \mathbf{J} = 0$). To model the dynamics along the field lines, SAMI3 solves the continuity and momentum equations for seven ion species and the temperature equation for three ion species (H^+, He^+, O^+) and for the electrons. SAMI3 includes 21 chemical reactions and photodeposition Huba et al. (2000). Cross-field transport is included as $\mathbf{E} \times \mathbf{B}$ drifts. The magnetic field is assumed to be dipole, aligned with the spin axis of the Earth so geographic and geomagnetic coordinates are the same. Using this assumption, the corotation potential can be easily incorporated into the model, although as this assumption is an idealization, seasonal/longitudinal effects are not captured. Huba and Krall (2013) incorporated the neutral wind dynamo potential, the corotation potential, and a time-dependent potential from Volland (1973), Stern (1975) and Maynard and Chen (1975) to model the convection potential for an idealized magnetic storm. Krall and Huba (2013) used SAMI3 also to simulate the plasmasphere refilling for an idealized storm and found that SAMI3 refilling rates are in agreement with empirical refilling rates generally within a factor of two, except at $L = 5$, where the difference reaches a factor of five. In that region, SAMI3 shows slower refilling rates than the empirical results. An overview of various other physics-based models of the plasmasphere based on the fluid and the kinetic approaches is given in Pierrard et al. (2009).

One important element of the physics-based models mentioned above is the electric field. The magnetospheric convection electric field plays a key role in existing theories of the plasmapause formation Pierrard et al. (2008). It is therefore important to have a reliable electric field model of the inner-magnetospheric region. Various empirical and mathematical models of the electric field have been developed in the past decades. One of the first models of the electric field that is still often used is that of Volland (1973) and Stern (1975). In this mathematical model, a uniform dawn-dusk convection electric potential distribution is applied across the magnetosphere. The magnetospheric electric field is derived from a scalar potential and there is no induced electric field resulting from time-dependent magnetic field variations. Stern (1975) derived the electric field from the Euler potential for the dipole magnetic field.

Maynard and Chen (1975) obtained the Kp dependence for this empirical model from the radius at the midnight of the equipotential shell passing through the dusk stagnation point, which was assumed as the plasmapause position. Plasmapause positions were determined from the OGO3 and OGO5 satellite observations. Another popular model is the E5D model built by McIlwain (1986) from ATS5 and ATS6 particle flux measurements at geosynchronous orbit. The E5D electric potential was adjusted to fit the dynamical energy spectra of electrons and protons in the range from 1 keV to 100 keV, injected after substorms and observed in the equatorial region. This model was constructed for the periods when Kp was smaller than 6 and remained nearly constant, and its main goal was to represent the distribution of the electric field right after a substorm injection. Similarly to the previous model, the E5D model also depends on the Kp index. Later, Weimer (1996) developed an electric field model for the high-latitude ionosphere based on ionospheric measurements from DE-2, ISEE-3, and IMP-8 satellites. The model does not include Kp dependence and instead is driven by the solar wind parameters, in particular by the interplanetary magnetic field magnitude, solar wind velocity, and also the dipole tilt angle. The electric potential of the Weimer model is given by an expansion of the polar cap potential in terms of spherical harmonics. Weimer (2001) extended the model by adding terms for solar wind electric field and dynamics pressure. The updated model was based on the same satellite measurements. Matsui et al. (2008) used data from the Cluster mission for more than five years to construct the innermagnetospheric electric field model. In addition to measurements from Cluster, they also used data from ground radars and an ionospheric spacecraft. This model depends on the interplanetary electric field (IEF) and is valid for $2 < L < 10$ and all magnetic local times. The data used to build this model were mostly obtained during periods of quiet or moderate geomagnetic activity. Matsui et al. (2013) extended this model by including Cluster electric field data from the Electron Drift (EDI) and the Electric Field and Wave (EFW) instruments during periods of high activity as well. This updated model is organized either by the IEF or by the Kp index.

Empirical models

A number of empirical models of plasma density have been developed in recent decades as well. Carpenter and Anderson (1992) used measurements of electron density derived from radio measurements made with the sweep frequency receiver (SFR) on

board the International Sun-Earth Explorer (ISEE-1) spacecraft and ground-based whistler measurements to develop an empirical model of plasma density. They selected 40 density profiles that corresponded to the relatively quiet conditions representing a "saturated" plasmasphere, and constructed piece-wise models describing the plasmasphere and plasmatrough densities and the plasmapause location. As such, this model is representative of the saturated plasmaspheric densities and was intended for applications in cases when global magnetic conditions have been slowly varying or relatively steady in the preceding ∼20 hours. The model covers the range from 2.25 to 8 in L-shell, and the interval of $0-15$ MLT (magnetic local time). It provides the mean value of density at different L-shells. Gallagher et al. (2000) developed the Global Core Plasma Model (GCPM), which combined several previously developed models (such as (Carpenter and Anderson, 1992) and (Gallagher et al., 1998)) by means of transition equations, in order to provide a more comprehensive description of the inner-magnetospheric plasma. Using density measurements obtained from the swept frequency receiver onboard the CRRES by identifying the upper hybrid resonance frequency, Sheeley et al. (2001) developed models of the plasmasphere and plasma trough. These models provide statistical averages of density based on those measurements and cover all local times and $3 \leq L \leq 7$. Moreover, Sheeley et al. (2001) also provide the standard deviation of density for both the plasmasphere and trough models to describe depleted and saturated density levels for various L-shells and MLT sectors for the trough.

The above-mentioned empirical density models provide statistically averaged values and generally treat the radial density profile and azimuthal shape of the plasmapause as gradually varying. However, the plasmasphere has been found to be highly irregular on a wide range of scale sizes. These models also do not account for the changing geomagnetic conditions, and therefore, are not able to produce reliable estimates of density during extreme geomagnetic events, whereas density is known to vary substantially during periods of strong geomagnetic activity (e.g., Park and Carpenter, 1970; Park, 1974; Moldwin et al., 1995). Therefore developing more advanced models of plasma density is still of continuing interest.

1.3 Modeling methodology

This section provides a general overview of the methodology employed in this dissertation. The description here intends to provide a general context around the methods used, while the implementation details of each particular method are given in the corresponding chapters. The description provided here may partially overlap with the description provided in the following chapters.

1.3.1 Brief background on neural networks

In this dissertation, we use artificial neural networks to solve a number of regression problems ranging from reconstruction of electron density from either satellite measurements or from solar wind parameters and geomagnetic indices, to prediction of the Kp index based on the recent time history of solar wind parameters.

Artificial neural networks are a family of mathematical models effective at solving problems of function approximation, pattern recognition, classification and clustering. They were inspired by biological neural networks in the brain in an attempt to mimic them in a very simplified manner (e.g., McCulloch and Pitts, 1943; Hebb et al., 1949; Marr et al., 1976).

Neural networks are composed of simple computational blocks called artificial neurons. An artificial neuron has a body, in which computations are performed, a number of inputs and one output, similar to a real biological neuron. Simply put, a neuron receives an input signal and then computes an output for it. Figure 1.8a shows a scheme of an artificial neuron with N inputs. Each input has a weight associated with it; the larger the weight, the greater impact the corresponding input has on the output. A neuron also has a bias, which for convenience can be considered as an additional input to the neuron, x_0. The corresponding input is always equal to 1 and has the weight identical to the value of the bias, $w_0 = b$. Additionally, a neuron has an activation, or transfer, function that defines a neuron's type. While there are numerous choices for the activation function, the most common ones are the sigmoid, the hyperbolic tangent, binary and linear functions. After the signal is applied to the neuron, it first computes the sum of inputs multiplied by their weights and then applies the activation function, f, to the resulting sum: $y = f(S),\ S = \sum_{j=1}^{N} w_j x_j + b$.

An artificial neuron is one of the first computational models developed in the research area of artificial neural networks (McCulloch and Pitts, 1943; Rosenblatt,

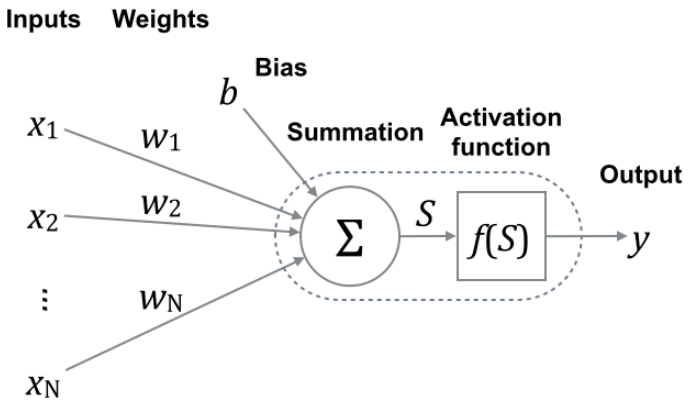

Figure 1.8: A scheme of an artificial neuron, a building block of a neural network. Artificial neurons can be used to solve linear problems.

1957). A single neuron can be used to solve a limited number of problems, such as linear regression and classification of two linearly separable subsets. However, a single neuron cannot be used to solve more complex, nonlinear problems. Fortunately, such problems can be solved by neural networks, which are composed of multiple neurons. The main concept of artificial neural networks is that an output signal from one neuron can be used as an input to other neurons.

The way neurons are connected into a network defines the topology or architecture of a neural network. Throughout this dissertation, we use a feedforward neural network architecture (FNN). An FNN is one of the most basic and widely used neural network architectures and is effective at solving multivariate nonlinear regression and classification problems. FNNs have displayed state-of-the-art performance in a number of applications (e.g., Salakhutdinov and Hinton, 2009; Krizhevsky and Hinton, 2011; Glorot et al., 2011; Mohamed et al., 2012). The topology of an FNN is shown in Figure 1.9a. Neurons in an FNN are arranged in layers. There are three types of layers: input, output, and hidden layers. The input layer is composed of inputs to the network and no computations are performed in this layer. Hidden layers are located between input and output layers and are composed of any number of neurons arranged in parallel. The network can have several hidden layers. The neurons of the same layer are not connected to each other but only to the neurons of the preceding and the subsequent layers; an output of one hidden layer serves as an input to the following layer. The output layer, and hence the network output, is formed by a weighted summation of the outputs of the last hidden layer. Neurons of one layer have the same activation function, while different layers can have different activation

functions. Formally, an FNN with L hidden layers can be defined as a superposition of L activation functions $f_1, ..., f_L$. The neural networks with at least one hidden layer can solve nonlinear regression or classification problems (Cybenko, 1989).

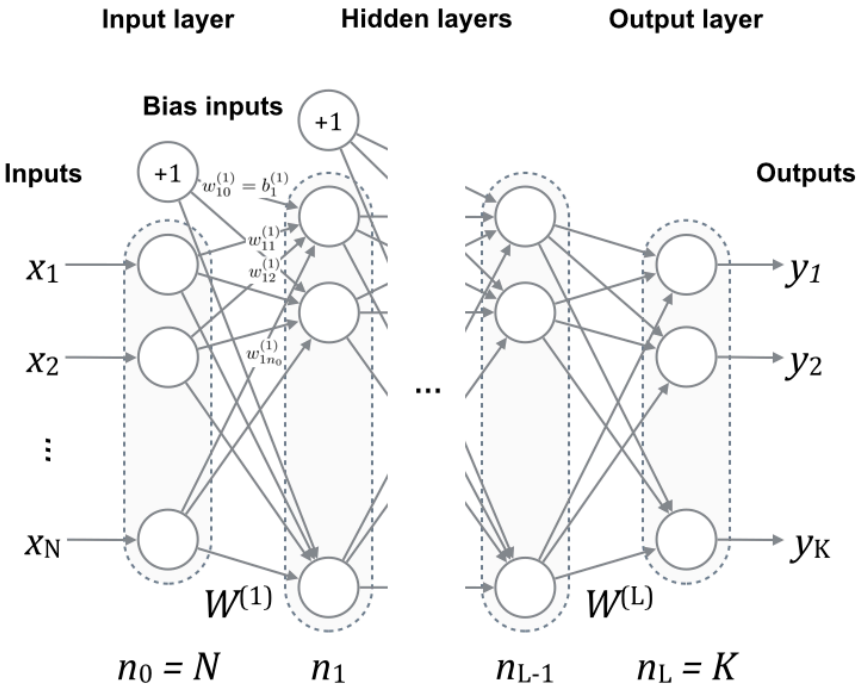

Figure 1.9: A scheme of a feedforward neural network. Circles denote artificial neurons. Feedforward neural networks arrange neurons in a layered configuration and can be used to solve nonlinear problems.

The description above assumes that weights and biases of the neural network are known. In practice, the weights and biases are not given, but they can be determined using a training data set, i.e., a set of data for which both inputs and outputs are known. Determining weights and biases is usually referred to as training and reduces to an optimization problem of minimizing a given cost function. The cost function is defined based on the type of a problem we are attempting to solve. Specifically for neural networks, a number of backpropagation algorithms are used to solve the optimization problem of determining the weights (Williams and Hinton, 1986).

1.3.2 Physics-based modeling of the plasmasphere dynamics

Another approach to model the dynamics of the plasmasphere that we employ in this dissertation is physics-based modeling. To be able to use this approach, a number of

physical processes needs to be taken into account, such as the configuration of the Earth's magnetic field and electric fields, refilling, etc. (see section 1.2).

The electron density in the plasmasphere can be modeled by assuming that cold particles in the plasmasphere undergo $\mathbf{E} \times \mathbf{B}$ drift in azimuthal and radial directions. The evolution of the electron density in the equatorial plane can then be described by the following equation:

$$\frac{\partial n}{\partial t} + v_\phi \frac{\partial n}{\partial \phi} + v_R \frac{\partial n}{\partial R} = S - L \tag{1.17}$$

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

To solve this equation, we need to make assumptions about the structure of magnetic and electric fields. In this dissertation, we assume that the Earth's magnetic field can be approximated by a dipole and that the electric field is a superposition of co-rotation, convection, and subauroral polarization stream (SAPS)-driven electric fields. Using these assumptions, we calculate $\mathbf{E} \times \mathbf{B}$ drift velocities. We use the Kp-dependent Volland-Stern electric field model (Maynard and Chen, 1975; Stern, 1975; Volland, 1973) to calculate the convection electric field, and the Kp-dependent model of the SAPS electric field developed by Goldstein et al. (2005a). We also account for refilling by using refilling rates of equatorial electron density from Denton et al. (2012). The use of these models, as well as the treatments of saturation of the plasmasphere density and loss of particles into the interplanetary medium, are described in detail in Chapter 4.

1.3.3 Data assimilation

A part of this dissertation is devoted to combining a physics-based and a neural network-based empirical model of plasma density in an optimal way. This is motivated by the need to improve the performance of the neural networks during strong geomagnetic storms as the latter are generally underrepresented in the available data.

The performance of neural networks strongly depends on the amount of training data, and the amount of data corresponding to storms is not sufficient for neural networks to efficiently learn from them. The physics-based model described above is employed to overcome this limitation, as it demonstrates a more reliable performance during such intervals. At the same time, as it is based on simplified assumptions, it has a lower accuracy compared to the data-driven neural network-based model. To combine the advantages of both models and blend them in an optimal way, we employ data assimilation.

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

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

We adapt the Kalman filter technique to optimally combine the predictions of two models of the plasmaspheric electron density, the neural network-based model developed in Chapter 3 and the physics-based model developed in Chapter 4. For this purpose, we consider the physics-based model as a model that propagates a state

of the system in time. The output of the neural network-based model, in turn, is used as "observations". Such a setup allows us to choose the "observations" on any desirable grid, which provides a huge advantage over typical situations when the data are very sparse. Here, we use the same grid for both the model and the "observations" in order to make the blending more efficient.

1.4 Contributions of the dissertation

The scope of this dissertation is a development of new tools and models aimed at the reconstruction of electron density in the plasmasphere from various sources ranging from satellite measurements, geomagnetic indices, and solar wind data to those based on first principles.

In Chapter 2, we develop an algorithm based on feedforward neural networks to automatically derive the upper-hybrid resonance frequency from electric field measurements obtained from the Van Allen Probes mission. The derived frequency is used to calculate electron density and to subsequently produce a publicly available data set of electron density covering nearly four years of measurements. The results presented in this chapter are based on the publication Zhelavskaya et al. (2018).

We employ this large electron density data set to develop a new global empirical model of plasma density parameterized by solar wind and geomagnetic indices in Chapter 3, also using feedforward neural networks. The model is extensively validated by comparing its output to in-situ density measurements from Van Allen Probes withheld from the training set, and also by comparing its global output to the images of He^+ distribution from the IMAGE EUV instrument. This chapter is based on the results from Zhelavskaya et al. (2017).

In Chapter 4, we explore the limitations of this global neural network-based model and develop an approach allowing to improve its performance during extreme geomagnetic storms, underrepresented in the data available for training. This is achieved by combining the neural network-based model with a physics-based model of plasma density in an optimal way by means of data assimilation. Such a blend allows achieving optimal performance for quiet, disturbed, and extreme geomagnetic conditions. The results presented in this chapter are based on the findings of Zhelavskaya et al. (2020a).

In Chapter 5, we extend the developed machine learning-based tools and apply

them to another important problem in the field of space weather, the prediction of the geomagnetic index Kp. Additionally, we investigate various machine learning and information theory methods for feature selection and apply them to select optimal inputs to a predictive model of Kp. This chapter is based on the Zhelavskaya et al. (2019) study.

The main contributions of this dissertation can be summarized as follows:

1. The plasmasphere:

    1. Developed the first automated neural network-based algorithm for reconstruction of plasma density from satellite electric field measurements.
    2. Developed a new global empirical model of plasmaspheric electron density based on geomagnetic and solar wind parameters capable of accurately capturing the dynamic evolution of the plasmasphere.
    3. Quantified the influence of different groups of solar wind parameters, geomagnetic indices and their times histories on the model performance.
    4. Identified the critical combination of geomagnetic and solar wind parameters and their time history that determine the evolution of the cold plasma.
    5. Developed a new approach for combining a neural network-based empirical model and a physics-based model of the plasmaspheric electron density in an optimal way, allowing optimal performance during quiet, disturbed, and extreme geomagnetic conditions.

2. The Kp index:

    1. Developed new tools for predicting the Kp geomagnetic index for up to 12 hours ahead.
    2. Determined the most important solar wind drivers to predict the Kp index by using feature selection algorithms.
    3. Analyzed the prediction accuracy that is set by using solar wind measurements at L1 as input to a Kp predictive model.

Chapter 2

Reconstruction of plasma electron density from satellite measurements via Artificial Neural Networks

Note

The following work is published as a book chapter in *Machine Learning Techniques for Space Weather*, Elsevier:
Zhelavskaya I. S., Y. Y. Shprits and M. Spasojević (2018). Chapter 12 - Reconstruction of Plasma Electron Density From Satellite Measurements Via Artificial Neural Networks, in Machine Learning Techniques for Space Weather, edited by Enrico Camporeale, Simon Wing and Jay R. Johnson, Elsevier, pp. 301-327, ISBN 9780128117880, `https://doi.org/10.1016/B978-0-12-811788-0.00012-3`.

Abstract

This chapter presents a new approach to reconstruction of plasma electron density from satellite measurements in an automated fashion using artificial neural networks. We design a feedforward neural network to derive the upper hybrid resonance frequency, f_{uhr}, from satellite measurements that is subsequently used to calculate the electron density. In previous studies, the reconstruction of f_{uhr} was either performed

manually or by semi-automated techniques. In this study, we use 2.5 years of electric and magnetic field measurements collected with the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) instrumentation suite of the Van Allen Probes mission to train, validate, and test a neural network. We then apply it to more than 4 years of EMFISIS data and produce the publicly available electron density data set. We describe the aspects of neural network design and implementation and perform analysis of the obtained electron density distribution.

2.1 Overview

Plasma electron density is a crucial parameter in space physics simulations and modeling and is important for predicting and preventing hazardous effects of space weather, such as satellite damage or even complete breakdown due to enhanced solar wind activity. Measuring plasma density accurately, however, has always been a challenge. One of the most accurate methods of measuring plasma electron density is to derive it from the satellite measurements of the upper-hybrid resonance frequency. The upper-hybrid resonance frequency is often associated with the most pronounced resonance band in dynamic spectrograms, which display electric power spectral density measured by a satellite as a function of frequency and time. In previous missions, upper-hybrid resonance bands were manually identified in the dynamic spectrograms, although such manual determination is a very tedious and time-consuming process. Moreover, as new satellites for scientific exploration are being launched and more data become available, manually identifying upper-hybrid resonance frequency becomes unfeasible. Research has been done in the past to automate the process of upper-hybrid resonance band extraction, but the developed algorithms still require significant manual intervention and correction.

In this chapter, we present an alternative approach for automated electron density determination based on artificial neural networks. The method employs feedforward neural networks to derive the upper-hybrid resonance bands from the dynamic spectrograms, and hence electron density, in an automated fashion. Neural networks are a powerful tool for finding the multivariate nonlinear mapping from input (in this case, dynamic spectrograms and other geophysical parameters) to output parameters (the upper-hybrid frequency). Neural networks inherently require a training data set, i.e.,

a data set for which both inputs and outputs are known. We use electric and magnetic field measurements produced by the two satellites of NASA's Van Allen Probes mission, currently the golden standard of measurements in space weather research, as input, and a large data set of upper-hybrid frequency measurements produced by another recently developed semi-automated technique as output in our training data set.

The chapter is organized as follows. First, we provide the necessary background on the space weather aspects of this study related to plasma density and its importance in space physics research and delve deeper into the motivation behind this application. Then, we give a brief overview of feedforward neural networks, the type of artificial neural networks used for this application, with a focus on the importance of model validation. Next, we describe the algorithm implementation in detail and demonstrate the results. We also discuss how the developed plasma electron density data set can be used to develop a global empirical plasma density model that does not depend on satellite measurements, also using neural networks.

2.1.1 Space weather-related aspects and motivation

Plasma density and the plasmasphere

Plasma electron density is a parameter characterizing a number of particles in a unit volume in space (measured in cm^{-3}) (Cohen, 2007). The electron density is a fundamental parameter of plasma. The focus of this study is the density of cold particles in the near-Earth space environment. These cold particles (of temperature ~ 1 eV) are trapped by the closed magnetic field lines of the Earth, forming a bubble-shaped region around the Earth. This region is called the plasmasphere (Lemaire and Gringauz, 1998). The plasmasphere is a relatively dense region of plasma compared to other regions in space (density of $10 - 10^4$ cm^{-3}). The plasmasphere extends from the topside ionosphere ($\sim$ 1000 km above the ground) out to a boundary called the plasmapause that ranges from 2 to 7 Earth radii, R_E ($1\ R_\mathrm{E} = 6,371$ km), depending on geomagnetic conditions (Gringauz, 1963; Carpenter, 1963; Grebowsky, 1970). The region outside the plasmapause is called the plasma trough, and it is a low-density region.

The plasmasphere is very dynamic, and its shape and size strongly depend on solar and geomagnetic conditions (O'Brien and Moldwin, 2003; Chappell et al., 1970b).

Two mechanisms, sunward convection and corotation with the Earth, determine the configuration of the plasmasphere (Darrouzet et al., 2009; Singh et al., 2011). The corotation regime dominates during quiet geomagnetic times; plasma material trapped inside the closed magnetic field lines corotates with the Earth (Carpenter, 1966). Meanwhile, the plasmasphere is refilled with ions from the topside ionosphere and expands up to $\sim 4 - 7\ R_{\mathrm{E}}$ (Goldstein et al., 2003b; Singh and Horwitz, 1992; Krall et al., 2008); its shape is roughly circular with a bulge on the dusk side (Nishida, 1966). In contrast, during periods of high geomagnetic activity, the sunward magnetospheric convection starts to dominate and erodes the plasmasphere: the closed magnetic field lines at the dayside magnetopause boundary are torn apart and the plasmaspheric material is carried sunward. Due to that, the outer layers of the plasmasphere are eroded and plasmapause contracts (Carpenter, 1970; Chappell et al., 1970a; Goldstein et al., 2003b). The stronger the disturbance, the more the plasmapause contracts (down to 2 R_{E} during severe geomagnetic storms).

Plasma density of the plasmasphere is a critical parameter in a number of important space weather applications such as GPS navigation (e.g., Mazzella, 2009; Xiong et al., 2016) and analysis of spacecraft anomalies due to spacecraft charging (e.g., Reeves et al., 2013). Plasma density is also a critical input parameter for quantifying wave-particle interactions necessary for modeling the formation and decay of Earth's radiation belt, a donut-shaped region around the Earth, hazardous for satellites electronics and crew in space (e.g., Spasojević et al., 2004; Thorne et al., 2013; Orlova et al., 2016; Shprits et al., 2016).

Plasma electron density can be measured on satellites using several methods. They include measuring the density directly with particle counters (e.g., Geiger and Müller, 1928), determining it using the spectral properties of waves (e.g., Trotignon et al., 2003) or deriving it from the spacecraft potential (e.g., Escoubet et al., 1997). A number of empirical models of plasma electron density have been developed using electron density measurements from previous missions.

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 of Carpenter and Anderson (1992) is based on electron density measurements derived from radio measurements made with the sweep frequency receiver (SFR) onboard the International Sun-Earth Explorer (ISEE-1) spacecraft and ground-based whistler measurements. This model presents the mean electron density

values for different L after several days of refilling, which means that it is applicable only for quiet geomagnetic activity. The model is valid for L shells from 2.25 to 8 R_E and local times between 0 and 15 MLT (here, L can be roughly considered as the distance from the center of the Earth, and MLT stands for magnetic local time and can be considered as an angular distance around the Earth from the local midnight). 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. (2000)), using transition equations in order to obtain a more comprehensive description of the plasma in the inner magnetosphere. The plasmasphere and plasma trough density models of Sheeley et al. (2001) present statistical density averages based on density measurements obtained using the Combined Release and Radiation Effects Satellite (CRRES) swept frequency receiver. The models are valid for L shells between 3 and 7 and all local times. The Sheeley et al. (2001) study provides the mean and the standard deviation of density in the plasmasphere and the trough to represent depleted and saturated density levels for different L (and MLT for the trough).

Despite the extensive use of these empirical density models in space physics simulations, they do not provide reliable electron density estimates during extreme events, such as geomagnetic storms, since they are parameterized only by static geomagnetic parameters such as L and MLT. The described models do not include the dynamic dependence of plasma density on geomagnetic and solar conditions, and plasma electron density is known to be highly variable during elevated geomagnetic activity (Park and Carpenter, 1970; Moldwin et al., 1995). Therefore, collecting reliable electron density measurements during varying geomagnetic conditions is still of continuing interest.

Determining the Electron Density from Upper-Hybrid Band Resonance Frequency

One of the most reliable techniques to measure the electron density is to use the upper-hybrid resonance frequency to derive it (Mosier et al., 1973). This method will be further employed in this study.

The upper-hybrid resonance frequency is a combination of the electron plasma frequency and the electron cyclotron frequency:

$$f_{\mathrm{uhr}} = \sqrt{\left(f_{\mathrm{ce}}^2 + f_{\mathrm{pe}}^2\right)}. \tag{2.1}$$

The electron plasma frequency and the electron cyclotron frequency are given as:

$$f_{\mathrm{pe}} = \frac{1}{2\pi}\sqrt{\frac{q_{\mathrm{e}}^2 n_{\mathrm{e}}}{m_{\mathrm{e}}\epsilon_0}}, f_{\mathrm{ce}} = \frac{|q_{\mathrm{e}}|\, B}{2\pi m_{\mathrm{e}}}, \tag{2.2}$$

where B is the magnetic field strength, n_{e} is electron density, q_{e} is the charge of electron, ϵ_0 is the permittivity of free space, and m_{e} is the mass of an electron. The upper boundary of the upper hybrid emission is generally the most pronounced feature in spacecraft plasma wave data. Mosier et al. (1973) found that the upper-hybrid resonance can often be observed visually as the brightest emission band in the dynamic spectrograms, displaying spectral properties of the electric field (see Figure 2.1 for more details). In this study, we use the plasma wave data measured on the Van Allen Probes satellites.

Van Allen Probes is a dual-spacecraft NASA mission launched in August 2012, and its scientific objective is to explore the dynamic evolution of the Van Allen radiation belts (Mauk et al., 2013). The satellites have a highly elliptical orbit in a near-equatorial plane (inclination 10 degrees) with the apogee $30,414$ km and perigee 618 km. A number of science instruments are deployed on the satellites. One of them is the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) (Kletzing et al., 2013). The instrument performs routine measurements of the electric field in the frequency range of 10 to 487 kHz, thus providing the capability to determine the upper hybrid resonance band, and hence to accurately estimate the electron density. An example of measurements from the High Frequency Receiver (HFR) of the EMFISIS instrument on board Van Allen Probe A for one orbital pass #1612 for 25 April 2014, is illustrated in Figure 2.1. Figure 2.1a shows the power spectral density of the electric field as a function of frequency and time, where color indicates the spectral density, as noted in the color bar. Such a plot is referred to as a dynamic spectrogram. Dynamic spectrograms are a tool to explore spectral properties of waves and resonances in plasma. Figure 2.1b shows the same spectrogram but with the upper-hybrid resonance frequency indicated with black. The upper hybrid resonance frequency is often associated with the most pronounced band in dynamic spectrograms, as found by Mosier et al. (1973). The background magnetic field strength, B, is measured by the magnetometer on board the spacecraft, and therefore, f_{ce} can be determined directly; it is indicated on the spectrogram as the dashed curve.

Several processes and phenomena that might at times impose challenges on the upper-hybrid resonance identification are denoted with white in Figure 2.1b. The plasmapause, the outer border of the plasmasphere, is usually identified by a sharp density gradient and can be monitored via the sharp drop in the upper-hybrid frequency. The electron density (hence, the upper-hybrid frequency) might have a complex structure in the vicinity of the plasmapause, and the density gradient is not always smooth but may sometimes have jumps. Accurately identifying those sudden changes in density may be challenging for some automated routines. Another phenomenon is the banded emissions, also referred to as "$(n + \frac{1}{2})f_{ce}$" emissions. These emissions are often observed between the harmonics of electron cyclotron frequency, f_{ce}, but not necessarily exactly in the middle between them (LaBelle et al., 1999), in the plasma trough (the low-density region). In some cases, emissions at f_{uhr} are not observed with the banded emissions (Benson et al., 2001); this brings uncertainty in the process of identification of the upper-hybrid resonance and presents challenges for making definite determination of f_{uhr} without performing an appropriate visual spectral interpretation. Instrument contaminations can also present challenges for automated routines of f_{uhr} identification. They are usually observed as horizontal lines of roughly identical spectral density and might have intersections with the upper-hybrid resonance.

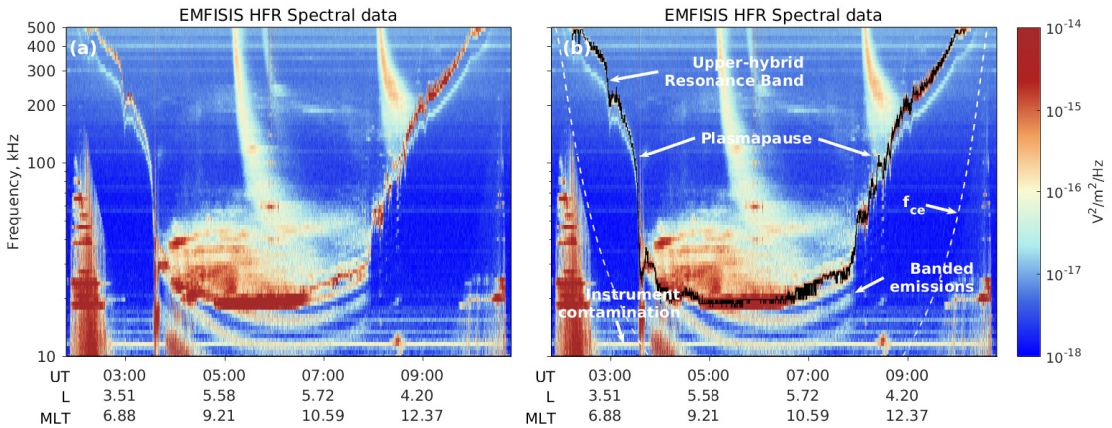

Figure 2.1: An example of the EMFISIS HFR spectral data for one orbit pass (a) with various parameters and phenomena denoted (b). Upper hybrid frequency is shown with the black curve; electron cyclotron frequency, f_{ce}, is shown with the white dashed curve.

Previously, the upper-hybrid resonance band has been manually derived from

dynamic spectrograms (e.g., LeDocq et al., 1994) and there have been several semi-automated techniques developed. Research on the development of semi-automated routines began in the ISEE-1 era (Trotignon et al., 1986) and still continues (Trotignon et al., 2010; Denton et al., 2012; Kurth et al., 2015). Determination of the electron density in Trotignon et al. (2010) was conducted using the active and passive wave spectra measured with the Waves of High frequency and Sounder for Probing of Electron density by Relaxation (WHISPER) instrument onboard the Cluster mission. The electron density in the work of Denton et al. (2012) was derived from the passive radio wave observations obtained from the Radio Plasma Imager (RPI) instrument onboard the IMAGE (Imager for Magnetopause-to-Aurora Global Exploration) spacecraft. The Automated Upper hybrid Resonance detection Algorithm (AURA) developed by Kurth et al. (2015) is based on the Van Allen Probes' EMFISIS HFR data and is a semi-automated algorithm to derive the upper hybrid resonance band in dynamic spectrograms. The algorithm searches a peak in the spectrum for every time step while assuming that each successive spectrum contains a peak associated with f_{uhr} near the previously determined peak. An operator then visually inspects the dynamic spectrogram for each orbital pass and corrects the resulting f_{uhr} profile identified by AURA where it is necessary. AURA significantly facilitates the processing of the HFR spectral data but still requires manual intervention.

In this chapter, we present an alternative algorithm for automated determination of f_{uhr}, hence electron density, from the satellite measurements using artificial neural networks, which is also described in Zhelavskaya et al. (2016). Neural networks are one of the most commonly used tools for a broad range of nonlinear approximations and mappings. In this application, we use feedforward neural networks (FNNs). FNNs are very efficient in solving nonlinear multivariate regression problems. The neural networks are "tuned" to a specific problem during the training using the training data set. The training data set is a data set for which both inputs (in our case, satellite measurements and geophysical parameters) and outputs (f_{uhr}) are known. We use electric (HFR spectra) and magnetic field (f_{ce}) measurements from Van Allen Probes as inputs in our training set, and the database of f_{uhr} measurements developed using AURA (courtesy of W. Kurth and the EMFISIS team) as output. The training data set covers $1,091$ orbital passes, i.e., represents a significant set of example data. After the neural network passes the training, validation and testing stages, it can be used in practice and be applied to a data set for which the f_{uhr} is not known. The neural

network is applied to a database of 3,750 orbital passes and its output is then used to produce a database of electron number density. The performance of the resulting neural network model is assessed by comparing the derived density to the density obtained in Kurth et al. (2015). The resulting density distribution is also analyzed and compared to the empirical density models of the plasmasphere and trough by Sheeley et al. (2001).

2.1.2 Brief background on neural networks

This part provides a brief overview of artificial neural networks and discussion on important aspects of training and validation. The notions introduced here might be useful for further understanding the application described in this chapter. For more details on neural networks, the reader should refer to the works cited herein.

Basic concepts related to neural networks

Artificial neural networks are a family of models effective at solving problems of function approximation, pattern recognition, classification, and clustering. Artificial neural networks were inspired by biological neural networks (in the brain) and are an attempt to mimic them in a very simplified manner (e.g., McCulloch and Pitts, 1943; Hebb et al., 1949; Marr et al., 1976).

Neural networks are composed of multiple simple computational blocks called artificial neurons. An artificial neuron has a body in which computations are performed, and a number of input channels and one output channel, similar to a real biological neuron. Simply put, a neuron receives an input signal and then computes an output on it. Figure 2.2a shows the construction of an artificial neuron with N inputs. Every input has a weight associated with it; the larger the weight, the more impact the corresponding input channel has on the output. A neuron also has a bias, which for convenience can be considered as an additional input to the neuron, x_0, that is equal to 1 and has the weight identical to the value of the bias, $w_0 = b$. Additionally, a neuron has a transfer or an activation function that defines the type of neuron. The activation function can be arbitrary; the most commonly used functions are sigmoid, hyperbolic tangent, binary, and linear. After the signal is applied to the neuron, it first computes the sum of inputs multiplied by their weights and then applies the transfer function to the resulting sum.

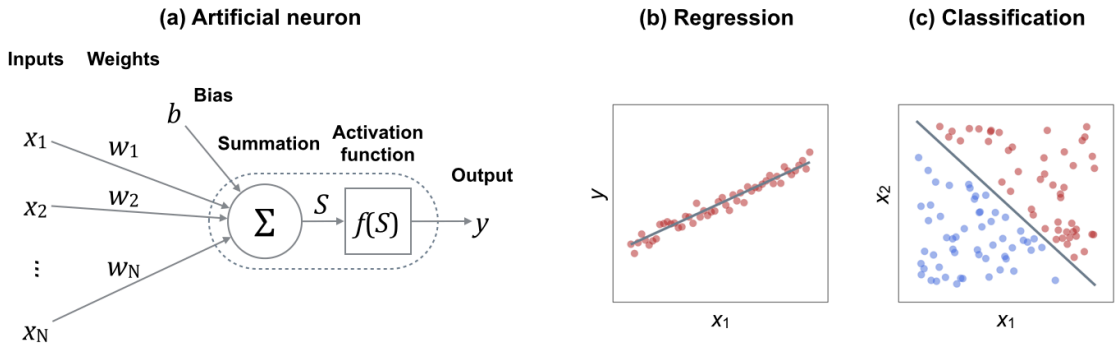

Figure 2.2: A scheme of an artificial neuron (on the left), a building block of a neural network. Artificial neurons can be used to solve linear problems. Simple examples of such problems for regression and classification are shown on the right.

An artificial neuron is one of the first computational models developed in the research area of artificial neural networks (McCulloch and Pitts, 1943; Rosenblatt, 1957). A single neuron can be used to solve a limited number of problems, such as linear regression and classification of two linearly separable subsets. Two toy examples related to these problems are shown in Figures 2.2b and 2.2c. Figure 2.2b illustrates a simple case of a linear regression problem with one dependent and one independent variable. The given data are plotted with red markers, where independent variable x_1 is plotted on the x-axis vs dependent variable y on the y-axis. The regression task consists in finding a mapping from x_1 to y. The gray line shows a linear fit to the data found by a single neuron. The neuron has a single input in this case. Figure 2.2c shows a simple case of the classification problem. Here, the given data consists of two classes shown with red and blue markers correspondingly. There are two independent variables, x_1 and x_2. They are plotted on x and y axes correspondingly, and the single dependent variable y is represented by color (numerically it can be represented as 0 and 1, or -1 and 1, for example). The classification task is to identify the class of each data point, given x_1 and x_2. This task can be solved by building a separation border between two classes. The gray line shows a separator found by a neuron. The neuron has two inputs in this case. These scenarios are intentionally oversimplified for demonstration purposes and can be expanded to more dimensions (hence, more inputs to the neuron). However, a single neuron cannot be used to solve more complex, nonlinear problems. Fortunately, such problems can be solved by neural networks, which are composed of multiple neurons. The main concept of artificial neural networks is that an output signal from one neuron can be used as an

input to other neurons.

The way neurons are connected into a network defines the topology or architecture of a neural network. In this work, we use a feedforward neural network architecture (FNN). An FNN is one of the most basic and widely used neural network architectures and is effective at solving multivariate nonlinear regression and classification problems. FNNs have displayed the state-of-the-art performance in a number of applications (e.g., Salakhutdinov and Hinton, 2009; Krizhevsky and Hinton, 2011; Glorot et al., 2011; Mohamed et al., 2012). The topology of an FNN is shown in Figure 2.3a. Neurons in an FNN are arranged in layers. Three types of layers exist: input, output, and hidden layers. The input layer is composed of inputs to the network and no computations are performed in this layer. Next, follow hidden layers that are composed of any number of neurons arranged in parallel. The network can have several hidden layers. The neurons of the same layer are not connected to each other but connected to the neurons of the preceding and the subsequent layers; an output of one hidden layer serves as an input to the following layer. The output layer, and hence the network output, is formed by a weighted summation of the outputs of the last hidden layer. Neurons of one layer have the same activation function. Different layers can have different activation functions. Formally, an FNN with L hidden layers can be defined as a superposition of L activation functions $f_1, ..., f_L$.

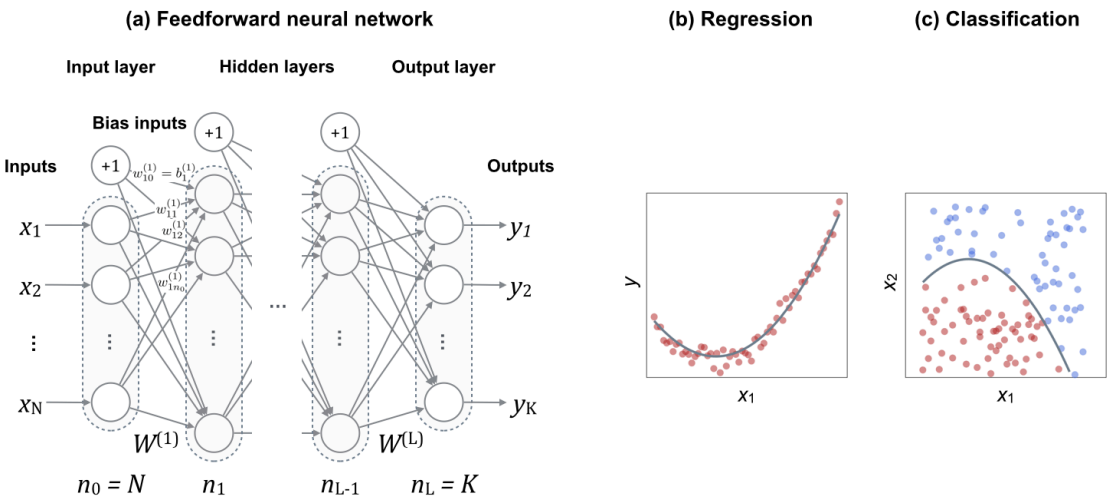

Figure 2.3: (a) A scheme of a feedforward neural network. Circles denote artificial neurons. Feedforward neural networks arrange neurons in a layered configuration and can be used to solve nonlinear problems. (b, c) Simple examples of such problems for regression and classification.

The neural network with at least one hidden layer can solve nonlinear regression or classification problems (Cybenko, 1989). Simple illustrations related to regression and classification are shown in Figures 2.3b and c. The x and y axes are the same as in Figures 2.2b and c, correspondingly. The neural network can fit the nonlinear function to given data (in the case of regression, Figure 2.3b) and determine a nonlinear separator between classes (in the case of classification, Figure 2.3c). The given examples are idealized and can be expanded to more dimensions.

The explanation above assumes that weights and biases of the neural network are known. In practice, the weights and biases are not given; however, they can be determined using a training data set, i.e., a set of data for which inputs and outputs are known. Determining weights and biases is usually referred to as training and reduces to an optimization problem of minimizing a given cost function. The cost function is defined based on the type of application or problem we are attempting to solve. Specifically for neural networks, a number of backpropagation algorithms are used to solve the optimization problem of determining the weights (Williams and Hinton, 1986). In this study, we employ the scaled conjugate gradient backpropagation algorithm (SCG) (Møller, 1993). This algorithm is based on conjugate gradient methods (CG), a class of optimization techniques. It works faster than most of the other algorithms for neural networks of large sizes and has relatively modest computer memory requirements (Mathworks.com, 2015).

Neural network design flow

The neural network design flow consists of three main stages: training, validation, and testing. Consequently, the set of data for which the output is known is split into three parts, training, validation, and test data sets, in a ratio defined by a user.

The goal of the training stage is to find the weights of the neural network, which minimize the difference between the output of the training data set and the output of the neural network. An optimization algorithm is run in order to determine such weights. Before performing the training procedure, the internal parameters of the neural network such as the number of hidden layers, the number of neurons in each hidden layer, and activation functions in each layer have to be selected. Typically, multiple neural networks with different internal parameters are trained, and in the validation stage, the model having the minimal validation error is selected for further use.

The main objective of the validation stage is to check the neural network ability to reconstruct the relation between inputs and outputs on the data it has not seen yet, i.e., data not used for training. This is usually called the generalization ability of a neural network. In this stage, we measure the performance of every neural network obtained during the training stage on the validation data set and select the model with the best performance (by comparing the output of the models to the known output of the validation set). The importance of the validation stage is described in the next subsection.

Finally, in the testing stage, the performance of the optimal model determined in the validation stage is assessed on the test set. The calculated error of the neural network is then treated as the resulting error of the model. After the neural network is tested and the obtained results are satisfactory, it can be used on the data for which the output is not known.

Importance of validation

Although neural networks are a powerful tool for building accurate multivariate nonlinear approximations, they are also inherently easy to overfit. Overfitting occurs when the model becomes too complex for the given task and therefore becomes capable of fitting the training data excellently but does not produce a reliable output on the unseen data. The more hidden layers and neurons the model has, the more complex it is. The ultimate goal of building a practical model is to find the optimal internal network parameters. The optimal internal network parameters are the parameters producing a model that generalizes well and avoids overfitting the training data, at times at the cost of a slightly reduced accuracy of the model. The optimal model has similar errors on both training and unseen data and is acceptable for the given application.

To demonstrate this, let us consider three possible scenarios that might occur in practice: overfitting, underfitting, and desirable model performance. Underfitting is directly opposite to overfitting: the model is too simple and cannot produce a reliable fit to the given data; its errors on the training and validation data are similar, but at the same time are very high. Figure 2.4 shows toy illustrations corresponding to these scenarios for the case of simple nonlinear regression problem. Here, similar to the illustrations in Figures 2.3b and c, the given data has one independent variable, x_1 (one input) and one dependent variable, y (one output), to be modeled. The variables

x_1 and y are plotted on the x and y axes correspondingly. The red markers denote the training data points, which are used to find the mapping from x_1 to y; the blue markers denote the validation data points that are not shown to the network during the training. The gray curve in each plot shows the resulting fit to the training data produced by the corresponding neural network. In the overfitting scenario (Figure 2.4a), the model is too complex and produces a fit that passes through every point of the training data set. Therefore, the model has a poor performance on the validation set and is unacceptable for use in practice. In contrast, in the case of underfitting (Figure 2.4b), the model is too simple and fits a straight line to the parabolic shape of the given data. The errors of the model both on the training and validation sets are unacceptably high, which is also not desirable in practice. Ideally, the model has an optimal complexity for the given data and learns the general dependency of the input, x_1, on the output, y, although it might have some error due to the presence of noise in data.

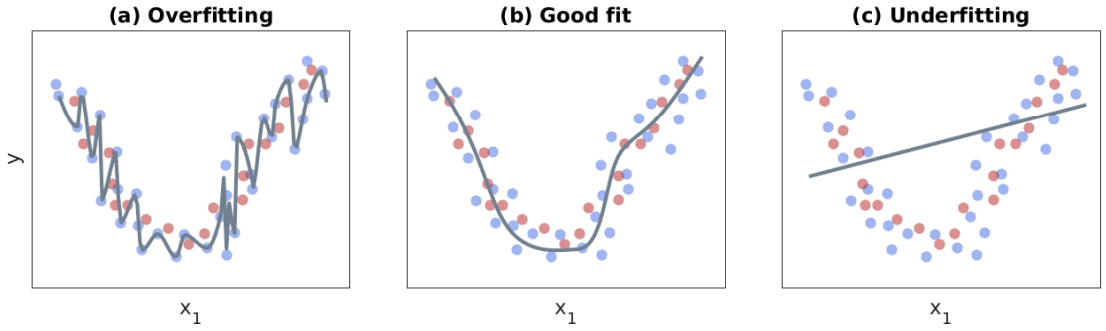

Figure 2.4: Toy illustrations of three different scenarios that can occur after the training: (a) overfitting, (b) ideal case, and (c) underfitting. Red points represent the data to be fitted with a function. Gray curves show the functions fitted to the data in these three different scenarios.

In practice, the number of inputs to the model can be much larger, reaching hundreds or even thousands of input parameters for some applications, as well as the size of the training set (reaching hundreds of thousands or millions data points). Visualization of the results in a manner shown in Figure 2.4 becomes impossible, and the quantitative assessment of the results is the only way of assessing the model performance. Such quantitative assessment is performed in the validation stage by measuring the error of the model on the training and validation sets and comparing them. If the training error is small while the validation error is large, then the model is likely to overfit the data. The contrary, when both errors are too large although

they are similar, might be a sign of underfitting. Ideally, the training and validation errors should be similar and acceptable for the given application. The same applies to classification.

2.2 Implementation of the algorithm

In this section, we describe the details of the application of feedforward neural networks to the plasma wave data. The inputs to the network include electric field spectral properties obtained from the satellite measurements, location of the satellite, and the parameter characterizing geomagnetic activity (Kp index). The network has a single output, the upper-hybrid frequency, from which plasma density can be easily derived (see formulae (2.1) and (2.2)). We use electric field measurements from Van Allen Probes for the input and database of f_{uhr} developed by Kurth et al. (2015) as the output of the training data set. The resulting neural network is capable of reconstructing the density from the satellite measurements, hence along the satellite orbit. The main concept of the algorithm and an example output for one orbital pass are shown in Figure 2.5. Further here, we describe the details of input and output data, neural network architecture, and validation. The developed algorithm is further referred to as the Neural-network-based Upper hybrid Resonance Determination (NURD) algorithm.

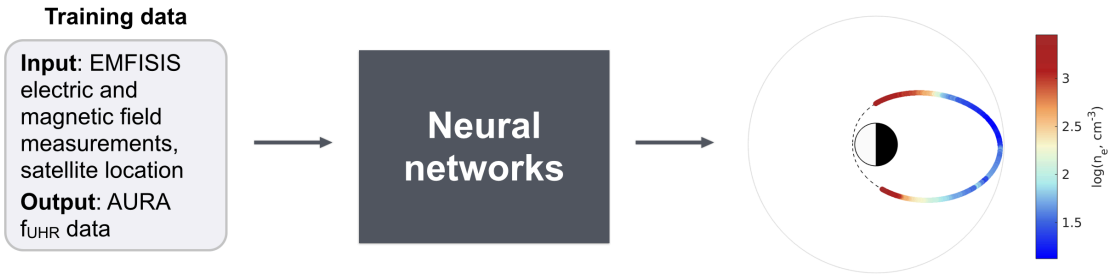

Figure 2.5: Schematic presentation of the NURD algorithm. We use EMFISIS electric and magnetic field measurements as input and f_{uhr} obtained with AURA as output to the neural network. The NURD algorithm is then able to reconstruct the upper hybrid frequency and hence density along the satellite orbit. An example of such a reconstruction is shown in the rightmost figure.

2.2.1 Training data set

Input data

The inputs to the neural network are listed in Table 2.1. A detailed explanation and rationale behind each input variable are given below.

The first 82 inputs to the model include the electric field measurements from the EMFISIS HFR instrument, specifically the electric field power spectral density produced in the survey mode of the instrument. The instrument produces measurements in the frequency range between 10 to 487 kHz binned into 82 logarithmically spaced bins with a 6-second temporal resolution. The frequency spacing provides approximately 5% spectral resolution, $\frac{\Delta f}{f}$, which defines the resulting density resolution, $\frac{\Delta n}{n}$. As previously discussed, n e is proportional to f_{pe}^2, which results in 10% density resolution. Furthermore, the upper frequency limit (487 kHz) restricts the maximum density that is possible to derive to ~ 2900 cm^{-3}, and the lower frequency limit (10 kHz) restricts the minimum density to ~ 1 cm^{-3}. The logarithm of the power spectral density of the electric field for the 82 frequency bins is used as input to the model.

The next input to the model is the logarithm of electron cyclotron frequency, f_{ce}. The electron cyclotron frequency can be directly derived from the background magnetic field, $|B|$: $f_{\mathrm{ce}} = 28|B|$, where $|B|$ is measured in nanotesla, and f_{ce} is measured in hertz. We use measurements of the background magnetic field, B, from the EMFISIS fluxgate magnetometer.

Since plasma density is known to vary spatially, we also use the spacecraft coordinates as input to the model (84 and 85 in the table). The Van Allen probes have low inclination orbits in the near-equatorial plane, therefore we consider two spatial coordinates corresponding to the position of the satellite in the equatorial plane. The L parameter reflects roughly the radial distance from the center of the Earth and is measured in units of Earth radii, R_{E}. More precisely, L denotes the distance at which the magnetic field line where the measurement was made crosses the equator. MLT stands for magnetic local time and represents the angular distance from the magnetic local midnight measured in decimal hours (0 to 24).

Plasma density dynamics also depends on the level of geomagnetic activity. Therefore, a parameter characterizing the global level of geomagnetic activity, the Kp index, is used as input to the model. The Kp index (originating from planetarische Kennziffer, German for planetary index) is derived from measurements of magnetic field on

Table 2.1: Inputs to the neural network.

#	Input name	Description	Units
1-82	$\log_{10}$ spectrum	Decimal logarithm of the spectrum	$\log_{10} V^2/\mathrm{m}^2/\mathrm{Hz}$
83	$\log_{10} f_{\mathrm{ce}}$	Decimal logarithm of electron cyclotron frequency	$\log_{10}$ Hz
84	L	Magnetic field line	Earth radii
85	MLT	Magnetic local time	Hours (024)
86	Kp index	Geomagnetic index	Unitless (09)
87	f_{binmax}	Frequency bin with the largest power spectral density of the electric field from the HFR spectrum	Unitless (1, 2, ... , 82)

the ground observatories located in different parts of the world and is an indicator of global geomagnetic disturbances. It has a 3-hour cadence and ranges from 0 (lowest geomagnetic activity) to 9 (severe geomagnetic storms) in discrete steps.

The last input, f_{binmax}, corresponds to the frequency bin with the largest spectral density in the HFR spectrum. We use this parameter as input since it can serve as a rough initial approximation of f_{uhr}. This assumption might introduce errors to the neural network since as shown in Benson et al. (2004) the upper hybrid frequency f_{uhr} is generally associated not with the maximum emission peak of the upper hybrid band but with its upper frequency edge. However, the analysis in Benson et al. (2004) was performed using active and passive observations measured by the IMAGE Radio Plasma Imager (RPI), which had a higher frequency resolution compared to EMFISIS HFR and therefore allowed for very accurate f_{uhr} determinations. The coarse frequency resolution of the EMFISIS HFR does not allow resolving the upper frequency edge precisely.

Output data

The neural network has a single output: the logarithm of the upper hybrid frequency, f_{uhr}. The data set of upper-hybrid frequencies derived with the AURA algorithm (Kurth et al., 2015) is used in the training. AURA is a semi-automated routine for determination of the f_{uhr} band in the dynamic spectrograms. AURA uses a restricted search approach for finding the maximum value of the spectrum at a specific time point. In this algorithm, the density derived from the spacecraft potential (using

EFW instrument) is used to weight the probability of a possible maximum peak occurrence in the spectrum and guide the search procedure. The weighting parameters are optimized using a bootstrapping method so that the determined spectral peaks are more likely to correspond to f_{uhr}. After AURA is run for one orbit, the operator visually detects where AURA has failed to identify the correct UHR band and manually corrects it. The produced f_{uhr} measurements have a high quality since the resulting f_{uhr} have been visually checked for each orbit. The accuracy of the AURA algorithm is $\approx 10\%$.

The UHR frequency data set derived by AURA and used for training was obtained from the EMFISIS instrument website. The available UHR data set derived by AURA at the moment of developing the algorithm consists of $1,091$ orbital passes and covers the period from October 2012 to March 2015. With a 6-second cadence and 9-hour orbital period, the data set comprises $\sim 5,900,000$ measurements. However, not all measurements can be used for training, since the UHR frequency may be greater than the upper limit of the HFR frequency range near the perigee. We exclude the out-of-frequency range measurements from the training data set be performing the following procedure. We consider the spectrograms for each orbital pass separately. Firstly, the left- and rightmost edges of the upper hybrid resonance line are identified by determining the bins with the highest values of the spectral density in the uppermost frequency spectral bin (at 487 kHz, near the perigee). We then apply the neural network only to the measurements within these boundaries. After the out-of-frequency range part of the data is excluded, the total number of measurements available for training, validation, and testing reaches $4,027,610$ measurements. An $87 \times 4,027,610$ matrix of input variables from Table 2.1 (usually referred to as design or feature matrix) and a $1 \times 4,027,610$ vector of output values (measurements of f_{uhr}) are constructed using this data set.

2.2.2 Neural network architecture

We use a feedforward neural network with a single hidden layer shown in Figure 2.6. The network has 87 input neurons as defined by the number of input variables (dimensions in the design matrix). The number of neurons in the hidden layer is initially not known and is determined during the validation procedure, as described further. For this application, 80 neurons in the hidden layer were found to be optimal

in the validation stage. The network has only one output neuron since we are to model only one variable, f_{uhr}.

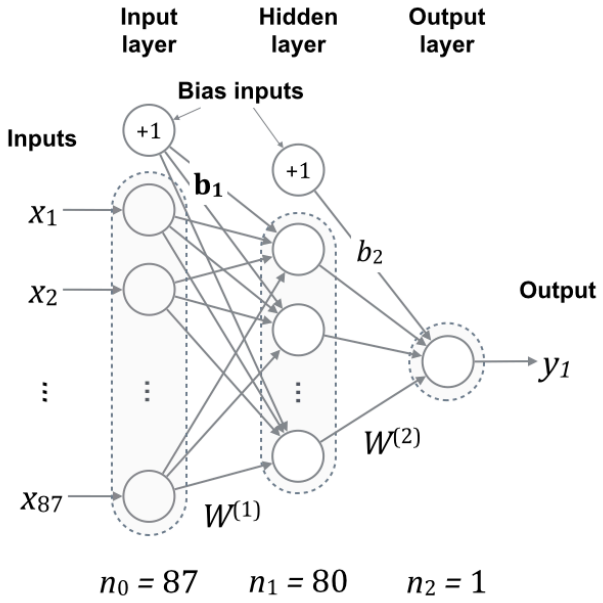

Figure 2.6: The optimal architecture of the feedforward neural network determined in the validation stage. The neural network has 87 inputs, one output, and 80 neurons in the hidden layer. The weights of the neural network can be considered as 80×87 $W^{(1)}$ and 1×80 $W^{(2)}$ matrices; the biases can be represented as an 80×1 vector $\mathbf{b}_1$ for the hidden layer and as a scalar b_2 for the output layer (adapted from Zhelavskaya et al. (2016)).

First, before being fed to the neural network, the input data is normalized to the common range (here, $[0; 1]$). Since the range of some variables could be much larger than others, large value inputs can dominate the input effect compared to small value inputs and influence the accuracy of the neural network (Li et al., 2000). Normalization ensures that inputs have identical initial importance to the neural network. Normalization is performed on the training set independently of validation and test data sets; the normalization parameters calculated on the training set are stored and applied afterward to the validation and test data sets (and not recalculated on these sets).

Input layer does not perform any computations; its function is to distribute the incoming signal to the neurons of the following hidden layer. The neurons of the hidden layer have the hyperbolic tangent activation function, which transforms the input to the range of $[-1; 1]$. The output of the hidden layer is then the following

vector:

$$\mathbf{a}_1 = \tanh(\mathbf{b}_1 + W^{(1)}\mathbf{x}), \tag{2.3}$$

where $\mathbf{x}$ is an input vector (87×1, since there are 87 inputs), $\mathbf{b}_1$ is the vector of biases (80×1, since there are 80 hidden neurons), $W^{(1)}$ is the matrix of weights connecting the input to the hidden layer (dimensionality: 80×87). The transfer function of the output neuron is linear. The output of the neural network is then given as:

$$a_2 = b_2 + W^{(2)}\mathbf{a}_1, \tag{2.4}$$

where $\mathbf{a}_1$ is the output vector of the hidden layer, b_2 is a bias (a scalar), and $W^{(2)}$ is the matrix of weights connecting the hidden to the output layer (dimensionality: 1×80).

2.2.3 Steps of the design flow

The design matrix comprising $4,027,610$ measurements is randomly divided into 3 parts: training, validation, and test sets, in a $34 : 33 : 33\%$ ratio, respectively. A large subset of the available data is allocated to the validation and testing. Such division allows for a more reliable test of the generalization ability of the network. The validation set is used to determine the optimal number of neurons in the hidden layer of the network. The test set is used to estimate the accuracy of the resulting neural network model. All three sets are kept separate during the respective stages, i.e., the validation and test sets are not used in training (and the test set is not used in training or in validation). It is worth noting that although in this work the division of data into training, validation, and test sets is performed in a random fashion, the more warranted way to perform division for the time series is to split data sequentially. The sequential split ensures independence of all three subsets, while random splitting might yield an optimistic evaluations on the validation and test sets for the events lying outside the time period covered by the data set. The network resulting from training performed on the random split, however, would still perform well for the reconstruction of the past events.

In order to determine the optimal number of neurons in the hidden layer, we train five neural networks with a different number of hidden neurons ranging from 40 to 120. The networks are trained on the training set of data using a conjugate gradient

backpropagation algorithm. After the networks are trained, their performance is measured on the validation set. We use the mean absolute percentage error (MAPE) to assess the performance of the networks:

$$MAPE = \frac{1}{M}\sum_{i=1}^{M}\frac{\left|f_i^{\mathrm{AURA}} - f_i^{\mathrm{NURD}}\right|}{f_i^{\mathrm{AURA}}} \times 100\%, \tag{2.5}$$

where f_i^{NURD} are f_{uhr} values predicted using the constructed neural network, f_i^{AURA} are f_{uhr} measurements provided by the team of EMFISIS (we refer to them as to ground truth values), and M is the size of the validation set.

Figure 2.7 displays the MAPE of the resulting neural networks plotted vs the number of neurons in their hidden layer. The blue solid line shows the MAPE on the training set, the red dashed line – on the validation set. The MAPE on the training set decreases as the number of neurons in the hidden layer increases. This happens due to the following. As the number of hidden neurons grows, the complexity of the network increases (it has more free parameters), and it becomes capable of fitting the training data better. On the other hand, as the complexity of the network increases, the MAPE on the validation set decreases until a certain point, and then it starts to increase. This is a sign of overfitting, which means that after that point, the model does not generalize well and does not produce a reliable output on the unseen data. In our case, we found that the optimal neural network contains 80 neurons in the hidden layer.

After the optimal number of hidden neurons is determined, the network with 80 neurons is retrained on the training and validation sets combined, and its performance is further examined on the test set. The MAPE of the resulting model on the test set is $\sim 8\%$.

2.2.4 Postprocessing step

After all the steps described above are completed, the resulting model can be used in practice and applied to data. The output of the system, f_{uhr}, must be binned to 82 logarithmically-spaced frequency bins since we are to derive the UHR band in the dynamic spectrograms with such spacing. However, the output of the neural network model is a real number by definition and is not bounded to those bins. Therefore, the raw output of the neural network must be processed and binned to the specific

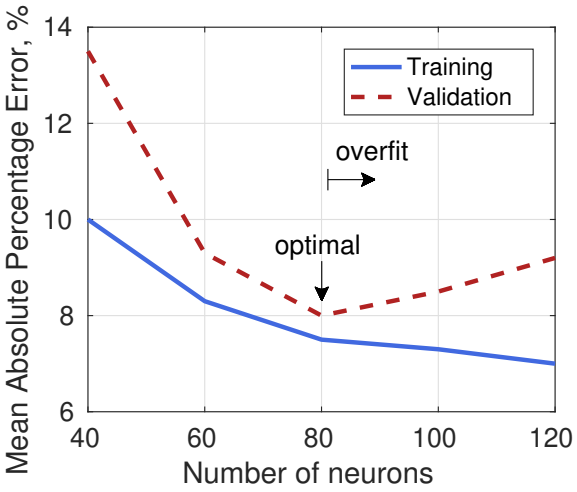

Figure 2.7: Mean absolute percentage error (MAPE) as a function of the number of hidden neurons of the neural network. As the complexity of the network increases, MAPE decreases on the training set. However, for the validation set, error stops decreasing after a certain point indicating that the network has been overfit to the training data. Here, a network with 80 neurons in the hidden layer is optimal (adapted from Zhelavskaya et al. (2016)).

frequency bins to be used in practice. The implemented postprocessing procedure is described below.

In the first step of the postprocessing procedure, the f_{uhr} value reconstructed by the neural network is rounded off to the closest EMFISIS HFR frequency bin. Next, we consider the determined bin and the two adjacent bins in the HFR spectrum (above and below). If the identified bin is the first or the last bin, then only one adjacent bin is considered (bin 2 or 81, correspondingly). The resulting f_{uhr} is then determined as the frequency corresponding to the bin containing the maximum power spectral density (of the three considered bins).

Figure 2.8 shows an example of the neural network output before ("raw" output of the neural network, blue curve) and after (black curve) the postprocessing procedure. It can be seen from the figure that the "raw" output of the neural network can slightly fluctuate from the true uhr band. The postprocessing procedure decreases these fluctuations. The derived upper hybrid line can still toggle between adjacent bins in the spectrum (see Figure 2.10 for the example), consequently introducing uncertainty in the density determination. Moreover, errors caused by using the frequency associated with the maximum emission intensity as f_{uhr} instead of the upper boundary of the upper hybrid band (Benson et al., 2004; Beghin et al., 1989) also should be taken into

account. As mentioned before, the coarser frequency resolution of the EMFISIS HFR instrument in comparison to the IMAGE RPI might not allow capturing this upper edge accurately. Thus, the introduced error can be estimated by evaluating the error in density when the derived f_{uhr} differs by one frequency bin from its real value. The following formula for this error was obtained after some derivations:

$$\frac{\Delta n}{n} \approx 2\frac{\Delta f}{f}\sqrt{\left(\frac{f_{\mathrm{pe}}}{f_{\mathrm{ce}}}\right)^{-2}+1}, \tag{2.6}$$

where $\frac{\Delta f}{f}$ is the frequency resolution of 1 bin that is $\approx 5\%$ as noted before. The expression under the square root is always less than 2 but greater than 1 since the $\frac{f_{\mathrm{pe}}}{f_{\mathrm{ce}}}$ ratio is always larger than 1. This makes the total expression always greater than $\approx 10\%$ and less than $\approx 14\%$. Therefore, the error introduced due to switching between two adjacent bins is between ~ 10 and 14%. It is worth noting, however, that this error should be taken into account in the trough most of the times since determination of the upper hybrid line for the plasmasphere is relatively straightforward.

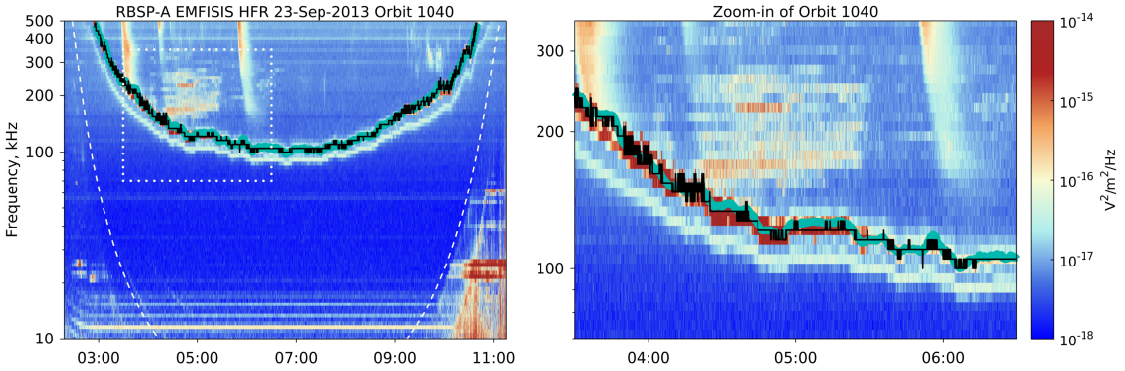

Figure 2.8: An example of the postprocessing procedure of the neural network output. Postprocessing assigns the output of the neural network to the most adjacent frequency bin and removes potential noise in the neural network output. The left panel shows the spectrogram for orbit pass #1040, and the right panel shows the zoomed-in part of the spectrogram (from 03:30 until 06:30 UT). The blue curve shows the raw neural network output, and the black curve shows the uhr after the postprocessing step (adapted from Zhelavskaya et al. (2016)).

Finally, after the f_{uhr} determined by the neural network is binned, the electron densities are derived using expressions (2.1) and (2.2).

2.3 Results

The NURD algorithm was applied to the Van Allen Probes measurements covering the period from October 1, 2012, to July 1, 2016, and a database of 33, 830, 887 electron density measurements (for both probes) was produced. The output of the NURD algorithm is compared to the density derived using AURA, and several examples of such comparisons for individual satellite passes are shown below. Furthermore, we perform analysis of the obtained electron density database and compare the resulting density distribution to the empirical trough and plasmasphere models by Sheeley et al. (2001).

2.3.1 Comparison with AURA and NURD performance

In order to compare the output of the NURD algorithm to the output provided by AURA and properly evaluate the NURD's performance, we should first introduce a classification of dynamic spectrograms into types as was done in Kurth et al. (2015). Kurth et al. (2015) classify dynamic spectrograms for each individual satellite pass (or orbit) into three types of complexity, A, B, and C, in terms of AURA performance:

1. Type A are the spectrograms in which less than 25% of the f_{uhr} points required manual correction. These are the spectrograms in which f_{uhr} is fairly clear and easy to identify. Type A constitutes 70% of the spectrograms processed by AURA.

2. Type B are the spectrograms in which 25 to 50% of data points must be corrected manually and comprises 20% of the spectrograms. These spectrograms are more difficult to process, as f_{uhr} is not always possible to identify unambiguously.

3. Type C are the spectrograms containing interpretational difficulties in finding the upper hybrid band. These spectrograms might be very contaminated. This type constitutes 10% of the spectrograms.

We perform a quantitative comparison between the NURD output and AURA's f_{uhr} by computing the MAPE for each orbit type as defined in (2.5). It is worth noting that such evaluation does not demonstrate the true error of either NURD or AURA since the ground truth densities are ambiguous, especially for type C. Thus, it is more

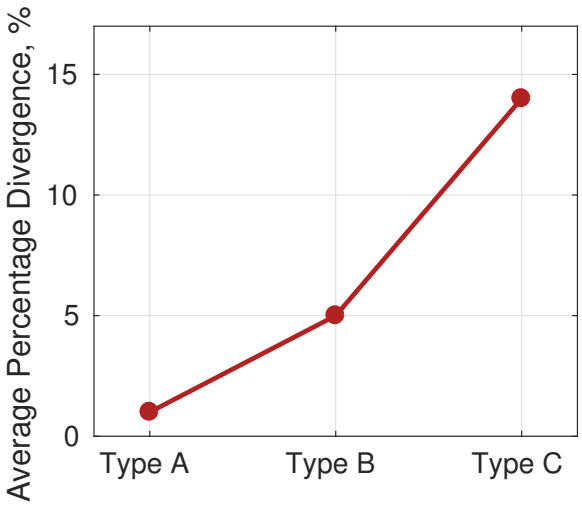

Figure 2.9: The average percentage divergence of the electron plasma density determined by the NURD algorithm from the electron density determined by the AURA algorithm calculated on the test set as a function of orbit types (adapted from Zhelavskaya et al. (2016)).

suitable to call this evaluation measure average percentage divergence rather than mean absolute percentage error. Figure 2.9 shows the average percentage divergence calculated for different orbit types on the test data set. The difference between AURA and NURD is not significant for orbits of type A, since the upper-hybrid resonance bands are clear and easy to identify for this type of orbit. For orbits of type B, the difference is slightly higher ($\sim 5\%$), and the difference for orbits of type C is $\sim 14\%$. Such deviation between the results might be caused by the ambiguity of the upper hybrid frequency determination during the geomagnetically active times and at times by the contaminated signal. During such periods, the plasma density in the trough can be very low and there might be strong electron cyclotron harmonic emissions present making it challenging to unambiguously identify the f_{uhr} profile both for the neural network and for AURA.

Figure 2.10 shows the typical examples for each type of orbit, where the f_{uhr} profiles obtained using AURA (red curve) and the resulting f_{uhr} profiles obtained with the NURD algorithm are indicated (black curve). The two f_{uhr} profiles produced by AURA and NURD for orbit of type A almost overlap. For more complex cases (types B and C), the upper hybrid band identified by NURD nearly overlaps with AURA's f_{uhr} in the plasmasphere (high-density region), but differs in the trough (low-density region). Such performance can be anticipated since determining the

density in the plasmasphere is relatively unambiguous. However, this is not the case for the trough, where the upper hybrid line not always can be observed. Thus, for such complicated types of orbits, the largest contribution to the average divergence can be attributed to differences in the trough region between two algorithms. In such cases, it may be hard to definitely conclude which of the algorithms performed better. For example, in Figure 2.10b,e (type B), NURD's f_{uhr} toggles between adjacent bins more than AURA's f_{uhr}. However, NURD's f_{uhr} tends to follow one resonance band that seems to be selected by it as the band corresponding to the intense emission extending beyond the banded emissions, while AURA's f_{uhr} might switch from band to band. It is very difficult to determine the correct diagnosis visually, and sometimes, both algorithms might be wrong. Most of such cases refer to parts of orbits of type C, where the electron density in the trough is extremely low and strong electron cyclotron harmonic emissions are observed. Figures 2.10c and 10f illustrate the case, when NURD and AURA may both provide wrong f_{uhr} determination, and the f_{uhr} band can be higher than the algorithms predicted (at 07:30 near the relatively strong emission observed near 40 kHz). More comprehensive analysis involving spectral interpretation of the emissions is needed to determine the actual value of f_{uhr}.

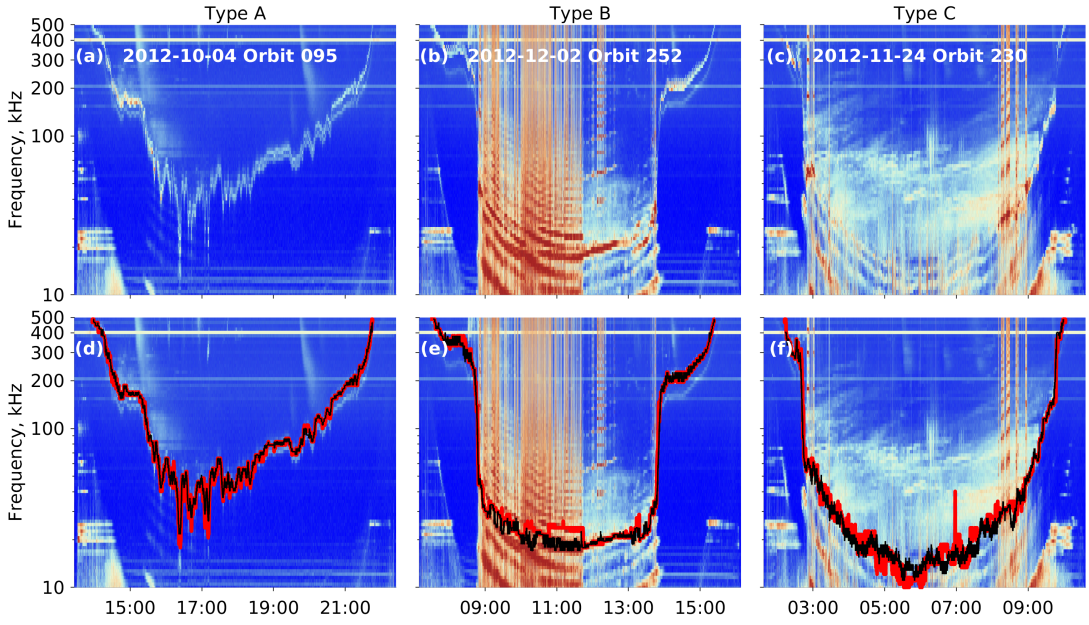

Figure 2.10: Examples of spectrograms of each type (the top panel) and the upper hybrid frequency identified by AURA, indicated by the red curve and by the NURD algorithm indicated by the black curve (the bottom panel) (adapted from Zhelavskaya et al. (2016)).

2.3.2 Comparison with empirical model of Sheeley et al. (2001)

Initial analysis of the obtained density database comprising 33, 830, 887 measurements is presented below. Here, we compare the derived plasma density distribution to the empirical density models of the trough and the plasmasphere developed by Sheeley et al. (2001) as functions of L and MLT. For the sake of comparison, we separate density values into plasmasphere- and trough-like data by applying the criteria used in Sheeley et al. (2001) for the threshold density (in cm^{-3}):

$$n_{\mathrm{b}} = 10 \left(\frac{6.6}{L} \right)^4 . \tag{2.7}$$

Density values at or above n_{b} for the given L shell are considered plasmasphere-like; values below n_{b} are considered trough-like.

Figure 2.11 shows the two-dimensional plots of normalized occurrence of the measurements as a function of plasma electron density and L (the top row) and as a function of electron density and MLT (the bottom row). The data set is divided into 26 bins in the logarithm of electron density and into 16 bins in $L/14$ bins in MLT. For the MLT plots, the data is limited to the range of $3 < L < 5$. The value in a particular bin is calculated as the number of measurements corresponding to n_{e} and L/MLT in that particular range divided by the number of measurements in that range of L/MLT. Hence, the color scale shows the normalized occurrence in different L/MLT bins. The minimum number of measurements in a bin required to calculate the normalized occurrence is 10; the mean number of measurements per bin is 77, 278 for L and 29, 220 for MLT plots.

Figures 2.11a and d show the normalized occurrence for all density measurements, Figures 2.11b and e – for the plasmasphere-like density, and Figures 2.11c and f – for the trough-like density, vs L and MLT, respectively. The intense red color indicates the regions where density measurements tend to be clustered, the blue color indicates that measurements are sparse in those regions. The black dashed line shows the threshold density, n_{b}. For the plots in the bottom row (vs MLT), the threshold density was calculated using the median of L for all data (both plasmasphere- and trough-like), which is approximately $L \approx 4.2$. The uppermost black curve correspond to the plasmasphere density model of Sheeley et al. (2001), and the lowermost black curve corresponds the trough density model for MLT = 0. Again, for the plots in the bottom row, the plasmasphere and the trough density values were calculated using

the median of L for measurements in the plasmasphere ($L \approx 4.1$) and the median of L for measurements in the trough ($L \approx 4.6$) correspondingly. The upper and lower white dotted curves show the mean of the logarithm of electron density obtained using NURD for plasmasphere and trough, correspondingly.

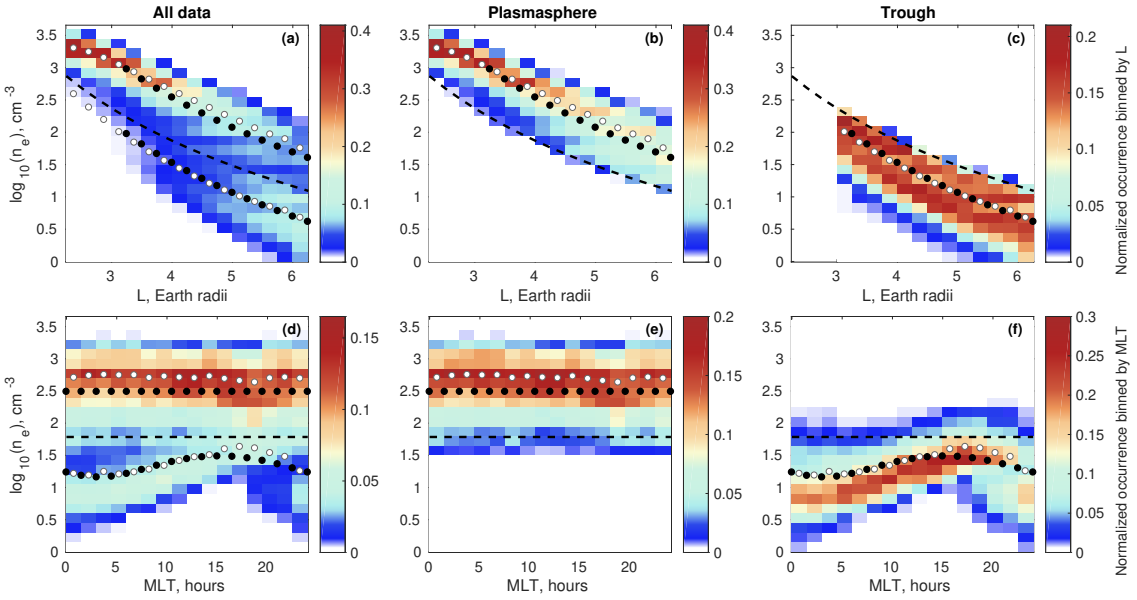

Figure 2.11: The occurrence of density measurements normalized by the number of data points in different L (the top row) or MLT (the bottom row) bins as a function of electron density on the logarithmic scale and L (top) or MLT (bottom). The dashed line indicates the separation between the trough-like and plasmasphere-like data as $n_{\mathrm{b}} = 10(\frac{6.6}{L})^4$ (the same as used in Sheeley et al. (2001)). The black dotted curves indicate the plasmasphere and trough density model by Sheeley et al. (2001). The white dotted curves indicate the mean of the log(ne) determined with the NURD algorithm for the plasmasphere and the trough, correspondingly (adapted from Zhelavskaya et al. (2016)).

Figure 2.11a shows that between $L = 2$ and 3.5, density measurements cluster around a narrow range of densities in the plasmasphere. At higher L, however, we see a bimodal structure with a distinct separation between plasmasphere and trough measurements. When we examine the plasmasphere and trough density regions separately (Figures 2.11b, c, and e, f) and compare those distributions to the Sheeley et al. (2001) model, we find good agreement with both the plasmasphere and the trough models. However, we also find that the peak of occurrence distributions of measurements produced by Van Allen Probes is slightly higher than the average density value measured by the CRRES satellite in the plasmasphere. The mean density

value is shifted by approximately 280 cm^{-3} ($\approx$ 0.14 on a logarithmic scale) on average. For the trough (Figure 2.11c), the peak of occurrence densities distributions is slightly lower than the empirical model. The shift in the the occurrence peak is approximately 5.32 cm^{-3} on average ($\approx$ 0.05 on a log scale). Such a shift may be caused by the fact that the CRRES mission operated during the peak phase of the solar cycle, while Van Allen Probes were launched at the increasing phase.

2.4 Discussion and future directions

Our results have shown that neural networks can be successfully applied to plasma wave data and produce reliable and accurate density determinations. Such performance of the developed algorithm was reached due to the proficient quality of the AURA density set used for training and to the proper neural network design process. However, it is worth acknowledging that although the electron density data set obtained using AURA and employed for training is currently the most reliable source of the electron plasma density measurements for the Van Allen Probes mission, there are still cases of uncertainties in density determination that cannot be resolved even by the manual inspection. Such uncertainties may be a source of errors in the training set, especially in the low-density trough region. As a result, the errors in the training set may influence the neural network performance. Due to that, the accuracy of both NURD and AURA may be lower in the trough. Therefore, while the use of density values marked questionable in the electron density data set produced by NURD is safe in statistical studies, the exact values with this flag should be used with caution.

At the moment, NURD is tuned specifically to the Van Allen Probes data, but using these data we have shown that neural networks and potentially other machine-learning algorithms are applicable and produce good results on such types of data. The created framework can be used to retrain the neural network model using other data sets, for which the input parameters can be adjusted if necessary.

Although feedforward neural networks (FNNs) were employed as a modeling technique in this work, it is worth mentioning alternative methods that are also effective at nonlinear multivariate regression problems. These methods include recurrent neural networks (RNNs) (e.g., Hochreiter and Schmidhuber, 1997; Goller and Kuchler, 1996), nonlinear least squares (e.g., Teunissen, 1990; Moré, 1978), regression trees

(e.g., Breiman et al., 1984), etc., which could be also potentially used for determining f_{uhr} from plasma wave measurements. These methods naturally have their own advantages and disadvantages, and a method that is more suitable for the data and problem at hand should be preferably selected. In the study described in this chapter, the FNNs were chosen over the methods mentioned above for the following reasons. Unlike nonlinear least squares, FNNs do not require to specify the form (model) of the nonlinear function to be modeled, which is not known in the problem at hand. They are also less susceptible to noise in data in comparison to regression trees if trained properly. FNNs were chosen over RNNs in this study with a goal to first employ a simpler network architecture without taking into account any temporal dependencies. In the future, however, the ability of RNNs to preserve the information about the past can be used to explore the influence of temporal dependencies on the NURDs performance.

The neural network-based methodology described in this chapter can also be employed in other space weather applications. A particular extension of this work is the development of the global plasmasphere density model driven by the solar and geomagnetic activity and therefore not bound by the satellite measurements. In this case, a neural network model will have the time history of solar wind and geomagnetic parameters and the satellite location as input, and a single output, plasma density. Such a model would be extremely useful in space physics, specifically radiation belt modeling, and also for the applications discussed in the overview section. As also mentioned in the overview, the existing empirical models tend to be oversimplified and are parameterized by location. The model driven by the solar wind and geomagnetic parameters would be a significant advancement in the field.

2.5 Conclusions

In this chapter, we present our algorithm for the automated determination of plasma electron density from electric and magnetic field measurements made by Van Allen Probe spacecraft, the Neural-network-based Upper-hybrid Resonance Determination algorithm (NURD). The algorithm uses a feedforward neural network to derive upper-hybrid frequency profiles from the dynamic spectrograms and magnetic field measurements obtained from the EMFISIS instrumentation suite. Plasma density is then derived from the inferred upper-hybrid frequency profiles. Plasma density database

developed using another semi-automated routine AURA (Kurth et al., 2015) is used to train the network. The NURD algorithm is valid for L-shells $1.25 < L < 6.2$ and all local time sectors.

The developed algorithm was applied to the available database of electric and magnetic field measurements from October 2012 until July 2016, comprising $3,750$ orbital passes of the Van Allen Probes satellites. Comparison with AURA showed that electron densities obtained using the proposed method are in good agreement with the densities inferred with AURA. To demonstrate this, we adopted the classification of the dynamic spectrograms into three levels of difficulty as introduced in Kurth et al. (2015): type A (upper-hybrid frequency is relatively straightforward to identify, 70% of all orbits); type B (interpretation from an expert is needed, 20% of orbits); type C (concealed signal, 10% of orbits). The mean average percentage divergence between the density values produced by the NURD and AURA was $\sim 1\%$ for type A, $\sim 5\%$ for type B, and $\sim 14\%$ for type C. The overall error of the derived density is $\sim 14\%$. However, the error may be larger in cases of high uncertainty in density determination, particularly in the low-density region and after the recently elevated geomagnetic activity. Indeed, the neural network-based approach does not fully remove the uncertainty in density determination. Nonetheless, it stills produces reasonable density estimates in the regions of high uncertainty that can be employed in statistical studies. Additionally, the proposed algorithm is automated meaning that it can remove a significant part of the manual aspect of the density determination.

The analysis of the resulting electron density data has shown an agreement with the plasmasphere and trough density models by Sheeley et al. (2001). On the other hand, a large variability in the density data was observed that cannot be reproduced by empirical models based on statistical averages. Using the NURD algorithm, the electron density can be determined to a much finer resolution than using the existing empirical models.

Acknowledgments

The work at GFZ was supported by Helmholtz Association Recruiting Initiative, NSF GEM AGS-1203747, NASA grant NNX12AE34G, NASA grant NNX16AF91G and

project PROGRESS funded by EU Horizon 2020 No 637302. The work at Stanford was supported by NASA award NNX15Al94G. The EMFISIS data were accessed through the official EMFISIS website hosted at the University of Iowa, and we thank the EMFISIS team, William Kurth and Principal Investigator, Craig Kletzing. The Kp index was provided by GFZ. John Wiley and Sons provided the license agreement to reproduce the figures from Zhelavskaya et al. (2016) (License Number 4114291495723). The electron density data set derived by the NURD algorithm is available at `ftp://rbm.epss.ucla.edu/ftpdisk1/NURD`.

Chapter 3

Empirical modeling of the plasmasphere dynamics using neural networks

Note

The following work is published in *Journal of Geophysical Research: Space Physics*: *Zhelavskaya I. S., Y. Y. Shprits, M. Spasojević (2017). Empirical modeling of the plasmasphere dynamics using neural networks. Journal of Geophysical Research: Space Physics, 122, 11,22711,244.* https://doi.org/10.1002/2017JA024406.

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 October 1, 2012 - July 1, 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-hour time history of Kp, AE, SYM-H, and F10.7 indices. The model successfully reproduces erosion of the plasmasphere on the night side 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.

3.1 Introduction

The plasmasphere is a toroidal region of cold (~1 eV) and relatively dense (~$10 - 10^4$ cm^{-3}) plasma surrounding the Earth (Lemaire and 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 (O'Brien and Moldwin, 2003; Chappell et al., 1970b) . 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 et al., 2003b; Singh and Horwitz, 1992; Krall et al., 2008); its shape is then roughly circular with a bulge on the dusk side (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., 1970a; Goldstein et al., 2003b). The stronger the disturbance, the more the plasmasphere contracts (down to or even lower than 2 R_E during severe geomagnetic storms (Baker et al., 2004; Goldstein et al., 2005c)). Features of various scales (plumes, channels, crenulations, shoulders, fingers, etc.) (Grebowsky, 1970; Sandel et al., 2001; Carpenter and Stone, 1967; LeDocq et al., 1994; Spasojević et al., 2003; Garcia et al., 2003; Goldstein et al., 2004; Foster and Burke, 2002) can be formed during that time and are transformed during the storm recovery phase by the

interplay between convection and co-rotating regimes (Dungey, 1961; Goldstein et al., 2002, 2005c). 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 (Gallagher et al., 2005; Darrouzet et al., 2009).

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., Spasojević et al., 2004; Orlova et al., 2016; Shprits et al., 2016). Eroded plasmaspheric material is transported sunward and is regularly observed near the dayside magnetopause boundary (e.g., Chen and 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., Borovsky and Denton, 2006; André et al., 2016). Storm-time plasmaspheric structures have been found to be strongly associated with ionospheric density features including ionospheric storm enhanced density (SED) and polar cap patches (Su et al., 2001a; Zhang et al., 2013).

One of the parameters that is used to quantify the plasmasphere dynamics is plasma density (Lemaire and Gringauz, 1998; Kotova, 2007, 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), particle counters (e.g., Geiger and 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 EMFISIS instrument (Kletzing et al., 2013) onboard 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); 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., Park and Carpenter, 1970; Park, 1974; Moldwin et al., 1995). Accurately calculating the evolving density distribution from first principles has also proven elusive due to the sheer number of physical processes involved (Huba and Krall, 2013; Marchaudon and 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) onboard the IMAGE satellite produced the first global images of the plasmasphere. Analysis of the images revealed the complex and dynamic evolution of the plasmasphere (Spasojević et al., 2003; Goldstein et al., 2005b) and highlighted the importance of the cold plasma distribution in controlling other magnetospheric processes (Spasojević et al., 2004; Baker 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 et al., 2003c). 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 (Bishop, 1995; Anderson, 1995; Haykin et al., 2009). An inherent part of neural networks includes finding the optimal inputs that best fits 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;

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 3.2, 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 3.3, we explain how the final model was selected and show examples of density determination for different disturbance intervals both for Van Allen Probes and IMAGE EUV data. In sections 3.4 and 3.5, we discuss the physical implications of the selected inputs, model performance, its advantages and limitations, and possible improvements.

3.2 Methodology

The methodology employed in this work is neural network-based empirical modeling. Neural networks are one of the best techniques for finding multivariate non-linear 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.

3.2.1 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., Le, 2013; Krizhevsky et al., 2012; Cireşan 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, i.e., will generalize well when applied to data that was not used to train the network. Here, we determine the optimal number of neurons using 5-fold cross validation combined with a learning curve approach. These techniques are described in section 3.2.4.

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 3.2.3. The plasma density data set used to train the neural networks is described in section 3.2.2.

3.2.2 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 NURD algorithm (Zhelavskaya et al., 2016) for the period October 1, 2012 to July 1, 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.

3.2.3 Inputs to the neural network

A wide variety of parameters might potentially be used to quantify the plasmasphere dynamics. They include:

1. Geomagnetic parameters (K_p, AE, AU, AL, D_st, SYM-H, ASYM-H),
2. Solar wind data ($n_\mathrm{p}, v, P_\mathrm{dyn}, B_\mathrm{z}, B_\mathrm{y}$),
3. Solar wind coupling functions ($vBs, vB_\mathrm{T}, vB_\mathrm{T}\sin\theta_c, d\Phi_\mathrm{MP}/dt$) (e.g., Newell et al., 2007),

4. Solar cycle indicators ($F10.7$, sunspot number, ionospheric IG index),

5. Time history of activity inputs (total duration (e.g., 24, 36, 48 hrs), resolution (linear, log),

6. Averaging technique (weighted avg, avg 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 towards 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, or

   (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?)

2. What is the memory of the magnetosphere? i.e. 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 3.1. 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 3.1). 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 hours of time history. We use averages of the time histories of activity parameters integrated

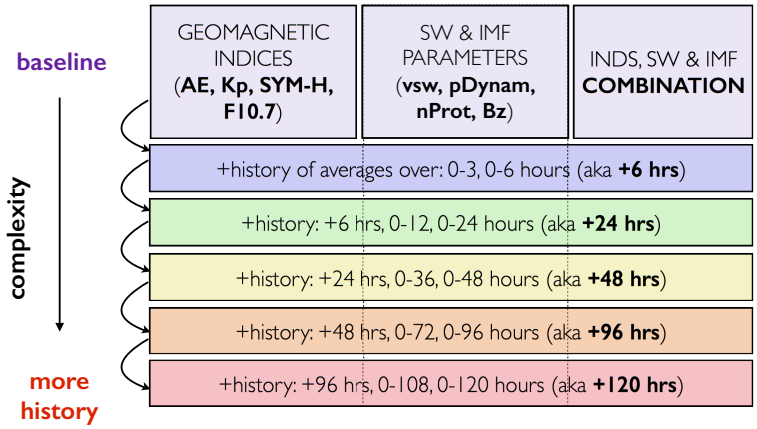

Figure 3.1: Eighteen combinations of input parameters to the neural network explored in this work. We start from simple models containing only instantaneous values of (1) geomagnetic indices, (2) solar wind and IMF parameters, and (3) their combinations, and then subsequently add a time history of these parameters to the models. Each subsequent model includes inputs of the previous model and several more inputs corresponding to more time history added.

from hour zero (for example, $0-3$ hr, $0-6$ hr, $0-12$ hr, etc.). 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 ($0-3$ hr, $3-6$ hr, $6-12$ hr, etc.). Overall, the lower the row in Figure 3.1, 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.

3.2.4 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 3.1. 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 (Levenberg, 1944; Marquardt, 1963; Hagan and Menhaj, 1994; Hagan et al., 1996) 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, i.e. 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);

2. Model's ability to reproduce the global dynamic evolution of the plasmasphere (global validation).

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.

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., Kohavi et al., 1995; Devijver and Kittler, 1982). 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., Meek et al., 2002; Cortes et al., 1994). Learning curves are also used to determine whether a model over- or under-fits (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 3.2. The data set is divided randomly into K subsets of approximately equal size ($K = 5$ in Figure 3.2). 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, sMAPE, etc.) 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.

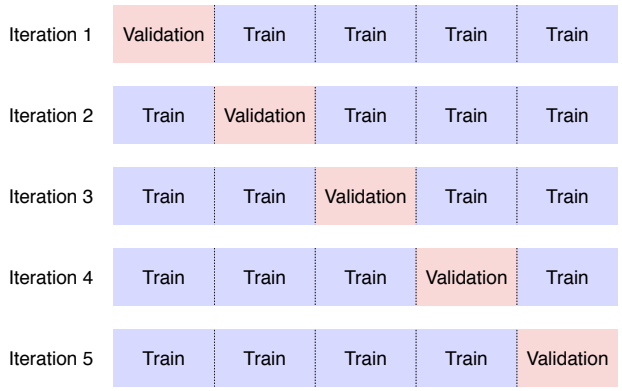

Figure 3.2: Schematic representation of K-fold cross validation (when K=5). The data set is split into 5 subsets of approximately equal size. At every iteration, a different subset is left out, and the rest of the data is used for training. Thus, we can estimate the error of the model on 5 different subsets unseen during the training and obtain the mean and standard deviation of error of the validation error.

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. 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 is 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: over-fitting, under-fitting, 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 is 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, i.e. 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 under-fits 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.

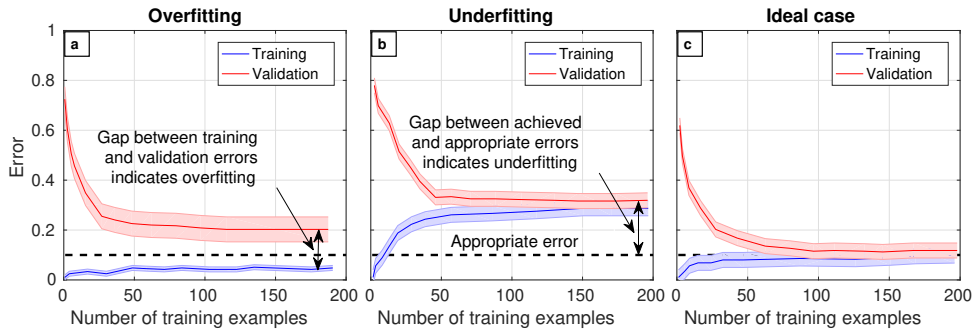

Figure 3.3: Toy illustration demonstrating learning curves in three different scenarios: (a) overfitting, (b) underfitting, and (c) ideal case. All three panels show plots of error vs. number of examples in the training set. In every panel, the blue solid curve denotes cross validation (CV) error on 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.

Figure 3.3 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 zero 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 3.3a), 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 3.3b), training and validation errors move closer to 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 3.3c), 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 3.3c, 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, i.e. performs similarly on different sets of data. This aspect is very important, since, if a model is not stable, i.e. 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 3.3 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 in consideration in the model selection procedure.

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 electrons/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 et al., 2003a). Although IMAGE EUV produced images of He^+ distribution and not of H^+ dominant in the plasmasphere, it was shown in the study of Goldstein et al. (2003a) that the sharp He^+ edge in the EUV images coincides with the actual plasmapause locations. Authors showed that 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

plasmapshere. 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.

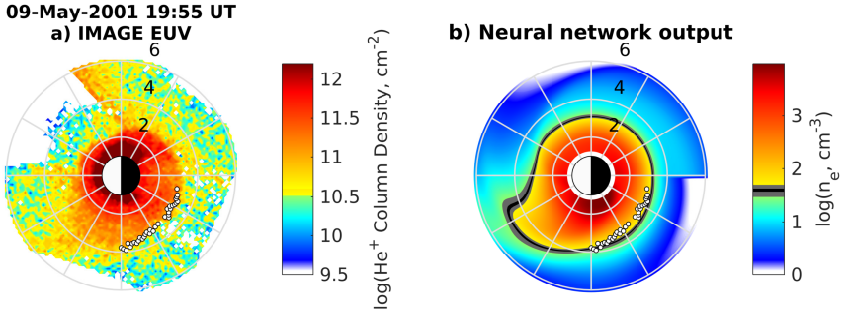

Figure 3.4: Example of comparison of the neural network output with EUV image. a) EUV image of the He^+ column density mapped to the equatorial plane with the manually selected plasmapause indicated (white circles). b) The output of a neural network model trained on 3.6 years of RBSP-A and RBSP-B measurements 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 D_{st} is -68 nT for this event, and maximum K_p is 5.3. The Sun is to the left.

Figure 3.4a shows an example EUV image from 9 May 2001, a moderate geomagnetic disturbance interval ($\min D_{st} = -68$ nT) examined in Spasojević et al. (2005). The white dots are the manually estimated plasmapause location. Figure 3.4b 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 F10.7 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 squared 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 3.3.2. 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 et al. (2003a) estimated that the lower sensitivity threshold of the EUV instrument was equivalent to 40 ± 10 cm^{-3}, which is indicated by the black and gray section of the color bar in Figure 3.4b. However, this threshold

depends on many factors, with the strong dependence on the variability in the EUV image background, which has contributions from (1) interstellar/interplanetary 30.4 nm light and (2) 58.4 nm light that leaks through the aluminum filter of the EUV instrument.

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 ± 10 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 night side.

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 over- or under-fit 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 October 1, 2012, to July 1, 2016) is 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.

3.3 Implementation

3.3.1 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 (Levenberg, 1944; Marquardt, 1963; Hagan and Menhaj, 1994; Hagan et al., 1996). 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-second cadence. The minimum resolution of the solar wind measurements from OMNIWeb is 1 minute. Therefore, we use 1-minute 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 2 parts in a $9:1$ ratio, as discussed above. The larger part, the design set, is used to select the optimal model by using 5-fold cross validation and learning-curve analysis. Eighteen neural networks models corresponding to 18 input combinations from Figure 3.1 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 hours 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. The following empirical formula was used to calculate the range of neurons to consider for each number of inputs: $\text{round}((x - \mod(x, 5)) * \text{coef}_i)$, 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 5 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 5 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 5 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 5 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 4 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 4 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 3.1) 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.

3.3.2 Optimal model selection

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 3.1). 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 3.5a shows the root mean squared 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 show the standard deviation of error. Red curves and shaded areas show the same but for the validation set correspondingly. Figure 3.5b is a zoomed-in fragment of Figure 3.5a 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 hours.

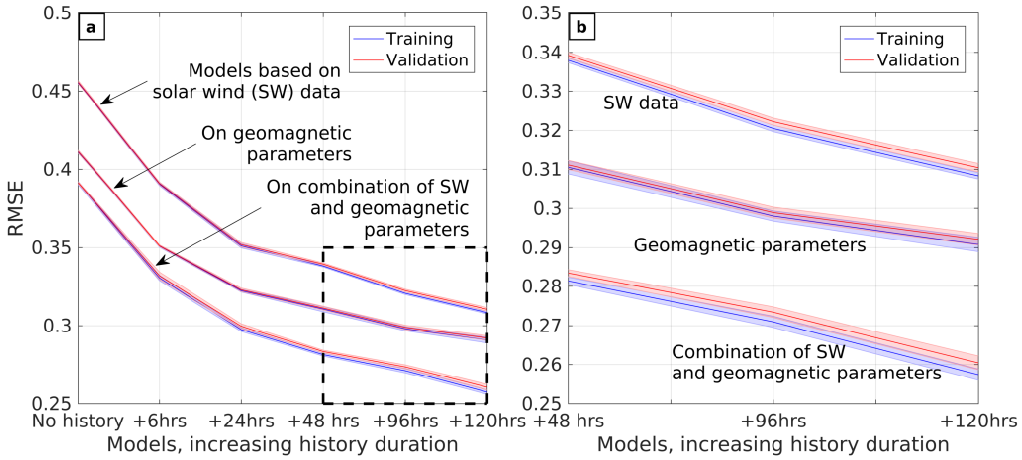

Figure 3.5: Root mean squared error (RMSE) on the y-axis vs. models having different time history included on the x-axis for (a) all models and (b) models having time histories of 48, 96, and 120 hours. The farther on the x-axis, the more time history is included into a model. In both panels, the blue 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.

Figure 3.5 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 3.5) 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.

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 ± 10 cm^{-3} (Goldstein et al., 2003a).

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 3.5. 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 3.5).

An example of such a comparison is shown in Figure 3.6. Figure 3.6a shows an example EUV image from 2 Jun 2001, a mild geomagnetic disturbance interval ($\min D_{\mathrm{st}} = -27$ nT) examined in Goldstein et al. (2003b). The white dots are the manually estimated plasmapause location. Figures 3.6b, 3.6c, 3.6d show the output of 3 neural network models correspondingly: b) based solely on geomagnetic indices, c) on solar wind parameters, and d) both geomagnetic and solar wind parameters.

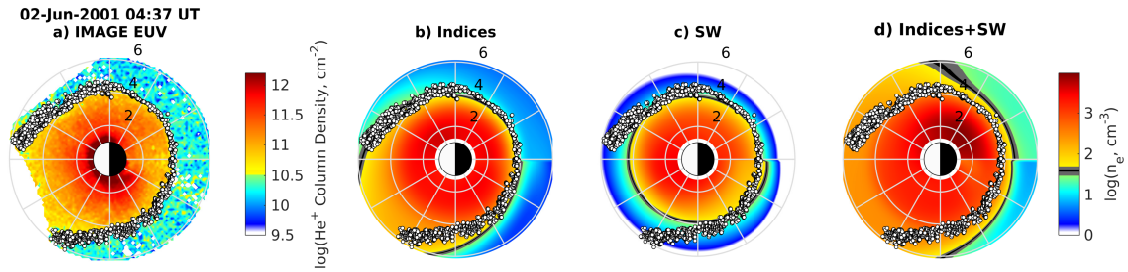

Figure 3.6: Example comparison of the EUV image of the He^+ column density mapped to the equatorial plane (a) with the output of 3 neural network models: (b) based solely on geomagnetic indices, (c) on solar wind parameters, and (d) both geomagnetic and solar wind parameters. White circles indicate the manually selected plasmapause. Minimum D_{st} for this event is -27 nT, maximum K_p is 5.3. The Sun is to the left.

The black and gray section of the color bar indicates a 40 ± 10 cm^{-3} density threshold. Inputs to all three networks contain 96 hours 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 b) 45, c) 45, and d) 81. It can be seen from Figure 3.6 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.

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 hours. A visual comparison with IMAGE EUV plasmapause data shows that there is no significant improvement in performance of the model that includes 120 hours of history. Therefore, the less complex model is selected. The optimal number of neurons in the hidden layer, i.e. 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-hour time history of the following indices: AE, K_{p}, SYM-H, $F10.7$, 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 Squared Error (RMSE) of the resulting model on the test set is 0.3015, 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.

3.4 Results

3.4.1 Examples of the resulting model output

Figure 3.7a and 3.7b 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 3.7c shows the K_{p} 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

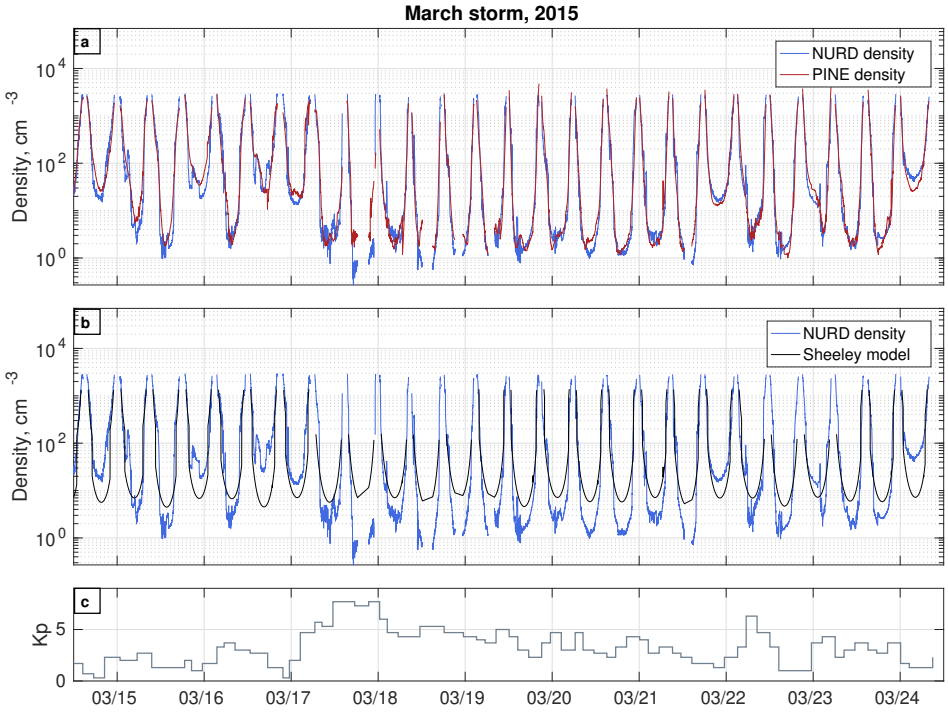

Figure 3.7: Example of density determination for the March storm in 2015. Pnales a and b plot density vs. time. Blue curves are the density derived from Van Allen Probes using NURD algorithm (Zhelavskaya et al., 2016) for Probe A. Red curve is the density determined by the PINE model. Black curve is the density calculated using the trough and plasmasphere density model by Sheeley et al. (2001). Panel c shows K_{p} geomagnetic activity index during this event.

et al. (2001).

The top row of Figure 3.8 shows an example of global density reconstruction for a series of time frames during $26-27$ June 2001, a minor geomagnetic disturbance interval ($\min D_{\mathrm{st}} = -21$ nT) examined in detail in Spasojević et al. (2003). The black and gray section of the color bar is 40 ± 10 electrons cm^{-3} and is equivalent to the estimated lower sensitivity threshold of the EUV instrument (Goldstein et al., 2003a). The K_{p} index for the duration of the event is shown in the bottom row of Figure 3.8, 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 $26-27$ June 2001. Several stages of this minor geomagnetic disturbance interval are shown. The first snapshot (a) shows the onset of the disturbance. The neural network captures very well the nightside edge of the plasmasphere. The next three panels (b, c, and d)

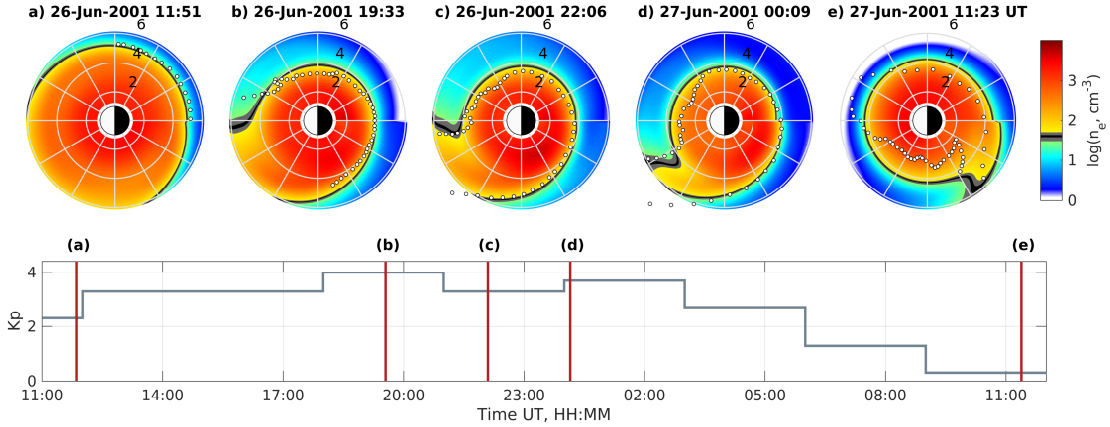

Figure 3.8: Example of global density reconstruction by the resulting neural network model during 26 – 27 June 2001, a minor geomagnetic disturbance interval ($\min D_{\mathrm{st}} = -21$ nT) examined in detail in Spasojević et al. (2003) (top row). The white dots denote the manually estimated plasmapause location. The bottom row shows K_{p} geomagnetic activity index in gray; red vertical lines correspond to the density snapshots in the upper panel as indicated by letters. The Sun is to the left.

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. The last panel (e) 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, K_{p} remained below 2 for over 40 hours, and then it dropped to 1- for over 30 hours immediately after the disturbance. The neural network model experiences difficulties in accurately reconstructing the plasmapause location in the recovery phase of the storm shown in panel (e). 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 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 panels (a-d).

Further examples of global density reconstruction are shown in Figure 3.9 for 4

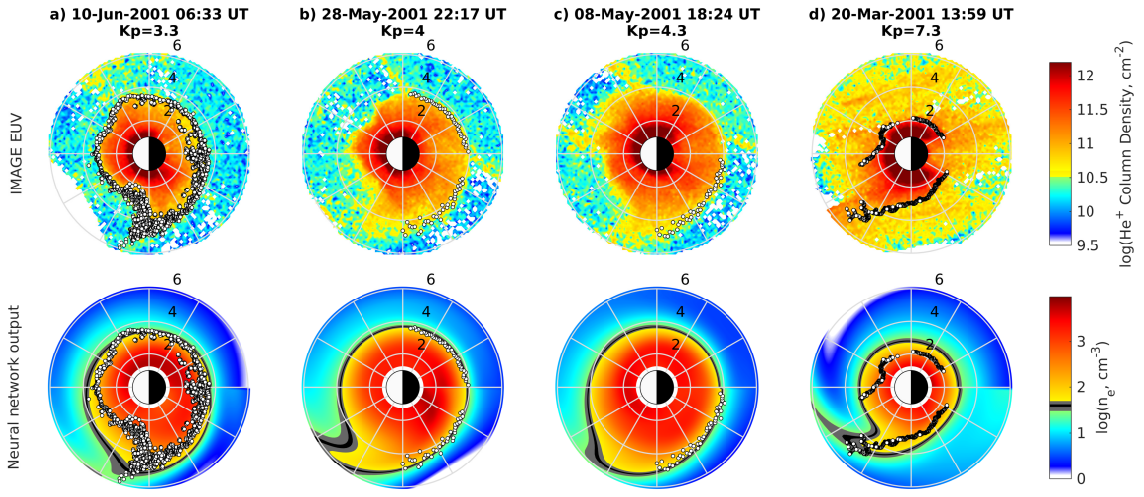

Figure 3.9: Examples of global density reconstruction by the resulting neural network model for 4 different events during the main phase plume formation. The top row shows the EUV images for the times indicated in the titles, and the bottom row shows the final model output for those times. Events are ordered from left to right according to K_p (from low to high). K_p is shown in the titles as well. The Sun is to the left.

different events during their respective main phase plume formations: a) 10 Jun 2001 06:33 UT, b) 28 May 2001 22:17 UT, c) 08 May 2001 18:24 UT, and d) 20 Mar 2001 13:59 UT (considered in Spasojević et al. (2003), Spasojević et al. (2005), Spasojević 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 K_p: from low to high K_p. 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 night side is very well captured for all 4 events. However, the model slightly overestimates the plasmapause boundary on the nightside for the time frame shown in Figure 3.9d, in the high geomagnetic disturbance interval where $K_\mathrm{p} = 7.3$. This is caused by the fact that the training density data obtained from the Van Allen Probes is underrepresented for high 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.

3.4.2 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`.

3.5 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 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 and 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-minute cadence). These results answer the first question stated in the introduction: the critical combination of solar wind and geomagnetic parameters that best quantify the plasmasphere dynamics includes the K_{p}, AE, SYM-H, and $F10.7$ 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 hours, 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 hours. By employing such a time history, it becomes possible to reconstruct density for interesting and complex events as the one shown in Figure 3.8. 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 hours.

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 is abundant. For our case, a large volume of data is available for the times of quiet geomagnetic activity. However, the available data is 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 ($K_{\mathrm{p}} > 7$), 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 $2000 - 2005$, whereas density data from Van Allen Probes is

measured between $2012 - 2016$. 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 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 have not been explored in this work and therefore are not included in the model (such as $AU, AL, B_{\mathrm{x}}, B_{\mathrm{y}}$, 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., Sandel et al., 2003; Spasojević et al., 2003; Goldstein et al., 2004; Darrouzet et al., 2009). 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.

3.6 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-hour 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 hours. 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 Association Recruiting Initiative, NSF GEM AGS-1203747, NASA grant NNX12AE34G, NASA grant NNX16AF91G and project PROGRESS funded by EU Horizon 2020 No 637302. The work at Stanford was supported by NASA award NNX15Al94G. 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`.

Chapter 4

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

Note

The following work is submitted to *Journal of Geophysical Research: Space Physics*: *Zhelavskaya I. S., N. Aseev, Y. Y. Shprits, M. Spasojević (2020). A combined neural network- and physics-based approach for modeling the plasmasphere dynamics. (submitted to Journal of Geophysical Research: Space Physics).*

Abstract

In recent years, feedforward neural networks (NNs) have been successfully applied to reconstruct global plasmasphere dynamics in the equatorial plane. These neural network-based models capture the large-scale dynamics of the plasmasphere, such as plume formation and erosion of the plasmasphere on the nightside. However, their performance depends strongly on the availability of training data. When the data coverage is limited or non-existent, as occurs during geomagnetic storms, the performance of NNs significantly decreases, as networks inherently cannot learn from the limited number of examples. This limitation can be overcome by employing physics-based modeling during strong geomagnetic storms. Physics-based models show a stable performance during periods of disturbed geomagnetic activity, if they

are correctly initialized and configured. In this study, we illustrate how to combine the neural network- and physics-based models of the plasmasphere in an optimal way by using the data assimilation Kalman filtering. The proposed approach utilizes advantages of both neural network- and physics-based modeling and produces global plasma density reconstructions for both quiet and disturbed geomagnetic activity, including extreme geomagnetic storms. We validate the models quantitatively by comparing their output to the in-situ density measurements from RBSP-A for an 18-month out-of-sample period from 30 June 2016 to 01 January 2018, and computing performance metrics. To validate the global density reconstructions qualitatively, we compare them to the IMAGE EUV images of the He^+ particle distribution in the Earth's plasmasphere for a number of events in the past, including the Halloween storm in 2003.

4.1 Introduction

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

a plasmaspheric plume in the dusk local time sector (e.g., Spasojević et al., 2003; Grebowsky, 1970).

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

A number of physics-based and empirical models have been developed in recent decades. The most commonly used empirical models are the Carpenter and Anderson (1992), Gallagher et al. (2000), and Sheeley et al. (2001) models. The Carpenter and Anderson (1992) model is based on measurements of electron density derived from radio measurements made with the sweep frequency receiver (SFR) on board the International Sun-Earth Explorer (ISEE-1) spacecraft and ground-based whistler measurements. It is a model of saturated density and, thus, represents the distribution of density after several days of refilling. The model covers the range from 2.25 to 8 in L-shell, and the interval of 0-15 MLT (magnetic local time). The model provides the mean value of density at different L-shells. Gallagher et al. (2000) developed the Global Core Plasma Model (GCPM), which combined several previously developed models (such as Carpenter and Anderson (1992) and Gallagher et al. (1998)) by means of transition equations, in order to provide a more comprehensive description of the inner-magnetospheric plasma. The models of plasmasphere and plasma trough developed by Sheeley et al. (2001) provide statistical averages of density based on density measurements obtained from the swept frequency receiver onboard the Combined Release and Radiation Effects Satellite (CRRES) by identifying the upper hybrid resonance frequency. The models cover all local times and $3 \leq L \leq 7$. Moreover,

Sheeley et al. (2001) also provide the standard deviation of density for both the plasmasphere and trough models to describe depleted and saturated density levels for various L-shells and MLT sectors for the trough.

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

This fact motivated the development of time-dependent models of plasma density. In recent years, a number of models utilizing neural networks and taking into account solar or geomagnetic conditions have been developed (Bortnik et al., 2016; Zhelavskaya et al., 2017; Chu et al., 2017a,b). In all these studies, the authors used feedforward neural networks with different architectures to model the plasma density in the equatorial plane or in 3D (in Chu et al. (2017b)). Feedforward neural networks are a powerful mathematical tool for finding nonlinear multivariate mappings from input to output variables, if such a mapping exists (Anderson, 1995; Bishop, 1995; Haykin et al., 2009). Bortnik et al. (2016) used density measurements inferred from the spacecraft potential (Li et al., 2010) on board the THEMIS (Time History of Events and Macroscale Interactions during Substorms) mission (Angelopoulos, 2009) to train their neural network model. They used a 5-hour time history of Sym-H index and location as an input to the model. The Chu et al. (2017a) model is based on the same density measurements. The inputs to the model were location and the time histories of Sym-H for the preceding 3 days, AL for 2 hours, and F10.7 for 3 days. Chu et al. (2017b) built upon those two studies and developed a three-dimensional electron density (DEN3D) model based on density measurements from the plasma wave experiment on board ISEE (Gurnett et al., 1978), the plasma wave experiment on board the CRRES (Anderson et al., 1992), the plasma wave instrument on board Polar (Gurnett et al., 1995), and the radio plasma imager (RPI) on board the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE). They used location and the time histories of Sym-H for the preceding 3 days, AL for 5 hours, and F10.7 for 3 days as inputs to their model. The model of Zhelavskaya et al. (2017), Plasma density in the Inner magnetosphere Neural network-based Empirical (PINE) model,

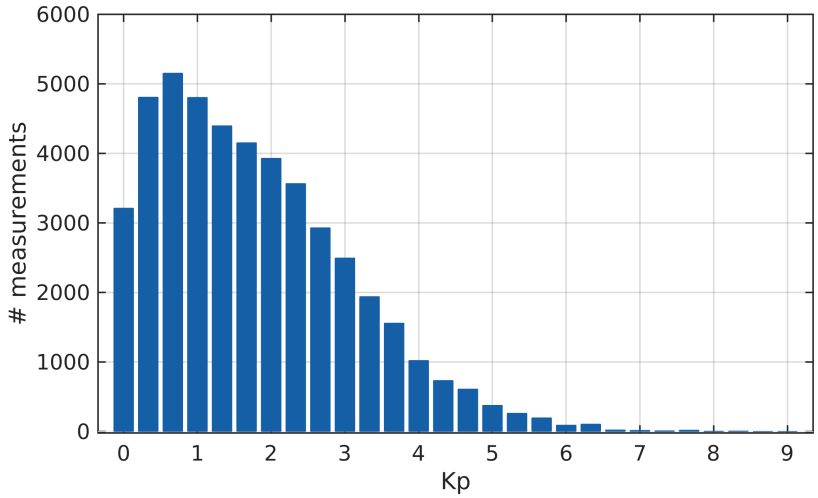

Figure 4.1: Distribution of Kp over October 2012 – June 2016 (the training time interval for the PINE model). The cadence of bins is 1/3, i.e., the same as the cadence of the Kp index.

is based on the density measurements obtained from the upper-hybrid resonance frequency measured with the EMFISIS instrument on board the Van Allen Probes. This technique is known to be one of the most reliable methods for obtaining plasma density (Mosier et al., 1973). The inputs to the model were the 96-hour time history of Kp, AE, Sym-H, and F10.7 indices and the location given by L and MLT. They showed that neural networks-based models can accurately reproduce the dynamics of the plasmasphere (with correlation coefficient ≈ 0.95), and can successfully reproduce the asymmetric shape of the plasmasphere, including plume formation and erosion on the nightside.

Neural networks learn from data and are very powerful when data are abundant. However, when the data are limited or lacking, their performance may significantly decrease (Priddy and Keller, 2005). This implies that neural networks can be difficult to apply to highly unbalanced regression problems and to predict rare events. Extreme geomagnetic storms are an example of such events. Figure 4.1 shows the distribution of the Kp index over the training period of the PINE model (Oct 2012-Jul 2016, Zhelavskaya et al. (2017)). As can be seen from the figure, its distribution is highly skewed. Observations for Kp > 7 are limited. In fact, there is not a single example of Kp $= 9$ during this period and, hence, in the training dataset. That implies that NNs may not be reliable during periods of high geomagnetic activity, which are the most interesting events.

One possible way to overcome this limitation is to employ a different approach to model the plasmasphere dynamics during disturbed geomagnetic conditions. In particular, physics-based modeling is a more stable approach than neural network-based modeling for high Kp, since it does not depend on data availability. A number of physics-based models have been developed in recent years. Pierrard and Stegen (2008) used the kinetic exospheric approach to model the dynamics of the plasmasphere. Jordanova et al. (2006) coupled their ring current-atmosphere interactions model (RAM) with a 3-D equilibrium code (SCB) and a cold plasma model (CPL). The RAM-SCB-CPL model calculates the cold electron density in the equatorial plane by following the motion of individual flux tubes, using a model of electric field which includes a corotation potential and a convection potential that is chosen from either semi-empirical models (Volland, 1973; Stern, 1975; Weimer, 2005), or self-consistently calculated electric potential (Yu et al., 2015), mapped to the equatorial plane along the SCB field lines. Krall et al. (2016) coupled this model with SAMI3 (Sami3 is Also a Model of the Ionosphere) to model the plasmasphere dynamics during two events in 2001. De Pascuale et al. (2018) used RAM-CPL to simulate equatorial plasmaspheric electron densities during two storm events in 2013, and compared them to in-situ measurements from the Van Allen Probes (Radiation Belt Storm Probes). **?** used the first-principles physics-based model SAMI3 to model the plasmasphere in 3D. They incorporated the neutral wind dynamo potential, the corotation potential, and a time-dependent potential from Volland (1973) and Maynard and Chen (1975) to model the convection potential for an idealized magnetic storm. An overview of various other physics-based models of the plasmasphere based on the fluid and the kinetic approaches is given in Pierrard et al. (2009).

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

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

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

The paper is structured as follows: In section 4.2, we describe the data used for training and validation of the models, i.e., in situ density measurements from the

Van Allen Probes and the plasmapause position derived from the IMAGE EUV. We describe the neural network, the physics-based VERB-CS code, and the Kalman filter and how it is used to develop the assimilative model in section 4.3. In section 4.4, we present the results obtained with the models for the Halloween storm in 2003 and several events from 2001, and also for a long-term density reconstruction. Finally, in sections 4.5 and 4.6, we discuss implications and possible improvements to the models developed in this study.

4.2 Data

All magnetic field, solar wind data, and geomagnetic indices have been downloaded from the OMNIWeb data service. We have used the density dataset obtained with the NURD (Neural-network-based Upper hybrid Resonance Determination) algorithm (Zhelavskaya et al., 2016) for the period from 01 October 2012 to 01 July 2016, to train the neural networks. Zhelavskaya et al. (2016) employed feedforward neural networks to identify the upper hybrid resonance bands in the dynamic spectrograms made with the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) suite (Kletzing et al., 2013) onboard the Van Allen Probes satellites and calculated the plasma density from the upper-hybrid resonance frequency. The electron density data set is publicly available at the GFZ Data Services (Zhelavskaya et al., 2020b). The Van Allen Probes provide electron density measurements for all local time sectors and $L \sim 2 - 6\ R_{\mathrm{E}}$. We use density measurements for a period of 30 June 2016 to 1 January 2018 (obtained with the same method) to quantify the performance of all the models developed in this study in section 4.4.3.

To validate the global output of our models, we use the plasmapause locations derived from the EUV instrument on board the IMAGE satellite (Sandel et al., 2000). The IMAGE EUV instrument provided the global images of the plasmasphere for the first time. These images can be used to derive the location of the plasmapause by considering the outermost sharp edge of He^+ (Goldstein et al., 2003c). Goldstein et al. (2003a) showed that the sharp edge of He^+ in the EUV images corresponds to the actual location of plasmapause. We use the density threshold of 40 ± 10 cm^{-3} as an approximation of the plasmapause position in the global reconstructions of density produced by the models, which corresponds to the lower sensitivity threshold of the IMAGE EUV instrument (Goldstein et al., 2003a). It is worth noting that the

IMAGE mission operated in 2000 – 2005, which was a different solar cycle compared to the one we used in the training of the neural network. Therefore, the IMAGE EUV images are the best available data source for validating the global evolution of the shape of the plasmasphere produced by the models developed in this study. The plasmapause database derived from the IMAGE EUV instrument was obtained from `http://enarc.space.swri.edu/EUV/`.

4.3 Methodology

4.3.1 The neural network-based model of plasma density PINE

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

In the original study, the authors used K-fold cross validation with K=5 to train and validate the model. They used this procedure also to find the optimal inputs to the model (for more details, please see Appendix A). The training and validation

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

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

We apply the K-fold cross validation procedure described above to the density measurements from the Van Allen Probes. The obtained results are described in detail in Appendix A. In summary, we confirm the findings of the original study of Zhelavskaya et al. (2017). The models based on the geomagnetic indices yield the best performance, compared to the models based only on solar wind or on both solar wind and geomagnetic indices. As discussed in the original study, the models based on the solar wind inputs are less accurate than models based on geomagnetic indices. At the

same time, the models based on both data sources tend to overfit the training data. We find that the optimal model is based on the 48-hour time history of geomagnetic indices AE, Kp, Sym-H, and also F10.7. The model also includes the location input given by L and MLT. This updated version of the PINE model is used in this study. Further in the paper, we apply the model to the out-of-sample period from 30 June 2016 to 01 January 2018 (which was not used neither in training or validation) and test the model performance during that interval.

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

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

$$\frac{\partial n}{\partial t} + v_\phi \frac{\partial n}{\partial \phi} + v_R \frac{\partial n}{\partial R} = S - L \tag{4.1}$$

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

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

$$\phi_{\mathrm{CR}} = -\frac{A_{\mathrm{CR}}}{r},\ A_{\mathrm{CR}} \approx 92\ \mathrm{kV/R_E} \tag{4.2}$$

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

$$\phi_{\mathrm{VS}} = -A_{\mathrm{MC}} r^2 \sin(\phi),\ A_{\mathrm{MC}} = \frac{0.045}{1 - 0.159\mathrm{Kp} + 0.0093\mathrm{Kp}^2} \tag{4.3}$$

We use a shielding parameter $\gamma = 1.8$ instead of standard $\gamma = 2$, as our experiments show that using $\gamma = 1.8$ agrees better with observations (more details are provided

in the Results and Discussion sections). We include the effect of SAPS in the model by using the Kp-dependent model of the SAPS electric field developed by Goldstein et al. (2005a). This model quantitatively includes the average properties of SAPS reported in Foster and Vo (2002). The SAPS has an effect on the location of the dusk side plasmapause and influences the shape and location of plasmaspheric plumes.

To account for refilling, we use long-term refilling rates of equatorial electron density from Denton et al. (2012). These rates were inferred from passive radio emissions measured by the IMAGE RPI instrument during quiet geomagnetic times and are valid for the range $L = 2$ to 9. The model provides median, mean, 1st and 3rd quartiles of the refilling rates. In our study, we have used the median refilling rate: $\log_{10}(\frac{dn_{\mathrm{e,eq}}}{dt}) = 2.22 - 0.006L - 0.0347L^2$ (in cm^{-3}/day). Our experiments show that VERB-CS achieves a better performance when refilling rates are multiplied by a factor of 1.75. The model without such increased refilling rates showed significant underestimation of density values on the night side. The increased refilling rates improved that aspect of the model and therefore in this study we use refilling rates from Denton et al. (2012) multiplied by a factor of 1.75. Additional experiments are shown in Supporting Information (Figures S12-S15 and description therein).

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

$$L = \frac{n}{\tau} \tag{4.4}$$

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

To solve equation (4.1) numerically, we employ the VERB-CS code (Aseev and Shprits, 2019). The VERB-CS code models electron transport in ambient electric and magnetic fields and loss due to interaction with plasma waves. The VERB-CS code solves the two-dimensional advection equation that describes the particle drift, and we have extended the code to solve equation (4.1) by introducing losses to the magnetopause and the source term S.

Equation (4.1) must be complemented by initial and boundary conditions. To specify the initial conditions, we use the empirical density model of Sheeley et al.

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

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

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

4.3.3 The assimilative model

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

The Kalman filter is a popular technique for data assimilation. It is commonly

used to adjust model predictions in accordance with available, typically sparse, observations, while taking into account uncertainties of both the model and observations (Kalman, 1960). In this study, we adapt the Kalman filter technique to combine the predictions of two models, namely the physics-based VERB-CS code and the neural network-based PINE model. For this purpose, we consider the VERB-CS code as a model that propagates a state of the system in time. The output of the data-driven PINE model, in turn, is used as observations.

The Kalman Filter

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

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

If equation (4.1) is linear, its discretized version can generally be written as

$$\mathbf{n}_k^f = \mathbf{M}_{k-1}\mathbf{n}_{k-1}^a, \tag{4.5}$$

where superscripts f and a denote forecast and analysis, respectively, and $\mathbf{M}_{k-1}$ is a matrix, also referred to as the model matrix or the model operator. This matrix can be obtained, for example, by applying a finite difference method to (4.1). At a given time t_{k-1}, the model matrix $\mathbf{M}_{k-1}$ propagates the current state of the system $\mathbf{n}_{k-1}^a$ to the next state in time $\mathbf{n}_k^f$. The analysis $\mathbf{n}_{k-1}^a$ is the best estimate of the state vector at time t_{k-1}, based on the model and the available observations. The evolution of plasma density can be modeled by sequentially solving equation (4.5) for $k = 1, 2, \ldots$.

If applied to a real ("true") state $\mathbf{n}_{k-1}^t$ of the system, the model matrix propagates $\mathbf{n}_{k-1}^t$ with some error ϵ_k^M. This error can originate from the uncertainties of the model,

such as errors due to missing physical processes in the model or numerical errors due to discretization of the continuous equation (4.1). It is also referred to as the model error:

$$\mathbf{n}_k^t = \mathbf{M}_{k-1}\mathbf{n}_{k-1}^t + \epsilon_k^M. \tag{4.6}$$

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

$$\mathbf{n}_k^{obs} = \mathbf{H}_k\mathbf{n}_k^t + \epsilon_k^{obs}, \tag{4.7}$$

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

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

Prediction step

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

$$\mathbf{n}_k^f = \mathbf{M}_{k-1}\mathbf{n}_{k-1}^a. \tag{4.8}$$

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

$$\mathbf{P}_k^f = \mathbf{M}_{k-1}\mathbf{P}_{k-1}^a\mathbf{M}_{k-1}^T + \mathbf{Q}_{k-1}, \tag{4.9}$$

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

Update step

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

$$\mathbf{n}_k^a = \mathbf{n}_k^f + \mathbf{K}_k\left(\mathbf{n}_k^{obs} - \mathbf{H}_k\mathbf{n}_k^f\right), \tag{4.10}$$

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

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}. \tag{4.11}$$

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

$$\mathbf{P}_k^a = \mathbf{P}_k^f - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^f. \tag{4.12}$$

This finishes the iteration k of the Kalman filter.

Details of implementation

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

The nonlinear term S in equation (4.1) does not allow us to write the discretization of the equation in the form (4.8). The non-linearity of the equation requires the extension of the Kalman filter equations (4.8) and (4.9) by linearizing the model operator. In order to simplify the implementation of the Kalman filter, we avoid the linearization of the model operator by running one step of the VERB-CS code instead of solving equation (4.8) to obtain the plasma density forecast $\mathbf{n}_k^f$ using a previous (optimal) state $\mathbf{n}_{k-1}^a$, obtained in the update step (eq. (4.10)), as initial condition for the VERB-CS code. The VERB-CS code solves the partial differential equation (4.1)

numerically by discretizing density n, drift velocities v_φ and v_R, sources S, losses L, and spatial and temporal derivatives $\frac{\partial}{\partial v_\varphi}$, $\frac{\partial}{\partial v_R}$, and $\frac{\partial}{\partial t}$. Discretization allows us to consider plasma density and other parameters at discrete times t_k, where k is an integer, on the grid consisting of discrete values of MLT and R.

To update the forecast error covariance matrix $\mathbf{P}_k^f$ in equation (4.9), we obtain the model matrix $\mathbf{M}_{k-1}$ by discretizing equation (4.1) without the source term S, which removes the nonlinearity from equation (4.1). For this, we use the first-order explicit upwind finite difference scheme with a time step that automatically adapts to changing drift velocities to satisfy the Courant stability condition. The forward model includes the refilling, therefore the refilling is not accounted for only in the calculation of the forecast error covariance matrix. Neglecting the refilling rates in the update of the forecast error covariance matrix $\mathbf{P}_k^f$ does not significantly affect the optimality of the Kalman filter, if the step of data assimilation is chosen to be much smaller than the characteristic time of the refilling (that is on the order of days, Denton et al. (2012) and references therein). In this study, the data assimilation is performed every 4 hours. Moreover, as the primary goal of this study is to demonstrate whether the Kalman filter can be applied to combine a neural network and a physics-based model of the plasmasphere and what limitations it has, we do not aim at the most accurate implementation of the Kalman filter here (for example, using the extended Kalman filter (Goodwin and Sin, 1984; Ide et al., 1997)).

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

Another aspect that should be noted is the implementation of the observation operator $\mathbf{H}_k$. This operator transforms the forecast of the model n_k^f from the model grid onto the grid of observations (see eq. (4.7)). In our case, the model grid is that of VERB-CS, and the observations grid is that of the PINE model. As discussed in the

previous section, the spatial grid of VERB-CS ranges from 0 to 24 hours with 0.5-hour grid step in MLT and from 1.75 to 10 R_{E} with 0.2 R_{E} grid step in radial distance. In order to obtain the global output using the PINE model (i.e., on the whole equatorial plane and not just at specific L and MLT), we need to assume a spatial grid, on which the output is produced. The PINE model is valid for all MLT sectors, and from 1.75 to 6.15 R_{E} in radial distance due to the use of density measurements from Van Allen Probes for training. Therefore, the lower and upper boundaries of the PINE grid are set at 1.75 and 6.15 R_{E} in radial distance, respectively. In order to simplify the implementation of the Kalman filter, we use a spatial grid of the same resolution as the VERB-CS, namely with 0.5-hour and 0.2-R_{E} grid steps in MLT and radial distance, respectively. Thus, the spatial grid of PINE is a subset of the VERB-CS grid, which makes it easier to assimilate the PINE output. $\mathbf{H}_k$ is then defined as a matrix consisting of zeros and ones, where 1 corresponds to an element of this matrix when the model's (VERB-CS') grid point coincides with the observation (PINE's) grid point, and 0 otherwise. The number of rows in $\mathbf{H}_k$ is the number of grid points of PINE, and the number of columns is the number of grid points of VERB-CS.

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

than the threshold are assumed to be inside the plasmasphere, otherwise – outside). We assign the model error inside the plasmapause $\alpha^m_{\text{inside}} = 0.397$, the model error outside the plasmapause $\alpha^m_{\text{outside}} = 0.463$, the observation error inside the plasmapause $\alpha^{obs}_{\text{inside}} = 0.335$, and the observation error outside the plasmapause $\alpha^{obs}_{\text{outside}} = 0.333$. The description of how these values were obtained is presented in Appendix B. Using such an approach, we obtain a better agreement between the assimilative model and observations, compared to using single constants for $\alpha^m(n^f)^2$ and $\alpha^{obs}(n^{obs})^2$, respectively. However, we choose the model and observation errors, α^m and α^{obs}.

4.4 Results

In this section, we perform several tests to compare the performance of the PINE, VERB-CS, and assimilative model, which is henceforth referred to as the Assimilative Magnetospheric Plasma density (AMP) model. We compare the outputs of the models during the 2003 Halloween storm and a number of events during March-June 2001. We validate the models by comparing the modeled and observed shape of the plasmasphere by using the plasmapause location obtained from the IMAGE EUV instrument. We also perform a long-term density reconstruction for the period of 30 June 2016 to 1 January 2018, using all the models. For the long-term run, we validate the models by comparing their output to the in-situ density measurements from RBSP-A. This period was not used in the training of the PINE model. The setup of all the models used in these tests is described in section 4.3.

4.4.1 Test 1: Halloween storm 2003

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

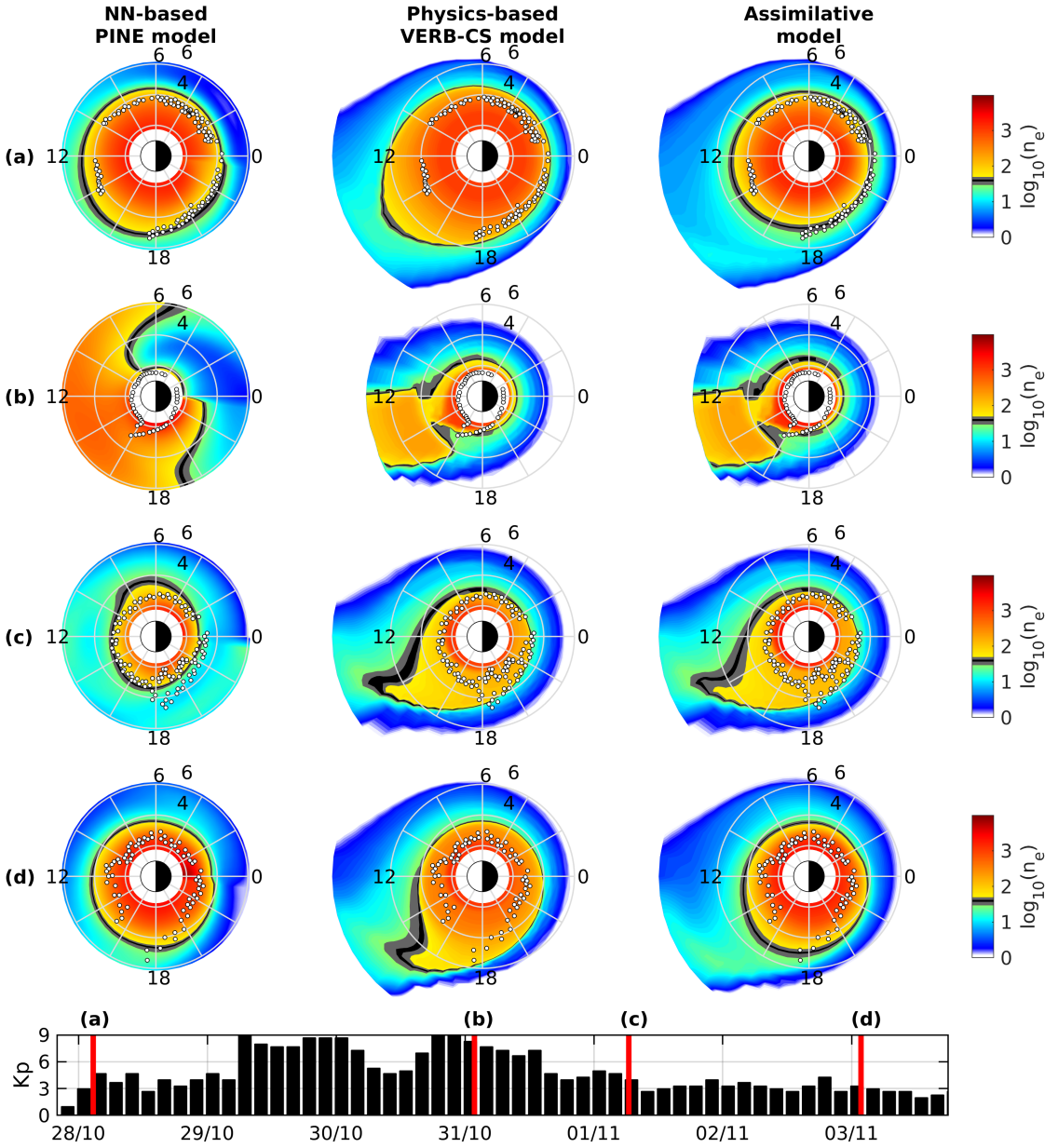

Figure 4.2: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the 2003 Halloween storm. The first four rows show the outputs of the models corresponding to the times marked with the red lines in the bottom panel showing the Kp index during the 2003 Halloween storm. The black-and-white dots show the location of the plasmapause derived from the IMAGE EUV images. The color in the first four rows indicates the logarithm of density (the scale of the colorbar is the same for all models and all times). The gray and black section of the colorbar indicates a density threshold of $40 \pm 10\,\mathrm{cm}^{-3}$ and can be considered a rough approximation of the plasmapause location for the sake of comparison to the observed plasmapause position obtained from IMAGE EUV (more details on that are given in section 4.2). The Sun is to the left. Row (a) corresponds to the time before the storm, (b) to the period during the storm (second CME), (c) and (d) to the recovery phase of the storm.

conditions.

Figure 4.2 shows the global electron density reconstruction during the Halloween storm 2003 using the PINE (left column), VERB-CS (middle column), and assimilative (right column) models. The first four rows show the global snapshots of density, and the bottom row shows the Kp index during the event. The first four rows correspond to the specific times during the event when the plasmapause from the IMAGE EUV instrument was available. These times are marked with the red lines and labels (a-d) in the bottom panel.

In order to obtain a global output using the PINE model, it was applied to each point on its spatial grid independently (described in section 4.3.3), and the smoothed global output shown in Figure 4.2 was obtained by interpolating between the points. The output of the VERB-CS code was obtained by running the model starting from quiet geomagnetic conditions (27-Oct-2003 20:00 UT, Kp = 1.7) with the setup described in section 4.3.2. The output of the assimilative model was obtained by running the model from the same time. Its setup is described in section 4.3.3. The time step of the simulations is 15 minutes for all the models. The time of the IMAGE plasmapause location is chosen to be as close as possible to the temporal grid of the models, but not further than 14 minutes away. Therefore, the time of the IMAGE plasmapause may not exactly coincide with the time of the simulation and may deviate from it by 14 minutes maximum. Although it is possible to set the exact time of the simulation for the PINE model, we choose to select the same temporal grid as in the VERB-CS and assimilative AMP models to ensure an equal comparison between all the models.

As seen in Figure 4.2, the PINE model agrees well with the plasmapause derived from the IMAGE EUV before the storm (row (a)), but produces unrealistic global density reconstruction during the main phase of the storm. As discussed in the introduction, the reason for that is the absence of training examples during extreme geomagnetic events (there is no single Kp = 9 in the training dataset of PINE). After the storm (row (c)), the size of the plasmasphere reproduced with PINE is in good agreement with the IMAGE observations. On the contrary, the VERB-CS model produces an overly extended plasmasphere during the quiet time before the storm, but reproduces the massive erosion of the plasmasphere (row (b)) observed in the IMAGE EUV observations as well. Several days after the storm (row (d)), VERB-CS produces lower densities inside the plasmasphere than those produced by PINE (this can be seen from the color in the density snapshots: yellow color in VERB-CS,

compared to the red color in the PINE output).

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

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

In the previous section, we showed that the assimilation of the neural network- and physics-based models demonstrated good agreement with the plasmapause observations during the 2003 Halloween storm, and performed better than either of the models separately. In this section, we test the models further by comparing their output for a number of events in March-June 2001. We have selected 5 events corresponding to different Kp levels, starting from quiet geomagnetic conditions (Kp = 2.7) and reaching disturbed ones (Kp = 8). The motivation behind this selection was to test how the models perform separately and when combined by means of data assimilation for different levels of geomagnetic disturbance. The output of the models throughout different phases of the selected events is provided and described in Supporting Information (Figures S1-S7).

Figure 4.3 shows snapshots of global density reconstructions using the PINE, VERB-CS, and assimilative models for 5 different events in 2001. The format is similar to Figure 4.2. The columns correspond to models, as labeled in the top row. The rows correspond to events. The times of the density snapshots and the

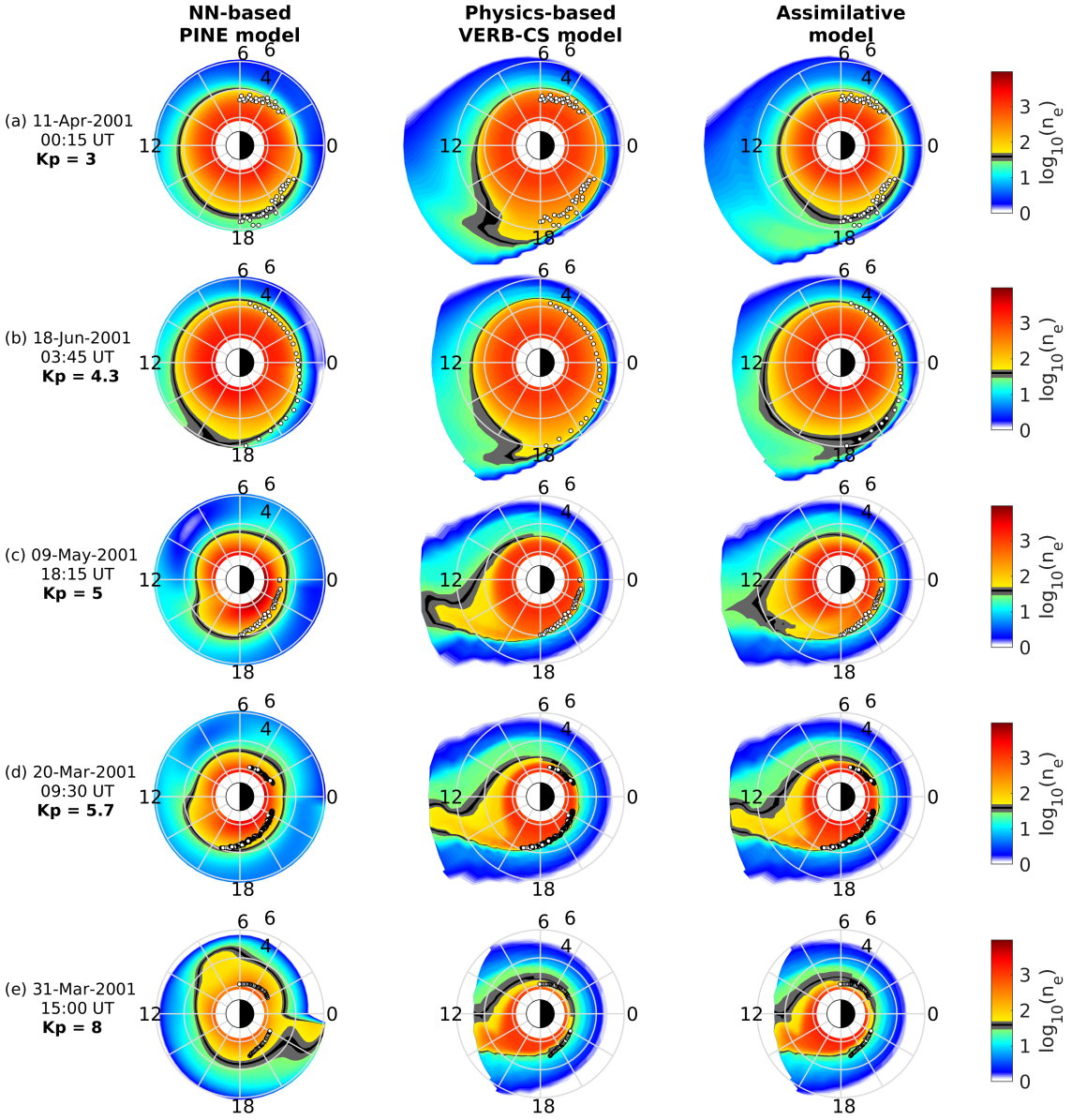

Figure 4.3: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during a series of events in 2001, as indicated in the labels on the left in each row. The format of the density snapshots is the same as in Figure 4.2.

corresponding Kp values are labeled in each row on the left. The events are ordered by increasing Kp index, rather than by time. The format of the density snapshots is the same as in the top four rows of Figure 4.2.

The global density reconstructions are obtained in the same fashion, as described in the previous section. Spatial and temporal grids of the models and their setup are also the same as used there. We note again that the time grid step is 15 minutes, and therefore may not exactly coincide with the timing of the plasmapause observations derived from IMAGE EUV (but is not farther than 10 minutes away). The simulations were run separately for 5 different events, each starting from quiet geomagnetic conditions. The starting time and Kp at the beginning of the events were: (a) 10 April 2001 03:00 UT (Kp = 1.7), (b) 17 June 2001 00:00 UT (Kp = 0.3), (c) 08 May 2001 03:00 UT (Kp = 0.7), (d) 18 March 2001 12:00 UT (Kp = 1), and (e) 30 March 2001 11:00 UT (Kp = 2).

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

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

4.4.3 Test 3: Long-term reconstruction of density

In the two previous sections, we have illustrated the performance of the assimilative, PINE, and VERB-CS models for different geomagnetic conditions, including quiet, moderate, and extreme conditions. The assimilative model demonstrated a better performance compared to PINE and VERB-CS used separately for all considered events. In this section, we test the performance of all models further by performing a long-term reconstruction of plasma density using all the models. We compare the modeled density with the in-situ density measurements from RBSP-A during the period between 30 June 2016 and 01 January 2018. We choose this particular interval for testing, as the density measurements during this time were mostly not used in the training of the PINE model. Therefore, this choice ensures a fair comparison between the performance of PINE, VERB-CS, and the assimilative model. Furthermore, RBSP-A crosses all MLT sectors during this interval, as shown in Figure 4.4, which allows us to evaluate the performance of the models in different MLT sectors. The setup of all three models is the same as in the two previous sections. Summary plots demonstrating performance metrics calculated during this period for all three models are shown at the end of this section. It is worth noting that the PINE model was trained on the interval 01 October 2012 – 01 July 2016 (Zhelavskaya et al., 2017), and therefore, we exclude the period 30 June – 01 July 2016 when calculating the performance metrics here. We choose 30 June 2016 as the start time of the simulation as the Kp index was smaller than on 02 July 2016 (0.3 vs. 0.7), and also since there was a minor disturbance (Kp = 3.3) between 30 June and 02 July, which could negatively influence the initial conditions for VERB-CS.

Figure 4.5 presents a comparison of the output of the neural network density model in-situ density measurements from RBSP-A from 30 June 2016 to 01 January 2018. Panel (a) shows the in-situ density observations from RBSP-A. Panel (b) shows the output of the PINE model. These two panels have the same format: the x-axis corresponds to time, the y-axis to the L-shell, and the color indicates the logarithm of electron density. The next two panels (c) and (d) show the difference between the observations and the output of the model. Panel (c) shows the absolute difference between logarithms of modeled and observed density. Panel (d) shows the difference itself. The blue color in this panel implies underestimation of density by the model (the modeled density is smaller than observed), the red color overestimation. We note the small data gaps that can be seen as white vertical stripes in these panels.

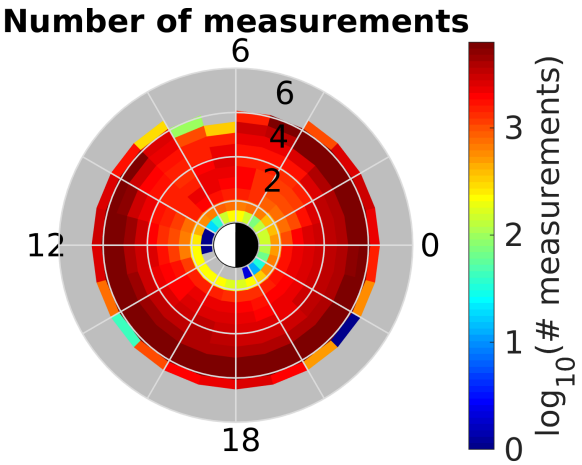

Figure 4.4: Coverage of RBSP-A during 30 June 2016 – 01 January 2018.

These are gaps in density measurements and the model output is not shown during these times as well since there are no measurements to compare with. Panel (e) shows the location of RBSP-A's apogee during this interval. The gray shaded area implies that the apogee of RBSP-A was at the nightside (i.e., from 18 to 6 MLT). It is worth noting that the apogee of RBSP-A is located in the night sector during approximately the first half of the interval, and therefore, the densities on the farther L-shells are smaller (dark blue color), compared to the second half of the interval (where the color is light green on the farther L-shells). During the second half of the interval, RBSP-A's apogee was located on the dayside, and therefore, the density is higher there due to plasmaspheric bulge and plume. The bottom panel shows the Kp index during this period.

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

It can be seen that the PINE model output is very similar to the observations. The model captures the expansion of the plasmasphere that occurs during periods of quiet

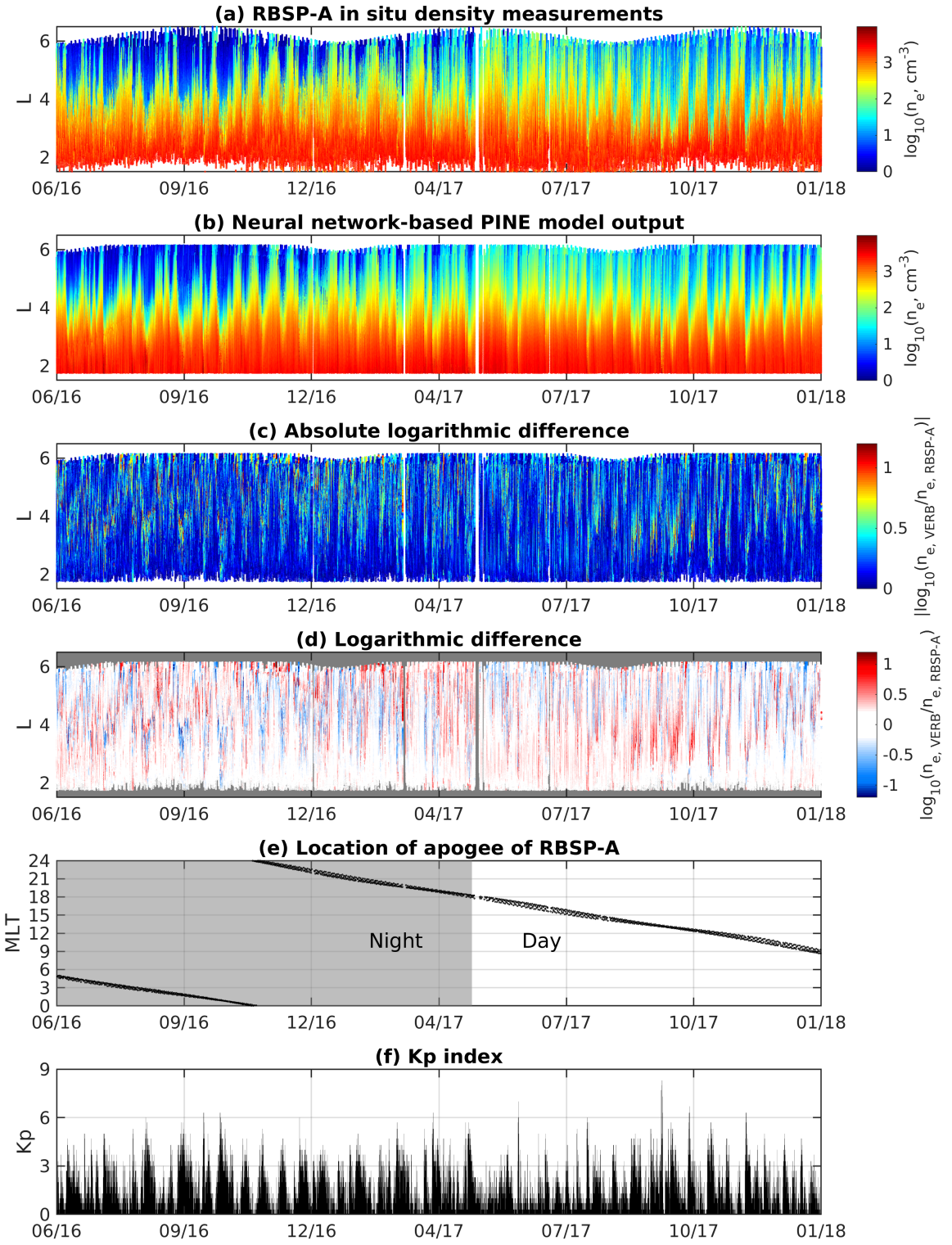

Figure 4.5: Long-term comparison of the PINE model and the RBSP-A density measurements during July 2016 – January 2018. Panels (a) and (b) show the RBSP-A density measurements and the output of the PINE model, respectively, where the L-shell is on the y-axis, time is on the x-axis, and the color indicates the log of density. Panels (c), (d), and (e) show the absolute difference, the sign of the difference, and the difference between log of the model and data, respectively.

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

Figure 4.6 shows the comparison between in-situ density from RBSP-A and the output of the physics-based VERB-CS code. The format of the figure is the same as in Figure 4.5, where panel (b) presents the output of VERB-CS, and panels (c) and (d) show the difference between the modeled and observed density in different formats. The model output at the coordinates of RBSP-A was obtained in the same fashion as in Figure 4.5: the model was first run on the full spatial grid, and then the virtual satellite was flown through the output at the L- and MLT-coordinates of RBSP-A.

It can be seen that the VERB-CS model captures the general dynamics of the plasmasphere, i.e., its erosion and expansion, well. Again, the model reproduces a massive erosion of the plasmasphere during the September 2017 storm. It can be seen, however, that the differences between observations and the output of VERB-CS shown in panel (c) are larger than those of the PINE model (shown in Figure 4.5). The positive and negative differences between the model output and observations (panel (d)) also seem to be spread randomly over the duration of the simulation, however, a slight overestimation can be seen during July-October 2017, when RBSP-A's apogee was in the 12-17 MLT sector.

Finally, Figure 4.7 presents the comparison between in-situ density measurements from RBSP-A and the output of the assimilative model. The format is the same as in Figures 4.5 and 4.6, where panel (b) shows the output of the assimilative model and panels (c) and (d) show the difference between the model output and observations in different formats (as described below in Figure 4.5). The output of the model was obtained in the same manner as for the other models. The assimilative model was first run on the full spatial grid, and then a virtual satellite was flown through the global output of the assimilative model along the RBSP-A coordinates.

It can be seen from the figure that the assimilative model successfully captures the general dynamics of the plasmasphere, i.e., erosion and expansion associated with

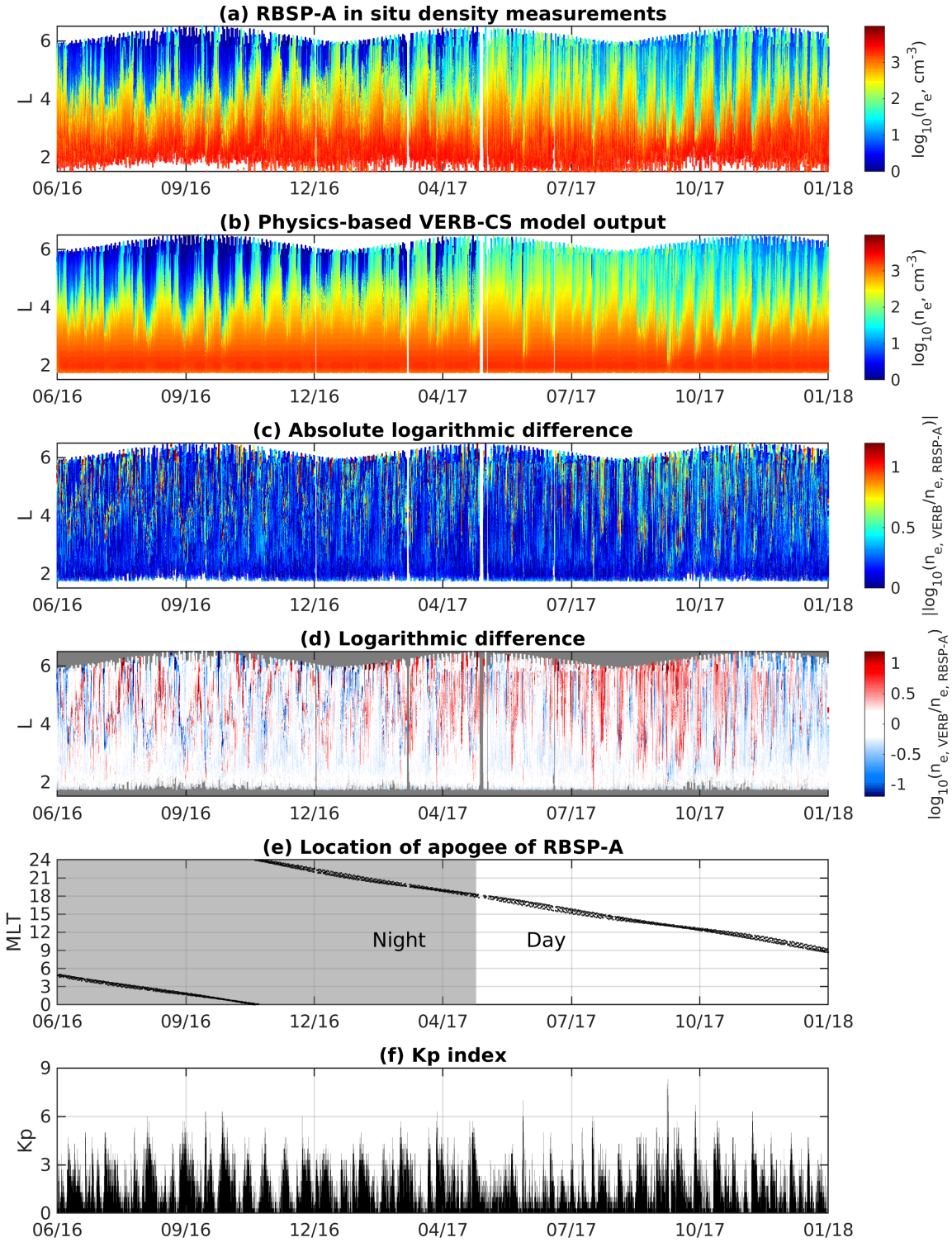

Figure 4.6: Long-term comparison of the physics-based model and the RBSP-A density measurements during July 2016 – January 2018. Panels (a) and (b) show the RBSP-A density measurements and the output of the physics-based model, respectively, where the L-shell is on the y-axis, time is on the x-axis, and the color shows the log of density. Panels (c), (d), and (e) show the absolute difference, the sign of the difference, and the difference between log of the model and data, respectively.

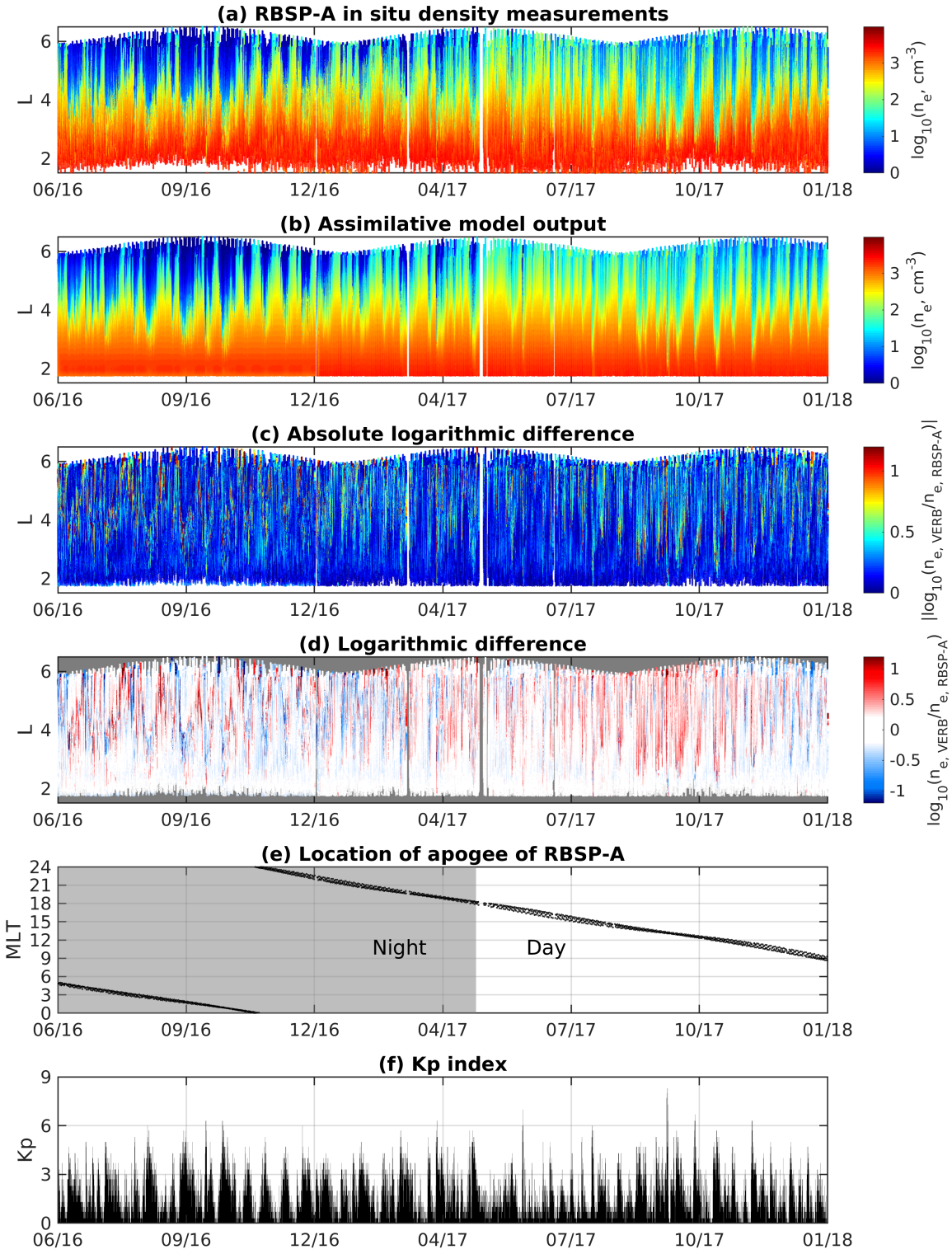

Figure 4.7: Long-term comparison of the assimilative model and the RBSP-A density measurements during July 2016 – January 2018. Panels (a) and (b) show the RBSP-A density measurements and the output of the assimilative model, respectively, where the L shell is on the y-axis, time is on the x-axis, and the color shows the log of density. Panels (c), (d), and (e) show the absolute difference, the sign of the difference, and the difference between log of the model and data, respectively.

corresponding geomagnetic conditions. Although, similarly to VERB-CS, it produces larger densities in the 12-17 MLT sector (in the second half of the interval), its output is in better agreement during the rest of the interval, compared to the VERB-CS model: the overestimation that was observed in the VERB-CS output is reduced. As can be seen from panel (c), the errors of the assimilative model are larger than those of PINE on the nightside but are comparable than those of PINE on the dayside. The densities inside the plasmasphere (at low L-shells) are lower compared to the observations from RBSP-A, which is caused by the use of the saturation density model (Carpenter and Anderson, 1992) in the assimilative model setup as well. Overall, the performance of the assimilative model improves on the dayside compared to VERB-CS and is similar to that of PINE. On the nightside, the assimilative model produces results closer to the VERB-CS.

Performance metrics

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

Figure 4.8 shows the RMSE (top row) and ME/bias (bottom row) of the PINE (left column), assimilative (middle), and VERB-CS (right) models calculated over the period from 02 July 2016 to 01 January 2018, used in the long-term simulations shown in Figures 4.5-4.7. It is worth mentioning again that the PINE model was trained on the interval 01 October 2012 – 01 July 2016 (Zhelavskaya et al., 2017), and therefore we exclude the period 30 June – 01 July 2016, when calculating the performance metrics here. In order to calculate the metrics, the data are binned in L and MLT, and the performance metrics are computed separately for each bin. The bins in L range from 1.5 to 6.5 with 0.5 bin size, and in MLT from 0 to 24 with bin size 1.

It can be seen that all the models have lower errors closer to the Earth (inside the plasmasphere), and that the errors increase with L. The errors of the PINE model are the lowest out of all models in terms of both RMSE and bias in all bins. The errors of the physics-based model are larger on the nightside and smaller for 6-14 MLT. The model has a larger bias in 12-18 MLT sector, however. On the nightside

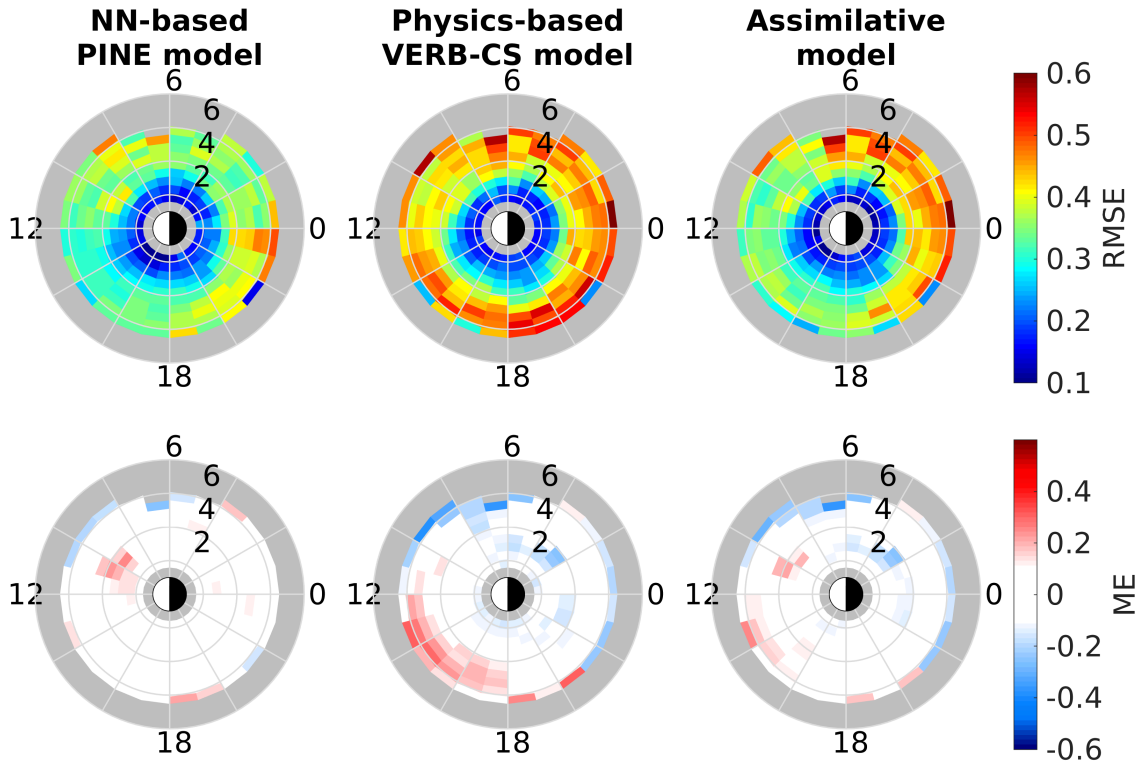

Figure 4.8: The root-mean-square error (top row) and the bias or mean error (bottom row) of the PINE, physics-based, and assimilative models for the 02 July 2016 – 01 January 2018 period. The Sun is to the left. The colorbar of each row shows the value of the corresponding metric (RMSE or ME). The colorbar limits are the same for all models in each row.

the bias is smaller and tends to be negative. This result is similar to the one shown in Figure 4.6, and implies that the VERB-CS model systematically produces slight overestimation of density in the 12-18 MLT sector, than is observed. This may be caused by the use of increased refilling rates (by the factor of 1.75). Our experiments (not shown here) demonstrated that the model without such increased refilling rates produces significant underestimation of density on the nightside, meaning that the code produces a more eroded plasmasphere on the nightside than is observed. After performing sensitivity tests to all the input parameters of the VERB-CS code (magnetic field, electric field, initial conditions, boundary conditions, etc.), we found that changes in the electric field have the most impact on this behavior (not shown here). Modifying the shielding parameter γ changes the extent of erosion significantly. From sensitivity tests (not shown), we found that using $\gamma = 1.8$ provides better agreement with observations than using the standard $\gamma = 2$ (Maynard and Chen, 1975). Therefore, we use $\gamma = 1.8$ in these simulations. This aspect of the VERB-CS code requires further investigation and testing, which we discuss in more detail in section 4.5.

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

This test still illustrates that the assimilative methodology provides quantitative improvement in performance compared to the performance of the VERB-CS model (in particular, on the dayside). The performance of the assimilative model is comparable to the PINE model performance on the dayside but is closer to the performance of

VERB-CS on the night side. It is clear that further improvement in the physics-based model (VERB-CS), or using a more advanced model, will improve the performance of the assimilative model. We note that although the VERB-CS code and assimilative model reproduce the location of the plasmapause during storm times better than the PINE model (Figures 4.2 and 4.3), they show larger errors in low-density regions outside the plasmasphere compared to PINE (Figure 4.8). These are two different aspects of the models: (1) ability to predict plasmapause correctly and (2) ability to reproduce density accurately.

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

4.5 Discussion

Our results show that the assimilative methodology employed in this study for combining the neural network PINE model and the physics-based VERB-CS code demonstrates great potential for combining advantages of both models. Namely, the assimilative model demonstrated good performance on a series of test events from the

IMAGE era for a variety of geomagnetic conditions: quiet, moderate, disturbed, and extreme geomagnetic storms. The output of the model showed better agreement with the plasmapause locations derived from IMAGE EUV than PINE or VERB-CS, when used separately (Figures 4.2 and 4.3). As discussed above, the performance of neural networks is limited by the training data. As there are no examples of extreme geomagnetic storms in the training dataset of PINE, its performance is reduced during such events. The assimilative methodology helps eliminate this limitation. The assimilative model also demonstrated good capabilities in combining quantitative advantages of models in terms of performance metrics. Long-term simulations of plasma density using all the models show that the current setup of the assimilative model allows for improving the performance of VERB-CS considerably in the dayside local time sector.

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

An important aspect of the assimilative approach employed here is the choice of model and observation errors α^m and α^{obs}. In this study, we employed an approach similar to Kondrashov et al. (2011), which was adjusted to use different constant values for model and observation errors α^m and α^{obs} inside and outside of the plasmasphere. We have compared the results obtained using such an approach to using constant values of errors throughout all radial extent of models (not shown here).

We found that using different values of errors for inside and outside the plasmapause works better in our case and provides better agreement with observations. It is worth noting that selecting the model and observation errors is one of the most difficult topics in data assimilation and is still an active area of research (e.g., Berry and Harlim, 2017; Bishop, 2019; Hamilton et al., 2019); the existing approaches to select them are mostly empirical. Investigating the selection of these errors in a systematic way and experimenting with the dependence of the errors on other parameters such as L, MLT, and/or geomagnetic activity should be the subject of future research.

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

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

It can also be seen from the results that, in some cases, the plasmasphere produced by the VERB-CS is more extended than was observed, in particular during geomagnetically quiet times (e.g., first row of Figure 4.2 and of Figure 4.3). This could be attributed to the refilling rates used. We have used long-term median refilling rates from Denton et al. (2012) (without accounting for solar-cycle dependence). The

model assumes that there is no significant dependence of the refilling rate on MLT. The same refilling rates are assumed for all geomagnetic conditions. At the moment, there still remain unsolved problems in the processes of plasmasphere ion refilling from the ionosphere (Gallagher and Comfort, 2016). Therefore, this topic should be explored further in regard to the physics-based modeling of the plasmasphere, and other approaches to parameterizing the refilling should be tested. For example, De Pascuale et al. (2018) used the approach of Rasmussen et al. (1993) to model the refilling using the Cold PLasma physics-based model CPL (Jordanova and Miyoshi, 2005; Jordanova et al., 2010b,a, 2014), where the approach of equatorial plasma densities toward equilibrium depends on the variation from the saturation level and a timescale. The timescale of refilling depends on the local time in addition to L, and was calculated from the differences in outgoing ion flux into the plasmasphere at hemispheric boundaries using empirical atmosphere models, including the MSIS-86 thermosphere model (Hedin, 1987), and IRI ionosphere model (Bilitza, 1986), in their study.

From long-term simulations (Figure 4.6), it can be seen that density inside the plasmasphere (on low L-shells) is slightly smaller on average than that of RBSP-A. This difference can be explained by the fact that the saturation model (Carpenter and Anderson, 1992) incorporated into the VERB-CS model provides lower saturation density on average than observed, using density measurements from the Van Allen Probes. Further investigation of other saturation density models or constructing a new saturation model that includes density measurements from the Van Allen Probes is required to improve the VERB-CS model performance.

It is also worth noting that in this study we inspect and compare the global outputs of the models and the observed and modeled plasmapause locations visually. It would be ideal to conduct a statistical analysis comparing the observed and modeled plasmapause locations. However, such analysis is a very challenging task. We discuss the challenges pertaining to it in detail in Supporting Information (please see Figures S8-S11 and their discussion). Due to the challenges discussed there, we perform the visual inspection of the global outputs of the models in this study. However, we are planning to develop tools for such comparisons in order to analyze the plasmapause locations statistically in the future.

The results obtained in this study illustrate that the assimilative methodology can be applied to combine both the qualitative and quantitative advantages of the

VERB-CS and PINE models. It is clear that further improvement of the mentioned models or use of more sophisticated models in the physics-based VERB-CS code will improve the performance of the assimilative model. The methodology developed in this study will be especially useful for modeling the plasmasphere dynamics during geomagnetic storms and extreme events, such as the Halloween storms, while also providing realistic density values during quiet and moderate geomagnetic conditions. The combined data assimilative model is not computationally expensive and can be used as a part of global models of the magnetosphere or coupled with ring current and radiation belt codes. The current study is intended to be a demonstration of the potential of the assimilative methodology and should be considered more as an illustration of how the physics-based and neural network models can be combined, rather than as a tool.

4.6 Summary and Conclusions

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

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

The tests conducted in this study show that the neural network model PINE has a good quantitative performance on average and reproduces the general dynamics of the plasmasphere well, such as erosion on the nightside and plume formation.

Its performance is limited, however, for Kp $>$ 7 due to the lack of training data. The physics-based VERB-CS code also reproduces the dynamics of the plasmasphere well, and is especially effective during high geomagnetic activity and extreme geomagnetic events. However, its quantitative performance is lower than PINE's. Using the Kalman filter technique of data assimilation, we were able to combine the advantageous features of both models. The assimilative model is capable of reproducing the dynamics of the plasmasphere well during both quiet and disturbed geomagnetic activity, including extreme geomagnetic events. Its quantitative performance is better than that of VERB-CS and is comparable to PINE's for the dayside local time sector.

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

4.A The updated version of the PINE model

As discussed in section 4.3.1, Zhelavskaya et al. (2017) used K-fold cross validation with random splitting of data into training and validation folds to validate the models. They also used this procedure to select optimal input variables to the model. They considered several different combinations of solar wind parameters and geomagnetic indices as potential inputs to the neural network. In particular, they considered models based solely on geomagnetic indices (Kp, AE, Sym-H, and also F10.7), solely on solar wind data (solar wind speed, dynamic pressure, proton density, and the interplanetary magnetic field (IMF) B_z), and on a combination of both. As the time history of previous conditions is important for the plasmasphere dynamics, they also considered different durations of time history of these parameters as inputs, starting with simple models based only on instantaneous values of activity parameters and subsequently adding more time history of the corresponding parameters to the networks, up to 120 h of time history. The time history was represented as averages of the time histories of activity parameters integrated from hour 0 (e.g., 0-3, 0-6, 0-12 h, etc.). Every neural network also included a location input, as given by L and MLT.

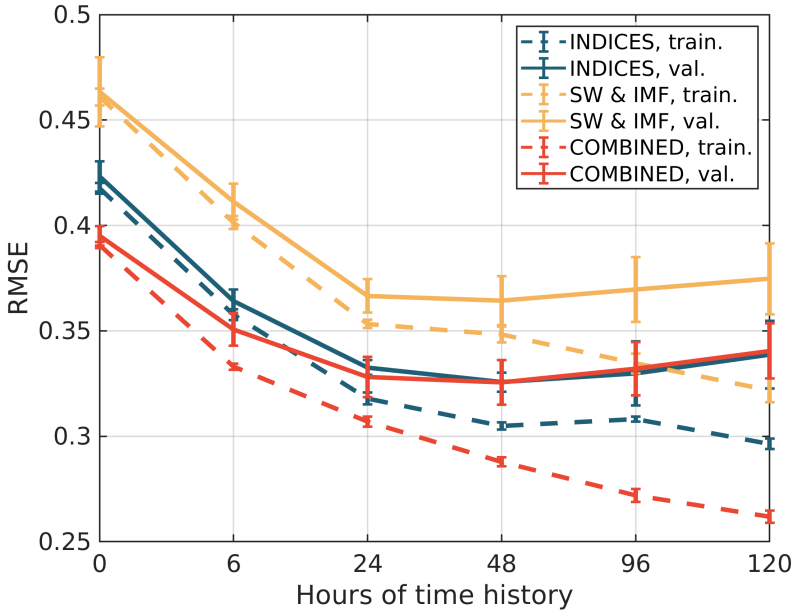

Figure A1: Root-mean-square error (RMSE) on the y-axis versus the hours of time history included in the models. The yellow color shows the errors of models based on solar wind, the blue color is for the models based on geomagnetic indices, and red is for the models based on both of them combined. Solid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

In this study, we extend this analysis using the K-fold cross validation procedure described in section 4.3.1. We consider the same combinations of input parameters to the neural networks. The neural networks are trained on the density measurements from both RBSP-A and RBSP-b during 01 October 2012 – 01 July 2016. We use cross validation to obtain validation and training errors, and the standard deviations of errors. As described in section 4.3.1, all available data for this time interval are split into 35-day blocks. At first, 10% of the data are left aside as a testing dataset. Then the remaining 35-day blocks are randomly assigned to the training or validation sets. This type of data split allows the sequentiality of data to be preserved, and also introduces randomness and representation of different geomagnetic conditions in both validation and training sets. The rest of the methodology is identical to that of Zhelavskaya et al. (2017).

Figure A1 shows the root-mean-square error (RMSE) plotted against the number of hours of time history included into the model. The yellow color indicates the errors of models based on solar wind, the blue color is for the models based on geomagnetic

indices, and red is for the models based on both of them combined. Solid lines show the validation errors and dashed lines the training errors. The error bars show the standard deviation of error obtained during the cross validation procedure. The validation error represents how well a model performs on the unseen data and is the error we aim to minimize. The difference between the training and validation errors indicates if a model overfits the data or not. When the difference is too large, this means that a model "learned" the training data too well – it memorized it essentially, and due to that performs poorly on the unseen data. As a consequence, it does not have good generalization capabilities.

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

4.B The model and observation error of the Kalman filter

The model and observation errors α^m and α^{obs} were obtained as outlined below. We use the results of the long-term density reconstruction obtained in section 4.4.3. There, PINE and VERB-CS were run for the period of 30 June 2016 – 01 January 2018, and their output was compared to the in-situ density measurements from RBSP-A during that period. The performance of both models was analysed using the RMSE performance metric. Here, we have also computed the RMSE of both models for locations inside and outside the plasmapause of the respective model separately for the period of 02 July 2016 – 01 January 2018. The plasmapause was calculated using the density threshold of 40 cm^{-3}, as described in section 4.2. Figure B1 (panels a-d) shows the histograms of error distributions during this period for both models at different locations. The RMSEs are labeled inside the respective panels. We have employed the RMSE values obtained in this analysis as model and observation errors α^m and α^{obs} inside and outside the plasmapause of each model.

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

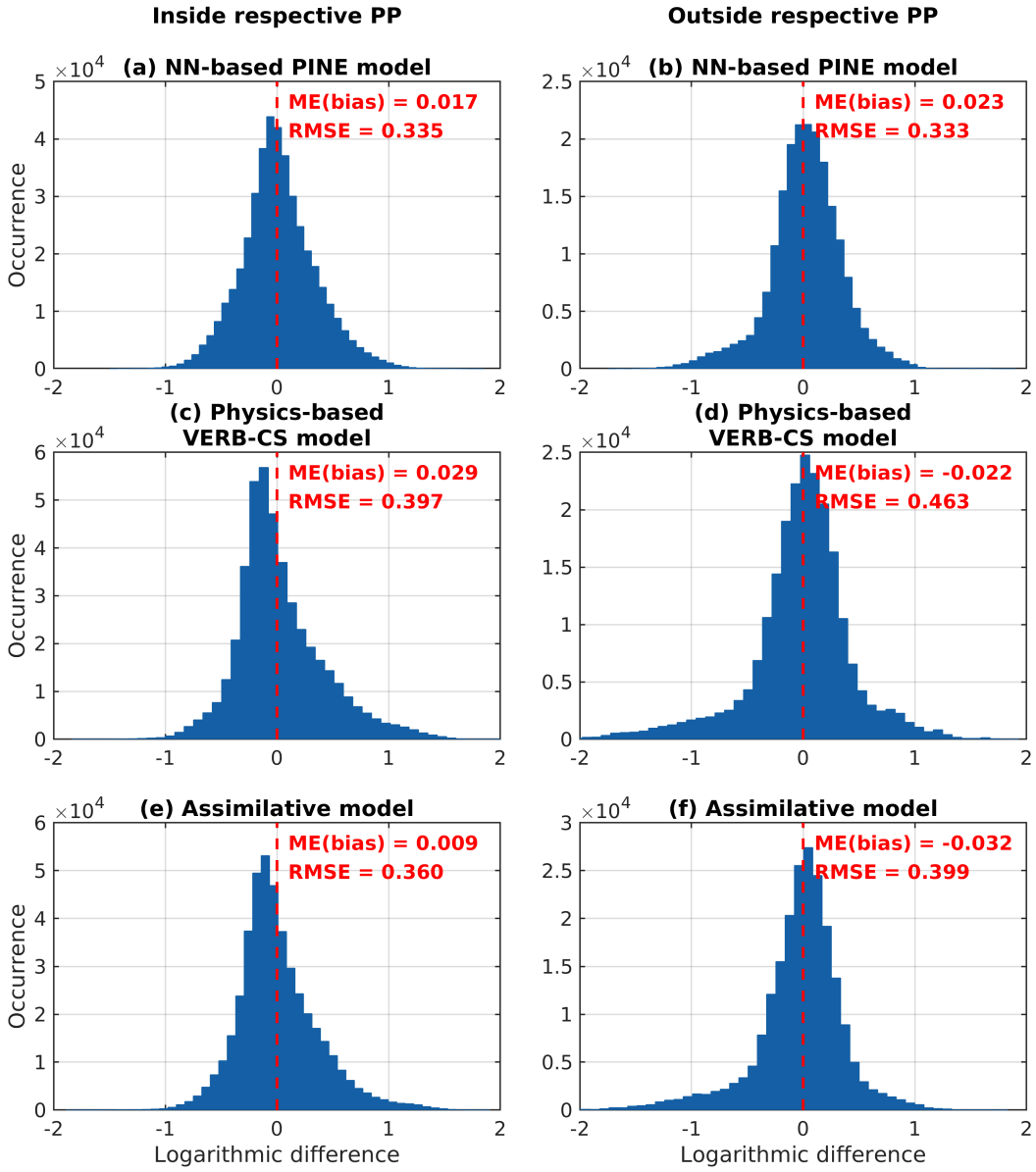

Figure B1: Distribution of errors of PINE, VERB-CS, and the assimilative model inside and outside the plasmapause during the out-of-sample period of 02 July 2016 – 01 January 2018 (compared to density measurements from RBSP-A). The respective RMSE and ME are given inside each panel.

Supporting Information

Contents of this file

1. Figures S1 to S7 show the sequences of the comparisons of the models throughout the storms presented in Figure 4.3 of the main paper. These sequences illustrate the predictions of the models at different storm phases.

2. Figures S8 to S11 demonstrate the issues with the statistical analysis of plasmapause locations.

3. Figures S12 to S15 show the comparison of models performance with modified refilling rates.

Introduction

In the main manuscript, we presented comparisons of outputs of neural network model PINE, physics-based model VERB-CS, and assimilative model AMP for a number of selected times during several storms (Figure 4.3). In this supporting information, we provide additional figures and information about the models output. Figures S1-S7 show the sequences of comparisons of model output throughout the storms presented in Figure 4.3. Figures S8-S12 and their description demonstrate the issues that may arise in statistical analysis comparing the plasmapause locations derived from models and from data. Figures S13-S16 show the comparison of the performance of the VERB-CS and assimilative models using different versions of refilling rates.

Models output at different storm phases

Figures S1-S7 illustrate the predictions of the models at different storm phases of the events presented in Figure 4.3 of the main manuscript. The format of Figures S1-S7 is the same as of Figure 4.2 of the main manuscript. The first four rows show the outputs of the models corresponding to the times marked with the red lines in the bottom panel showing the Kp index during the 2003 Halloween storm. The black-and-white dots show the location of the plasmapause derived from the IMAGE EUV images. The color in the first four rows indicates the logarithm of density (the scale of the colorbar is the same for all models and all times). The gray and black section of the colorbar indicates a density threshold of $40 \pm 10\,\mathrm{cm}^{-3}$ and can be considered a

rough approximation of the plasmapause location for the sake of comparison to the observed plasmapause position obtained from IMAGE EUV (more details on that are given in section 2). The Sun is to the left.

Figure S1 shows the comparison of the models for the storm on April 11, 2001, corresponding to event (a) in Figure 4.3. The plasmapause derived from the IMAGE EUV instrument spanned from 11-Apr-2001 00:04:00 to 14-Apr-2001 15:07:00.

Figure S2 shows the comparison of the models for the storm on June 18, 2001, corresponding to event (b) in Figure 4.3. The plasmapause derived from the IMAGE EUV instrument spanned from 18-Jun-2001 00:10:00 to 18-Jun-2001 20:26:00.

Figure S3 shows the comparison of the models for the storm on May 8-9, 2001, corresponding to event (c) in Figure 4.3. The plasmapause derived from the IMAGE EUV instrument spanned from 08-May-2001 00:01:00 to 14-May-2001 03:54:00.

Figure S4 shows the comparison of the models for the storm on March 20, 2001, corresponding to event (d) in Figure 4.3. The plasmapause derived from the IMAGE EUV instrument spanned from 20-Mar-2001 09:33:00 to 20-Mar-2001 17:44:00.

Figure S5-S7 show the comparison of the models for the storm on March 30-April 2, 2001, corresponding to event (e) in Figure 4.3, during initial, main, and recovery phases of the storm, respectively. The plasmapause derived from the IMAGE EUV instrument spanned from 30-Mar-2001 00:05:00 to 02-Apr-2001 19:13:00.

Comparison of the modeled and observed plasmapause locations

Below we outline a number of challenges pertaining to comparison of the modeled and observed plasmapause locations in a statistical manner.

First of all, comparison of the real plasmapause and plasmapause derived from the models is not straightforward. In particular, it is not clear whether different plasmapause locations should be compared for the same MLT sector or in some other way. Figures S8-S11 illustrate several possible scenarios of comparison between modeled and observed plasmapause. In these figures, we generated synthetic "observed" and "modeled" plasmapause locations to demonstrate this point.

Figure S8 shows the scenario, when the predicted and observed shapes of the plasmasphere are both circular. Here, we introduced a slight shift of the plasmapause by 0.5 Re in -x and -y directions into the "model". In this case, the RMSE represents

the error of the model well.

Figure S9 and S10 illustrate the case, when the predicted shape of the plasmasphere is correct, however, the modeled plasmapause is slightly rotated compared to the observed one. The "modeled" plasmapause is the absolutely the same as the "observed" plasmapause, but just rotated by 10 degrees in Figure S9, and by 25 degrees in the Figure S10. In this case, the differences between plasmapause locations are large for the same MLT sectors in the dusk sector. At the same time, one can say that the predicted shapes are very similar except for the slight rotation (at least for the case of rotation by 10 degrees).

Figure S11 shows the case, when the predicted shape of the plasmasphere is not correct in the dusk sector (no plume is predicted), but the size of the main body of the plasmasphere is almost correct. The RMSE is similar to the scenario presented in Figure S10 (with rotation by 25 degrees). However, in reality the shape of the plasmasphere is not predicted correctly, since the plume is not reproduced. The same applies when the model predicts a plume, when in reality no plume was observed.

Figures S10 and S11 illustrate that while the errors of the model may be similar, the behavior of the model can be quite different in reality. Thus, comparing the modeled and observed plasmapause locations using a single error metric, such as RMSE, for different MLT sectors may not adequately explain the physical reasons for the differences between modeled and observed plasmapause.

In addition to the challenges outlined above, there are other complications in comparing the plasmapause locations. The plasmapause identified in IMAGE is manually derived from the gradients in He^+ ion distribution imaged by the EUV instrument. Sometimes the plasmapause may be diffuse (not sharp), particularly on the dayside, which results in multiple dots for the same MLT selected as plasmapause in the IMAGE EUV images (for example, as in Figure S1b). This may further complicate the comparison between modeled and observed plasmapause. The definition of the plasmapause itself is ambiguous, as some authors use a density threshold and some use density gradients to define it (in this study, we use density threshold for comparison with IMAGE EUV).

Due to these factors, the statistical analysis comparing plasmapause locations may not reveal the real errors of the models. Ideally it should be performed in a way in which we have the ability to quantify the differences in the rotation of the plasmasphere and in the preservation of the shape (plume, or no plume) in the model

and IMAGE observations.

Influence of refilling rates

Our experiments have shown that a better performance is achieved by VERB-CS when refilling rates of Denton et al. (2012) are multiplied by a factor of 1.75. The VERB-CS model incorporating unchanged refilling rates by Denton et al. (2012) has a negative bias and systematically underestimates density on the nightside. The increased refilling rates improve that aspect of the model. The comparison between two setups is shown in the figures below (in Figure S12, the setup with unchanged refilling rates is in the left column and the setup with refilling rates multiplied by the factor of 1.75 is in the right column; in Figure S13, setup with unchanged refilling rates - top row, setup with refilling rates multiplied by 1.75 - bottom row).

The results of the assimilative model with increased refilling rates are consequently improved compared to the model with unchanged refilling rates. Figures S14 and S15 show the comparison between these two setups. The format of these figures is the same as Figures S12 and S13. The model and observation errors are the values of RMSEs presented in Figure S14 for each setup, respectively.

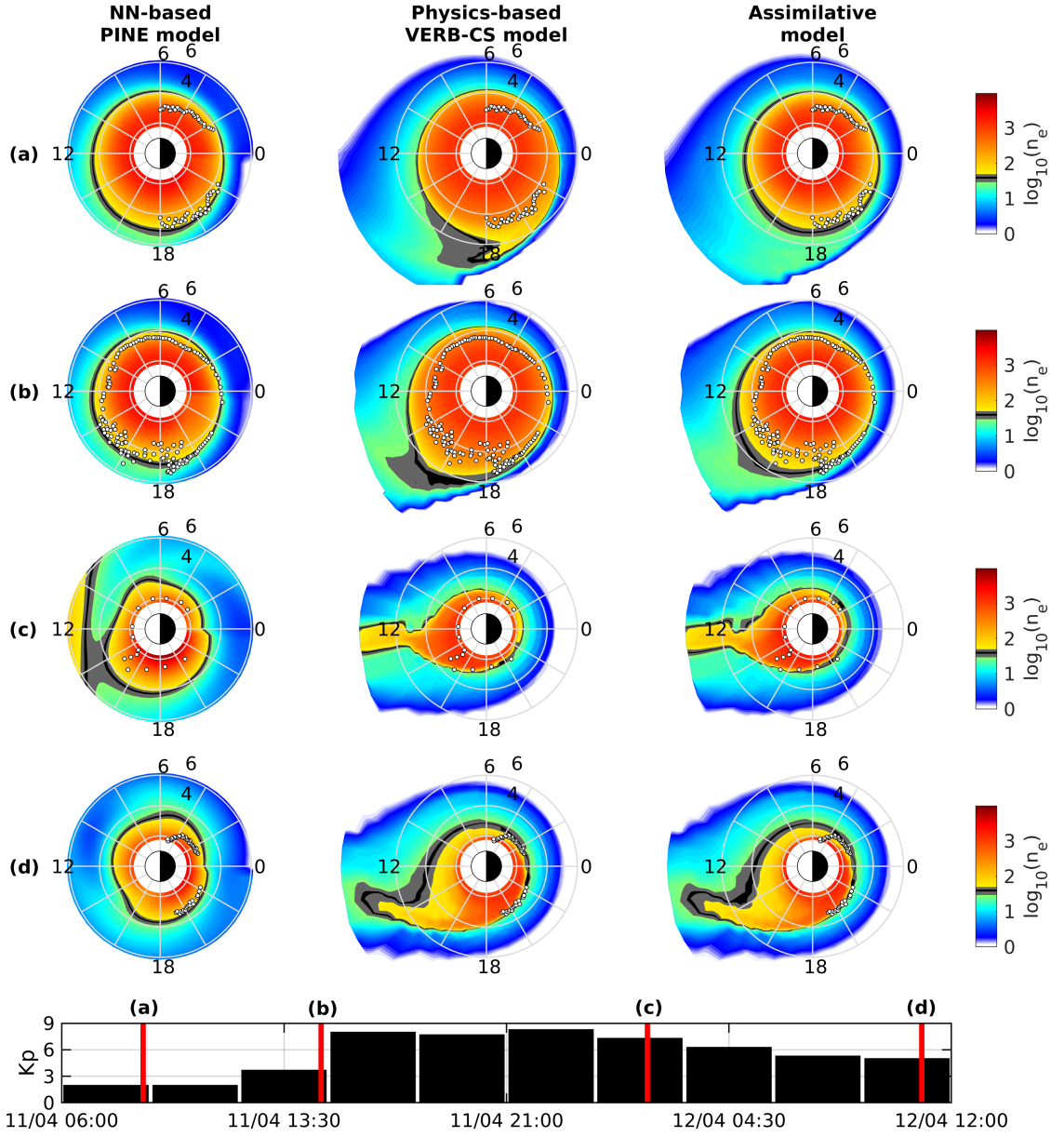

Figure S1: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the storm on April 11, 2001, corresponding to event (a) in Figure 4.3 of the paper. The format of the figure is described in the paper and above.

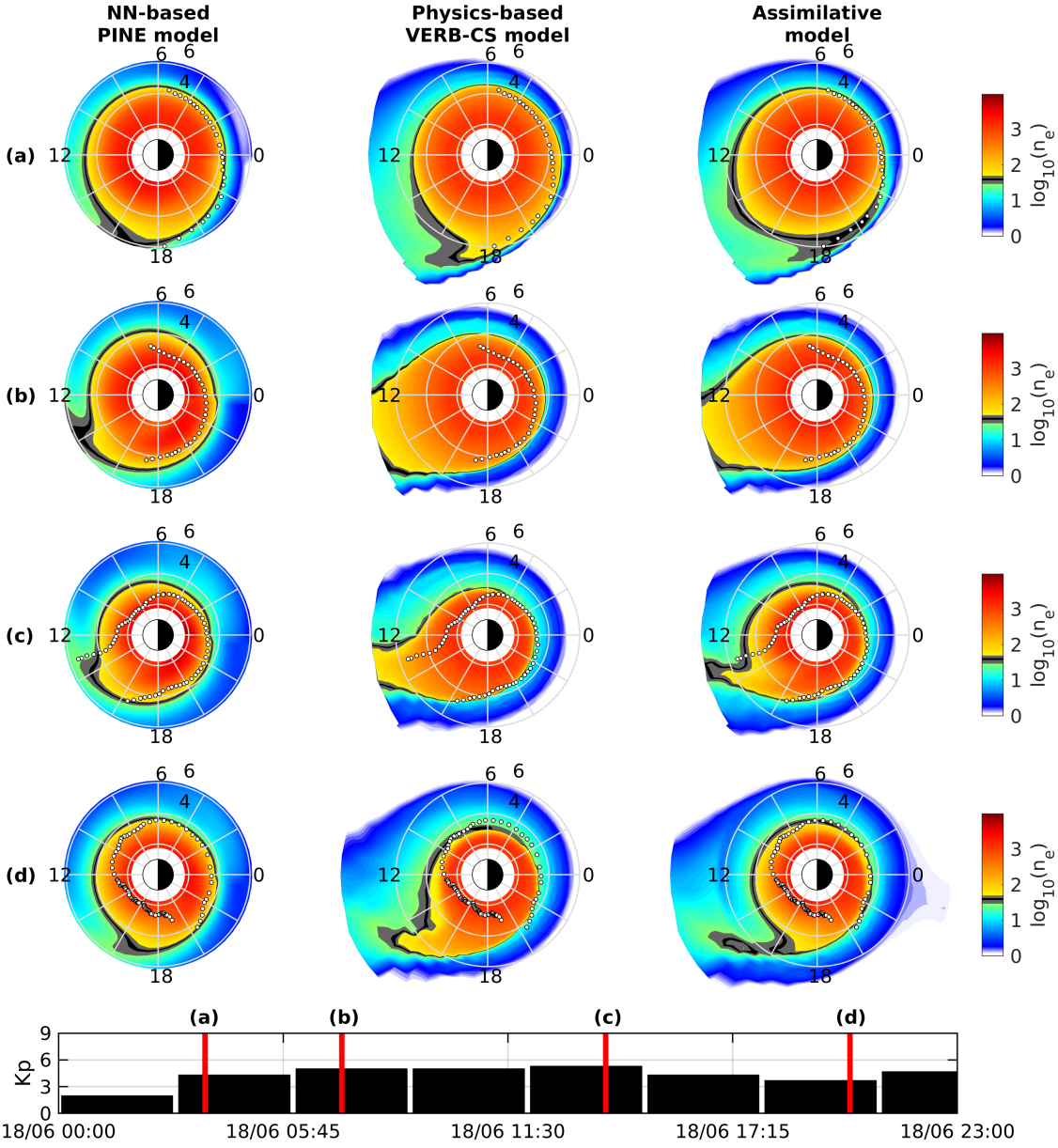

Figure S2: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the storm on June 18, 2001, corresponding to event (b) in Figure 4.3 of the paper. The format of the figure is described in the paper and above.

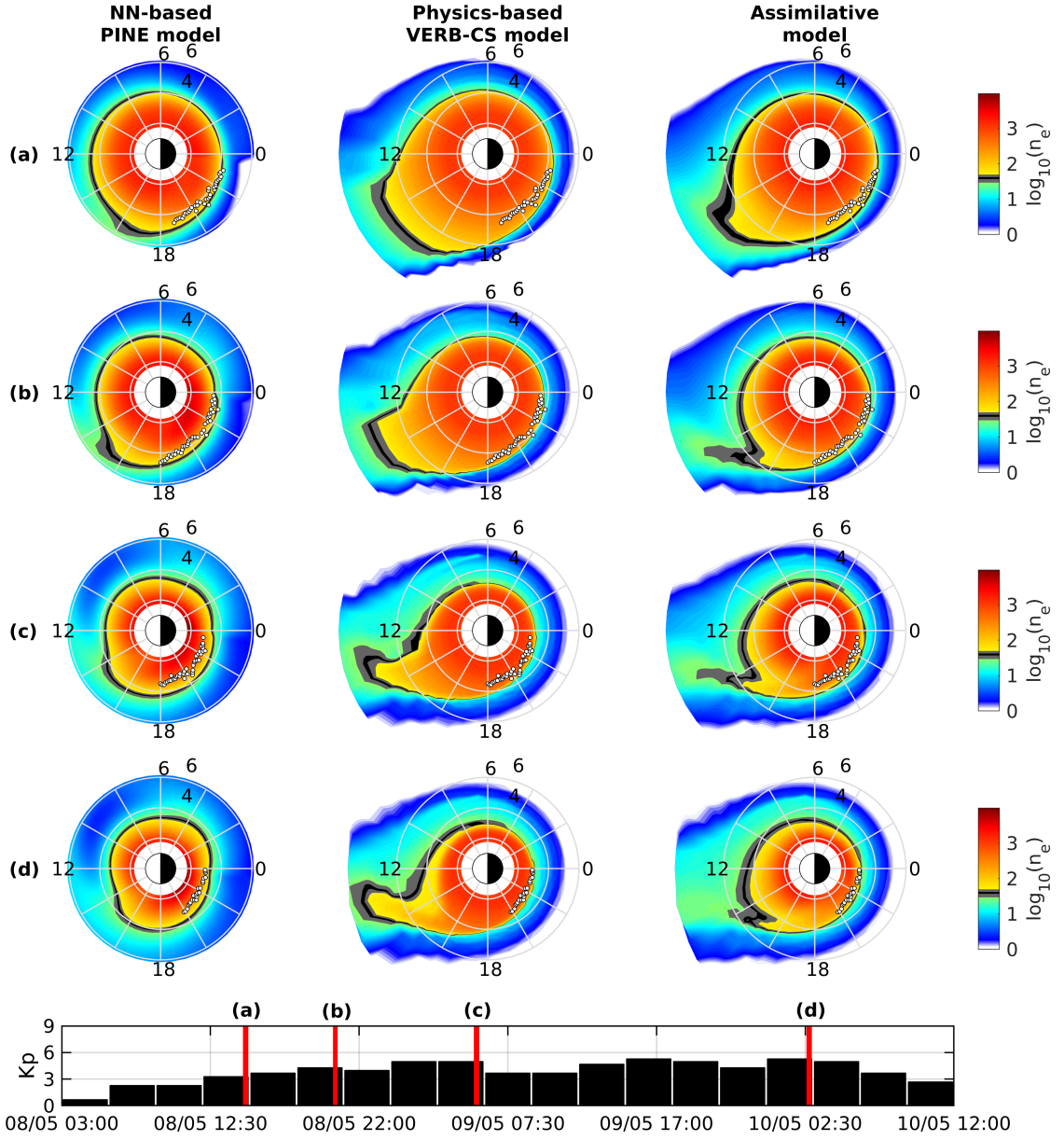

Figure S3: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the storm on May 8-9, 2001, corresponding to event (c) in Figure 4.3 of the paper. The format of the figure is described in the paper and above.

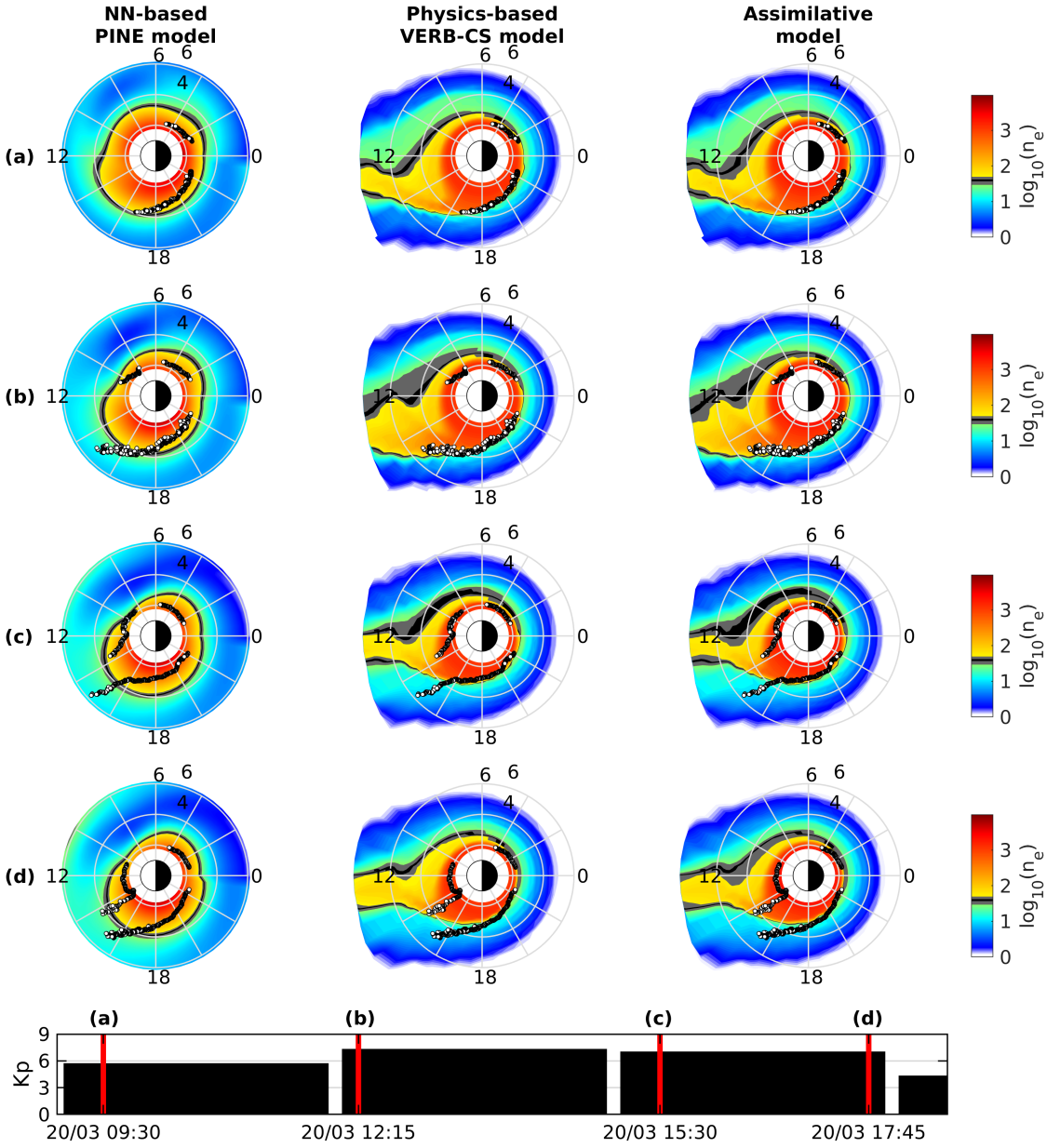

Figure S4: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the storm on March 20, 2001, corresponding to event (d) in Figure 4.3 of the paper. The format of the figure is described in the paper and above.

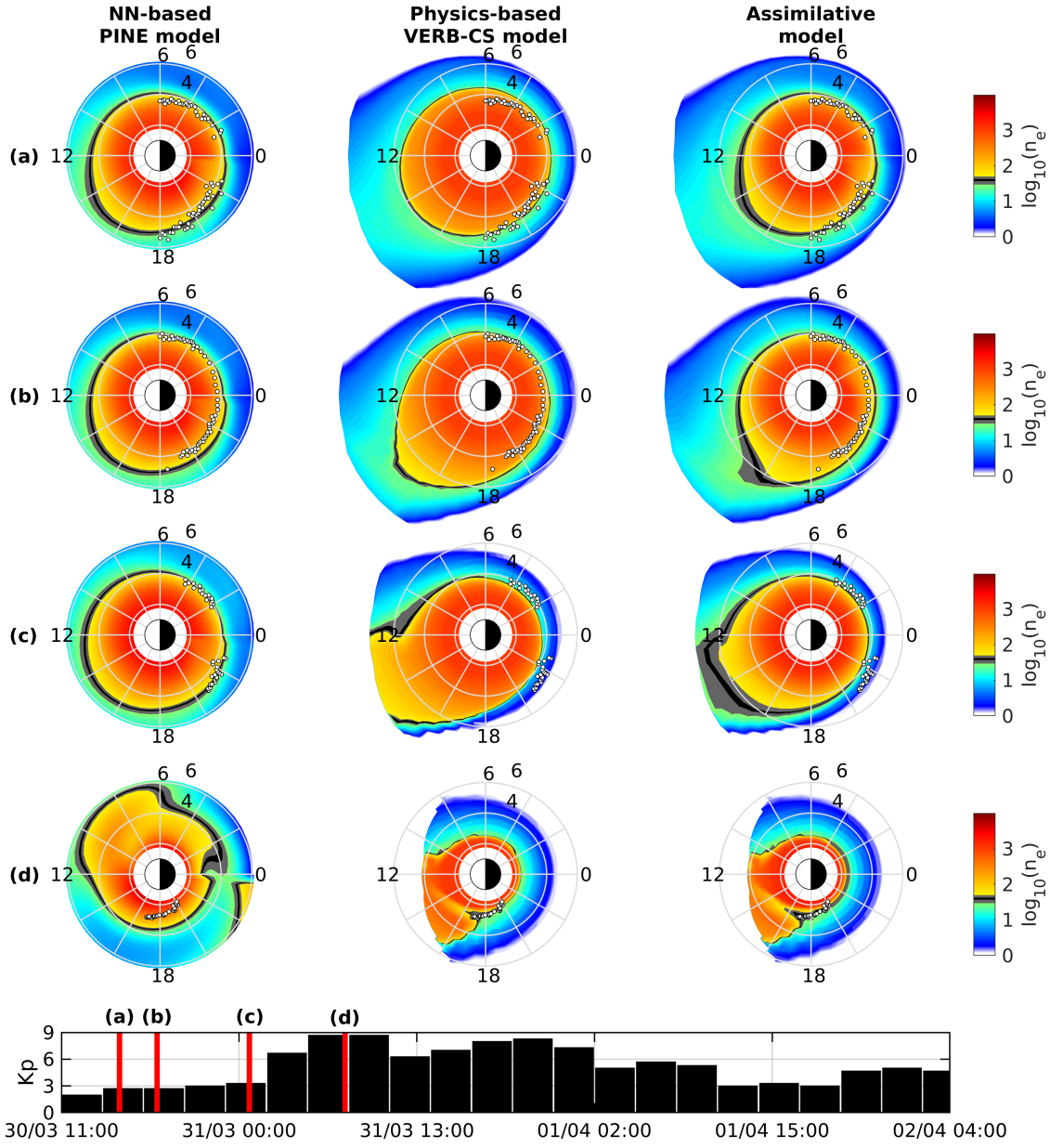

Figure S5: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the initial phase of the storm on March 30-April 2, 2001, corresponding to event (e) in Figure 4.3 of the paper. The format of the figure is described in the paper and above.

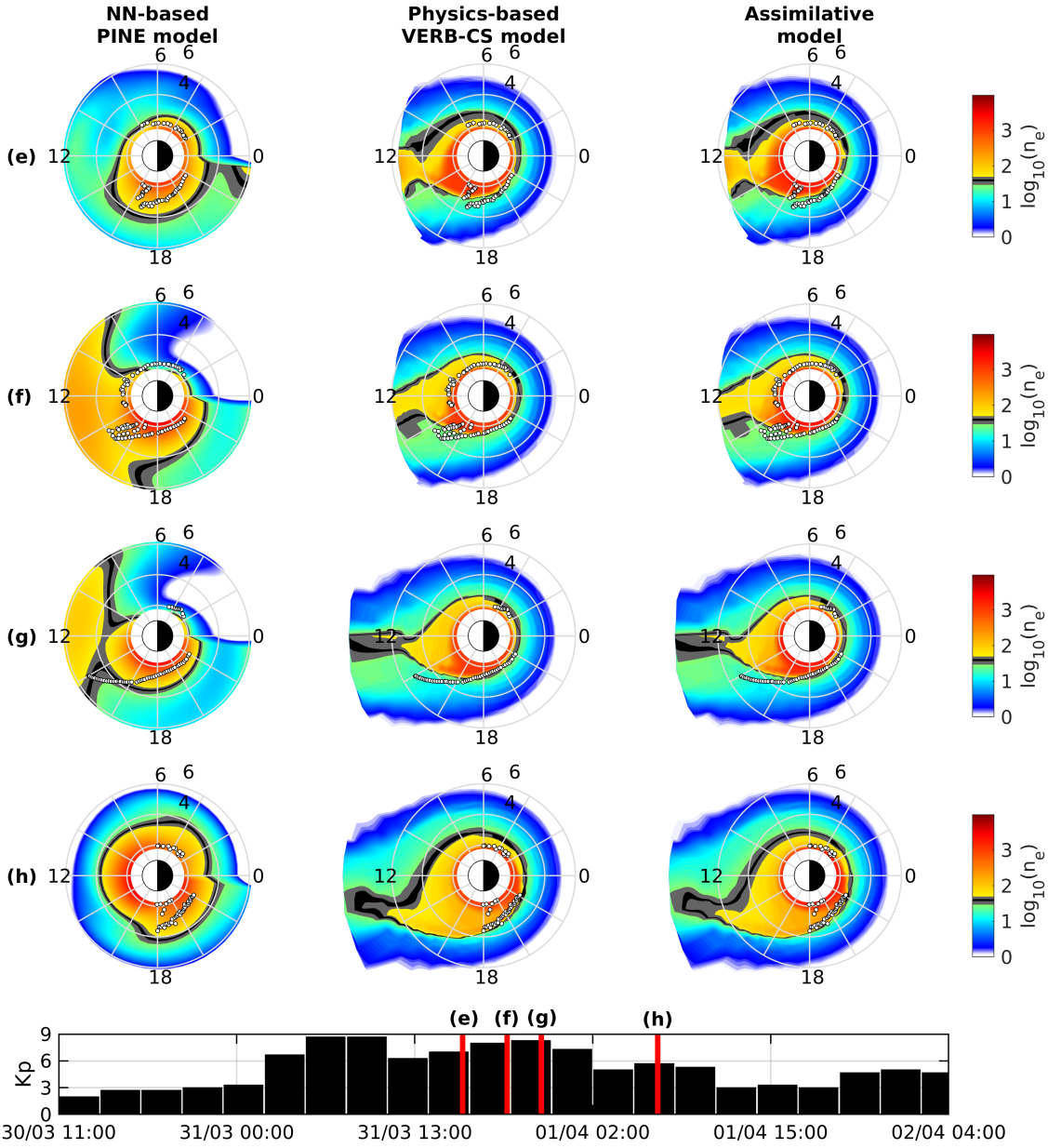

Figure S6: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the main phase of the storm on March 30-April 2, 2001, corresponding to event (e) in Figure 4.3 of the paper. The format of the figure is described in the paper and above.

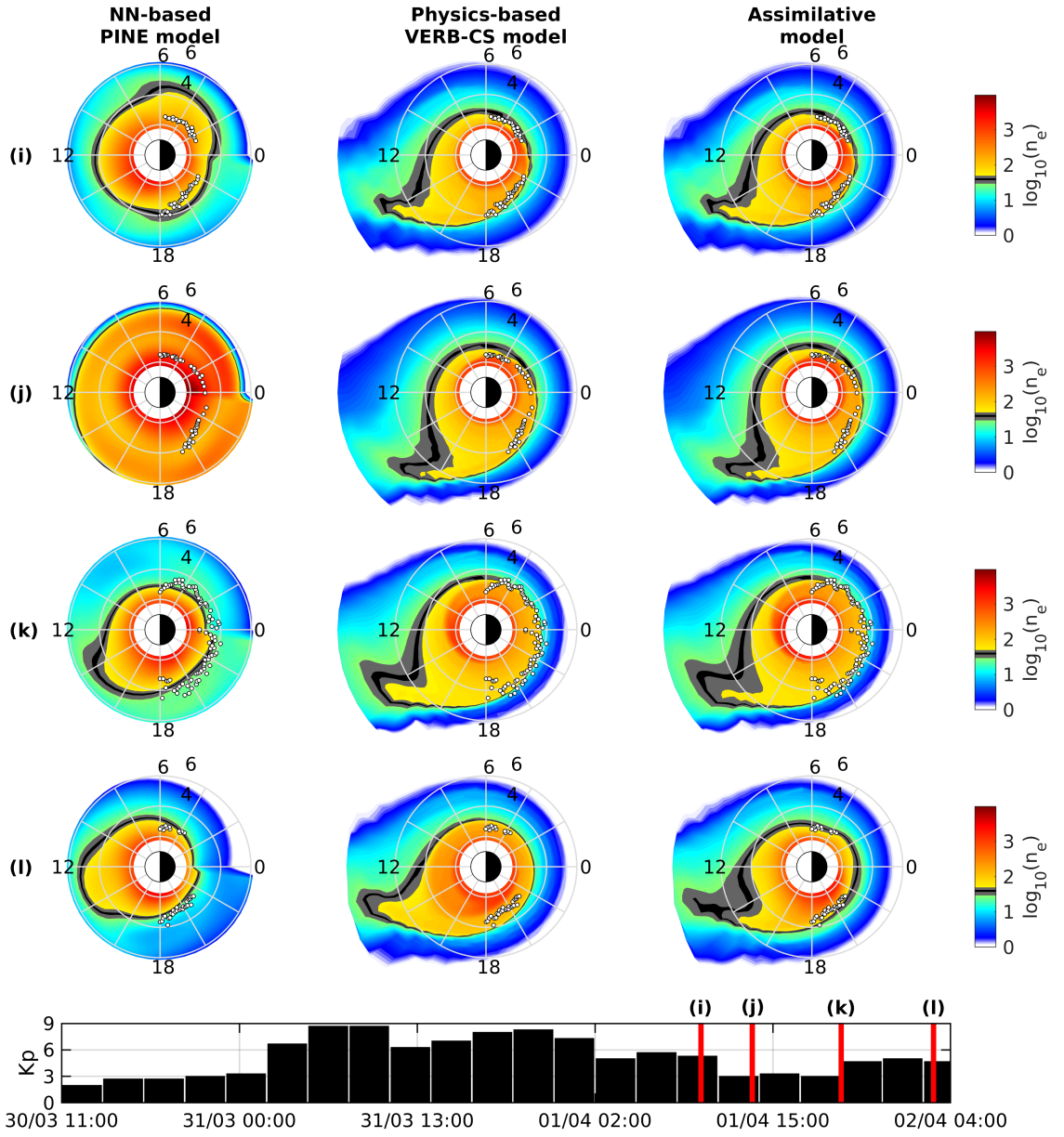

Figure S7: Comparison of the PINE (left), VERB-CS (middle), and assimilative model (right) outputs during the recovery phase of the storm on March 30-April 2, 2001, corresponding to event (e) in Figure 4.3 of the paper. The format of the figure is described in the paper and above.

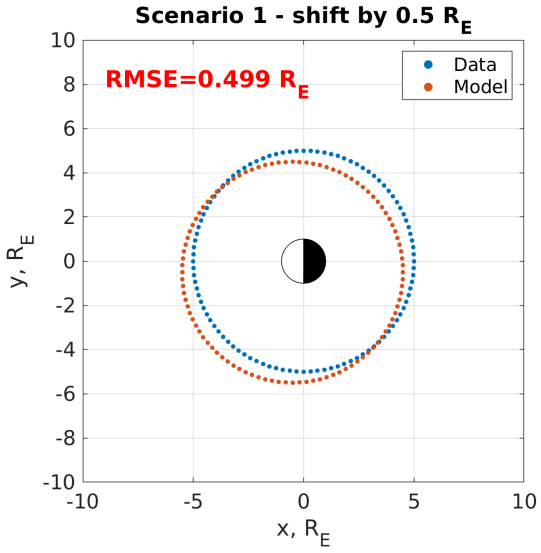

Figure S8: The predicted and observed shapes of the plasmasphere are both circular. Here, we introduced a slight shift of the plasmapause by 0.5 Re in -x and -y directions into the "model". For further details, please refer to the text.

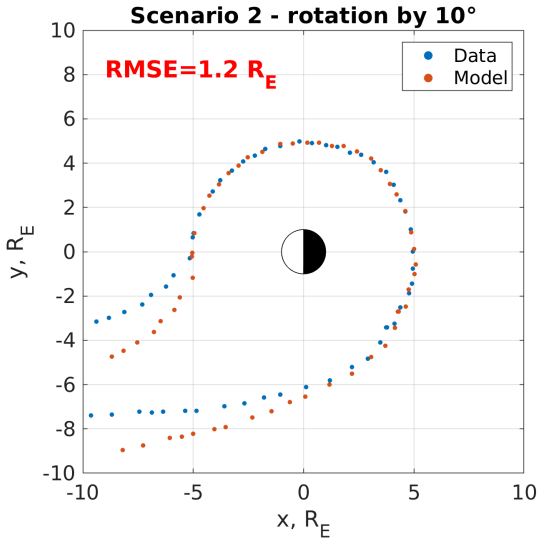

Figure S9: The predicted shape of the plasmasphere is correct, however, the modeled plasmapause is slightly rotated compared to the observed one. The "modeled" plasmapause is the absolutely the same as the "observed" plasmapause, but just rotated by 10 degrees. For further details, please refer to the text.

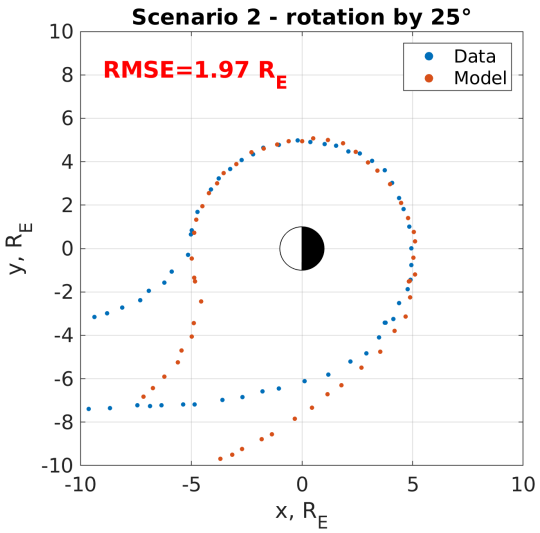

Figure S10: The predicted shape of the plasmasphere is correct, however, the modeled plasmapause is slightly rotated compared to the observed one. The "modeled" plasmapause is the absolutely the same as the "observed" plasmapause, but just rotated by 25 degrees. For further details, please refer to the text.

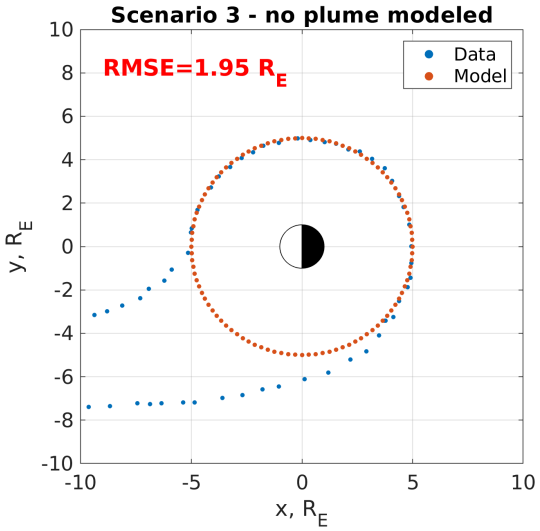

Figure S11: The predicted shape of the plasmasphere is not correct in the dusk sector (no plume is predicted), but the size of the main body of the plasmasphere is almost correct. For further details, please refer to the text.

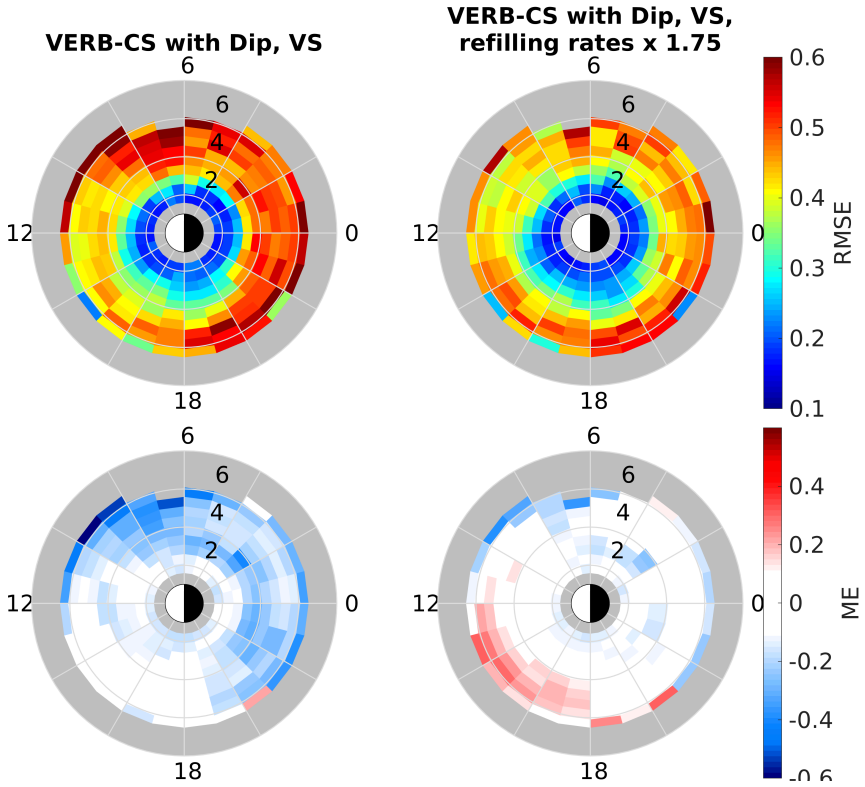

Figure S12: The root-mean-square error (top row) and the bias or mean error (bottom row) of the VERB-CS model with the dipole magnetic field, the Volland-Stern electric field model and refilling rates from Denton et al. (2012) (left), and with the same setup but refilling rates multiplied by 1.75 (right).

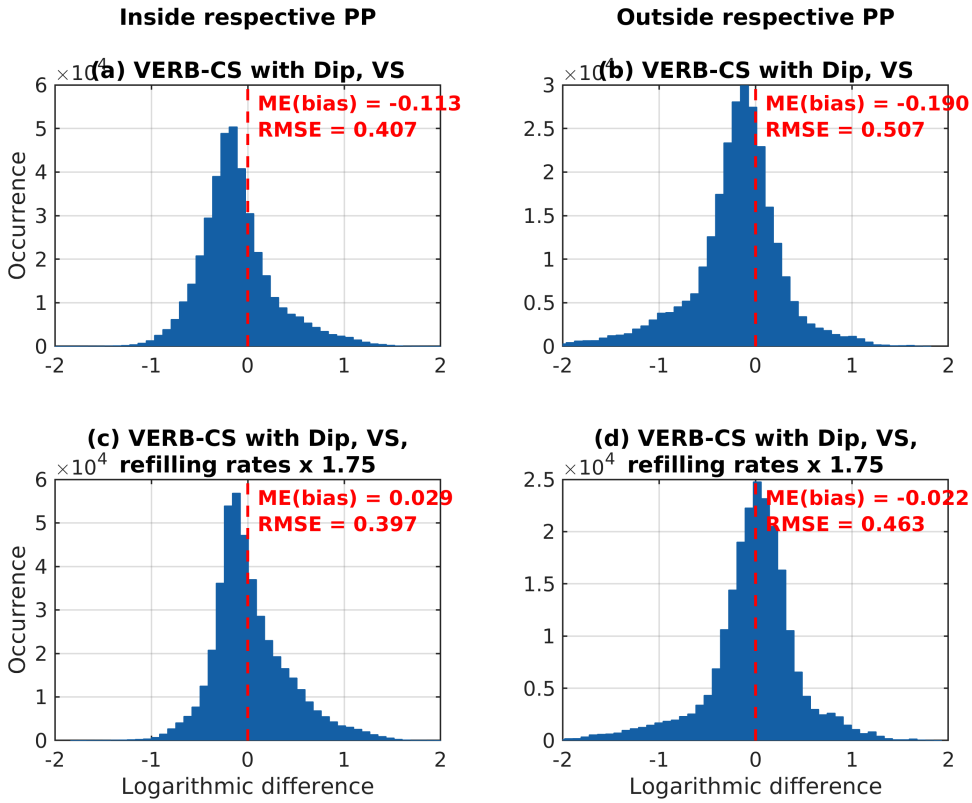

Figure S13: Distribution of errors of the VERB-CS model with the dipole magnetic field, the Volland-Stern electric field model and refilling rates from Denton et al. (2012) (top row) and with the same setup but refilling rates multiplied by 1.75 (bottom) model inside and outside the plasmapause.

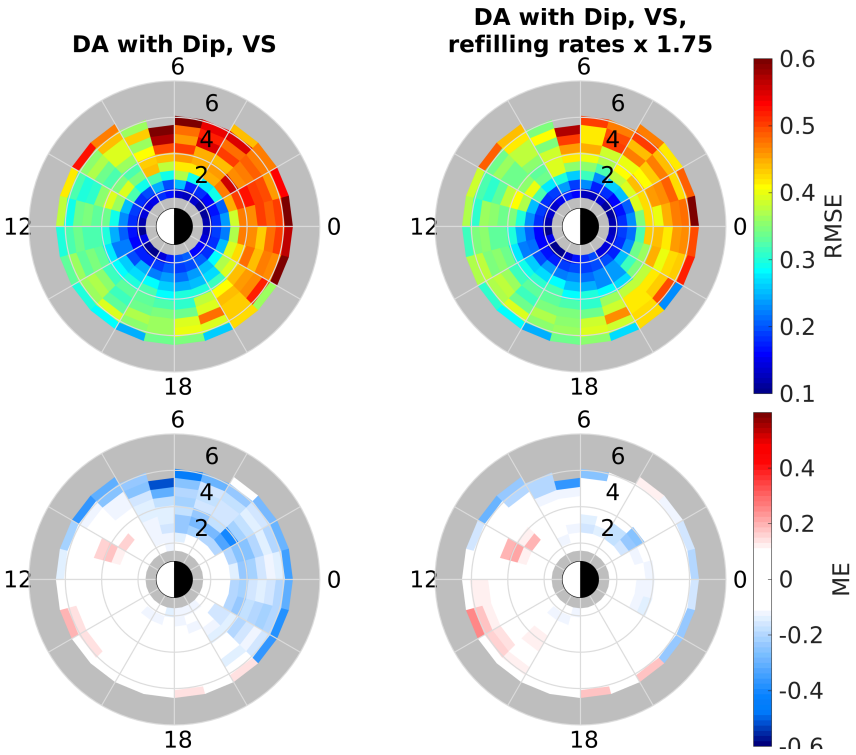

Figure S14: The root-mean-square error (top row) and the bias or mean error (bottom row) of the assimilative model (DA) with the dipole magnetic field, the Volland-Stern electric field model and refilling rates from Denton et al. (2012) (left), and with the same setup but refilling rates multiplied by 1.75 (right).

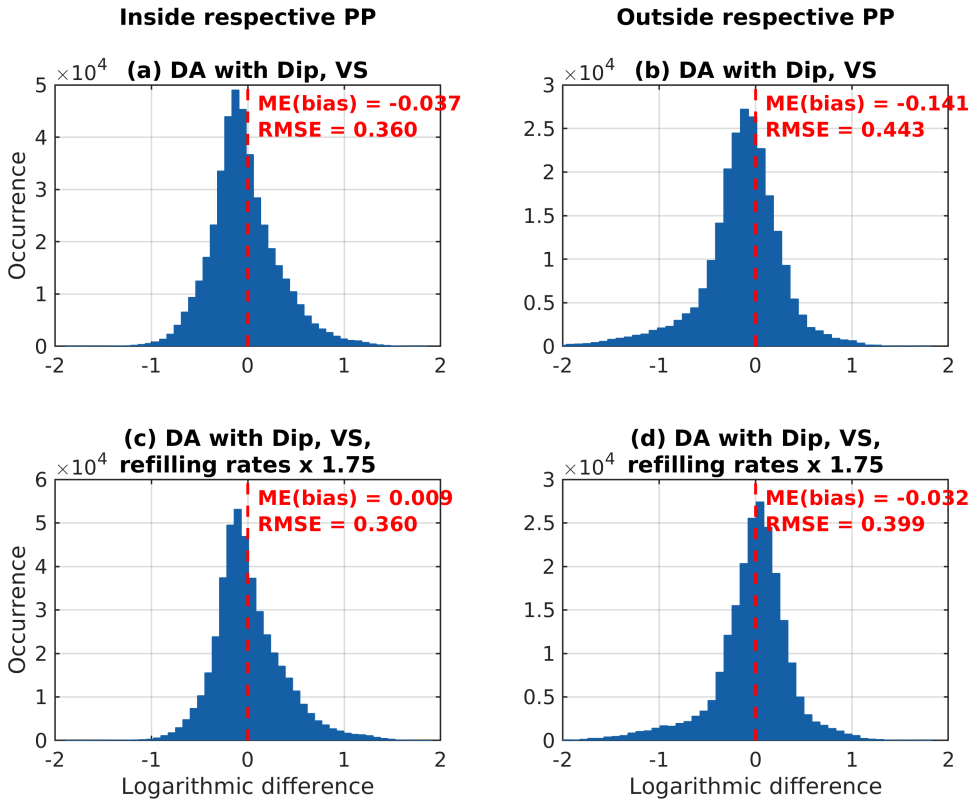

Figure S15: Distribution of errors of the assimilative model (DA) with the dipole magnetic field, the Volland-Stern electric field model and refilling rates from Denton et al. (2012) (top row) and with the same setup but refilling rates multiplied by 1.75 (bottom) model inside and outside the plasmapause.

Acknowledgments

The authors used geomagnetic indices and solar wind data provided by OMNIWeb (`http://omniweb.gsfc.nasa.gov/form/dx1.html`). IMAGE EUV plasmapause database was obtained from `http://enarc.space.swri.edu/EUV/`. This work was funded by the NASA Heliophysics Guest Investigator Program under NASA grant NNX07AG48G, P.I. Jerry Goldstein. The electron density data set was derived using the NURD algorithm and is available at the GFZ Data Services: `https://dataservices.gfz-potsdam.de/panmetaworks/showshort.php?id=escidoc:5098892` (Zhelavskaya et al., 2020b). I. Z. was supported by Geo.X, the Research Network for Geosciences in Berlin and Potsdam, under Grant No SO_087_GeoX. This work was supported by the pilot project "MAP" which is a part of the 'Helmholtz Pilot projects Information & Data Science II' funded by the Initiative and Networking Fund of the Helmholtz Association, and by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870452 (PAGER). The research has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through grant SFB 1294. We would like to thank Maria Spasojević for her contribution to the development of the PINE model and the NURD algorithm. I. Z. is grateful to Jerry Goldstein, T. Paul O'Brien, Alexander Drozdov, and Hayley Allison for helpful discussions. We also thank Jerry Goldstein for extensive advise on the physics-based model, in particular regarding the refilling rates.

Chapter 5

Systematic Analysis of Machine Learning and Feature Selection Techniques for Prediction of the Kp Index

Note

The following work is published in *Space Weather*:
Zhelavskaya I. S., R. Vasile, Y. Y. Shprits, C. Stolle, J. Matzka (2019). Systematic analysis of machine learning techniques for Kp prediction, Space Weather, 17. `https://doi.org/10.1029/2019SW002271`.

Abstract

The Kp index is a measure of the mid-latitude global geomagnetic activity and represents short-term magnetic variations driven by solar wind plasma and IMF. The Kp index is one of the most widely used indicators for space weather alerts and serves as input to various models, such as for the thermosphere and the radiation belts. It is therefore crucial to predict the Kp index accurately. Previous work in this area has mostly employed artificial neural networks to nowcast Kp, based their inferences on the recent history of Kp and on solar wind measurements at L1. In this study, we systematically test how different machine learning techniques perform on the task of

nowcasting and forecasting Kp for prediction horizons of up to 12 hours. Additionally, we investigate different methods of machine learning and information theory for selecting the optimal inputs to a predictive model. We illustrate how these methods can be applied to select the most important inputs to a predictive model of Kp and to significantly reduce input dimensionality. We compare our best performing models based on a reduced set of optimal inputs with the existing models of Kp, using different test intervals and show how this selection can affect model performance.

5.1 Introduction

The Kp index is one of the most widely used global measures of geomagnetic activity. It is used as an input to many scientific applications, including the parameterization of ionospheric ion outflow (Yau et al., 2011) and aurora particle precipitation (Emery et al., 2008) in the ionosphere, thermosphere (Bruinsma et al., 2018), hot plasma particle density (Korth et al., 1999; Denton et al., 2016), cold plasma density in the plasmasphere (Maynard and Chen, 1975; Pierrard et al., 2009; Goldstein et al., 2014; Zhelavskaya et al., 2017), plasmapause location (Carpenter and Anderson, 1992), and radiation belt models and wave parameterizations (Brautigam and Albert, 2000; Shprits et al., 2007b; Orlova et al., 2014; Ozeke et al., 2014; Agapitov et al., 2015) in magnetospheric physics, among others. It is therefore important to predict the Kp index accurately in order to produce most reliable forecasts in the aforementioned areas.

A number of models for Kp index prediction have been developed in the past decades. All these models use solar wind parameters measured at L1 as an input and the Kp index is their only output. Various methods were employed to develop these models. The first two models predicting Kp, Costello (1998) and Boberg et al. (2000), used feedforward neural networks (FNNs), a type of artificial neural networks (ANNs) often used for solving regression, classification and clusterization problems. Wing et al. (2005) employed FNNs and recurrent neural networks to develop a predictive model of Kp, and have shown that both types of networks have a similar performance. Recent studies by Bala and Reiff (2012) and Wintoft et al. (2017) have also employed feedforward neural networks for the Kp prediction. Tan et al. (2018) have used long-short-term memory (LSTM), an artificial recurrent neural network

architecture used in the field of deep learning that is powerful for processing sequential data such as sound, natural language, or other complex time series. Alternative methodologies, such as the nonlinear autoregressive moving average with exogenous inputs (NARMAX) algorithm and support vector machines, were employed in Balikhin et al. (2001); Boaghe et al. (2001); Ji et al. (2013); and Wang et al. (2015).

The models listed above were developed using different machine learning techniques. They were trained and tested using data from different time intervals (depending on the data availability but also on the choice of the training and testing time intervals made by the authors). The inputs to the models were also constructed differently in different studies. This makes it difficult to compare models and objectively evaluate progress. Thus, it is also unclear whether another new modeling technique or a different way of constructing the inputs can improve the quality of the predictions. It has also not been systematically investigated whether it is the use of solar wind measurements at L1 as input that sets a limit to the prediction accuracy, since the single point observations around L1 cannot fully capture the complex solar wind – magnetosphere coupling.

In this study, we investigate what brings the most improvement to the model accuracy and whether there is a limit to the prediction accuracy set by using solar wind measurements at L1 as input to a model. We perform such an analysis by applying different machine learning modeling techniques to develop predictive models of Kp for different prediction horizons up to 12 hours and comparing their performance. We focus our analysis on three algorithms: Linear Regression (LR), artificial Feedforward Neural Networks (FNN) (Bishop, 2006; Goodfellow et al., 2016), and Gradient Boosting (GB) (Friedman, 2001). We use linear regression as a benchmark for comparison with nonlinear (FNN) and ensemble-based (GB) models. We use the same validation technique and the same time intervals to train and validate the models, and in doing so create an unbiased technique to validate and compare models. This analysis step helps determine to what extent the chosen modeling approach affects the accuracy of predictions for different prediction horizons.

We also compare how different approaches for constructing input variables to the models affect the accuracy of predictions. Additionally, we test different machine learning and information theoretical methods for optimal input selection. The motivation to explore methods for optimal input selection (also called *feature selection* in machine learning) is to identify the most important solar wind drivers to predict

the Kp index. As the number of inputs to a model (or features) grows very quickly when their cadence is increased or when more time history is included, it becomes difficult to interpret the physical importance of each input variable and can also increase the training time. Feature selection methods allow us to find a subset of the most important inputs that contain a sufficient amount of information to model the target variable (here, Kp) and, at the same time, to achieve good accuracy of the predictions. We investigate feature selection procedures based on the following methods: Fast Function Extraction (FFX) (McConaghy, 2011), Random Forest (RF) (Ho, 1995), Mutual Information Maximization (MIM) (Bollacker and Ghosh, 1996), and Maximum Relevancy Minimum Redundancy (MRMR) (Peng and Ding, 2005; Ding and Peng, 2005). The Random Forest algorithm is often used for feature selection and is implemented in many machine learning packages. The feature selection method based on the FFX algorithm makes use of the intrinsic feature selection of FFX and the K-fold cross-validation, and is developed in this work. The last two methods are based on the concept of mutual information.

Finally, we compare the best performing models to the previous predictive models of Kp using different time intervals for testing to illustrate the importance of choice of testing interval and how it can affect the model performance.

The structure of the paper is as follows: In section 5.2, we provide a brief description of the machine learning algorithms and feature selection methods used in this work. In section 5.3, we describe the data, the training and validation methodology, and the hyperparameter selection for all considered methods. In section 5.4, we present the comparison of the machine learning methods and results of the selection of input variables. We also present the comparison of our best performing models to the existing ones. Finally, we summarize the main results of the paper in section 5.5.

5.2 Machine learning background

This section provides a brief description of the algorithms used in this work. The description is intended to provide the reader with an overview of these methods, while more details of each method can be found in the references therein. Section 5.2.1 provides an overview of the algorithms used in this study to develop the predictive model of the Kp index, namely Gradient Boosting (GB), Feedforward Neural Networks (FNNs), and Linear Regression (LR). Section 5.2.2 describes the procedures

for optimal feature selection that we employ in this work based on Fast Function Extraction (FFX), Random Forest (RF), and information theoretical methods, namely Mutual Information Maximization (MIM) and Maximum Relevancy Minimum Redundancy (MRMR).

5.2.1 Machine learning algorithms for model development

Feedforward Neural Networks (FNN). FNNs are a type of artificial neural network inspired by the way biological neural networks in our brain process information (Bishop, 2006; Goodfellow et al., 2016). FNNs are used for solving regression, classification, and clusterization problems. In regression problems, they are used to find multivariate nonlinear relationships between the input and the output variables. An FNN consists of an input layer, an output layer, and a number of hidden layers. Its schematic representation is shown in Figure 5.1a for the case of a network with one hidden layer. Each node in the layer is a neuron, which can be thought of as the basic processing unit of a neural network. In the FNN, each neuron is connected to all neurons in the preceding and succeeding layer; neurons of the same layer are not connected. Each connection between two neurons in an FNN has a weight associated to it. The information in the FNN moves only forward, from the input to the output with no feedback connections or loops. An FNN is applied to solve a specific problem after it is trained on a set of data pertaining to this problem. Training is an optimization procedure, in which the weights (the internal parameters of the network) are tuned using the training set of data so that the difference between the network output and the actual target variable is minimal. A description of FNNs applied to space physics problems can be found in Zhelavskaya et al. (2017); Chu et al. (2017a,b); and Zhelavskaya et al. (2018). In this work we use the MATLAB Deep Learning Toolbox to train neural networks (`https://mathworks.com/products/deep-learning.html`).

Gradient Boosting (GB). Gradient boosting is an ensemble machine learning algorithm for solving classification and regression problems. It combines the outputs of many simple prediction models to obtain a more accurate prediction (Friedman, 2001). Boosting iteratively produces a hierarchy of these models, as shown in Figure 5.1b. These models are referred to as weak classifiers/regressors. In GB, each weak model is typically a shallow decision tree (Breiman et al., 1984). In a regression problem, the first model is trained to fit the actual output, and then every new model

is trained to fit the residual between the actual target variable and the prediction value given by the previous model. The prediction is then given by a weighted linear combination of outputs of each weak model. The weights of that linear combination and internal parameters of each decision tree of the ensemble, such as the maximum tree depth, are determined during the training phase. In Gradient Boosting, training is performed using gradient descent minimization of the target cost function in a functional space. We use the python xgboost library to implement the gradient boosting algorithm (`https://xgboost.readthedocs.io/en/latest/`).

Linear regression (LR). Linear regression is a linear approach to modeling the relationship between a scalar-dependent variable and one or more explanatory variables. The fit to the data is obtained using a maximum likelihood estimator (Bishop, 2006). Linear regression algorithms may perform worse than other nonlinear methods (e.g., neural networks) in practice, since they are only able to model a linear relationship between input and output variables. On the other hand, especially in regression problems, they provide an analytic expression and allow for the interpretability of the result. We use the python sklearn library for linear regression (`https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html`).

5.2.2 Feature selection procedures

Feature selection based on Fast Function eXtraction (FFX). The Fast Function Extraction algorithm is a deterministic scalable algorithm for symbolic regression problems (McConaghy, 2011). Symbolic regression is a type of regression, in which the model is constructed by searching the space of mathematical expressions for the optimal combination of expressions, i.e., the combination of mathematical blocks that best fits to a given data set in terms of accuracy and simplicity. These mathematical blocks are usually represented by a set of basis functions (e.g., polynomial functions, exponential functions, etc.) that is used and combined iteratively to obtain more complex functions of the input variables. The FFX algorithm uses a deterministic procedure to build new regression functions in each iteration and allows for faster training times and prototyping, in comparison to other more general symbolic regression schemes, such as genetic programming (Koza, 1992). The choice of the basis

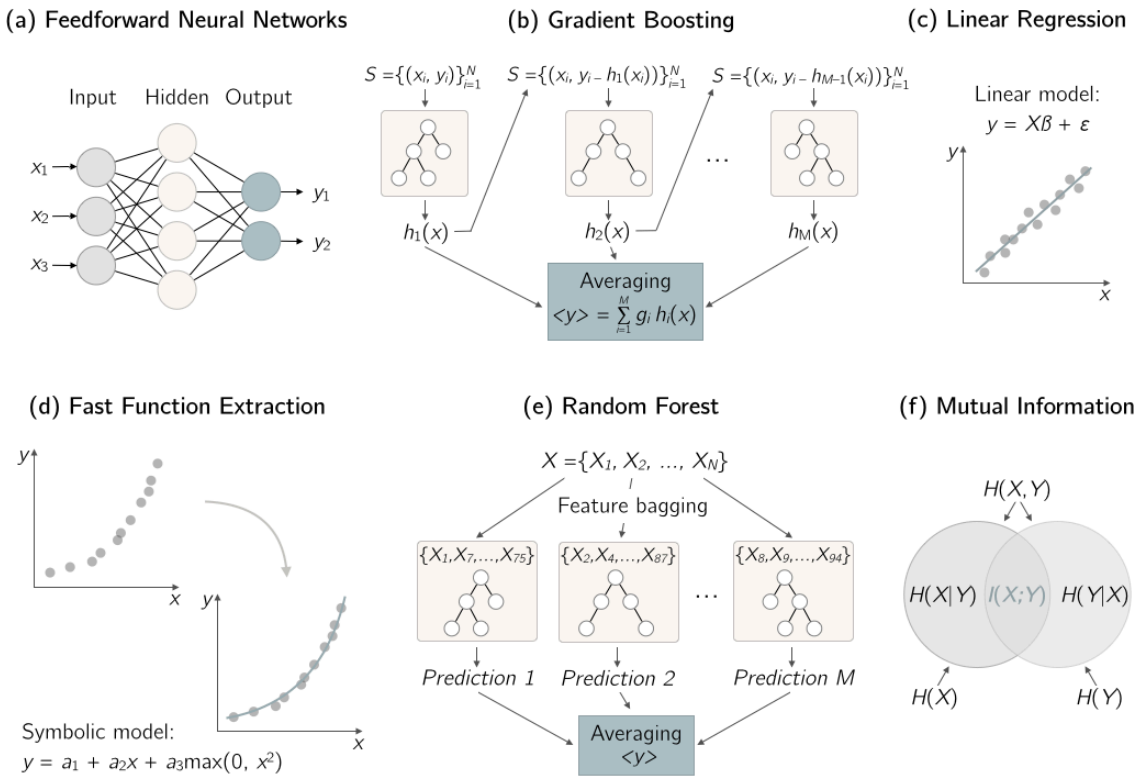

Figure 5.1: A schematic representation of a) Feedforward Neural Network, b) Gradient Boosting, c) Linear Regression, d) Fast Function Extraction, e) Random Forest, f) Mutual Information methods for model construction and feature selection used in this work.

functions determines the class of solutions for a specific problem. Typically, a polynomial basis augmented by nonlinear basis elements (such as max or min functions) is used to model quasi linear systems and still obtain much better performance than simpler linear regression schemes. In that sense, symbolic regression constitutes a natural nonlinear extension to linear regression models, by allowing an interpretation of the result and, at the same time, improving the accuracy by adding nonlinear elements to the solution.

The FFX algorithm performs an internal feature selection by building symbolic expressions using only a subset of input variables. As the input data set changes, different variables may appear in the expression of the final trained model. In the context of a K-fold cross-validation procedure (described in section 5.3.3), a number of different models is obtained by training on different training/validation partitions. Given this set of models, we develop the following feature selection procedure. We fix a threshold integer value $k \in [1...N]$, where N is the total number of trained model instances ($N = 5 \times 10$ for 5-fold CV with 10 repetitions). Only those input variables that appear in at least k of the trained model instances are extracted. As k approaches N, the set of extracted variables is reduced: For $k = N$, only those variables that appear in every trained model, are selected. Thus, the FFX algorithm provides a definitive number of selected features for each threshold k. Contrary to the RF feature selection procedure described below, no ranking among the selected variables is obtained (they are considered to be equally important). The library used for this algorithm can be found at `https://github.com/natekupp/ffx`.

Feature selection based on Random Forest (RF). The forest of random trees, or more commonly, the Random Forest algorithm, is an ensemble machine learning algorithm for classification and regression problems that can also be used for the input selection (Ho, 1995; Breiman, 2001). RF is based on decision trees (Breiman et al., 1984), similarly to GB, but in RF, each decision tree is fitted directly to the target variable, thus all trees can be trained in parallel. Two points should be noted regarding the way the trees in the RF ensemble are constructed. First, each decision tree in the ensemble is built using the bootstrapped version of the initial training data set. Bootstrapping is an algorithm that produces replicas of a data set by performing random sampling with replacement. Therefore, each decision tree is built using a slightly modified version of the initial data set. Second, each decision tree is trained using a subset of inputs drawn randomly from the whole set of inputs, which makes

the resulting trees uncorrelated to each other (Breiman, 2001). The output of the final model is the average of predictions made by all decision trees in the ensemble. Its schematic representation is shown in Figure 5.1e.

The Random Forest algorithm can be used to rank the importance of variables in a regression or classification problem (Breiman, 2001), and perform optimal feature selection and therefore reduce the input dimensionality. The method is referred to as feature importance extraction and is based on the Gini importance or mean decrease impurity concept (Ho, 1995). The method returns an ordered list of input features according to the value of the mean decrease impurity, with the sum of the mean impurities over all variables being unity. To select a reduced number of input features a threshold value T, $0 < T < 1$, is fixed, and only those variables, the ordered cumulative sum of which is as close as possible to the threshold T, are selected. The selection of this threshold is usually performed empirically. We used the sklearn library to implement the RF regression algorithm (`https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html`).

Feature selection based on Mutual Information (MI). Mutual information is a concept of information theory that can be used to study the relationships between different variables (usually input and target variables of a model). It is not a machine learning algorithm by itself, but can be used for feature selection and, therefore, we present a brief description of it. MI between two variables X and Y is a measure that quantifies the amount of information obtained about one variable through the other variable. It is defined as:

$$I(X,Y) = \int_X \int_Y p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) dxdy, \tag{5.1}$$

where $p(x,y)$ is the joint probability density function of X and Y, and $p(x)$ and $p(y)$ are the marginal density functions. Mutual information determines how similar the joint distribution $p(x,y)$ is to the product of the marginal distributions $p(x)$ and $p(y)$. If X and Y are independent, then $p(x,y)$ is equal to $p(x)p(y)$, and the integral in (1) is equal to zero. In practice, the probability distribution functions can be obtained by discretizing variables X and Y and using an alternative definition of MI utilizing the concept of Shannon entropy:

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \tag{5.2}$$

where $H(X)$ and $H(Y)$ are entropies of X and Y respectively, and $H(X,Y)$ is their joint entropy. Entropy is a measure of uncertainty of a variable, and entropies of discrete variables are defined as:

$$\begin{gathered} H(X) = -\sum p(x)\log p(x),\ H(Y) = -\sum p(y)\log p(y), \\ H(X,Y) = -\sum\sum p(x,y)\log p(x,y) \end{gathered} \tag{5.3}$$

A more detailed description of MI and other concepts of information theory and their application in space physics can be found in Wing and Johnson (2019). Below, two methods based on the MI, that are used in this work, are described.

Mutual Information Maximization (MIM). The MIM feature selection algorithm employs the concept of mutual information and selects input features that maximize the mutual information between them and the target variable. Formally, if $S_{t-1} = \left\{X_{f_1}, \ldots, X_{f_{t-1}}\right\}$ is the set of selected features at time step $t-1$, where f_i is the input feature selected at time step i, MIM selects the next input feature f_t by solving the following optimization problem:

$$f_t = \arg\max_{i \notin S_{t-1}} I(X_i, Y), \tag{5.4}$$

where X_i is the input variable that is not yet included in the set of selected features S_{t-1}, and Y is the target variable. Simply stated, MIM selects input variables that have the largest mutual information with the target variable and ranks them according to their MI with the target variable. MIM makes the following assumptions:

- Assumption 1: The selected features X_S are independent and are also class-conditionally independent, given the unselected feature under consideration X_k (i.e., the knowledge of the unselected feature does not give additional knowledge of the selected features). Here, X_S is the reduced data set containing selected features.

- Assumption 2: All features are pairwise class-conditionally independent, i.e., $p(X_i, X_j|Y) = p(X_i|Y)p(X_j|Y)$, meaning that $\sum I(X_j, X_k|Y)$ is zero.

- Assumption 3: All features are pairwise independent, i.e., $p(X_i, X_j) = p(X_i)p(X_j)$, meaning that $\sum I(X_j, X_k)$ is zero.

These assumptions may not always hold in practice and therefore, MIM is not

widely used as it can have a poor performance. Nevertheless, we include it for comparison and illustrative purposes.

Maximum Relevancy Minimum Redundancy (MRMR). The MRMR algorithm also employs the concept of MI, similarly to MIM, but adds another term to the optimization problem in selecting the next input feature:

$$f_t = \arg\max_{i \notin S_{t-1}} I(X_i, Y) - \alpha \sum_{k=1}^{t-1} I(X_{f_k}, X_i), \quad \alpha = \frac{1}{(t-1)}, \tag{5.5}$$

where $I(X_{f_k}, X_i)$ denotes MI between already selected input variables X_{f_k} in S_{t-1} and candidates for the new input variable. Therefore, this method accounts not only for the "relevancy" of features, as in MIM, but also for the "redundancy" of information brought by new input features, that is expressed by the last term in (5). Even if MI of a new input variable with the target variable is large, it may be strongly correlated with already selected input variables in X_S and may therefore not bring in any new information. Therefore, other variables that minimize the redundancy factor with already selected inputs will be selected by MRMR. MRMR makes the assumptions 1 and 2 described above.

Both the MIM and MRMR feature selection algorithms provide a ranking of the the variables in the order of importance. The number of variables to be selected is usually chosen empirically.

5.3 Data and methodology

5.3.1 Data

We use solar wind and interplanetary magnetic field (IMF) data from NASA's OMNIWeb data service during the time period of 1998–2017, to construct the input to our machine learning models. Specifically, we use 1-minute resolution measurements of solar wind speed, proton density, total interplanetary magnetic field B, and interplanetary magnetic field components B_x, B_y, B_z in the GSM coordinate frame. Using solar wind data with 1-minute resolution allows better capturing spikes and minima in the solar wind parameter, compared to data with 5-minute or 1-hour resolutions. The Kp index, i.e., the target variable or the model output, is obtained from the GFZ Potsdam website (`https://www.gfz-potsdam.de/en/kp-index/`), and has a

Table 5.1: Considered inputs for all prediction horizons.

Solar wind and IMF parameter	$B, B_x, B_y, B_z, V_{sw}, nProt$
Aggregate functions	avg, min, max
Time windows, hours (here, current time is hour 0)	0-1, 1-2, ..., 8-9
Other inputs	$\sin(2\pi T/24), \cos(2\pi T/24)$, $\sin(2\pi D/365), \cos(2\pi D/365)$

three-hour cadence.

5.3.2 Model inputs

A variety of solar wind parameters and their derivatives, such as their time history or solar wind coupling functions (Newell et al., 2007), can be used as inputs to predictive models of Kp. Here, we consider a limited number of solar wind parameters and their time history as inputs. Specifically, we consider the solar wind speed (V_{sw}), proton density ($nProt$), IMF components B_x, B_y, B_z, and the total IMF magnitude B. We then construct minimum, maximum and average values of these variables over 1-hour time windows (starting from 0-1 to 8-9 hours previous to the current time) and use them as inputs to our models. In some of the previous studies, 3-hour time windows were used to construct the inputs (Wintoft et al., 2017; Bala and Reiff, 2012, e.g.). We test how the window size affects the model performance by comparing the performance of models based on 1-hour and 3-hour inputs in section 5.4.2 (Figure 5.4).

To take into account possible seasonal variation, we also include indicators of day of the year and time of the day represented by $\sin(2\pi T/24)$, $\cos(2\pi T/24)$, $\sin(2\pi D/365)$, $\cos(2\pi D/365)$, where T is the UT hour of the day and D is day of the year, following Wintoft et al. (2017).

In total, the considered variables comprise 166 input features. The output of the model is the Kp index with 3-hour cadence. One data sample corresponds to one Kp value in our data set (we do not interpolate Kp to 1-minute cadence of the solar wind data) and therefore, the number of data samples in the full data set for the 1998–2017 period is 58439, and corresponds to the number of Kp values over this period. Please refer to Table 5.1 for a summary of the inputs.

5.3.3 Training and validation setup

We use a repeated K-fold cross-validation procedure with $K = 5$ folds and 10 repetitions in order to evaluate the performance of different models (i.e., obtain estimates of the mean and standard deviation of the training and validation errors) and compare them with each other.

In K-fold cross validation, the available data set is split into K folds, where one fold is used for validation and the remaining $K - 1$ folds are used for training (validation data are not shown to the model during training). This procedure is implemented K times with different validation fold each time. In a repeated K-fold cross-validation (CV), this procedure is repeated multiple times (here we choose 10), where in each repetition the folds are split differently. After each repetition, model assessment metrics are computed (e.g., RMSE, linear correlation coefficient, etc.), and then the scores from all repetitions are averaged to obtain the final model assessment score. In a 5-fold CV with 10 repetitions, we train 50 model instances and compute/obtain 50 values of training and validation errors, and use them to calculate the mean of model assessment scores. Such a repeated CV procedure produces a more robust assessment score than if CV is performed only once, and especially if only one hold-out test set is used instead of the CV procedure. A repeated CV also gives an idea about the variance of a model by examining the standard deviation of the model error. A high standard deviation indicates that the model produces different results when trained on different data splits and, therefore, has a high variance. Models with low standard deviation of error are desirable, since they perform similarly on different data splits and are therefore robust. Additionally, we withhold a separate test set comprising 10% of all data before the start of the CV procedure for the final model evaluation.

Since the data under consideration is a time series, the neighbouring data points may be strongly correlated. Consequently, random splitting into training and validation sets may lead to correlations between these two sets. At the same time, if the data are split into CV folds sequentially, the distribution of the target variable in different validation folds can be significantly different, e.g., it may occur that high Kp values are present only in the validation or only in the training data set. To avoid both of these unwanted scenarios, we implement an intermediate procedure. We first split all data into 35-day blocks sequential in time, and then assign these 35-day blocks randomly to the CV folds. The reason for using a 35-day block length is to avoid the possible effect of 27-day recurrence caused by the rotation of the sun.

Each block contains 280 measurements, and in total, we obtain 209 blocks using this procedure. Finally, the sizes of the training and validation sets for different CV splits comprise 41918 and 10640 data samples respectively, and the test set 5880.

5.3.4 Hyperparameter optimization

Hyperparameter optimization is an essential step of the optimal model selection. It is performed to find the model complexity that is appropriate for a specific regression or classification problem, so that the model does not underfit (its complexity is too low) or overfit (the complexity is too high) the training data. Each of the machine learning algorithms described in section 5.2 has a specific set of hyperparameters that can be tuned.

Feedforward Neural Networks. We employ a single hidden layer neural network. One hidden layer is typically sufficient to approximate a continuous function (Cybenko, 1989). The number of neurons in the input layer is equal to the number of inputs, and only one neuron is present in the output layer, which outputs the predicted value of the Kp index. Therefore, the main hyperparameter to tune is the number of neurons in the hidden layer. The choice of the number of neurons is made using a grid search, and an optimal number of 19 neurons was determined when using all the 166 input variables for all horizons. Due to the shallowness of the network, we employ a second order optimization method based on the Levenberg-Marquardt algorithm to train the networks (provided by the Matlab Deep Learning toolbox).

Gradient Boosting. As given by the xgboost python library (`https://xgboost.readthedocs.io/en/latest/`), the gradient boosting regression algorithm has 15 hyperparameters. Since the sensitivity of the results on many of them is negligible, we focus on tuning the three most important parameters: the number of estimators (i.e., the number of trees in the ensemble), the learning rate (regulates the step of the gradient descent), and the maximum depth of each tree (controls the complexity of the model and therefore affects overfitting). Using the grid search, we find that 100 estimators and a learning rate of 0.08 lead to the best performance on the validation set for all prediction horizons. Regarding the max depth of trees, we find that a max depth of 5 provides a good performance on the validation set and, at the same time, limits overfitting.

Linear Regression. The linear regression algorithm does not require hyperparameter optimization. The model complexity is low enough to avoid overfitting and, therefore, no regularization is necessary. The bias of the model is relatively high, however, due to the insufficiency of the model to perform well on the nonlinear task, and cannot be mitigated unless opting for more complex, nonlinear models, or using more sophisticated input features constructed from the solar wind parameters (i.e., solar wind coupling functions).

FFX. The FFX algorithm has a number of hyperparameters to optimize in the FFX python package (5 parameters corresponding to the choice of the basis functions that can be selected by a user, and 8 other default parameters that were optimized by McConaghy (2011)). The most important is the choice of the basis functions. We obtained the best model performance by including single variables, the interaction terms that allow building low-order polynomials in the input variables (e.g., $x_1 \times x_2$ or x_1^2), and hinge functions (i.e., max and min functions) to introduce non-linear thresholds.

Random Forest. The hyperparameter search for the random forest algorithm is similar to that of gradient boosting. RF has 12 hyperparameters to optimize. We focus on the two main parameters that control the complexity of the model: the number of estimators and the maximum depth of each element of the ensemble. Using the grid search, we find that the optimal values are 30 estimators in the ensemble and a maximum depth of 7.

MIM and MRMR. These methods do not have hyperparameters as the methods described above, but there is one factor that may influence how MI is calculated, and that is how the variables are discretized. There are several ways in which this can be done, and this is currently an active area of research (Ali et al., 2015; Jiang and Wang, 2016; Gao et al., 2017, e.g.). One way is to bin variables uniformly using bins of a predefined size for each variable (Wing and Johnson, 2019; Wing et al., 2016, e.g.). Sturges (1926) proposed that the optimal bin size for a normal distribution is $n_b = \log_2(n) + 1$ and bin width $w = range/n_b$, where n is the total number of measurements in the data set and range is the maximum - minimum value of a variable. We have explored a number of different bin sizes and found that $n_b = 20$ is optimal for our task.

5.4 Results

5.4.1 Comparison of the ML methods for model development

In this section, we compare the performance of the gradient boosting (GB), neural network (NN), and linear regression (LR) based models on the task of predicting the Kp index for prediction horizons from 0 to 12 hours ahead. All models have the same inputs described in section 5.3.2, and are trained and validated using the repeated K-fold cross validation procedure described in section 5.3.3.

Figures 5.2a and 5.2b show the cross-validation root mean square error (RMSE) and correlation coefficient (CC) of the different models as a function of prediction horizon intervals, respectively. For reference, errors of the persistence model and of the averaged Kp model are shown as well. The persistence model is a model that always predicts the most recent known value of Kp, and we defined the averaged Kp model as a model that predicts the mean value of the Kp index over the previous 15 days. The values of RMSE and CC for all the methods are also listed in Table 5.5 in Appendix 5.A for reference. As can be seen from Figure 5.2, the averaged Kp model, $< \text{Kp} >$, has the highest RMSE and the lowest CC (shown in light gray), as expected. The persistence model (dark gray) has the second highest error and lowest CC, and its performance decreases for the longer prediction horizons. For the LR, GB, and NN methods, the bars in the plot are arranged in pairs, where the color corresponds to a particular method. Solid (darker) colors show the mean validation error and faded (lighter) colors show the mean training error, both obtained in the cross validation procedure. It can be seen that the error of the LR model (shown in yellow) is much lower than that of the persistence and $< \text{Kp} >$ models, but is higher compared to the NN- and GB-based models, especially for prediction horizons 0 and 3 hours ahead. For longer prediction horizons, the errors of the LR-, NN- and GB-based models are comparable. This confirms the existence of a nonlinear component in the Kp prediction problem for the short-term prediction horizons that can be modeled only with nonlinear methods. The GB-based (blue) and NN-based (red) models yield similar validation errors for prediction horizons 0 and 3 hours ahead. The validation error of the GB model is slightly smaller for the longer prediction horizons. The standard deviation of error (shown with error bars) is less than 0.05 for both the GB- and the NN-based models, which is significantly less than the discretization of Kp levels, indicating that the models are quite robust. It should be also noted that the

difference between the training (faded blue) and the validation (solid blue) errors of the GB model is larger than the one of the NN model, which may indicate that the GB-based model is overfit. It is therefore difficult to make a definitive conclusion about which model is suited better for the problem of Kp prediction since the difference between their performance is small. Due to the fact that the overfitting behavior is not desirable for a model, we choose the NN-based model as a benchmark model for further use and comparison to the previous studies. More information on overfitting and why it is not desirable can be found in Zhelavskaya et al. (2017).

5.4.2 Comparison of feature selection methods

In this section, we compare the performance of the input selection algorithms described in section 5.2.2 for the same prediction horizons (0 to 12 hours ahead). We also test how the size of a time window used for constructing the input variables to the models affect the accuracy of the predictions.

A feature selection procedure provides a list of the most important input variables, and a model based on these input variables needs to be constructed and trained in order to assess the quality of the selected inputs (and hence, the feature selection method). In the previous section, we have selected the NN-based model as a benchmark model for further use and comparison to the previous studies. Hence, we also use neural networks as a benchmark algorithm to test and compare different feature selection procedures. For each prediction horizon, we train and compare neural networks with five different configurations of input variables: (1) all input variables, as in section 5.4.1 (166 variables listed in Table 5.1), (2) inputs selected by the FFX algorithm, (3) inputs selected by the RF algorithm, and inputs selected by (4) the MIM and (5) the MRMR feature selection algorithms. Since only FFX provides the definitive number of selected inputs, as discussed in section 5.2.2, the number of selected inputs for the RF, MIM, and MRMR algorithms is chosen to be equal to the number of inputs selected by the FFX algorithm. This allows an objective comparison between the feature selection methods, as the number of inputs is the same for all methods. The number of selected inputs for different prediction horizons is listed in Table 5.2. The full list of optimal input variables selected by different algorithms is provided in Appendix 5.B.

Figures 5.3a and 5.3b show the RMSE and correlation coefficients of the neural

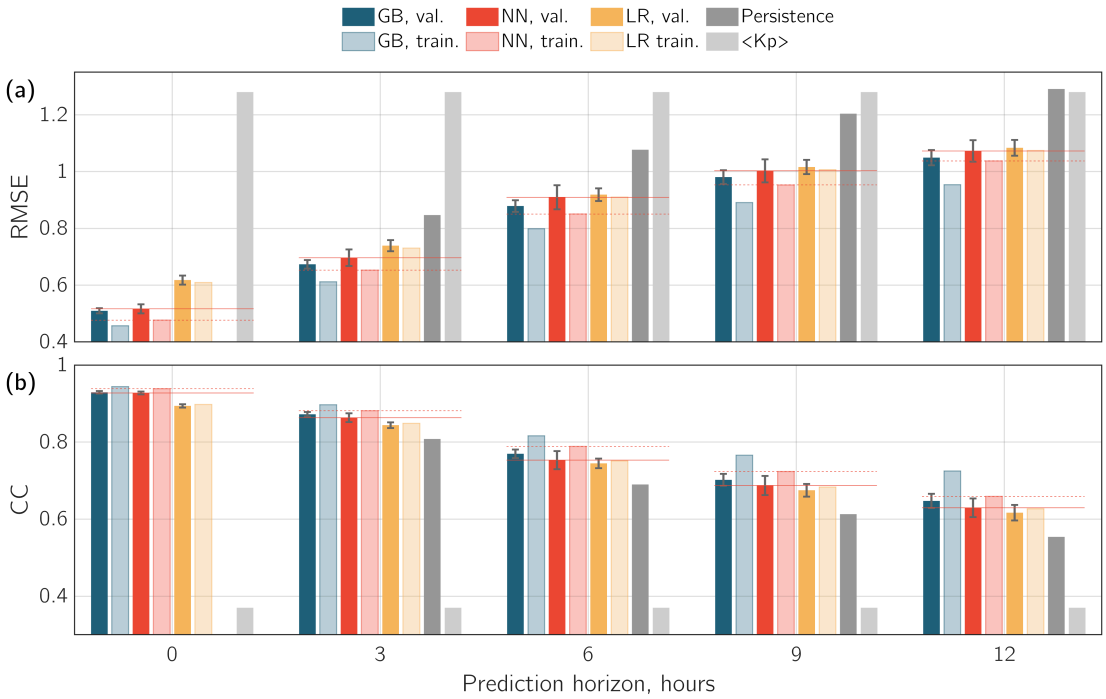

Figure 5.2: a) Root mean square error (RMSE) of the ML methods used for model development as a function of prediction horizon in hours: for gradient boosting (blue), neural networks (red), linear regression (orange), persistence model (dark gray), and averaged Kp over the 15 previous days (light gray). The bars are arranged in pairs for the first three methods, and the solid (darker) colors show the error on the validation set, the faded (lighter) colors show the error on the training set. The error bars indicate the standard deviation of error obtained from 5-fold cross validation with 10 repetitions. The horizontal red solid and dashed lines are help lines for a more convenient comparison between different methods (the solid lines correspond to the validation error of the NN-based models, the dashed lines to the training error). b) Same as in the top panel, but using the Pearson correlation coefficient (CC) as a performance assessment metric.

Table 5.2: Number of features selected using the feature selection algorithm based on FFX with threshold $k = 50$.

Prediction horizon, hours	**0**	**3**	**6**	**9**	**12**
Number of selected inputs (out of 166 available inputs)	52	14	13	12	10

networks with the inputs selected by different feature selection algorithms vs. the prediction horizon. The neural network containing all 166 input variables is denoted as NN (shown with the red bars) and is the same as in Figure 5.2; all other models are denoted as NN-X, where X corresponds to the algorithm used to select the optimal inputs (FFX, RF, MIM, or MRMR). Only the error on the validation set is shown, since the training error is very close to it for all the methods. The values of RMSE and CC for all the methods are also listed in Table 5.6 in Appendix 5.A for reference, for both validation and training sets. Neural networks having the inputs selected by the MIM algorithm show the poorest performance, while all other neural networks have comparable validation errors. This confirms that the Mutual Information of input variables with the target variable used in isolation from other input variables is not a good indicator of the importance of features when the input variables are correlated with each other. It should therefore not be used in such cases. Models with the inputs selected by other methods (FFX, RF, MRMR) have small differences in the resulting errors and perform similarly to the model containing all 166 inputs. Compared to all other models, NN-FFX shows the best performance for all horizons as well as a reduced variance, i.e., lower standard deviation of the validation error. Despite having fewer inputs, it also shows a slightly better performance compared to the neural network containing all 166 inputs. This result indicates that using these methods, particularly FFX, we can significantly reduce the number of input parameters and select the optimal ones containing a sufficient amount of information to model the target variable, i.e., the Kp index. We can use the obtained set of optimal input variables to gain a better understanding of what solar wind drivers are most significant to predict the Kp index for different horizons.

To identify how the construction of inputs affects the performance of the models, we perform the same analysis for the inputs constructed using 3-hour time windows (for computing averages, min, and max of solar wind parameters). The performance of different feature selection algorithms is similar to the results shown in Figure 5.3, and for brevity, we only present the results of the best performing models. Figure

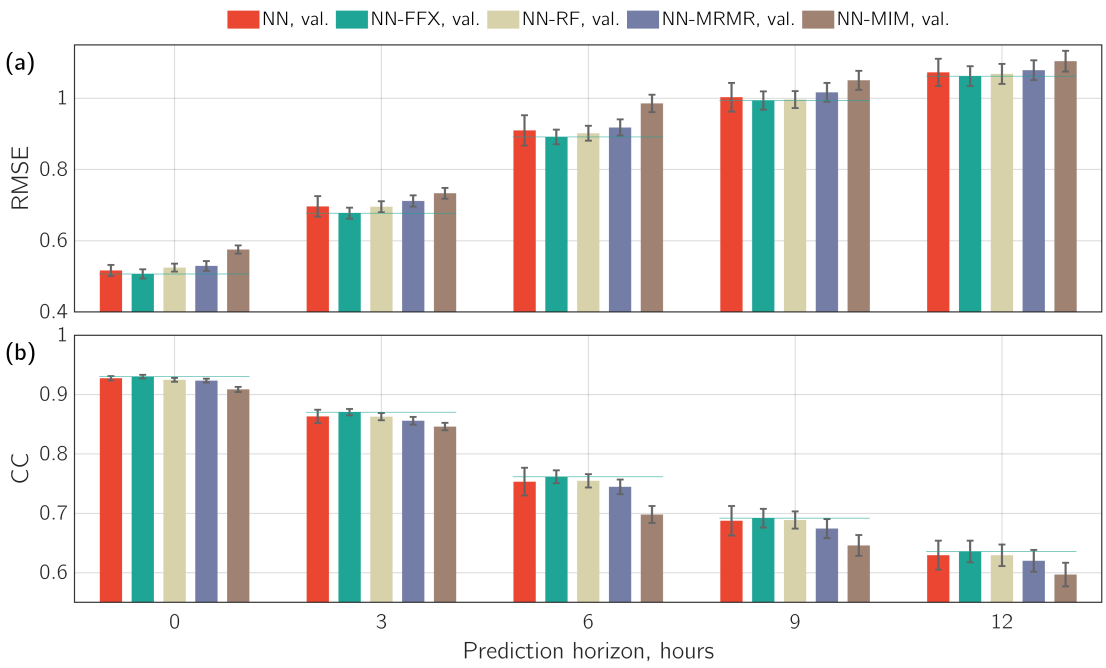

Figure 5.3: a) Root mean square error (RMSE) and b) correlation coefficient (CC) of the trained models used for input selection as a function of prediction horizon for neural networks trained on the whole set of inputs replotted from Figure 5.2 (red), neural networks trained on inputs selected using FFX with threshold $k = 50$ (green), neural network trained on inputs selected using random forest (light yellow), and neural networks trained on inputs selected using MRMR (violet) and MIM (brown); The bars show the average validation a) RMSE and b) correlation coefficient. The error on the training set is not shown in this figure. The error bars indicate the standard deviation of the error obtained from 5-fold cross validation with 10 repetitions. The horizontal green lines are help lines for a more convenient comparison between different methods (they correspond to the validation error of the NN-FFX models).

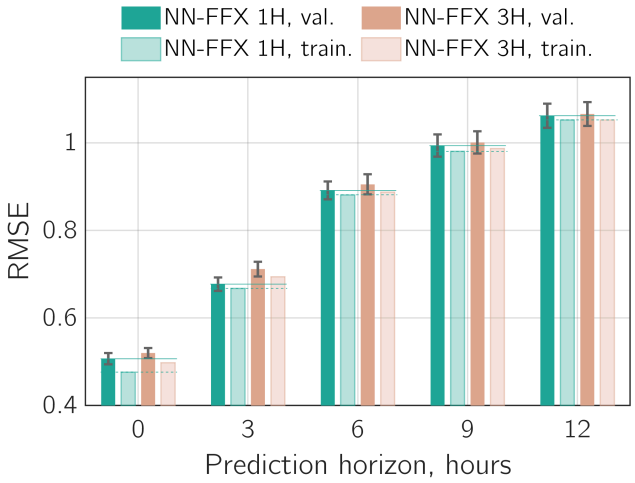

Figure 5.4: Comparison between 1-hour and 3-hour cadence inputs. Format is the same as in Figure 5.2. The horizontal green solid and dashed lines are help lines for a more convenient comparison between the models (the solid lines correspond to the validation error of the NN-FFX 1H models, the dashed lines to the training error).

5.4 shows the RMSE of the NN-FFX model with inputs constructed using 1-hour time windows (shown with green) and with inputs constructed using 3-hour time windows (pink) vs. prediction horizon. The values of RMSE and CC for all the methods are also listed in Table 5.7 in Appendix 5.A for reference. The model with the inputs constructed using 1-hour time windows shows a slightly lower validation error than the model based on 3-hour inputs, however, the improvement is marginal, and the differences in the model errors are very small. This potentially indicates that a further increase of the cadence of inputs may not lead to significant improvements, when using solar wind measurements at L1 as input to the predictive model of Kp.

5.4.3 Resulting models

Based on the results obtained in the previous sections, we select the optimal models for each prediction horizon, i.e., the ones that do not overfit, show the lowest validation error and lowest standard deviation of error, for further use and comparison with existing models. These are the neural network-based models with the input variables selected by the FFX feature selection algorithm (NN-FFX).

In our further analysis, we apply these models to all data combined (training, validation and test sets, described in section 5.3.3) for all prediction horizons, to produce the normalized occurrence maps (presented in this section, Figure 5.6) and

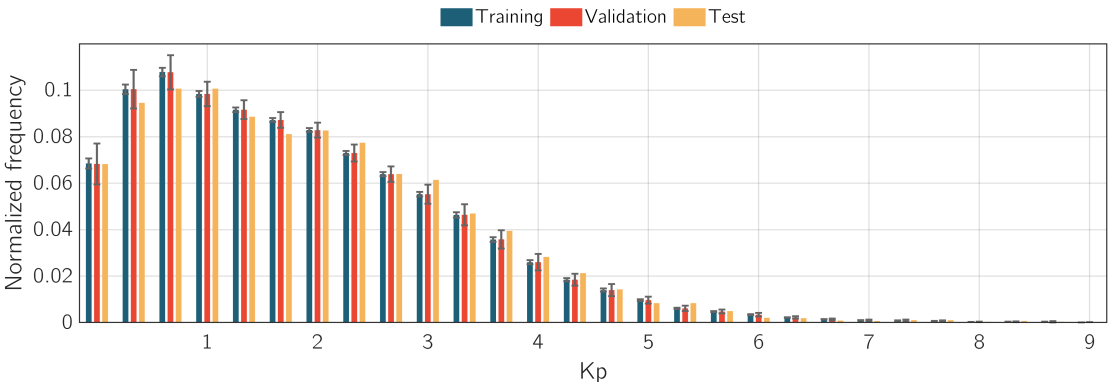

Figure 5.5: Histogram of normalized distribution of training, validation and test sets. The gray error bars show the spread of the normalized number of measurements in the training and validation sets in different Kp bins as per different splits produced in the cross validation procedure.

to compute the accuracy metrics (section 5.4.4, Table 5.3). We use the whole data set to do that in order to maximize the coverage and the number of measurements in the comparison (the test set provides only 5880 measurements). Also, the results for the training, validation, and test sets separately are similar to the ones produced using the full data set (please see Appendix 5.C for the accuracy metrics computed on the training, validation, and test sets separately). The normalized distribution of the training, validation, and test sets are shown in Figure 5.5, in blue, red, and yellow, respectively. Due to the use of cross validation procedure, the splitting into training and validation sets was performed for 10×5 times (5-fold CV with 10 repetitions). Therefore, we compute the mean value and the standard deviation of the normalized frequency of measurements in the training and validation sets over different splits. The standard deviation of the normalized frequency is shown with the error bars. No error bars are associated with the test set, since it is the same for all splits. It can be seen that the distributions of measurements in the training, validation, and test sets are similar to each other (the coverage of measurements in different sets is similar for different Kp bins), which supports the use of the combined data set (i.e., training, validation, and test) for further analysis.

Figure 5.6 shows the normalized occurrence of the observed vs. predicted Kp. The occurrence is normalized by the number of measurements of the observed Kp, i.e., the color of each bin denotes the number of measurements in that bin divided by the total number of measurements in that bin of observed Kp. Bins containing 4 or fewer measurements are not taken into account. The gray dashed lines show the

ideal fit to observations and the blue dashed lines show the obtained linear fits. The correlation coefficient (CC) and the root mean square error (RMSE) of the models for each prediction horizon are indicated as well. It can be seen that the models for prediction horizons 0 (nowcast) and 3 hours ahead perform well: most of the measurements are clustered close to the diagonal and the spread of the observed vs. predicted Kp values is quite small. However, the performance decreases for longer prediction horizons as the models tend to underestimate the Kp index for higher Kp values, i.e., the bias of the models for high Kp values increases as the prediction horizon increases. This indicates that either the information from the solar wind measurements at L1 is not sufficient to predict the elevated geomagnetic activity for the longer prediction horizons, or that it cannot be predicted accurately due to the lack of observations of high Kp in our training data set. The second option is less likely, since influence of the lack of observations of high Kp values would manifest in all the models. Indeed, the models for predicting Kp for 0 and 3 hour ahead are capable of predicting those events (Figures 5.6a, b and 5.7a, b). This suggests that such a decrease in performance is due to the lack of information in the solar wind measurements at L1. This behavior of predictive models based on the solar wind measurements at L1 is also noted by Shprits et al. (2019), and is discussed and analyzed there in more detail.

Figure 5.7 shows examples of the Kp index prediction for different prediction horizons during the May-June 2005 period, that is, an event from the test set (not used in the training). Again, the models for prediction horizons for 0 and 3 hours ahead perform better than models for longer prediction horizons: the latter do not capture the arrival of the storm nor its magnitude. Also, the models underestimate high Kp values for longer prediction horizons (the bias of the models for high Kp values increases). Again, this potentially indicates that the information in the solar wind measurements at L1 is not sufficient for long-term predictions, especially for predicting the elevated geomagnetic activity.

5.4.4 Benchmarking

Following Liemohn et al. (2018), we calculate the standard assessment metrics for comparison with previous and future studies. The standard assessment metrics proposed in Liemohn et al. (2018) are listed in Table 5.3 and are the following: linear

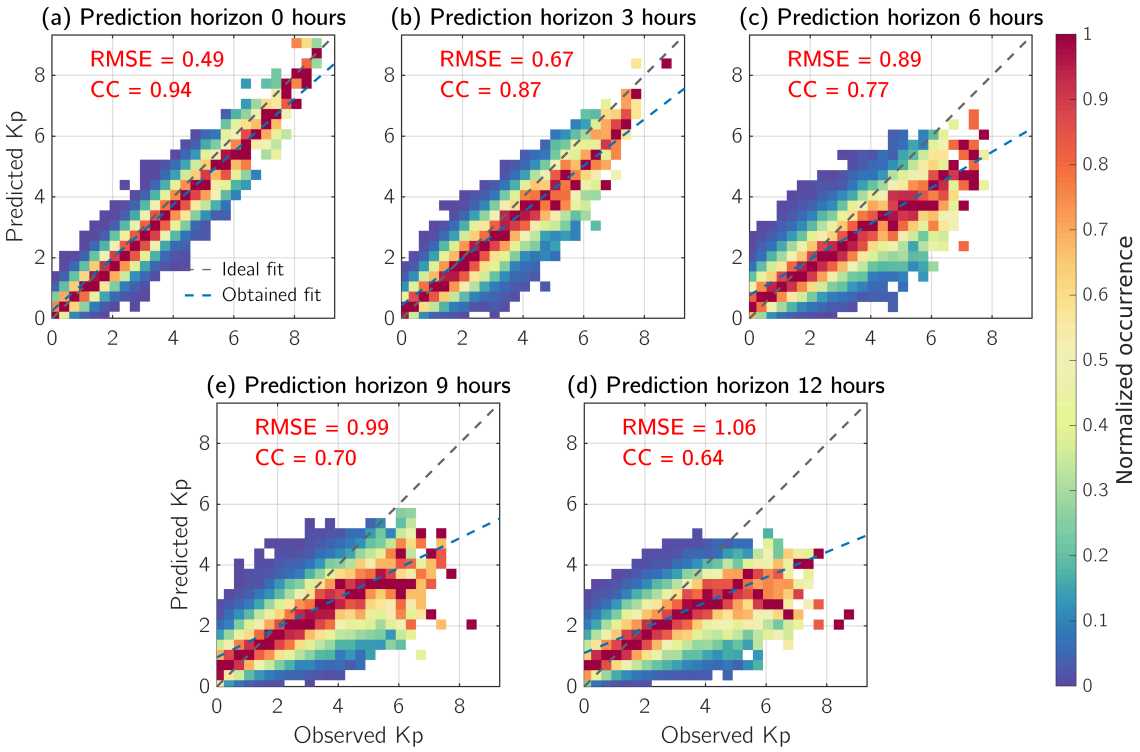

Figure 5.6: Correlation between the observed and predicted Kp values by the neural network for all data (combined training, validation, and test sets) for prediction horizons a) 0 (nowcast), b) 3, c) 6, d) 9, and e) 12 hours ahead. The gray dashed lines indicate the perfect fit and the blue dashed lines indicate the obtained fit.

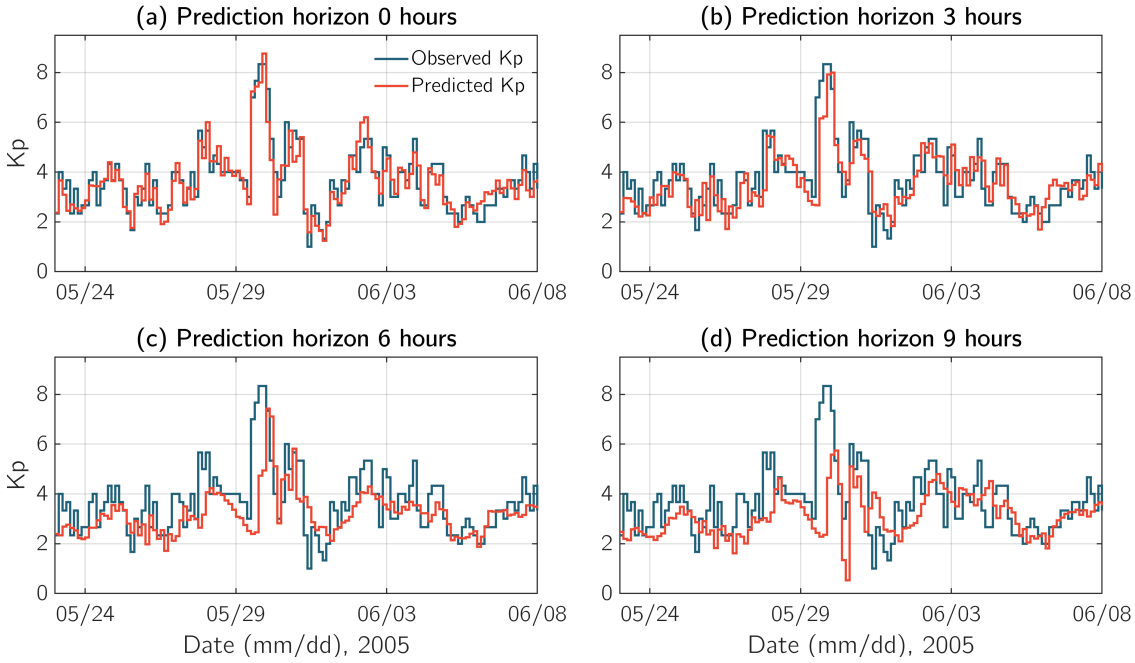

Figure 5.7: Examples of Kp prediction for prediction horizons a) 0, b) 3, c) 6, and d) 9 hours ahead on the event from the test set. The blue lines show the observed Kp index, the red lines show the predicted Kp index.

fit intercept and slope, Pearson correlation coefficient, root mean square error, mean absolute error, mean error, and prediction efficiency. We calculate these metrics using the whole data set, as discussed in the previous section (Figure 5.5). It can be seen that RMSE increases (and correlation coefficient decreases) for the longer prediction horizons, as shown and discussed before. MAE and PE have a similar behavior, indicating that the model performance is better for short-term predictions than for long-term. It is interesting to note that ME, or bias, is close to zero for all prediction horizons. The slope of the linear fit decreases with prediction horizon, however, indicating that the model underpredict Kp as the prediction horizon increases. But, as seen in Figure 5.6, the model tends to underpredict high Kp values and overpredict low Kp values (which is also reflected in the values of the intercept of the linear fit in Table 5.3). Since there are many more low Kp values in our data set than high Kp values, the differences tend to cancel out, and as a result, ME is close to zero for all horizons.

We also compare our best performing models (NN-FFX) to the existing predictive models of Kp. The results of this comparison are presented in Table 5.4. We only

Table 5.3: Fit performance statistics of the NN-FFX models for different prediction horizons computed on all data.

Prediction horizon, hours	**0**	**3**	**6**	**9**	**12**
Number of values in comparison	58439	58439	58439	58439	58439
Intercept of the linear fit	0.2624	0.4444	0.7724	0.9653	1.1073
Slope of the linear fit	0.8712	0.7641	0.5888	0.4894	0.4152
Pearson correlation coefficient (R)	0.9361	0.8742	0.7668	0.7002	0.6430
Root mean square error (RMSE)	0.4865	0.6698	0.8853	0.9856	1.0565
Mean absolute error (MAE)	0.3786	0.5081	0.6734	0.7495	0.8042
Mean error (ME, or bias)	0.0179	-7.26e-04	-0.0030	-0.0016	0.0078
Prediction efficiency (PE)	0.8761	0.7643	0.5880	0.4903	0.4134

consider models that can be compared to our models in terms of prediction horizons. We have therefore not included the following studies into this comparison, as their prediction horizons cannot be directly compared to the ones used in this study (nowcast, 3, 6, 9, and 12 hours ahead): 1 hour ahead in Costello (1998); Balikhin et al. (2001); Boaghe et al. (2001); Ji et al. (2013); 1 and 4 hours ahead in Wing et al. (2005); results for 1-hour predictions by Bala and Reiff (2012). To allow direct comparison between the models, we use the same training, validation, and test time periods as in those studies. We use the input variables to the models that are found by the FFX feature selection algorithm. The comparison, when done in such a way, also helps illustrate that the training interval, as well as validation and test intervals, can affect the performance of the resulting model. The accuracy metrics in Table 5.4 are computed on the test set of the corresponding study. Overall, the resulting accuracy of our models is comparable or slightly better than that of the listed studies. There are no published studies for prediction horizons of 9 and 12 hours ahead to compare. The table also reflects that model errors change depending on the chosen training and test sets (e.g., prediction for 3 hours ahead) and also differ from those listed in Table 5.3. This means that selecting only a specific time interval for testing or validation can affect the resulting model performance. It also demonstrates that assessing a model error using just a specific time interval may not reflect the actual performance of a model. The cross validation procedure described in this study attempts to overcome these issues by including an element of randomness and ensuring that the distributions of the training, validation and test sets are representative and similar to each other.

Table 5.4: Comparison with existing Kp predictive models for prediction horizons 0, 3, and 6 hours ahead. Numbers in bold indicate the best performance within one row (one model for one prediction horizon).

Prediction horizon	**Model**	**RMSE**	**CC**	**RMSE NN-FFX model**	**CC NN-FFX model**	**Test period**
0 (now-cast)	Wintoft et al. (2017)	0.55	0.92	**0.51**	**0.93**	2001, 2011
3 hours	Boberg et al. (2000)	0.98	0.77	n/a	n/a	1986–1996
	Bala and Reiff (2012)	0.65	0.86	**0.62**	**0.88**	2001/04, 2006/01–2007/12
	Tan et al. (2018)	**0.64**	**0.81**	0.65	0.80	2013/12–2014/9
6 hours	Bala and Reiff (2012)	0.85	0.76	**0.82**	**0.78**	2001/04, 2006/01–2007/12

5.5 Discussion

The results obtained in the previous sections demonstrate that machine learning-based models driven by the solar wind measurements at L1 can produce accurate short-term Kp predictions, but the accuracy is reduced for long-term prediction horizons ($>$ 3 hours ahead). This is observed for all machine learning methods considered in this study and for the previous studies as well (Table 5.4). The models cannot capture the storm onset times for long-term horizons accurately and tend to underpredict high Kp and overpredict low Kp values. This indicates that the information contained in solar wind measurements at L1 is not sufficient for accurate long-term predictions of Kp ($>$ 3 hours ahead). Other information sources, such as images of the Sun or features derived from them, should be incorporated into the model to produce accurate long-term forecasts.

Nonlinear machine learning methods, such as gradient boosting and neural networks, perform significantly better than linear regression for short-term prediction horizons, but their performance becomes comparable as the prediction horizon increases (Figure 5.2). This implies that the relation between Kp and solar wind measurements at L1 is non-linear for short-term prediction horizons, but there is little to no gain in using non-linear methods for longer prediction horizons ($>$ 3 hours ahead) when using solar wind measurements at L1 as input to the model.

It should also be noted that all models listed in Table 5.4, as well as ours, have a similar performance (for the same horizon) independent of the inputs or the modeling techniques/methods used. This implies that the usage of another new method/ML technique or a different way of constructing Fthe inputs will probably not bring much more improvement in the performance of a model that uses solar wind measurements at L1 as an input. Moreover, it should be noted that the average RMSE of the nowcast models considered here is ~ 0.5 (during disturbed times, Kp > 4, the average RMSE is ~ 0.8 (Shprits et al., 2019)), which is higher than the cadence of Kp, that is 1/3. Ideally, if the solar wind measurements at L1 contained sufficient information for the prediction of Kp at the current moment, the maximum RMSE of a model would be equal 1/3, since a model could be wrong only by one Kp bin (due to the discretization of Kp; the average RMSE of such a model would be much lower than 1/3, but the RMSE during storms or onset of storms could be larger). However, the fact that the average RMSE of the models is greater than 1/3, independent of the way the inputs are constructed or the method used to develop a model, implies that there is a stochastic component of the magnetosphere system that is not captured in the solar wind measurements and cannot be modeled properly, assuming that our models optimally utilize all the data. Since Kp reflects the geomagnetic disturbance at the Earth's surface due to electric currents in the ionosphere and magnetosphere, there is an uncertainty in the direct relation between solar wind and Kp associated with the coupling between ionosphere and magnetosphere, which is a complex stochastic process. It is also possible that the magnetosphere operates in different regimes depending on the type of incoming solar wind and this, in turn, affects the Kp index. This can be further investigated by defining different types of solar wind and training models separately for different types of solar wind. Types of solar wind can be defined depending on the charge state composition of the solar wind, solar wind speed, proton temperature, proton density, etc. (for example, as done in Heidrich-Meisner and Wimmer-Schweingruber (2018) or Xu and Borovsky (2015)).

Feature selection methods, such as FFX, RF, or MRMR, the use of which was demonstrated in this paper, showed their capabilities to select the most important/significant inputs to the model. Using these methods, the number of inputs to the model was reduced from 166 to 52 for nowcast, and from 166 to 10 for prediction 12 hours ahead (Table 5.2). At the same time, the performance of the models based on the reduced input set remained the same or even slightly improved (when

using the FFX feature selection algorithm) compared to the model based on all 166 inputs.

In addition to the significant reduction of input dimensionality, the selected input variables can be analysed to understand the main drivers for the Kp predictive models. Different methods select slightly different variables as the optimal ones, but we can consider the best performing model, which is based on the variables selected by the FFX method. For the nowcast of Kp, minimum of B_z over the previous 7 hours and maximum of solar wind speed V_{sw} over the previous hour are ones of the selected variables, as well as average and minumum of V_{sw}, average and maximum B_z, average, minimum, and maximum of B (their time history over the varying time intervals). It is interesting to note that the FFX algorithm also selects the minimum and maximum of B_y and maximum of B_x components, as well as minimum and maximum of proton density over the previous 4 hours and the indicators of seasonal variability. For the 3 hours ahead prediction, the dimensionality is reduced even more, down to 14 input variables. The subset of the same variables is selected (min and avg B_z; max and min V_{sw}; max and min B, max B_y, max and min of proton density, and seasonal indicators), except that they are selected for only the several previous hours. At the same time, the performance of the model containing only these 14 inputs is even slightly better compared to the model containing 166 variables, indicating that 14 variables are sufficient and encompass the necessary information to predict Kp 3 hours ahead. The selected input variables for all prediction horizons and all feature selection methods considered are displayed in Appendix 5.B.

5.6 Conclusions

In this study, we explore how different machine learning algorithms, namely Gradient Boosting, Feedforward Neural Networks, and Linear Regression, perform on the task of predicting the Kp index for prediction horizons of 0, 3, 6, 9, and 12 hours ahead using solar wind measurements at L1 as an input. We also illustrate how different feature selection methods can be applied to select the optimal inputs to the predictive model of the Kp index. In particular, we assess the performance of four feature selection procedures based on the Fast Function Extraction (FFX), Random Forest (RF), Mutual Information Maximisation (MIM), and Maximum Relevancy Minimum Redundancy (MRMR) algorithms. We have found that:

1) The models trained using neural networks (NNs) and gradient boosting (GB) notably outperform the models constructed using linear regression (LR) for the short-term prediction horizons. This implies the existence of a nonlinear component in the Kp prediction problem for short-term predictions that cannot be modeled using linear methods alone.

2) The performance of all considered methods decreases as the prediction horizon increases. This likely means that the information in the solar wind measurements at L1 is not sufficient to produce accurate long-term predictions (e.g., > 3 hours), especially for high Kp values, and is not sufficient to accurately capture the arrival time of geomagnetic storms for the long-term prediction horizons.

3) The proposed FFX feature selection algorithm (i.e., a procedure for finding optimal input variables to a model) outperforms other feature selection algorithms considered in this study. It provides a significant reduction of the number of input variables sufficient to model the Kp index (starting from more than 3-fold for the nowcast to more than 16-fold for the prediction of 12 hours ahead).

4) Despite having fewer inputs, the models based on the reduced set of input variables obtained with the FFX algorithm have a slightly better performance than the models based on the full input set. This implies that, using the FFX feature selection algorithm, we can significantly reduce the input dimensionality, obtain a set of the most significant input variables sufficient for predicting the Kp index for different prediction horizons, and, at the same time, improve the model performance.

The obtained sets of optimal input variables can be used to gain an understanding of what inputs are the most important and physically meaningful for predicting the Kp index. Moreover, the models can be trained faster and have less tendency to overfit using such a reduced set of inputs. The feature selection methods described in this work can also be applied to other problems in space physics in order to significantly reduce the input dimensionality and identify the most important inputs that contain sufficient information to produce accurate predictions.

5.A RMSE and CC of all methods considered in the paper

Tables below contain values of RMSE and CC on the validation and training sets for all the methods considered in this study. Table 5.5 contains RMSE and CC of the methods presented in Figure 5.2: Gradient Boosting (GB), Neural Networks (NN), Linear Regression (LR), Persistence, and <Kp>. Table 5.6 contains RMSE and CC of the methods presented in Figure 5.3: Neural Networks with all 166 inputs (NN), NN with inputs selected by Fast Function Extraction (NN-FFX), Random Forest (NN-RF), Maximum Relevancy Minimum Redundancy (NN-MRMR), and Mutual Information Maximization (NN-MIM) feature selection procedures. Table 5.7 contains RMSE and CC of the methods presented in Figure 5.4: Neural Networks with inputs selected by Fast Function Extraction feature selection procedure constructed using 1-hour intervals (NN-FFX 1H) and 3-hour intervals (NN-FFX 3H).

Table 5.5: RMSE and CC of the methods presented in Figure 5.2.

Prediction horizon, hours	Method	RMSE		CC	
		Validation	Training	Validation	Training
0	GB	0.5098	0.4573	0.9291	0.9437
	NN	0.5167	0.4769	0.9274	0.9386
	LR	0.6181	0.6090	0.8941	0.8974
	Persistence	-		-	
	<Kp>	1.2800		0.3700	
3	GB	0.6736	0.6118	0.8721	0.8966
	NN	0.6964	0.6530	0.8632	0.8813
	LR	0.7392	0.7303	0.8439	0.8484
	Persistence	0.8470		0.8080	
	<Kp>	1.2800		0.3700	
6	GB	0.8789	0.7990	0.7696	0.8163
	NN	0.9097	0.8505	0.7532	0.7886
	LR	0.9189	0.9102	0.7448	0.7513
	Persistence	1.0770		0.6900	
	<Kp>	1.2800		0.3700	
9	GB	0.9807	0.8906	0.7020	0.7658
	NN	1.0029	0.9532	0.6875	0.7239
	LR	1.0163	1.0070	0.6751	0.6836
	Persistence	1.2040		0.6130	
	<Kp>	1.2800		0.3700	
12	GB	1.0492	0.9539	0.6473	0.7252
	NN	1.0726	1.0374	0.6295	0.6588
	LR	1.0839	1.0742	0.6168	0.6271
	Persistence	1.2910		0.5540	
	<Kp>	1.2800		0.3700	

Table 5.6: RMSE and CC of the methods presented in Figure 5.3.

Prediction horizon hours	Method	RMSE		CC	
		Validation	Training	Validation	Training
0	NN	0.5167	0.4769	0.9274	0.9386
	NN-FFX	0.5068	0.4764	0.9301	0.9386
	NN-RF	0.5247	0.5004	0.9248	0.9320
	NN-MRMR	0.5295	0.4994	0.9234	0.9324
	NN-MIM	0.5755	0.5642	0.9087	0.9127
3	NN	0.6964	0.6530	0.8632	0.8813
	NN-FFX	0.6774	0.6678	0.8703	0.8747
	NN-RF	0.6956	0.6892	0.8627	0.8658
	NN-MRMR	0.7117	0.7033	0.8559	0.8602
	NN-MIM	0.7334	0.7263	0.8460	0.8497
6	NN	0.9097	0.8505	0.7532	0.7886
	NN-FFX	0.8913	0.8814	0.7615	0.7685
	NN-RF	0.9017	0.8953	0.7547	0.7596
	NN-MRMR	0.9178	0.9091	0.7446	0.7514
	NN-MIM	0.9854	0.9788	0.6979	0.7039
9	NN	1.0029	0.9532	0.6875	0.7239
	NN-FFX	0.9936	0.9804	0.6919	0.7031
	NN-RF	0.9963	0.9857	0.6887	0.6979
	NN-MRMR	1.0164	1.0008	0.6744	0.6879
	NN-MIM	1.0504	1.0430	0.6459	0.6535
12	NN	1.0726	1.0374	0.6295	0.6588
	NN-FFX	1.0618	1.0523	0.6357	0.6453
	NN-RF	1.0683	1.0573	0.6294	0.6404
	NN-MRMR	1.0786	1.0700	0.6201	0.6296
	NN-MIM	1.1039	1.0993	0.5970	0.6029

Table 5.7: RMSE and CC of the methods presented in Figure 5.4.

Prediction horizon, hours	Method	RMSE		CC	
		Validation	Training	Validation	Training
0	NN-FFX 1H	0.5068	0.4764	0.9301	0.9386
	NN-FFX 3H	0.5201	0.4978	0.9254	0.9320
3	NN-FFX 1H	0.6774	0.6678	0.8703	0.8747
	NN-FFX 3H	0.7118	0.6944	0.8551	0.8631
6	NN-FFX 1H	0.8913	0.8814	0.7615	0.7685
	NN-FFX 3H	0.9051	0.8878	0.7523	0.7640
9	NN-FFX 1H	0.9936	0.9804	0.6919	0.7031
	NN-FFX 3H	1.0007	0.9870	0.6850	0.6967
12	NN-FFX 1H	1.0618	1.0523	0.6357	0.6453
	NN-FFX 3H	1.0659	1.0521	0.6307	0.6442

5.B Optimal inputs selected by feature selection algorithms

Tables below contain reduced sets of input variables selected by the feature selection algorithms described in this paper (section 5.2.2). Results for all prediction horizons considered in this work (from nowcast to 12 hours ahead) are presented for both 1-hour and 3-hour based inputs. The format of the inputs is the following: the subscript denotes the operation which is performed to obtain the input (taking max, min, or averaging) and the numbers in the brackets (e.g., $(-3;\ -2)$ in $B_{z\min(-3;\,-2)}$) denote the time interval over which max, min, or average is taken. T is the UT hour of the day, and D is day of the year.

Table 5.8: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 0$ (nowcast) with a 1-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

$h = 0$ (nowcast)			
FFX (1H)	**RF (1H)**	**MRMR (1H)**	**MIM (1H)**
$B_{z\min(-3;\,-2)}$	$B_{\mathrm{avg}(-1;\,0)}$	$B_{z\min(-3;\,-2)}$	$B_{z\min(-3;\,-2)}$

$B_{z\min(-4;\,-3)}$	$B_{z\min(-1;\,0)}$	$V_{sw\max(-1;\,0)}$	$V_{sw\max(-1;\,0)}$
$B_{z\min(-5;\,-4)}$	$B_{z\min(-2;\,-1)}$	$V_{sw\max(-2;\,-1)}$	$B_{z\min(-4;\,-3)}$
$B_{z\min(-7;\,-6)}$	$B_{z\min(-3;\,-2)}$	$V_{sw\text{avg}(-2;\,-1)}$	$B_{z\min(-2;\,-1)}$
$V_{sw\text{avg}(-2;\,-1)}$	$B_{z\min(-4;\,-3)}$	$V_{sw\min(-1;\,0)}$	$V_{sw\text{avg}(-1;\,0)}$
$V_{sw\max(-1;\,0)}$	$B_{z\min(-5;\,-4)}$	$B_{z\min(-1;\,0)}$	$V_{sw\max(-2;\,-1)}$
$V_{sw\min(-1;\,0)}$	$B_{z\min(-6;\,-5)}$	$B_{z\min(-5;\,-4)}$	$V_{sw\text{avg}(-2;\,-1)}$
$V_{sw\min(-3;\,-2)}$	$B_{\text{avg}(-7;\,-6)}$	$B_{\max(-3;\,-2)}$	$V_{sw\min(-1;\,0)}$
$B_{\text{avg}(-9;\,-8)}$	$V_{sw\text{avg}(-2;\,-1)}$	$B_{z\min(-4;\,-3)}$	$V_{sw\max(-3;\,-2)}$
$V_{sw\min(-5;\,-4)}$	$V_{sw\text{avg}(-3;\,-2)}$	$B_{z\min(-2;\,-1)}$	$V_{sw\min(-2;\,-1)}$
$V_{sw\min(-9;\,-8)}$	$V_{sw\text{avg}(-4;\,-3)}$	$B_{\max(-9;\,-8)}$	$V_{sw\text{avg}(-3;\,-2)}$
$B_{\max(-1;\,0)}$	$V_{sw\text{avg}(-5;\,-4)}$	$V_{sw\max(-3;\,-2)}$	$V_{sw\max(-4;\,-3)}$
$nProt_{\max(-1;\,0)}$	$V_{sw\text{avg}(-6;\,-5)}$	$B_{z\min(-7;\,-6)}$	$B_{\max(-3;\,-2)}$
$nProt_{\max(-2;\,-1)}$	$V_{sw\text{avg}(-7;\,-6)}$	$B_{z\min(-6;\,-5)}$	$B_{z\min(-1;\,0)}$
$nProt_{\max(-3;\,-2)}$	$V_{sw\max(-1;\,0)}$	$B_{\max(-1;\,0)}$	$V_{sw\min(-3;\,-2)}$
$nProt_{\max(-4;\,-3)}$	$V_{sw\max(-2;\,-1)}$	$B_{z\text{avg}(-3;\,-2)}$	$B_{\max(-4;\,-3)}$
$B_{\max(-2;\,-1)}$	$V_{sw\max(-3;\,-2)}$	$B_{z\min(-9;\,-8)}$	$B_{\max(-2;\,-1)}$
$nProt_{\min(-2;\,-1)}$	$V_{sw\max(-4;\,-3)}$	$V_{sw\max(-5;\,-4)}$	$V_{sw\text{avg}(-4;\,-3)}$
$nProt_{\min(-3;\,-2)}$	$V_{sw\max(-5;\,-4)}$	$B_{\max(-2;\,-1)}$	$B_{z\min(-5;\,-4)}$
$\sin(2\pi T/24)$	$V_{sw\max(-6;\,-5)}$	$B_{z\min(-8;\,-7)}$	$B_{\max(-1;\,0)}$
$\cos(2\pi T/24)$	$V_{sw\max(-7;\,-6)}$	$B_{z\text{avg}(-2;\,-1)}$	$V_{sw\max(-5;\,-4)}$
$\sin(2\pi D/365)$	$V_{sw\min(-1;\,0)}$	$B_{x\min(-7;\,-6)}$	$B_{\max(-5;\,-4)}$
$\cos(2\pi D/365)$	$V_{sw\min(-2;\,-1)}$	$B_{z\text{avg}(-4;\,-3)}$	$B_{\max(-6;\,-5)}$
$B_{\max(-6;\,-5)}$	$V_{sw\min(-3;\,-2)}$	$\cos(2\pi D/365)$	$B_{\text{avg}(-3;\,-2)}$
$B_{\max(-9;\,-8)}$	$V_{sw\min(-4;\,-3)}$	$B_{\max(-4;\,-3)}$	$V_{sw\min(-4;\,-3)}$
$B_{\min(-6;\,-5)}$	$V_{sw\min(-5;\,-4)}$	$V_{sw\min(-2;\,-1)}$	$B_{\text{avg}(-4;\,-3)}$
$B_{z\text{avg}(-1;\,0)}$	$V_{sw\min(-6;\,-5)}$	$B_{y_{\max(-3;\,-2)}}$	$V_{sw\text{avg}(-5;\,-4)}$
$V_{sw\text{avg}(-1;\,0)}$	$V_{sw\min(-7;\,-6)}$	$B_{z\text{avg}(-1;\,0)}$	$B_{\text{avg}(-2;\,-1)}$
$B_{x\max(-2;\,-1)}$	$V_{sw\min(-8;\,-7)}$	$B_{\max(-7;\,-6)}$	$V_{sw\max(-6;\,-5)}$
$B_{x\min(-1;\,0)}$	$V_{sw\min(-9;\,-8)}$	$\sin(2\pi T/24)$	$B_{\max(-7;\,-6)}$
$B_{x\min(-5;\,-4)}$	$B_{\max(-1;\,0)}$	$B_{x\min(-1;\,0)}$	$B_{\text{avg}(-5;\,-4)}$
$B_{y_{\max(-1;\,0)}}$	$nProt_{\max(-1;\,0)}$	$B_{y_{\min(-1;\,0)}}$	$B_{\text{avg}(-1;\,0)}$
$B_{y_{\max(-2;\,-1)}}$	$nProt_{\max(-2;\,-1)}$	$V_{sw\max(-4;\,-3)}$	$B_{\text{avg}(-6;\,-5)}$
$B_{y_{\max(-3;\,-2)}}$	$nProt_{\max(-3;\,-2)}$	$B_{\max(-5;\,-4)}$	$V_{sw\min(-5;\,-4)}$

$B_{y\max(-4;\,-3)}$	$nProt_{\max(-4;\,-3)}$	$B_{z\text{avg}(-5;\,-4)}$	$V_{sw\text{avg}(-6;\,-5)}$
$B_{y\max(-5;\,-4)}$	$B_{\max(-2;\,-1)}$	$B_{\max(-8;\,-7)}$	$B_{\max(-8;\,-7)}$
$B_{y\min(-1;\,0)}$	$B_{\max(-3;\,-2)}$	$B_{x\min(-9;\,-8)}$	$B_{\max(-9;\,-8)}$
$B_{y\min(-2;\,-1)}$	$\sin(2\pi D/365)$	$B_{z\max(-9;\,-8)}$	$V_{sw\max(-7;\,-6)}$
$B_{y\min(-3;\,-2)}$	$\cos(2\pi D/365)$	$B_{y\max(-1;\,0)}$	$B_{\text{avg}(-7;\,-6)}$
$B_{y\min(-4;\,-3)}$	$B_{\max(-4;\,-3)}$	$V_{sw\max(-9;\,-8)}$	$V_{sw\min(-6;\,-5)}$
$B_{z\text{avg}(-2;\,-1)}$	$B_{\max(-5;\,-4)}$	$B_{\text{avg}(-1;\,0)}$	$B_{\text{avg}(-8;\,-7)}$
$B_{z\text{avg}(-3;\,-2)}$	$B_{\max(-6;\,-5)}$	$B_{z\text{avg}(-6;\,-5)}$	$V_{sw\text{avg}(-7;\,-6)}$
$B_{z\text{avg}(-4;\,-3)}$	$B_{\max(-7;\,-6)}$	$B_{y\max(-9;\,-8)}$	$B_{z\min(-6;\,-5)}$
$B_{z\text{avg}(-5;\,-4)}$	$B_{z\text{avg}(-1;\,0)}$	$B_{x\max(-3;\,-2)}$	$V_{sw\max(-8;\,-7)}$
$B_{z\text{avg}(-6;\,-5)}$	$V_{sw\text{avg}(-1;\,0)}$	$V_{sw\text{avg}(-3;\,-2)}$	$B_{\text{avg}(-9;\,-8)}$
$B_{z\text{avg}(-7;\,-6)}$	$nProt_{\text{avg}(-1;\,0)}$	$B_{\max(-6;\,-5)}$	$V_{sw\min(-7;\,-6)}$
$B_{z\text{avg}(-8;\,-7)}$	$B_{y\max(-5;\,-4)}$	$B_{z\max(-8;\,-7)}$	$V_{sw\text{avg}(-8;\,-7)}$
$B_{z\text{avg}(-9;\,-8)}$	$B_{z\text{avg}(-2;\,-1)}$	$B_{z\max(-3;\,-2)}$	$V_{sw\max(-9;\,-8)}$
$B_{z\max(-3;\,-2)}$	$B_{z\text{avg}(-3;\,-2)}$	$B_{y\min(-4;\,-3)}$	$V_{sw\min(-8;\,-7)}$
$B_{z\max(-4;\,-3)}$	$B_{z\text{avg}(-4;\,-3)}$	$\cos(2\pi T/24)$	$V_{sw\text{avg}(-9;\,-8)}$
$B_{z\max(-5;\,-4)}$	$B_{z\text{avg}(-5;\,-4)}$	$B_{\text{avg}(-2;\,-1)}$	$B_{z\min(-7;\,-6)}$
$B_{z\max(-6;\,-5)}$	$B_{z\text{avg}(-7;\,-6)}$	$V_{sw\text{avg}(-1;\,0)}$	$V_{sw\min(-9;\,-8)}$

Table 5.9: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 0$ (nowcast) with a 3-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

$h = 0$ (nowcast)

FFX (3H)	**RF (3H)**	**MRMR (3H)**	**MIM (3H)**
$B_{\text{avg}(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$	$B_{z\min(-3;\,0)}$	$B_{z\min(-3;\,0)}$
$B_{x\text{avg}(-3;\,0)}$	$B_{\max(-9;\,-6)}$	$V_{sw\max(-3;\,0)}$	$B_{z\min(-6;\,-3)}$
$B_{\min(-9;\,-6)}$	$B_{\min(-3;\,0)}$	$B_{z\min(-6;\,-3)}$	$V_{sw\max(-3;\,0)}$
$B_{x\text{avg}(-6;\,-3)}$	$B_{y\max(-3;\,0)}$	$B_{\max(-9;\,-6)}$	$V_{sw\text{avg}(-3;\,0)}$
$B_{x\max(-3;\,0)}$	$B_{y\max(-6;\,-3)}$	$\cos(2\pi D/365)$	$B_{\max(-3;\,0)}$
$B_{x\max(-9;\,-6)}$	$B_{y\min(-9;\,-6)}$	$B_{y\max(-3;\,0)}$	$B_{\max(-6;\,-3)}$
$B_{y\text{avg}(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$V_{sw\min(-3;\,0)}$
$B_{y\text{avg}(-6;\,-3)}$	$B_{z\text{avg}(-6;\,-3)}$	$B_{z\min(-9;\,-6)}$	$V_{sw\max(-6;\,-3)}$

$B_{y\max(-3;\,0)}$	$B_{z\text{avg}(-9;\,-6)}$	$B_{x\max(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$
$B_{y\max(-6;\,-3)}$	$B_{z\max(-3;\,0)}$	$V_{sw\min(-3;\,0)}$	$B_{\text{avg}(-6;\,-3)}$
$B_{y\min(-3;\,0)}$	$B_{z\min(-3;\,0)}$	$B_{\max(-3;\,0)}$	$V_{sw\text{avg}(-6;\,-3)}$
$B_{y\min(-6;\,-3)}$	$B_{z\min(-6;\,-3)}$	$B_{z\text{avg}(-6;\,-3)}$	$B_{\max(-9;\,-6)}$
$B_{z\text{avg}(-3;\,0)}$	$B_{z\min(-9;\,-6)}$	$\sin(2\pi T/24)$	$B_{\text{avg}(-9;\,-6)}$
$B_{z\text{avg}(-6;\,-3)}$	$V_{sw\text{avg}(-6;\,-3)}$	$B_{x\min(-9;\,-6)}$	$V_{sw\min(-6;\,-3)}$
$B_{z\text{avg}(-9;\,-6)}$	$V_{sw\text{avg}(-9;\,-6)}$	$B_{y\min(-3;\,0)}$	$B_{z\min(-9;\,-6)}$
$B_{z\max(-3;\,0)}$	$V_{sw\text{avg}(-3;\,0)}$	$B_{\max(-6;\,-3)}$	$V_{sw\max(-9;\,-6)}$
$B_{z\max(-6;\,-3)}$	$V_{sw\max(-3;\,0)}$	$V_{sw\max(-9;\,-6)}$	$V_{sw\text{avg}(-9;\,-6)}$
$B_{z\max(-9;\,-6)}$	$V_{sw\max(-6;\,-3)}$	$B_{x\min(-3;\,0)}$	$V_{sw\min(-9;\,-6)}$
$B_{z\min(-3;\,0)}$	$V_{sw\max(-9;\,-6)}$	$B_{z\max(-9;\,-6)}$	$B_{z\text{avg}(-3;\,0)}$
$V_{sw\text{avg}(-3;\,0)}$	$V_{sw\min(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$	$B_{y\max(-3;\,0)}$
$V_{sw\max(-3;\,0)}$	$V_{sw\min(-6;\,-3)}$	$B_{z\text{avg}(-9;\,-6)}$	$B_{y\max(-6;\,-3)}$
$V_{sw\min(-3;\,0)}$	$V_{sw\min(-9;\,-6)}$	$B_{y\min(-9;\,-6)}$	$B_{\min(-6;\,-3)}$
$V_{sw\min(-6;\,-3)}$	$nProt_{\text{avg}(-6;\,-3)}$	$V_{sw\text{avg}(-3;\,0)}$	$B_{\min(-9;\,-6)}$
$V_{sw\min(-9;\,-6)}$	$nProt_{\max(-3;\,0)}$	$B_{y\max(-6;\,-3)}$	$B_{\min(-3;\,0)}$
$nProt_{\max(-3;\,0)}$	$nProt_{\max(-6;\,-3)}$	$B_{z\max(-3;\,0)}$	$B_{z\text{avg}(-6;\,-3)}$
$nProt_{\max(-6;\,-3)}$	$nProt_{\text{avg}(-3;\,0)}$	$B_{x\max(-9;\,-6)}$	$B_{y\max(-9;\,-6)}$
$nProt_{\min(-3;\,0)}$	$nProt_{\max(-9;\,-6)}$	$B_{\text{avg}(-9;\,-6)}$	$B_{y\min(-3;\,0)}$
$nProt_{\min(-6;\,-3)}$	$nProt_{\min(-3;\,0)}$	$B_{y\max(-9;\,-6)}$	$B_{x\min(-3;\,0)}$
$nProt_{\min(-9;\,-6)}$	$\sin(2\pi D/365)$	$V_{sw\max(-6;\,-3)}$	$B_{y\min(-6;\,-3)}$
$\sin(2\pi T/24)$	$\cos(2\pi D/365)$	$B_{x\min(-6;\,-3)}$	$B_{x\min(-6;\,-3)}$
$\cos(2\pi T/24)$	$B_{\text{avg}(-6;\,-3)}$	$nProt_{\max(-3;\,0)}$	$B_{z\max(-9;\,-6)}$
$\sin(2\pi D/365)$	$B_{\text{avg}(-9;\,-6)}$	$B_{y\min(-6;\,-3)}$	$B_{z\max(-6;\,-3)}$
$\cos(2\pi D/365)$	$B_{\max(-3;\,0)}$	$B_{z\max(-6;\,-3)}$	$B_{x\min(-9;\,-6)}$
$B_{\max(-3;\,0)}$	$B_{\max(-6;\,-3)}$	$\cos(2\pi T/24)$	$B_{z\max(-3;\,0)}$

Table 5.10: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 3$ hours ahead with a 1-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

$\boldsymbol{h = 3}$ **hours ahead**			

FFX (1H)	RF (1H)	MRMR (1H)	MIM (1H)
$B_{z\min(-1;\,0)}$	$B_{z\min(-1;\,0)}$	$B_{z\min(-1;\,0)}$	$B_{z\min(-1;\,0)}$
$B_{z\min(-2;\,-1)}$	$B_{z\min(-2;\,-1)}$	$V_{sw\max(-1;\,0)}$	$V_{sw\max(-1;\,0)}$
$V_{sw\max(-1;\,0)}$	$B_{z\min(-3;\,-2)}$	$B_{\max(-3;\,-2)}$	$B_{\max(-1;\,0)}$
$V_{sw\min(-1;\,0)}$	$B_{z\min(-4;\,-3)}$	$B_{z\min(-9;\,-8)}$	$V_{sw\text{avg}(-1;\,0)}$
$B_{\max(-1;\,0)}$	$V_{sw\text{avg}(-2;\,-1)}$	$B_{z\min(-2;\,-1)}$	$B_{z\min(-2;\,-1)}$
$nProt_{\max(-1;\,0)}$	$V_{sw\max(-1;\,0)}$	$\cos(2\pi D/365)$	$V_{sw\max(-2;\,-1)}$
$nProt_{\min(-1;\,0)}$	$V_{sw\max(-2;\,-1)}$	$B_{z\min(-4;\,-3)}$	$B_{\max(-2;\,-1)}$
$\sin(2\pi T/24)$	$V_{sw\min(-1;\,0)}$	$B_{x\min(-9;\,-8)}$	$B_{\max(-3;\,-2)}$
$\cos(2\pi D/365)$	$V_{sw\min(-2;\,-1)}$	$B_{\max(-1;\,0)}$	$V_{sw\min(-1;\,0)}$
$B_{z\text{avg}(-1;\,0)}$	$B_{\max(-1;\,0)}$	$B_{z\min(-6;\,-5)}$	$B_{\text{avg}(-1;\,0)}$
$B_{\min(-9;\,-8)}$	$nProt_{\max(-1;\,0)}$	$V_{sw\max(-3;\,-2)}$	$V_{sw\text{avg}(-2;\,-1)}$
$B_{y\max(-1;\,0)}$	$B_{z\text{avg}(-1;\,0)}$	$B_{z\min(-3;\,-2)}$	$V_{sw\max(-3;\,-2)}$
$B_{z\text{avg}(-2;\,-1)}$	$V_{sw\text{avg}(-1;\,0)}$	$B_{z\text{avg}(-1;\,0)}$	$B_{\max(-4;\,-3)}$
$B_{z\text{avg}(-5;\,-4)}$	$nProt_{\text{avg}(-1;\,0)}$	$\sin(2\pi T/24)$	$B_{\text{avg}(-2;\,-1)}$

Table 5.11: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 3$ hours ahead with a 3-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

h = 3 hours ahead

FFX (3H)	RF (3H)	MRMR (3H)	MIM (3H)
$B_{\text{avg}(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$	$B_{z\min(-3;\,0)}$	$B_{z\min(-3;\,0)}$
$B_{x\text{avg}(-3;\,0)}$	$B_{\min(-3;\,0)}$	$V_{sw\max(-3;\,0)}$	$B_{\max(-3;\,0)}$
$B_{\min(-3;\,0)}$	$B_{x\min(-3;\,0)}$	$B_{\max(-3;\,0)}$	$V_{sw\max(-3;\,0)}$
$B_{\min(-9;\,-6)}$	$B_{x\min(-6;\,-3)}$	$\cos(2\pi D/365)$	$B_{\text{avg}(-3;\,0)}$
$B_{x\max(-9;\,-6)}$	$B_{y\max(-3;\,0)}$	$B_{z\min(-9;\,-6)}$	$V_{sw\text{avg}(-3;\,0)}$
$B_{y\text{avg}(-3;\,0)}$	$B_{y\min(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$B_{\max(-6;\,-3)}$
$B_{y\max(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$B_{x\min(-3;\,0)}$	$B_{\text{avg}(-6;\,-3)}$
$B_{y\max(-9;\,-6)}$	$B_{z\max(-3;\,0)}$	$B_{z\min(-6;\,-3)}$	$V_{sw\min(-3;\,0)}$
$B_{y\min(-3;\,0)}$	$B_{z\min(-3;\,0)}$	$\sin(2\pi T/24)$	$B_{z\min(-6;\,-3)}$
$B_{y\min(-9;\,-6)}$	$B_{z\min(-6;\,-3)}$	$B_{y\min(-3;\,0)}$	$V_{sw\max(-6;\,-3)}$
$B_{z\text{avg}(-3;\,0)}$	$B_{z\min(-9;\,-6)}$	$B_{\max(-9;\,-6)}$	$B_{\max(-9;\,-6)}$

$B_{z\text{avg}(-9;\,-6)}$	$V_{sw\text{avg}(-6;\,-3)}$	$V_{sw\text{min}(-3;\,0)}$	$V_{sw\text{avg}(-6;\,-3)}$
$B_{z\text{max}(-3;\,0)}$	$V_{sw\text{avg}(-9;\,-6)}$	$B_{y_{\text{max}(-3;\,0)}}$	$B_{\text{avg}(-9;\,-6)}$
$B_{z\text{max}(-6;\,-3)}$	$V_{sw\text{avg}(-3;\,0)}$	$B_{z\text{max}(-9;\,-6)}$	$V_{sw\text{min}(-6;\,-3)}$
$B_{z\text{min}(-3;\,0)}$	$V_{sw\text{max}(-3;\,0)}$	$B_{x\text{max}(-3;\,0)}$	$V_{sw\text{max}(-9;\,-6)}$
$V_{sw\text{max}(-3;\,0)}$	$V_{sw\text{max}(-6;\,-3)}$	$B_{\text{avg}(-3;\,0)}$	$B_{z\text{min}(-9;\,-6)}$
$V_{sw\text{min}(-3;\,0)}$	$V_{sw\text{max}(-9;\,-6)}$	$B_{x\text{min}(-9;\,-6)}$	$V_{sw\text{avg}(-9;\,-6)}$
$nProt_{\text{max}(-3;\,0)}$	$V_{sw\text{min}(-3;\,0)}$	$B_{z\text{avg}(-6;\,-3)}$	$V_{sw\text{min}(-9;\,-6)}$
$nProt_{\text{max}(-6;\,-3)}$	$V_{sw\text{min}(-6;\,-3)}$	$V_{sw\text{max}(-9;\,-6)}$	$B_{y_{\text{max}(-3;\,0)}}$
$nProt_{\text{avg}(-3;\,0)}$	$V_{sw\text{min}(-9;\,-6)}$	$B_{y_{\text{min}(-9;\,-6)}}$	$B_{\text{min}(-3;\,0)}$
$nProt_{\text{max}(-9;\,-6)}$	$nProt_{\text{avg}(-6;\,-3)}$	$B_{z\text{max}(-3;\,0)}$	$B_{\text{min}(-6;\,-3)}$
$nProt_{\text{min}(-3;\,0)}$	$nProt_{\text{max}(-3;\,0)}$	$B_{y_{\text{max}(-6;\,-3)}}$	$B_{z\text{avg}(-3;\,0)}$
$nProt_{\text{min}(-6;\,-3)}$	$nProt_{\text{max}(-6;\,-3)}$	$B_{z\text{avg}(-9;\,-6)}$	$B_{y_{\text{max}(-6;\,-3)}}$
$\sin(2\pi T/24)$	$nProt_{\text{avg}(-3;\,0)}$	$nProt_{\text{max}(-3;\,0)}$	$B_{\text{min}(-9;\,-6)}$
$\cos(2\pi T/24)$	$nProt_{\text{min}(-3;\,0)}$	$B_{\text{max}(-6;\,-3)}$	$B_{y_{\text{min}(-3;\,0)}}$
$\sin(2\pi D/365)$	$\sin(2\pi D/365)$	$B_{x\text{max}(-9;\,-6)}$	$B_{y_{\text{max}(-9;\,-6)}}$
$\cos(2\pi D/365)$	$\cos(2\pi D/365)$	$V_{sw\text{avg}(-3;\,0)}$	$B_{x\text{min}(-3;\,0)}$
$B_{\text{avg}(-9;\,-6)}$	$B_{\text{avg}(-6;\,-3)}$	$B_{z\text{max}(-6;\,-3)}$	$B_{z\text{max}(-6;\,-3)}$
$B_{\text{max}(-3;\,0)}$	$B_{\text{max}(-3;\,0)}$	$B_{x\text{min}(-6;\,-3)}$	$B_{z\text{max}(-3;\,0)}$
$B_{\text{max}(-6;\,-3)}$	$B_{\text{max}(-6;\,-3)}$	$B_{y_{\text{max}(-9;\,-6)}}$	$B_{x\text{min}(-6;\,-3)}$

Table 5.12: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 6$ hours ahead with a 1-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

$\boldsymbol{h}$ **= 6 hours ahead**			
FFX (1H)	**RF (1H)**	**MRMR (1H)**	**MIM (1H)**
$B_{z\text{min}(-1;\,0)}$	$B_{\text{avg}(-1;\,0)}$	$B_{\text{max}(-1;\,0)}$	$B_{\text{max}(-1;\,0)}$
$B_{z\text{min}(-4;\,-3)}$	$B_{z\text{min}(-1;\,0)}$	$V_{sw\text{max}(-1;\,0)}$	$B_{\text{max}(-2;\,-1)}$
$V_{sw\text{max}(-1;\,0)}$	$B_{z\text{min}(-2;\,-1)}$	$B_{z\text{min}(-1;\,0)}$	$B_{\text{max}(-3;\,-2)}$
$V_{sw\text{min}(-1;\,0)}$	$V_{sw\text{max}(-1;\,0)}$	$\cos(2\pi D/365)$	$V_{sw\text{max}(-1;\,0)}$
$B_{\text{max}(-1;\,0)}$	$V_{sw\text{max}(-2;\,-1)}$	$B_{z\text{min}(-8;\,-7)}$	$B_{\text{avg}(-1;\,0)}$
$nProt_{\text{max}(-1;\,0)}$	$V_{sw\text{min}(-1;\,0)}$	$B_{z\text{min}(-3;\,-2)}$	$B_{\text{avg}(-2;\,-1)}$
$nProt_{\text{min}(-1;\,0)}$	$V_{sw\text{min}(-2;\,-1)}$	$B_{x\text{min}(-7;\,-6)}$	$V_{sw\text{avg}(-1;\,0)}$

$\sin(2\pi T/24)$	$B_{\max(-1;\,0)}$	$\sin(2\pi T/24)$	$B_{\max(-4;\,-3)}$
$\cos(2\pi D/365)$	$nProt_{\max(-1;\,0)}$	$B_{z\min(-5;\,-4)}$	$V_{sw\max(-2;\,-1)}$
$B_{\min(-1;\,0)}$	$B_{\max(-2;\,-1)}$	$B_{\max(-9;\,-8)}$	$B_{\text{avg}(-3;\,-2)}$
$B_{z\text{avg}(-1;\,0)}$	$B_{z\text{avg}(-1;\,0)}$	$B_{z\min(-2;\,-1)}$	$V_{sw\min(-1;\,0)}$
$B_{\min(-9;\,-8)}$	$V_{sw\text{avg}(-1;\,0)}$	$V_{sw\max(-7;\,-6)}$	$V_{sw\text{avg}(-2;\,-1)}$
$B_{z\text{avg}(-5;\,-4)}$	$nProt_{\text{avg}(-1;\,0)}$	$B_{y_{\max(-1;\,0)}}$	$B_{\max(-5;\,-4)}$

Table 5.13: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 6$ hours ahead with a 3-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

h = 6 hours ahead

FFX (3H)	**RF (3H)**	**MRMR (3H)**	**MIM (3H)**
$B_{\text{avg}(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$	$B_{\max(-3;\,0)}$	$B_{\max(-3;\,0)}$
$B_{\min(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$V_{sw\max(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$
$B_{\min(-9;\,-6)}$	$B_{z\max(-3;\,0)}$	$B_{z\min(-3;\,0)}$	$B_{z\min(-3;\,0)}$
$B_{x\min(-3;\,0)}$	$B_{z\min(-3;\,0)}$	$\cos(2\pi D/365)$	$V_{sw\max(-3;\,0)}$
$B_{y_{\max(-3;\,0)}}$	$B_{z\min(-9;\,-6)}$	$B_{z\min(-9;\,-6)}$	$B_{\max(-6;\,-3)}$
$B_{z\text{avg}(-3;\,0)}$	$V_{sw\text{avg}(-6;\,-3)}$	$\sin(2\pi T/24)$	$V_{sw\text{avg}(-3;\,0)}$
$B_{z\text{avg}(-6;\,-3)}$	$V_{sw\text{avg}(-9;\,-6)}$	$B_{x\min(-3;\,0)}$	$B_{\text{avg}(-6;\,-3)}$
$B_{z\text{avg}(-9;\,-6)}$	$V_{sw\text{avg}(-3;\,0)}$	$B_{z\min(-6;\,-3)}$	$B_{\max(-9;\,-6)}$
$B_{z\min(-3;\,0)}$	$V_{sw\max(-3;\,0)}$	$B_{y_{\min(-3;\,0)}}$	$V_{sw\min(-3;\,0)}$
$B_{z\min(-9;\,-6)}$	$V_{sw\max(-6;\,-3)}$	$B_{z\max(-9;\,-6)}$	$V_{sw\max(-6;\,-3)}$
$V_{sw\max(-3;\,0)}$	$V_{sw\min(-3;\,0)}$	$B_{z\max(-3;\,0)}$	$B_{\text{avg}(-9;\,-6)}$
$V_{sw\min(-3;\,0)}$	$V_{sw\min(-6;\,-3)}$	$V_{sw\max(-9;\,-6)}$	$B_{z\min(-6;\,-3)}$
$V_{sw\min(-9;\,-6)}$	$V_{sw\min(-9;\,-6)}$	$B_{x\min(-9;\,-6)}$	$V_{sw\text{avg}(-6;\,-3)}$
$nProt_{\max(-3;\,0)}$	$nProt_{\max(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$V_{sw\min(-6;\,-3)}$
$nProt_{\text{avg}(-3;\,0)}$	$nProt_{\text{avg}(-3;\,0)}$	$B_{y_{\max(-3;\,0)}}$	$V_{sw\max(-9;\,-6)}$
$nProt_{\min(-3;\,0)}$	$nProt_{\min(-3;\,0)}$	$B_{x\max(-3;\,0)}$	$V_{sw\text{avg}(-9;\,-6)}$
$\sin(2\pi T/24)$	$\sin(2\pi D/365)$	$B_{\max(-9;\,-6)}$	$B_{z\min(-9;\,-6)}$
$\cos(2\pi D/365)$	$\cos(2\pi D/365)$	$nProt_{\min(-3;\,0)}$	$B_{\min(-3;\,0)}$
$B_{\max(-3;\,0)}$	$B_{\max(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$	$V_{sw\min(-9;\,-6)}$

Table 5.14: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 9$ hours ahead with a 1-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

$h = 9$ hours ahead			
FFX (1H)	**RF (1H)**	**MRMR (1H)**	**MIM (1H)**
$B_{z\min(-1;\,0)}$	$B_{\text{avg}(-1;\,0)}$	$B_{\max(-1;\,0)}$	$B_{\max(-1;\,0)}$
$B_{z\min(-4;\,-3)}$	$B_{z\min(-1;\,0)}$	$V_{sw\max(-2;\,-1)}$	$B_{\max(-2;\,-1)}$
$B_{z\min(-7;\,-6)}$	$B_{z\min(-2;\,-1)}$	$\cos(2\pi D/365)$	$B_{\max(-3;\,-2)}$
$B_{z\min(-9;\,-8)}$	$V_{sw\max(-1;\,0)}$	$B_{z\min(-3;\,-2)}$	$B_{\text{avg}(-1;\,0)}$
$V_{sw\max(-1;\,0)}$	$V_{sw\max(-2;\,-1)}$	$B_{z\min(-9;\,-8)}$	$B_{\text{avg}(-2;\,-1)}$
$V_{sw\min(-9;\,-8)}$	$V_{sw\min(-1;\,0)}$	$\sin(2\pi T/24)$	$B_{\max(-4;\,-3)}$
$B_{\max(-1;\,0)}$	$V_{sw\min(-2;\,-1)}$	$B_{z\min(-1;\,0)}$	$B_{\text{avg}(-3;\,-2)}$
$nProt_{\min(-1;\,0)}$	$B_{\max(-1;\,0)}$	$B_{x\min(-7;\,-6)}$	$V_{sw\max(-1;\,0)}$
$\sin(2\pi T/24)$	$nProt_{\max(-1;\,0)}$	$B_{z\min(-6;\,-5)}$	$B_{\max(-5;\,-4)}$
$\cos(2\pi D/365)$	$B_{\max(-2;\,-1)}$	$B_{y\min(-1;\,0)}$	$V_{sw\text{avg}(-1;\,0)}$
$B_{\min(-1;\,0)}$	$nProt_{\min(-1;\,0)}$	$B_{x\min(-1;\,0)}$	$B_{\text{avg}(-4;\,-3)}$
$B_{z\text{avg}(-1;\,0)}$	$\cos(2\pi D/365)$	$B_{z\max(-9;\,-8)}$	$V_{sw\max(-2;\,-1)}$
$B_{\min(-8;\,-7)}$	$B_{z\text{avg}(-1;\,0)}$	$B_{z\min(-2;\,-1)}$	$B_{\max(-6;\,-5)}$
$B_{\min(-9;\,-8)}$	$V_{sw\text{avg}(-1;\,0)}$	$V_{sw\max(-7;\,-6)}$	$B_{\text{avg}(-5;\,-4)}$
$B_{z\text{avg}(-5;\,-4)}$	$nProt_{\text{avg}(-1;\,0)}$	$B_{\max(-5;\,-4)}$	$V_{sw\text{avg}(-2;\,-1)}$

Table 5.15: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 9$ hours ahead with a 3-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

$h = 9$ hours ahead			
FFX (3H)	**RF (3H)**	**MRMR (3H)**	**MIM (3H)**
$B_{\text{avg}(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$	$B_{\max(-3;\,0)}$	$B_{\max(-3;\,0)}$
$B_{\min(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$V_{sw\max(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$
$B_{\min(-9;\,-6)}$	$B_{z\min(-3;\,0)}$	$\cos(2\pi D/365)$	$B_{\max(-6;\,-3)}$
$B_{y\max(-3;\,0)}$	$V_{sw\text{avg}(-9;\,-6)}$	$B_{z\min(-3;\,0)}$	$V_{sw\max(-3;\,0)}$
$B_{z\text{avg}(-3;\,0)}$	$V_{sw\text{avg}(-3;\,0)}$	$\sin(2\pi T/24)$	$B_{\text{avg}(-6;\,-3)}$

$B_{z\text{avg}(-6;\,-3)}$	$V_{sw\max(-3;\,0)}$	$B_{z\min(-9;\,-6)}$	$B_{z\min(-3;\,0)}$
$B_{z\text{avg}(-9;\,-6)}$	$V_{sw\max(-6;\,-3)}$	$B_{x\min(-3;\,0)}$	$V_{sw\text{avg}(-3;\,0)}$
$B_{z\min(-3;\,0)}$	$V_{sw\min(-3;\,0)}$	$B_{z\max(-9;\,-6)}$	$B_{\max(-9;\,-6)}$
$B_{z\min(-9;\,-6)}$	$V_{sw\min(-6;\,-3)}$	$B_{y_{\min(-3;\,0)}}$	$B_{\text{avg}(-9;\,-6)}$
$V_{sw\max(-3;\,0)}$	$V_{sw\min(-9;\,-6)}$	$B_{z\max(-3;\,0)}$	$V_{sw\min(-3;\,0)}$
$V_{sw\min(-9;\,-6)}$	$nProt_{\max(-3;\,0)}$	$B_{z\min(-6;\,-3)}$	$V_{sw\max(-6;\,-3)}$
$nProt_{\max(-3;\,0)}$	$nProt_{\text{avg}(-3;\,0)}$	$B_{x\min(-9;\,-6)}$	$V_{sw\text{avg}(-6;\,-3)}$
$nProt_{\min(-3;\,0)}$	$nProt_{\min(-3;\,0)}$	$nProt_{\min(-3;\,0)}$	$B_{z\min(-6;\,-3)}$
$\sin(2\pi T/24)$	$\sin(2\pi D/365)$	$B_{x\max(-9;\,-6)}$	$V_{sw\max(-9;\,-6)}$
$\cos(2\pi D/365)$	$\cos(2\pi D/365)$	$B_{y_{\max(-3;\,0)}}$	$V_{sw\min(-6;\,-3)}$
$B_{\max(-3;\,0)}$	$B_{\max(-3;\,0)}$	$V_{sw\max(-9;\,-6)}$	$B_{\min(-3;\,0)}$

Table 5.16: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 12$ hours ahead with a 1-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

h = 12 hours ahead

FFX (1H)	**RF (1H)**	**MRMR (1H)**	**MIM (1H)**
$B_{\text{avg}(-1;\,0)}$	$B_{\text{avg}(-1;\,0)}$	$B_{\max(-1;\,0)}$	$B_{\max(-1;\,0)}$
$B_{z\min(-1;\,0)}$	$B_{z\min(-1;\,0)}$	$V_{sw\max(-2;\,-1)}$	$B_{\max(-2;\,-1)}$
$B_{z\min(-4;\,-3)}$	$V_{sw\max(-1;\,0)}$	$\cos(2\pi D/365)$	$B_{\text{avg}(-1;\,0)}$
$B_{z\min(-9;\,-8)}$	$V_{sw\max(-2;\,-1)}$	$B_{z\min(-9;\,-8)}$	$B_{\max(-3;\,-2)}$
$V_{sw\max(-1;\,0)}$	$V_{sw\min(-1;\,0)}$	$\sin(2\pi T/24)$	$B_{\text{avg}(-2;\,-1)}$
$B_{\max(-1;\,0)}$	$B_{\max(-1;\,0)}$	$B_{z\min(-1;\,0)}$	$B_{\text{avg}(-3;\,-2)}$
$nProt_{\min(-1;\,0)}$	$B_{\max(-2;\,-1)}$	$B_{x\text{avg}(-9;\,-8)}$	$B_{\max(-4;\,-3)}$
$\sin(2\pi T/24)$	$nProt_{\min(-1;\,0)}$	$B_{z\min(-5;\,-4)}$	$B_{\max(-5;\,-4)}$
$\cos(2\pi D/365)$	$\sin(2\pi D/365)$	$B_{x\min(-1;\,0)}$	$B_{\text{avg}(-4;\,-3)}$
$B_{\min(-1;\,0)}$	$\cos(2\pi D/365)$	$B_{z\min(-3;\,-2)}$	$V_{sw\max(-1;\,0)}$
$B_{z\text{avg}(-5;\,-4)}$	$V_{sw\text{avg}(-1;\,0)}$	$B_{z\max(-9;\,-8)}$	$B_{\max(-6;\,-5)}$

Table 5.17: Features selected using FFX, RF, MRMR, and MIM feature selection algorithms for the prediction horizon $h = 12$ hours ahead with a 3-hour time window used to construct input features. RF, MRMR, and MIM provide the ordered list of variables, with the most important variable at the top of the list. FFX does not provide the feature importance ranking.

$h = 12$ hours ahead			
FFX (3H)	**RF (3H)**	**MRMR (3H)**	**MIM (3H)**
$B_{\text{avg}(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$	$B_{\text{max}(-3;\,0)}$	$B_{\text{max}(-3;\,0)}$
$B_{\text{min}(-3;\,0)}$	$B_{z\text{avg}(-3;\,0)}$	$V_{sw\text{min}(-3;\,0)}$	$B_{\text{avg}(-3;\,0)}$
$B_{\text{min}(-9;\,-6)}$	$B_{z\text{min}(-3;\,0)}$	$\cos(2\pi D/365)$	$B_{\text{max}(-6;\,-3)}$
$B_{y\text{max}(-3;\,0)}$	$B_{z\text{min}(-6;\,-3)}$	$\sin(2\pi T/24)$	$B_{\text{avg}(-6;\,-3)}$
$B_{z\text{avg}(-3;\,0)}$	$V_{sw\text{avg}(-3;\,0)}$	$B_{z\text{min}(-3;\,0)}$	$V_{sw\text{max}(-3;\,0)}$
$B_{z\text{avg}(-6;\,-3)}$	$V_{sw\text{max}(-3;\,0)}$	$B_{x\text{min}(-9;\,-6)}$	$B_{\text{max}(-9;\,-6)}$
$B_{z\text{avg}(-9;\,-6)}$	$V_{sw\text{max}(-6;\,-3)}$	$B_{z\text{min}(-9;\,-6)}$	$B_{\text{avg}(-9;\,-6)}$
$B_{z\text{min}(-3;\,0)}$	$V_{sw\text{min}(-3;\,0)}$	$B_{z\text{max}(-3;\,0)}$	$V_{sw\text{avg}(-3;\,0)}$
$B_{z\text{min}(-9;\,-6)}$	$V_{sw\text{min}(-6;\,-3)}$	$B_{x\text{max}(-3;\,0)}$	$B_{z\text{min}(-3;\,0)}$
$V_{sw\text{max}(-3;\,0)}$	$V_{sw\text{min}(-9;\,-6)}$	$B_{z\text{max}(-9;\,-6)}$	$V_{sw\text{max}(-6;\,-3)}$
$V_{sw\text{min}(-9;\,-6)}$	$nProt_{\text{max}(-3;\,0)}$	$nProt_{\text{min}(-3;\,0)}$	$V_{sw\text{min}(-3;\,0)}$
$nProt_{\text{max}(-3;\,0)}$	$nProt_{\text{avg}(-3;\,0)}$	$B_{y\text{max}(-3;\,0)}$	$V_{sw\text{avg}(-6;\,-3)}$
$nProt_{\text{min}(-3;\,0)}$	$nProt_{\text{min}(-3;\,0)}$	$B_{z\text{min}(-6;\,-3)}$	$B_{z\text{min}(-6;\,-3)}$
$\sin(2\pi T/24)$	$\sin(2\pi D/365)$	$B_{y\text{min}(-9;\,-6)}$	$B_{\text{min}(-3;\,0)}$
$\cos(2\pi D/365)$	$\cos(2\pi D/365)$	$B_{x\text{min}(-3;\,0)}$	$V_{sw\text{max}(-9;\,-6)}$
$B_{\text{max}(-3;\,0)}$	$B_{\text{max}(-3;\,0)}$	$V_{sw\text{max}(-3;\,0)}$	$V_{sw\text{min}(-6;\,-3)}$

5.C Fit performance statistics

Table 5.18: Fit performance statistics of the NN-FFX models for different prediction horizons computed on the training set.

Prediction horizon, hours	**0**	**3**	**6**	**9**	**12**
Number of values in comparison	42479	42479	42479	42479	42479
Intercept of the linear fit	0.2597	0.4363	0.7578	0.9467	1.0866
Slope of the linear fit	0.8743	0.7671	0.5935	0.4951	0.4225
Pearson correlation coefficient (R)	0.9376	0.8758	0.7696	0.7048	0.6493
Root mean square error (RMSE)	0.4822	0.6668	0.8827	0.9817	1.0511
Mean absolute error (MAE)	0.3758	0.5062	0.6714	0.7473	0.8007
Mean error (ME, or bias)	0.0238	0.0020	-9.0616e-05	0.0040	0.0126
Prediction efficiency (PE)	0.8787	0.7670	0.5923	0.4967	0.4216

Table 5.19: Fit performance statistics of the NN-FFX models for different prediction horizons computed on the validation set.

Prediction horizon, hours	**0**	**3**	**6**	**9**	**12**
Number of values in comparison	10080	10080	10080	10080	10080
Intercept of the linear fit	0.2780	0.4870	0.8496	1.0579	1.2035
Slope of the linear fit	0.8618	0.7503	0.5670	0.4659	0.3872
Pearson correlation coefficient (R)	0.9303	0.8666	0.7518	0.6758	0.6105
Root mean square error (RMSE)	0.4987	0.6792	0.8936	0.9990	1.0755
Mean absolute error (MAE)	0.3853	0.5150	0.6835	0.7596	0.8197
Mean error (ME, or bias)	0.0062	-0.0012	0.0054	0.0180	0.0141
Prediction efficiency (PE)	0.8654	0.7509	0.5652	0.4563	0.3721

Table 5.20: Fit performance statistics of the NN-FFX models for different prediction horizons computed on the test set.

Prediction horizon, hours	**0**	**3**	**6**	**9**	**12**
Number of values in comparison	5880	5880	5880	5880	5880
Intercept of the linear fit	0.2570	0.4359	0.7569	0.9542	1.1052
Slope of the linear fit	0.8653	0.7642	0.5884	0.4836	0.4061
Pearson correlation coefficient (R)	0.9350	0.8745	0.7695	0.7056	0.6477
Root mean square error (RMSE)	0.4964	0.6754	0.8898	0.9909	1.0632
Mean absolute error (MAE)	0.3872	0.5101	0.6710	0.7480	0.8029
Mean error (ME, or bias)	-0.0059	-0.0199	-0.0387	-0.0436	-0.0374
Prediction efficiency (PE)	0.8740	0.7646	0.5914	0.4965	0.4184

Acknowledgments

We acknowledge support by the European Unions Horizon 2020 research and innovation program under grant agreement No 776287 SWAMI. IZ was supported by Geo.X, the Research Network for Geosciences in Berlin and Potsdam, under Grant No SO_087_GeoX. Solar wind data and geomagnetic indices were obtained from `http://omniweb.gsfc.nasa.gov/form/dx1.html`. Kp index of geomagnetic activity was obtained from the GFZ Potsdam website (`https://www.gfz-potsdam.de/en/kp-index/`). IZ thanks Nikita Aseev, Alexander Drozdov, Hayley Allison, and Dominika Boneberg for helpful discussions.

Chapter 6

Summary and future work

6.1 Summary

In this dissertation, we demonstrated how one can utilize sparse, single-point satellite measurements to reconstruct the global space environment by using neural networks. We showed that neural networks are an efficient technique for building complex models based on sparse space observations. This research began with deriving electron density from satellite measurements along the satellite orbit, continued to reconstructing the global plasmasphere environment from sparse density measurements employing solar wind, geomagnetic indices and the location as inputs, and concluded with improving the performance of the developed neural network for periods of enhanced geomagnetic activity by combining it with the physics-based model of plasma density. In addition to that, the developed machine learning-based tools were extended and applied to predict the geomagnetic Kp index. The summary and main findings of this research are presented below.

We applied feedforward neural networks to automatically derive the upper-hybrid resonance frequency from the electric field measurements made with the EMFISIS instrument onboard the Van Allen Probes spacecraft (Zhelavskaya et al., 2018). The derived upper-hybrid frequency was then used to calculate the electron density. The neural network was trained on the 2.5-year plasma density database developed using another semi-automated routine AURA (Kurth et al., 2015). We demonstrated that the developed Neural-network-based Upper hybrid Resonance Determination (NURD) algorithm was in a good agreement with AURA. Specifically, the densities derived by NURD deviated from the ones derived by AURA by 1-5% for the

cases when the upper-hybrid band was relatively easy to identify or partially obscured (which comprised 90% of all the data), and by 14% for the difficult cases. The NURD algorithm was further applied to the available database of electric and magnetic field measurements from October 2012 until July 2016. Via a statistical analysis, we demonstrated that the resulting electron density was in a good agreement with the empirical plasmasphere and trough density models by Sheeley et al. (2001), with only a slight shift of the mean density values, which may be due to the different part of the solar cycle used for constructing the Sheeley et al. (2001) model. The resulting database of electron densities was made publicly available (Zhelavskaya et al., 2020b).

Using this large database of electron density measurements along the orbit of the Van Allen Probes satellites, we developed a new empirical model of the plasma density, the Plasma density in the Inner magnetosphere Neural network-based Empirical (PINE) model (Zhelavskaya et al., 2017). We used a similar approach based on feed-forward neural networks, and in this case the inputs to the neural network were the time histories of solar wind parameters and geomagnetic indices, and the location. The neural networks were able to learn the global distribution of cold plasma from sparse density measurements and the developed model can be applied to reconstruct the cold plasma density distribution globally at any time, when the inputs are known. Furthermore, we assessed the influence of different drivers on the performance of the model, in particular the geomagnetic and solar wind parameters and their time history. To determine the critical combination of activity parameters, we trained several neural networks with different combinations of input parameters and compared them to each other. The performance of the networks was estimated both quantitatively, by means of cross validation on the data from Van Allen Probes, and qualitatively, by comparing the modeled plasmapause locations with the ones observed with the IMAGE EUV instrument. We found that using geomagnetic indices alone is sufficient to model the plasmasphere dynamics accurately (with a correlation coefficient of $\sim$0.95 and an RMSE of 0.30 on the testing data set). The inclusion of solar wind confused the model and caused overfitting, while using solar wind parameters for modeling alone did not provide a sufficient model accuracy. The critical duration of the time history of parameters, after which no significant improvement is observed, was found to be 48 hours. The model is capable of globally reconstructing the large-scale plasmasphere dynamics and successfully captures plume formation and evolution and the

storm-time erosion of the plasmasphere.

As a next step, we explored the performance of the developed neural network-based model during extreme geomagnetic storms. The performance of neural networks is known to strongly depend on the availability of training data. In our case, the data during strong geomagnetic storms are very limited, and consequently the performance of neural networks is also reduced for such events. In order to mitigate this issue, we employed a physics-based modeling approach during intervals of enhanced geomagnetic activity. We then developed an approach to optimally combine the empirical neural network and physics-based models of the plasmasphere electron density by means of data assimilation (Zhelavskaya et al., 2020a). Data assimilation is typically used to combine sparse observations and a background model, but in this study we used this approach for the first time to combine two models. In particular, we used the Kalman filter technique to optimally blend the developed neural network model and the physics-based VERB-CS code (Aseev and Shprits, 2019), adjusted to model the plasmasphere dynamics. We evaluated the performance of all the models by (1) comparing the model-predicted global evolution of plasma density to the global images of the He^+ distribution from the IMAGE EUV for a number of events in the past including the 2003 Halloween storm, and (2) by comparing the output of the models for a 18-month out-of-sample interval to the in-situ density obtained with the RBSP-A spacecraft and computing performance metrics. We demonstrated that the new assimilative model is capable of reproducing the dynamics of the plasmasphere well during both quiet and disturbed geomagnetic activity, including extreme geomagnetic events. Its quantitative performance is better than that of VERB-CS and is comparable to the neural network performance for the dayside local time sector.

Finally, we applied the developed neural network-based modeling tools to another important problem in the field of space weather, the prediction of the geomagnetic index Kp (Zhelavskaya et al., 2019). Previous studies used various machine learning methods and algorithms to predict the Kp index, basing their inferences on the recent history of Kp and on solar wind measurements at L1. Our goal was to understand what brings the most improvement to the model accuracy (i.e., a machine learning method, on which a model is based, or a method for constructing inputs to the model) and whether there is a limit to the prediction accuracy set by using solar wind measurements at L1 as input to a model. To answer those questions, we systematically tested a number of machine learning algorithms, namely Gradient Boosting,

Feedforward Neural Networks, and Linear Regression. We used the same validation technique and the same time intervals to train and validate the models, and in doing so created an unbiased technique to validate and compare models. We have found that models constructed using nonlinear methods (neural networks and gradient boosting) outperform models constructed using linear regression for the short-term prediction horizons, which implies the existence of a nonlinear component in the Kp prediction problem for short-term predictions. We also found that the performance of all considered methods decreases as the prediction horizon increases, which may indicate that the information in the solar wind measurements at L1 is not sufficient to produce accurate long-term predictions (e.g., $>$ 3 hours), especially for high Kp values, and is not sufficient to accurately capture the arrival time of geomagnetic storms for the long-term prediction horizons. Additionally, we investigated different methods of machine learning and information theory for optimal input selection, and used these methods to identify the most important inputs to a predictive model of Kp. Models based on the selected inputs, using the proposed feature selection method, showed an improvement in performance compared to models based on the full input set. The obtained optimal input variables can be used to gain an understanding of what inputs are the most important and physically meaningful to predict the Kp index. The feature selection methods described in this work can also be applied to other problems in space physics in order to significantly reduce the input dimensionality and identify the most important inputs that contain sufficient information to produce accurate predictions.

6.2 Future work

The results of this dissertation are planned to be extended in the future. In this section, we outline directions of the planned future work.

The developed neural network model of plasma density PINE is based on the electron density measurements from the Van Allen Probes mission. The training data set for this model covered $\sim$4 years of data, which did not coincide with substantial geomagnetic activity. The performance of the PINE model can be further improved by enlarging the training data set and accounting for extreme geomagnetic activity. We will expand the training data set by using in-situ density measurements from other missions, such as IMAGE (using the RPI instrument), CRRES (Anderson et al.,

1992), Cluster (Escoubet et al., 2001), ISEE-1 (Gurnett et al., 1978), and Polar (Gurnett et al., 1995), and inter-calibrate them. This will allow us to include more storm-time events into the training data set and improve the model performance for such intervals. We will also investigate different machine learning techniques for unbalanced data sets, such as rebalancing or weighting of the cost function.

In addition to that, we will use the output of the developed assimilative model during the past storm-time intervals as a part of the training data set. This will allow increasing the number of storm-time cases in the training data set and eventually producing a balanced training data set, i.e., with a uniform coverage in terms of geomagnetic activity. Weights can be assigned to these additional surrogate data, indicating a lower trust level than for the actual in-situ observations. We will test if the addition of such surrogate data aids the model performance or not, and which of the methods mentioned above produce the optimal results.

We will test different and more realistic electric field (e.g., Matsui et al. (2013)), refilling, and saturation density models in order to improve the physics-based VERB-CS model of plasma density. More work will be done regarding the selection of model and observation error in the Kalman filter setup of the assimilative model. The assimilative model can be further extended by assimilating in-situ density measurements in the model (e.g., from Van Allen Probes, IMAGE RPI, or other sources, depending on the time period), in addition to the output of the neural network model PINE.

Furthermore, we will extend the neural network PINE model to lower altitudes by including electron density measurements from those altitudes. This will allow us to model the plasma density starting from the ionospheric heights all the way up to the plasmasphere. In particular, we will employ the in-situ density measurements and ionospheric density reconstructions from the following missions: CHAllenging Minisatellite Payload (CHAMP), Swarm, Constellation Observing System for Meteorology, Ionosphere, and Climate (FormoSat-3/COSMIC), Gravity Recovery and Climate Experiment (GRACE), and MetOP. We will employ a methodology similar to the PINE modeling, based on feedforward neural networks. In addition, we will apply the feature selection techniques employed in this dissertation for the Kp prediction to select the most important drivers of the plasmasphere and ionosphere dynamics. We will investigate whether the electron density in the ionosphere and plasmasphere can be accurately modeled using a single model, or whether different models for different regions need to be developed. Similar methods may be applied

to model other regions, such as the ionosphere or ring current. They can also be applied to model maps of vertical total electron content (VTEC) used for applications in navigation.

The developed models of the plasmaspheric electron density can be used to improve existing and to develop more advanced parameterizations of the diffusion coefficients used for radiation belt modeling. Currently, diffusion coefficients are parameterized using statistical density models, such as Sheeley et al. (2001) or Carpenter and Anderson (1992) (e.g., Orlova et al., 2012; Ni et al., 2008). These models do not account for the asymmetric plasmapause and plumes in the plasmasphere, which is important for separating the regions of hiss and chorus waves. Usage of more advanced models, such as PINE or the assimilative model, will allow us to account for these effects and explore how the plasmasphere dynamics influence the radiation belts. It will be especially interesting to analyse the influence of plasma density on the radiation belt dynamics during extreme events, such as the Halloween storm in 2003.

Finally, we will apply the developed plasma density models to predict the plasmasphere dynamics in real-time and complement the system for forecasting of the Earth's radiation belt dynamics operating at GFZ.

Bibliography

Adrian, M. L., Gallagher, D. L., and Avanov, L. A. (2004). Image euv observation of radially bifurcated plasmaspheric features: First observations of a possible standing ulf waveform in the inner magnetosphere. *Journal of Geophysical Research: Space Physics*, 109(A1).

Agapitov, O., Artemyev, A., Mourenas, D., Mozer, F., and Krasnoselskikh, V. (2015). Empirical model of lower band chorus wave distribution in the outer radiation belt. *Journal of Geophysical Research: Space Physics*, 120(12):10–425.

Akasofu, S.-I. (1964). The development of the auroral substorm. *Planetary and Space Science*, 12(4):273–282.

Ali, R., Siddiqi, M. H., and Lee, S. (2015). Rough set-based approaches for discretization: a compact review. *Artificial Intelligence Review*, 44(2):235–263.

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

Anderson, P., Carpenter, D., Tsuruda, K., Mukai, T., and Rich, F. (2001). Multisatellite observations of rapid subauroral ion drifts (said). *Journal of Geophysical Research: Space Physics*, 106(A12):29585–29599.

Anderson, R. R., Gurnett, D. A., and Odem, D. L. (1992). CRRES plasma wave experiment. *Journal of Spacecraft and Rockets*, 29(4):570–573.

André, N., M., Li, W., Toledo-Redondo, S., Khotyaintsev, Y. V., Vaivads, A., Graham, D. B., Norgren, C., Burch, J., Lindqvist, P.-A., Marklund, G., Ergun, R., Torbert, R., Magnes, W., Russell, C. T., Giles, B., Moore, T. E., Chandler, M. O., Pollock, C., Young, D. T., Avanov, L. A., Dorelli, J. C., Gershman, D. J., Paterson, W. R., Lavraud, B., and Saito, Y. (2016). Magnetic reconnection and modification

of the Hall physics due to cold ions at the magnetopause. *Geophysical Research Letters*, 43(13):6705–6712. 2016GL069665.

Angelopoulos, V. (2009). The THEMIS mission. In *The THEMIS mission*, pages 5–34. Springer.

Angerami, J. and Carpenter, D. L. (1966). Whistler studies of the plasmapause in the magnetosphere: 2. electron density and total tube electron content near the knee in magnetospheric ionization. *Journal of Geophysical Research*, 71(3):711–725.

Appleton, E. V. and Barnett, M. (1925). On some direct evidence for downward atmospheric reflection of electric rays. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 109(752):621–641.

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

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

Axford, W. I. and Hines, C. O. (1961). A unifying theory of high-latitude geophysical phenomena and geomagnetic storms. *Canadian Journal of Physics*, 39(10):1433–1464.

Baker, D., Kanekal, S., Li, X., Monk, S., Goldstein, J., and 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.

Bala, R. and Reiff, P. (2012). Improvements in short-term forecasting of geomagnetic activity. *Space Weather*, 10(6).

Balikhin, M., Boaghe, O., Billings, S., and Alleyne, H. S. C. (2001). Terrestrial magnetosphere as a nonlinear resonator. *Geophysical research letters*, 28(6):1123–1126.

Bartels, J. (1949). The standardized index, Ks, and the planetary index, Kp. *IATME Bull. 12b*, 97(2010):2021.

Bartels, J., Heck, N., and Johnston, H. (1939). Main literature on the magnetic three-hour-range index (kennziffer). *Terr. Mag.*, 44:411–454.

Bauer, S. J. (2012). *Physics of planetary ionospheres*, volume 6. Springer Science & Business Media.

Baumjohann, W. (1993). The near-Earth plasma sheet: An AMPTE/IRM perspective. *Space science reviews*, 64(1-2):141–163.

Baumjohann, W. and Treumann, R. A. (1997). *Basic space plasma physics*. World Scientific.

Beghin, C., Rauch, J., and Bosqued, J. (1989). Electrostatic plasma waves and HF auroral hiss generated at low altitude. *Journal of Geophysical Research: Space Physics*, 94(A2):1359–1378.

Benson, R., Webb, P., Green, J., Garcia, L., and Reinisch, B. (2004). Magnetospheric electron densities inferred from upper-hybrid band emissions. *Geophysical research letters*, 31(20).

Benson, R. F., Osherovich, V. A., Fainberg, J., F-Vinas, A., and Ruppert, D. (2001). An interpretation of banded magnetospheric radio emissions. *Journal of Geophysical Research: Space Physics*, 106(A7):13179–13190.

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

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

Bilitza, D. and Reinisch, B. W. (2008). International reference ionosphere 2007: improvements and new parameters. *Advances in space research*, 42(4):599–609.

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

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

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1987). Occam's razor. *Information processing letters*, 24(6):377–380.

Boaghe, O., Balikhin, M., Billings, S., and Alleyne, H. (2001). Identification of nonlinear processes in the magnetospheric dynamics and forecasting of dst index. *Journal of Geophysical Research: Space Physics*, 106(A12):30047–30066.

Boberg, F., Wintoft, P., and Lundstedt, H. (2000). Real time kp predictions from solar wind data using neural networks. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, 25(4):275–280.

Bollacker, K. D. and Ghosh, J. (1996). Linear feature extractors based on mutual information. In *Proceedings of 13th International Conference on Pattern Recognition*, volume 2, pages 720–724. IEEE.

Borovsky, J. E. and Denton, M. H. (2006). Effect of plasmaspheric drainage plumes on solar-wind/magnetosphere coupling. *Geophysical Research Letters*, 33(20):n/a–n/a. L20101.

Borovsky, J. E., Welling, D. T., Thomsen, M. F., and Denton, M. H. (2014). Long-lived plasmaspheric drainage plumes: Where does the plasma come from? *Journal of Geophysical Research: Space Physics*, 119(8):6496–6520.

Bortnik, J., Li, W., Thorne, R. M., and Angelopoulos, V. (2016). A unified approach to inner magnetospheric state prediction. *Journal of Geophysical Research: Space Physics*, 121(3):2423–2430. 2015JA021733.

Brautigam, D. and Albert, J. (2000). Radial diffusion analysis of outer radiation belt electrons during the october 9, 1990, magnetic storm. *Journal of Geophysical Research: Space Physics*, 105(A1):291–309.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees. wadsworth int. *Group*, 37(15):237–251.

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

Brice, N. M. (1967). Bulk motion of the magnetosphere. *Journal of Geophysical Research*, 72(21):5193–5211.

Bruinsma, S., Sutton, E., Solomon, S., Fuller-Rowell, T., and Fedrizzi, M. (2018). Space weather modeling capabilities assessment: Neutral density for orbit determination at low earth orbit. *Space Weather*, 16(11):1806–1816.

Burch, J. (2000). IMAGE mission overview. *Space Science Reviews*, 91(1-2):1–14.

Burch, J., Mende, S., Mitchell, D., Moore, T., Pollock, C., Reinisch, B., Sandel, B., Fuselier, S., Gallagher, D., Green, J., et al. (2001a). Views of earth's magnetosphere with the image satellite. *Science*, 291(5504):619–624.

Burch, J. L., Goldstein, J., and Sandel, B. R. (2004). Cause of plasmasphere corotation lag. *Geophysical Research Letters*, 31(5).

Burch, J. L., Mitchell, D. G., Sandel, B. R., Brandt, P. C., and West, M. (2001b). Global dynamics of the plasmasphere and ring current during magnetic storms. *Geophysical Research Letters*, 28(6):1159–1162.

Burke, W., Maynard, N., Hagan, M., Wolf, R., Wilson, G., Gentile, L., Gussenhoven, M., Huang, C., Garner, T., and Rich, F. (1998). Electrodynamics of the inner magnetosphere observed in the dusk sector by crres and dmsp during the magnetic storm of june 4–6, 1991. *Journal of Geophysical Research: Space Physics*, 103(A12):29399–29418.

Burton, R., McPherron, R., and Russell, C. (1975). The terrestrial magnetosphere: A half-wave rectifier of the interplanetary electric field. *Science*, 189(4204):717–718.

Canuto, V., Axford, W. I., and Gordon, C. W. (1978). *The Earth, 1: The Upper Atmosphere, Ionosphere, and Magnetosphere*, volume 1. CRC Press.

Carpenter, D. and Smith, A. (2001). The study of bulk plasma motions and associated electric fields in the plasmasphere by means of whistler-mode signals. *Journal of Atmospheric and Solar-Terrestrial Physics*, 63(11):1117–1132.

Carpenter, D. L. (1962). The magnetosphere during magnetic storms: A whistler analysis.

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

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

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

Carpenter, D. L., Anderson, R., Calvert, W., and Moldwin, M. (2000). Crres observations of density cavities inside the plasmasphere. *Journal of Geophysical Research: Space Physics*, 105(A10):23323–23338.

Carpenter, D. L. and Anderson, R. R. (1992). An ISEE/whistler model of equatorial electron density in the magnetosphere. *Journal of Geophysical Research: Space Physics*, 97(A2):1097–1108.

Carpenter, D. L., Giles, B. L., Chappell, C. R., Dcrau, P. M. E., Anderson, R. R., Persoon, A. M., Smith, A. J., Corcuff, Y., and Canu, P. (1993). Plasmasphere dynamics in the duskside bulge region: A new look at an old topic. *Journal of Geophysical Research: Space Physics*, 98(A11):19243–19271.

Carpenter, D. L. and Smith, R. (1964). Whistler measurements of electron density in the magnetosphere. *Reviews of Geophysics*, 2(3):415–441.

Carpenter, D. L., Spasojevic, M. A., Bell, T. F., Inan, U. S., Reinisch, B. W., Galkin, I. A., Benson, R. F., Green, J. L., Fung, S. F., and Boardsen, S. A. (2002). Small-scale field-aligned plasmaspheric density structures inferred from the radio plasma imager on image. *Journal of Geophysical Research: Space Physics*, 107(A9):SMP 22–1–SMP 22–19.

Carpenter, D. L. and 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.

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

Chamberlain, J. W. (1961). Spectral photometry of the nightglow. *Physics of the Aurora and Airglow, Academic Press Inc., New York & London.*

Chapman, S. and Bartels, J. (1940). Geomagnetism, vol. ii: Analysis of the data, and physical theories. *Geomagnetism, Vol. II: Analysis of the Data, and Physical Theories, by S. Chapman and J. Bartels. London: Oxford Univ. Press, 1940*, 2.

Chapman, S. and Ferraro, V. C. (1931). A new theory of magnetic storms. *Terrestrial Magnetism and Atmospheric Electricity*, 36(2):77–97.

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

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

Chappell, C., Harris, K., and Sharp, G. (1971). The dayside of the plasmasphere. *Journal of Geophysical Research*, 76(31):7632–7647.

Chen, A. and Grebowsky, J. (1974). Plasma tail interpretations of pronounced detached plasma regions measured by ogo 5. *Journal of Geophysical Research*, 79(25):3851–3855.

Chen, A. and Wolf, R. (1972). Effects on the plasmasphere of a time-varying convection electric field. *Planetary and Space Science*, 20(4):483–509.

Chen, S.-H. and Moore, T. E. (2006). Magnetospheric convection and thermal ions in the dayside outer magnetosphere. *Journal of Geophysical Research: Space Physics*, 111(A3):n/a–n/a. A03215.

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

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

Cireşan, D., Meier, U., Masci, J., and Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32:333–338.

Cohen, E. R. (2007). *Quantities, units and symbols in physical chemistry*. Royal Society of Chemistry.

Comfort, R. (1986). Plasmasphere thermal structure as measured by isee-1 and de-1. *Advances in space research*, 6(3):31–40.

Comfort, R., Newberry, I., and Chappell, C. (1988). Preliminary statistical survey of plasmaspheric ion properties from observations by de 1/rims. *Modeling Magnetospheric Plasma, Geophys. Monogr. Ser*, 44:107–114.

Cortes, C., Jackel, L. D., Solla, S. A., Vapnik, V., and Denker, J. S. (1994). Learning curves: Asymptotic values and rate of convergence. *Advances in Neural Information Processing Systems*, pages 327–327.

Costello, K. A. (1998). *Moving the Rice MSFM into a real-time forecast mode using solar wind driven forecast modules*. PhD thesis, Rice University.

Cowley, S. W. (1996). A beginner's guide to the earth's magnetosphere. *Earth in Space*, 8:9–13.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

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

Daglis, I. A., Thorne, R. M., Baumjohann, W., and Orsini, S. (1999). The terrestrial ring current: Origin, formation, and decay. *Reviews of Geophysics*, 37(4):407–438.

Darrouzet, F., Gallagher, D. L., André, N., Carpenter, D. L., Dandouras, I., Décréau, P. M., De Keyser, J., Denton, R. E., Foster, J. C., Goldstein, J., et al. (2009). Plasmaspheric density structures and dynamics: Properties observed by the CLUSTER and IMAGE missions. *Space science reviews*, 145(1-2):55–106.

Davis, T. N. and Sugiura, M. (1966). Auroral electrojet activity index ae and its universal time variations. *Journal of Geophysical Research*, 71(3):785–801.

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

Denton, M., Henderson, M. G., Jordanova, V. K., Thomsen, M. F., Borovsky, J. E., Woodroffe, J., Hartley, D., and Pitchford, D. (2016). An improved empirical model of electron and ion fluxes at geosynchronous orbit based on upstream solar wind conditions. *Space Weather*, 14(7):511–523.

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

Dessler, A. J. and Parker, E. N. (1959). Hydromagnetic theory of geomagnetic storms. *Journal of Geophysical Research*, 64(12):2239–2252.

Devijver, P. A. and Kittler, J. (1982). *Pattern recognition: A statistical approach*. Prentice hall.

Ding, C. and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205.

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

Dungey, J. (1954). *Electrodynamics of the outer atmosphere: report to National Science Foundation on work carried on under grant NSF-G450*, volume 69. Pennsylvania State University, Ionosphere Research Laboratory.

Dungey, J. (1955). The physics of the ionosphere. *Nature*.

Dungey, J. (1963). The structure of the exosphere, or adventures in velocity space. *Geophysics, The Earth's Environment*.

Dungey, J. W. (1961). Interplanetary magnetic field and the auroral zones. *Physical Review Letters*, 6(2):47.

Emery, B. A., Coumans, V., Evans, D. S., Germany, G. A., Greer, M. S., Holeman, E., Kadinsky-Cade, K., Rich, F. J., and Xu, W. (2008). Seasonal, kp, solar wind, and solar flux variations in long-term single-pass satellite estimates of electron and ion auroral hemispheric power. *Journal of Geophysical Research: Space Physics*, 113(A6).

Escoubet, C., Fehringer, M., and Goldstein, M. (2001). Introduction the cluster mission.

Escoubet, C. P., Pedersen, A., Schmidt, R., and Lindqvist, P. A. (1997). Density in the magnetosphere inferred from isee 1 spacecraft potential. *Journal of Geophysical Research: Space Physics*, 102(A8):17595–17609.

Fennell, J. F., Claudepierre, S. G., Blake, J. B., O'Brien, T. P., Clemmons, J. H., Baker, D. N., Spence, H. E., and Reeves, G. D. (2015). Van allen probes show that the inner radiation zone contains no mev electrons: Ect/mageis data. *Geophysical Research Letters*, 42(5):1283–1289.

Foster, J. and Burke, W. (2002). SAPS: A new categorization for sub-auroral electric fields. *Eos, Transactions American Geophysical Union*, 83(36):393–394.

Foster, J. and Vo, H. (2002). Average characteristics and activity dependence of the subauroral polarization stream. *Journal of Geophysical Research: Space Physics*, 107(A12):SIA–16.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Gallagher, D., Adrian, M., and Liemohn, M. (2005). Origin and evolution of deep plasmaspheric notches. *Journal of Geophysical Research: Space Physics*, 110(A9).

Gallagher, D. and Adrian, M. L. (2007). Two-dimensional drift velocities from the image euv plasmaspheric imager. *Journal of atmospheric and solar-terrestrial physics*, 69(3):341–350.

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

Gallagher, D. L. and Comfort, R. H. (2016). Unsolved problems in plasmasphere refilling. *Journal of Geophysical Research: Space Physics*, 121(2):1447–1451.

Gallagher, D. L., Craven, P. D., and Comfort, R. H. (2000). Global core plasma model. *Journal of Geophysical Research: Space Physics*, 105(A8):18819–18833.

Galvan, D. A., Moldwin, M. B., and Sandel, B. R. (2008). Diurnal variation in plasmaspheric he+ inferred from extreme ultraviolet images. *Journal of Geophysical Research: Space Physics*, 113(A9).

Gao, W., Kannan, S., Oh, S., and Viswanath, P. (2017). Estimating mutual information for discrete-continuous mixtures. In *Advances in Neural Information Processing Systems*, pages 5986–5997.

Garcia, L. N., Fung, S. F., Green, J. L., Boardsen, S. A., Sandel, B. R., and Reinisch, B. W. (2003). Observations of the latitudinal structure of plasmaspheric convection plumes by IMAGE-RPI and EUV. *Journal of Geophysical Research: Space Physics*, 108(A8).

Geiger, H. and Müller, W. (1928). Elektronenzählrohr zur messung schwächster aktivitäten. *Naturwissenschaften*, 16(31):617–618.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 513–520.

Gold, T. (1959). Motions in the magnetosphere of the earth. *Journal of Geophysical Research*, 64(9):1219–1224.

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

Goldstein, J., Burch, J., Sandel, B., Mende, S., Hairston, M., et al. (2005b). Coupled response of the inner magnetosphere and ionosphere on 17 April 2002. *Journal of Geophysical Research: Space Physics*, 110(A3).

Goldstein, J., Gallagher, D., Craven, P. D., Comfort, R. H., Genestreti, K. J., Mouikis, C., Spence, H., Kurth, W., Wygant, J., Skoug, R. M., et al. (2019). Temperature dependence of plasmaspheric ion composition. *Journal of Geophysical Research: Space Physics*, 124(8):6585–6595.

Goldstein, J., Kanekal, S., Baker, D., and Sandel, B. (2005c). Dynamic relationship between the outer radiation belt and the plasmapause during March–May 2001. *Geophysical research letters*, 32(15).

Goldstein, J., Pascuale, S. D., Kletzing, C., Kurth, W., Genestreti, K., Skoug, R., Larsen, B., Kistler, L. M., Mouikis, C., and Spence, H. (2014). Simulation of van allen probes plasmapause encounters. *Journal of Geophysical Research: Space Physics*, 119(9):7464–7484.

Goldstein, J. and Sandel, B. (2005). The global pattern of evolution of plasmaspheric drainage plumes. *GEOPHYSICAL MONOGRAPH-AMERICAN GEOPHYSICAL UNION*, 159:1.

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

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

Goldstein, J., Sandel, B., Thomsen, M., Spasojević, M., and Reiff, P. (2004). Simultaneous remote sensing and in situ observations of plasmaspheric drainage plumes. *Journal of Geophysical Research: Space Physics*, 109(A3).

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

Goldstein, J., Spiro, R., Reiff, P., Wolf, R., Sandel, B., Freeman, J., and Lambour, R. (2002). IMF-driven overshielding electric field and the origin of the plasmaspheric shoulder of May 24, 2000. *Geophysical research letters*, 29(16).

Goller, C. and Kuchler, A. (1996). Learning task-dependent distributed representations by backpropagation through structure. In *Neural Networks, 1996., IEEE International Conference on*, volume 1, pages 347–352. IEEE.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Goodwin, G. and Sin, K. S. (1984). Adaptive filtering prediction and control. prentice hall. *New Jersey*.

Gopalswamy, N. (2006). Highlights of the october-november 2003 extreme events. *Solar Extreme Events: Fundamental Science and Applied Aspects. Yerevan: Alikhanyan Physics Institute*, 24.

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

Grewal, M. and Andrews, A. (2010). How good is your gyro [ask the experts]. *IEEE Control Systems Magazine*, 30(1):12–86.

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

Gringauz, K., Kurt, V., Moroz, V., and Shklovskii, I. (1961). Results of observations of charged particles observed out to r= 100, 000 km, with the aid of charged-particle traps on soviet space rockets. *Soviet Astronomy*, 4:680.

Gurgiolo, C., Sandel, B., Perez, J., Mitchell, D., Pollock, C., and Larsen, B. (2005). Overlap of the plasmasphere and ring current: Relation to subauroral ionospheric heating. *Journal of Geophysical Research: Space Physics*, 110(A12).

Gurnett, D., Persoon, A., Randall, R., Odem, D., Remington, S., Averkamp, T., Debower, M., Hospodarsky, G., Huff, R., Kirchner, D., et al. (1995). The polar plasma wave instrument. *Space Science Reviews*, 71(1-4):597–622.

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

Hagan, M. T., Demuth, H. B., Beale, M. H., et al. (1996). Neural network design, PWS Pub. *Co., Boston*, 3632.

Hagan, M. T. and Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*, 5(6):989–993.

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

Hanson, W. and Patterson, T. (1963). Diurnal variation of the hydrogen concentration in the exosphere. *Planetary and Space Science*, 11(9):1035–1052.

Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. MIT press.

Haykin, S. (1994). Neural networks: A comprehensive foundation, 1MacMillan. *New York*, 19942:1.

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

Hebb, D. O. et al. (1949). The organization of behavior: A neuropsychological theory.

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

Heidrich-Meisner, V. and Wimmer-Schweingruber, R. F. (2018). Solar wind classification via k-means clustering algorithm. In *Machine learning techniques for space weather*, pages 397–424. Elsevier.

Helliwell, R. A. (1965). *Whistlers and related ionospheric phenomena*, volume 50. Stanford University Press Stanford, Calif.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE.

Hochreiter, S., Bengio, Y., Frasconi, P., and Schmidhuber, J. (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Horne, R. B. (2002). The contribution of wave-particle interactions to electron loss and acceleration in the earth's radiation belts during geomagnetic storms.

Horne, R. B., Thorne, R. M., Shprits, Y. Y., Meredith, N. P., Glauert, S. A., Smith, A. J., Kanekal, S. G., Baker, D. N., Engebretson, M. J., Posch, J. L., et al. (2005). Wave acceleration of electrons in the van allen radiation belts. *Nature*, 437(7056):227–230.

Huba, J., Joyce, G., and Fedder, J. (2000). Sami2 is another model of the ionosphere (sami2): A new low-latitude ionosphere model. *Journal of Geophysical Research: Space Physics*, 105(A10):23035–23053.

Huba, J., Joyce, G., and Krall, J. (2008). Three-dimensional equatorial spread f modeling. *Geophysical Research Letters*, 35(10).

Huba, J. and Krall, J. (2013). Modeling the plasmasphere with SAMI3. *Geophysical Research Letters*, 40(1):6–10.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C. (1997). Unified notation for data assimilation: Operational, sequential and variational (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):181–189.

Jaggi, R. K. and Wolf, R. (1973). Self-consistent calculation of the motion of a sheet of ions in the magnetosphere. *Journal of Geophysical Research*, 78(16):2852–2866.

Ji, E.-Y., Moon, Y.-J., Park, J., Lee, J.-Y., and Lee, D.-H. (2013). Comparison of neural network and support vector machine methods for kp forecasting. *Journal of Geophysical Research: Space Physics*, 118(8):5109–5117.

Jiang, S.-y. and Wang, L.-x. (2016). Efficient feature selection based on correlation measure between continuous and discrete features. *Information Processing Letters*, 116(2):203–215.

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

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

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

Jordanova, V. K. and Miyoshi, Y. (2005). Relativistic model of ring current and radiation belt ions and electrons: Initial results. *Geophysical Research Letters*, 32(14).

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

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

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

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

Kelley, M., Fejer, B. G., and Gonzales, C. (1979). An explanation for anomalous equatorial ionospheric electric fields associated with a northward turning of the interplanetary magnetic field. *Geophysical Research Letters*, 6(4):301–304.

Kim, K.-H., Goldstein, J., and Berube, D. (2007). Plasmaspheric drainage plume observed by the polar satellite in the prenoon sector and the image satellite during the magnetic storm of 11 april 2001. *Journal of Geophysical Research: Space Physics*, 112(A6).

Kivelson, M. G. and Russell, C. T. (1995). *Introduction to space physics*. Cambridge university press.

Kletzing, C., Kurth, W., Acuna, M., MacDowall, R., Torbert, R., Averkamp, T., Bodet, D., Bounds, S., Chutter, M., Connerney, J., et al. (2013). The electric and magnetic field instrument suite and integrated science (EMFISIS) on RBSP. *Space Science Reviews*, 179(1-4):127–181.

Knipp, D. J., McQuade, M. K., and Kirkpatrick, D. (2011). *Understanding Space Weather and the Physics behind it*. Learning Solutions.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA.

Kondrashov, D., Ghil, M., and Shprits, Y. (2011). Lognormal kalman filter for assimilating phase space density data in the radiation belts. *Space Weather*, 9(11).

Kondrashov, D., Shprits, Y., and Ghil, M. (2010). Gap filling of solar wind data by singular spectrum analysis. *Geophysical research letters*, 37(15).

Korth, H., Thomsen, M. F., Borovsky, J. E., and McComas, D. J. (1999). Plasma sheet access to geosynchronous orbit. *Journal of Geophysical Research: Space Physics*, 104(A11):25047–25061.

Kotova, G. (2007). The earth's plasmasphere: State of studies (a review). *Geomagnetism and Aeronomy*, 47(4):409.

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA.

Kozyra, J., Jordanova, V., Home, R., and Thorne, R. (1997). Modeling of the contribution of electromagnetic ion cyclotron (emic) waves to stormtime ring current erosion. *Magnetic storms*, 98:187–202.

Krall, J. and Huba, J. (2013). Sami3 simulation of plasmasphere refilling. *Geophysical Research Letters*, 40(11):2484–2488.

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

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

Krizhevsky, A. and Hinton, G. E. (2011). Using very deep autoencoders for content-based image retrieval. In *ESANN*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Kurita, K., Hayakawa, M., et al. (1985). Evaluation of the effectiveness of theoretical model calculation in determining the plasmapause structure. *Journal of Geophysics— IF 32.18*, 57(1):130–135.

Kurth, W., De Pascuale, S., Faden, J., Kletzing, C., Hospodarsky, G., Thaller, S., and 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(2):904–914.

LaBelle, J., Ruppert, D., and Treumann, R. (1999). A statistical study of banded magnetospheric emissions. *Journal of Geophysical Research: Space Physics*, 104(A1):293–303.

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

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

Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8595–8598. IEEE.

LeDocq, M., Gurnett, D., and Anderson, R. (1994). Electron number density fluctuations near the plasmapause observed by the CRRES spacecraft. *JOURNAL OF GEOPHYSICAL RESEARCH-ALL SERIES-*, 99:23–661.

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

Lemaire, J. (1989). Plasma distribution models in a rotating magnetic dipole and refilling of plasmaspheric flux tubes. *Physics of Fluids B: Plasma Physics*, 1(7):1519–1525.

Lemaire, J. and Kowalkowski, L. (1981). The role of plasma interchange motion for the formation of a plasmapause. *Planetary and Space Science*, 29(4):469 – 478.

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

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168.

Li, H., Chen, C. P., and Huang, H.-P. (2000). *Fuzzy neural intelligent systems: Mathematical foundation and the applications in engineering*. CRC Press.

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

Lichtenberger, J., Ferencz, C., Bodnár, L., Hamar, D., and Steinbach, P. (2008). Automatic whistler detector and analyzer system: Automatic whistler detector. *Journal of Geophysical Research: Space Physics*, 113(A12).

Lichtenberger, J., Ferencz, C., Hamar, D., Steinbach, P., Rodger, C., Clilverd, M., and Collier, A. (2011). Automatic retrieval of plasmaspheric electron densities: first results form automatic whistler detector and analyzer network. In *2011 XXXth URSI General Assembly and Scientific Symposium*, pages 1–4. IEEE.

Lichtenberger, J., Ferencz, C., Hamar, D., Steinbach, P., Rodger, C. J., Clilverd, M. A., and Collier, A. B. (2010). Automatic whistler detector and analyzer system: Implementation of the analyzer algorithm. *Journal of Geophysical Research: Space Physics*, 115(A12).

Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., Tobiska, W. K., Wintoft, P., Ganushkina, N. Y., Welling, D. T., Bingham, S., Balikhin, M. A., Opgenoorth, H. J., Engel, M. A., Weigel, R. S., Singer, H. J., Buresova, D., Bruinsma, S., Zhelavskaya, I. S., Shprits, Y. Y., and Vasile, R. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, 16(12):2079–2102.

Linoff, G. S. and Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

Liu, S., Chen, M., Roeder, J., Lyons, L., and Schulz, M. (2005). Relative contribution of electrons to the stormtime total ring current energy content. *Geophysical research letters*, 32(3).

Lui, A. T., McEntire, R., and Krimigis, S. (1987). Evolution of the ring current during two geomagnetic storms. *Journal of Geophysical Research: Space Physics*, 92(A7):7459–7470.

Lyons, L. and Williams, D. (1984). Quantitative aspects of. *Mag. netospheric Physics, D. Reidel, Hingham, Mass.*

Marchaudon, A. and Blelly, P.-L. (2015). A new interhemispheric 16-moment model of the plasmasphere-ionosphere system: IPIM. *Journal of Geophysical Research: Space Physics*, 120(7):5728–5745.

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.

Marr, D., Poggio, T., et al. (1976). Cooperative computation of stereo disparity. *From the Retina to the Neocortex*, pages 239–243.

Mathworks.com (2015). Choose a multilayer neural network training function - MATLAB & Simulink. *http://www.mathworks.com/help/nnet/ug/choose-a-multilayer-neural-network-training-function.html*.

Matsui, H., Puhl-Quinn, P., Jordanova, V., Khotyaintsev, Y., Lindqvist, P.-A., and Torbert, R. (2008). Derivation of inner magnetospheric electric field (unh-imef) model using cluster data set. In *Annales Geophysicae*, volume 26, pages 2887–2898. Copernicus GmbH.

Matsui, H., Torbert, R. B., Spence, H. E., Khotyaintsev, Y. V., and Lindqvist, P.-A. (2013). Revision of empirical electric field modeling in the inner magnetosphere using cluster data. *Journal of Geophysical Research: Space Physics*, 118(7):4119–4134.

Mauk, B., Fox, N. J., Kanekal, S., Kessel, R., Sibeck, D., and Ukhorskiy, A. (2013). Science objectives and rationale for the radiation belt storm probes mission. *Space Science Reviews*, 179(1-4):3–27.

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

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

McConaghy, T. (2011). *FFX: Fast, Scalable, Deterministic Symbolic Regression Technology*, pages 235–260. Springer New York, New York, NY.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

McIlwain, C. (1986). A kp dependent equatorial electric field model. *Advances in Space Research*, 6(3):187–197.

Meek, C., Thiesson, B., and Heckerman, D. (2002). The learning-curve sampling method applied to model-based clustering. *Journal of Machine Learning Research*, 2(Feb):397–418.

Meyer-Vernet, N. and Perche, C. (1989). Tool kit for antennae and thermal noise near the plasma frequency. *Journal of Geophysical Research: Space Physics*, 94(A3):2405–2415.

Mohamed, A.-r., Dahl, G. E., and Hinton, G. (2012). Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):14–22.

Moldwin, M. B., Downward, L., Rassoul, H., Amin, R., and Anderson, R. (2002). A new model of the location of the plasmapause: Crres results. *Journal of Geophysical Research: Space Physics*, 107(A11):SMP–2.

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

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 6(4):525–533.

Moré, J. J. (1978). The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer.

Mosier, S. R., Kaiser, M. L., and 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.

Murayama, T. (1982). Coupling function between solar wind parameters and geomagnetic indices. *Reviews of Geophysics*, 20(3):623–629.

Nakano, S., Fok, M.-C., Brandt, P. C., and 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(5):3724–3740.

Neugebauer, M. and Snyder, C. W. (1966). Mariner 2 observations of the solar wind: 1. average properties. *Journal of Geophysical Research*, 71(19):4469–4484.

Newberry, I. T., Comfort, R. H., Richards, P. G., and Chappell, C. R. (1989). Thermal he+ in the plasmasphere: Comparison of observations with numerical calculations. *Journal of Geophysical Research: Space Physics*, 94(A11):15265–15276.

Newell, P., Sotirelis, T., Liou, K., Meng, C.-I., and Rich, F. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research: Space Physics*, 112(A1).

Ni, B., Thorne, R. M., Shprits, Y. Y., and Bortnik, J. (2008). Resonant scattering of plasma sheet electrons by whistler-mode chorus: Contribution to diffuse auroral precipitation. *Geophysical Research Letters*, 35(11).

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.

Northrop, T. G. (1963). Adiabatic charged-particle motion. *Reviews of Geophysics*, 1(3):283–304.

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

Orlova, K., Shprits, Y., and Spasojević, M. (2016). New global loss model of energetic and relativistic electrons based on Van Allen Probes measurements. *Journal of Geophysical Research: Space Physics*, 121(2):1308–1314.

Orlova, K., Spasojević, M., and Shprits, Y. (2014). Activity-dependent global model of electron loss inside the plasmasphere. *Geophysical Research Letters*, 41(11):3744–3751.

Orlova, K. G., Shprits, Y. Y., and Ni, B. (2012). Bounce-averaged diffusion coefficients due to resonant interaction of the outer radiation belt electrons with oblique chorus waves computed in a realistic magnetic field model. *Journal of Geophysical Research: Space Physics*, 117(A7).

Ozeke, L. G., Mann, I. R., Murphy, K. R., Jonathan Rae, I., and Milling, D. K. (2014). Analytic expressions for ulf wave radiation belt radial diffusion coefficients. *Journal of Geophysical Research: Space Physics*, 119(3):1587–1605.

Park, C. (1972). Methods of determining electron concentrations in the magnetosphere from nose whistlers.

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

Park, C. G. (1971). Westward electric fields as the cause of nighttime enhancements in electron concentrations in midlatitude f region. *Journal of Geophysical Research*, 76(19):4560–4568.

Parker, E. N. (1963). Interplanetary dynamical processes. *New York, Interscience Publishers, 1963.*

Peng, H., L. F. and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238.

Perlich, C., Provost, F., and Simonoff, J. S. (2003). Tree induction vs. logistic regression: A learning-curve analysis. *Journal of Machine Learning Research*, 4(Jun):211–255.

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

Pierrard, V., Khazanov, G., Cabrera, J., and Lemaire, J. (2008). Influence of the convection electric field models on predicted plasmapause positions during magnetic storms. *Journal of Geophysical Research: Space Physics*, 113(A8).

Pierrard, V. and Lemaire, J. F. (2004). Development of shoulders and plumes in the frame of the interchange instability mechanism for plasmapause formation. *Geophysical Research Letters*, 31(5).

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

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

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

Reeves, G., Spence, H. E., Henderson, M., Morley, S., Friedel, R., Funsten, H., Baker, D., Kanekal, S., Blake, J., Fennell, J., et al. (2013). Electron acceleration in the heart of the van allen radiation belts. *Science*, 341(6149):991–994.

Reinisch, B. W., Huang, X., Song, P., Green, J. L., Fung, S. F., Vasyliunas, V. M., Gallagher, D. L., and Sandel, B. R. (2004). Plasmaspheric mass loss and refilling as a result of a magnetic storm. *Journal of Geophysical Research: Space Physics*, 109(A1).

Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory.

Salakhutdinov, R. and Hinton, G. (2009). Semantic hashing. *International Journal of Approximate Reasoning*, 50(7):969–978.

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

Sandel, B., Goldstein, J., Gallagher, D., and Spasojević, M. (2003). Extreme ultraviolet imager observations of the structure and dynamics of the plasmasphere. *Space Science Reviews*, 109(1):25–46.

Sandel, B. R., Broadfoot, A. L., Curtis, C., King, R., Stone, T., Hill, R., Chen, J., Siegmund, O., Raffanti, R., Allred, D. D., et al. (2000). The extreme ultraviolet imager investigation for the IMAGE mission. In *The IMAGE Mission*, pages 197–242. Springer.

Sandel, B. R., King, R. A., Forrester, W., Gallagher, D. L., Broadfoot, A. L., and Curtis, C. (2001). Initial results from the IMAGE extreme ultraviolet imager. *Geophysical research letters*, 28(8):1439–1442.

Sckopke, N. (1966). A general relation between the energy of trapped particles and the disturbance field near the earth. *Journal of Geophysical Research*, 71(13):3125–3130.

Selesnick, R., Baker, D., Jaynes, A., Li, X., Kanekal, S., Hudson, M., and Kress, B. (2014). Observations of the inner radiation belt: Crand and trapped solar protons. *Journal of Geophysical Research: Space Physics*, 119(8):6541–6552.

Selesnick, R., Looper, M., and Mewaldt, R. (2007). A theoretical model of the inner proton radiation belt. *Space weather*, 5(4).

Selesnick, R. S. (2015). High-energy radiation belt electrons from crand. *Journal of Geophysical Research: Space Physics*, 120(4):2912–2917.

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

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

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

Shprits, Y. Y., Drozdov, A. Y., Spasojević, M., Kellerman, A. C., Usanova, M. E., Engebretson, M. J., Agapitov, O. V., Zhelavskaya, I. S., Raita, T. J., Spence, H. E., et al. (2016). Wave-induced loss of ultra-relativistic electrons in the Van Allen radiation belts. *Nature Communications*, 7:12883.

Shprits, Y. Y., Elkington, S. R., Meredith, N. P., and Subbotin, D. A. (2008a). Review of modeling of losses and sources of relativistic electrons in the outer radiation belt I: Radial transport. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(14):1679–1693.

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

Shprits, Y. Y., Meredith, N. P., and Thorne, R. M. (2007b). Parameterization of radiation belt electron loss timescales due to interactions with chorus waves. *Geophysical Research Letters*, 34(11).

Shprits, Y. Y., Subbotin, D. A., Meredith, N. P., and Elkington, S. R. (2008b). Review of modeling of losses and sources of relativistic electrons in the outer radiation belt II: Local acceleration and loss. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(14):1694–1713.

Shprits, Y. Y., Vasile, R., and Zhelavskaya, I. S. (2019). Nowcasting and predicting the Kp index using historical values and real-time observations. *Space Weather*, 17(8):1219–1229.

Shue, J.-H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., Vaisberg, O. L., Kokubun, S., Singer, H. J., Detman, T. R., and Kawano, H. (1998). Magnetopause location under extreme solar wind conditions. *Journal of Geophysical Research: Space Physics*, 103(A8):17691–17700.

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

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

Smith, R. and Carpenter, D. (1961). Extension of nose whistler analysis. *Journal of Geophysical Research*, 66(8):2582–2586.

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

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

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

Spasojević, M. and Sandel, B. (2010). Global estimates of plasmaspheric losses during moderate disturbance intervals. In *Annales Geophysicae*, volume 28, pages 27–36. Copernicus GmbH.

Spasojević, M., Thomsen, M., Chi, P., and Sandel, B. (2005). Afternoon subauroral proton precipitation resulting from ring current–Plasmasphere interaction. *Inner Magnetosphere Interactions: New Perspectives from Imaging*, pages 85–99.

Stern, D. (1974). Models of the earth's electric field.

Stern, D. P. (1973). A study of the electric field in an open magnetospheric model. *Journal of Geophysical Research*, 78(31):7292–7305.

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

Storey, L. (1953). An investigation of whistling atmospherics. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 246(908):113–141.

Storey, O. (1958). Protons outside the earth's atmosphere. In *Annales de Geophysique*, volume 14, page 144.

Sturges, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66.

Su, Y.-J., Thomsen, M. F., Borovsky, J. E., and Foster, J. C. (2001a). Linkage between polar patches and plasmaspheric drainage plumes. *Geophysical research letters*, 28(1):111–113.

Su, Y.-J., Thomsen, M. F., Borovsky, J. E., and Lawrence, D. J. (2001b). A comprehensive survey of plasmasphere refilling at geosynchronous orbit. *Journal of Geophysical Research: Space Physics*, 106(A11):25615–25629.

Sugiura, M., Kertz, W., Price, A., and Stone, D. (1964). *P. 1. Hourly Values of Equatorial Dst for the IGY.* Pergamon Press.

Tan, Y., Hu, Q., Wang, Z., and Zhong, Q. (2018). Geomagnetic index kp forecasting with lstm. *Space Weather*, 16(4):406–416.

Teunissen, P. (1990). Nonlinear least squares.

Thomsen, M. (2004). Why kp is such a good measure of magnetospheric convection. *Space Weather*, 2(11).

Thorne, R., Li, W., Ni, B., Ma, Q., Bortnik, J., Chen, L., Baker, D., Spence, H. E., Reeves, G., Henderson, M., et al. (2013). Rapid local acceleration of relativistic radiation-belt electrons by magnetospheric chorus. *Nature*, 504(7480):411.

Trotignon, J., Décréau, P., Rauch, J., Vallières, X., Rochel, A., Kougblénou, S., Lointier, G., Facskó, G., Canu, P., Darrouzet, F., et al. (2010). The whisper relaxation sounder and the Cluster active archive. In *The Cluster Active Archive*, pages 185–208. Springer.

Trotignon, J., Etcheto, J., and Thouvenin, J. (1986). Automatic determination of the electron density measured by the relaxation sounder on board isee 1. *Journal of Geophysical Research: Space Physics*, 91(A4):4302–4320.

Trotignon, J., Rauch, J., Décréu, P., Canu, P., and 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.

Van Allen, J. A. and Frank, L. A. (1959). Radiation around the earth to a radial distance of 107,400 km. *Nature*, 183(4659):430–434.

Vernov, S. and Chudakov, A. (1960). Investigation of radiation in outer space. In *International Cosmic Ray Conference*, volume 3, page 19.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and 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(Dec):3371–3408.

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

Walt, M. (1994). *The geomagnetic field*, page 2535. Cambridge Atmospheric and Space Science Series. Cambridge University Press.

Wang, J., Zhong, Q., Liu, S., Miao, J., Liu, F., Li, Z., and Tang, W. (2015). Statistical analysis and verification of 3-hourly geomagnetic activity probability predictions. *Space Weather*, 13(12):831–852.

Weimer, D. (1996). A flexible, imf dependent model of high-latitude electric potentials having space weather applications. *Geophysical Research Letters*, 23(18):2549–2552.

Weimer, D. (2001). An improved model of ionospheric electric potentials including substorm perturbations and application to the geospace environment modeling november 24, 1996, event. *Journal of Geophysical Research: Space Physics*, 106(A1):407–416.

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

Williams, D. and Hinton, G. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–538.

Williams, D. J. (1987). Ring current and radiation belts. *Reviews of Geophysics*, 25(3):570–578.

Wing, S. and Johnson, J. R. (2019). Applications of information theory in solar and space physics. *Entropy*, 21(2):140.

Wing, S., Johnson, J. R., Camporeale, E., and Reeves, G. D. (2016). Information theoretical approach to discovering solar wind drivers of the outer radiation belt. *Journal of Geophysical Research: Space Physics*, 121(10):9378–9399.

Wing, S., Johnson, J. R., Jen, J., Meng, C.-I., Sibeck, D. G., Bechtold, K., Freeman, J., Costello, K., Balikhin, M., and Takahashi, K. (2005). Kp forecast models. *Journal of Geophysical Research: Space Physics*, 110(A4).

Wintoft, Wik, Magnus, Matzka, Jürgen, and Shprits, Yuri (2017). Forecasting kp from solar wind data: input parameter study using 3-hour averages and 3-hour range values. *J. Space Weather Space Clim.*, 7:A29.

Xiao-Ting, S., Gendrin, R., and Caudal, G. (1988). Refilling process in the plasmasphere and its relation to magnetic activity. *Journal of Atmospheric and Terrestrial Physics*, 50(3):185 – 195.

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

Xu, F. and Borovsky, J. E. (2015). A new four-plasma categorization scheme for the solar wind. *Journal of Geophysical Research: Space Physics*, 120(1):70–100.

Yau, A. W., Peterson, W., and Abe, T. (2011). *Influences of the Ionosphere, Thermosphere and Magnetosphere on Ion Outflows*, pages 283–314. Springer Netherlands, Dordrecht.

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

Zhang, Q.-H., Zhang, B.-C., Lockwood, M., Hu, H.-Q., Moen, J., Ruohoniemi, J. M., Thomas, E. G., Zhang, S.-R., Yang, H.-G., Liu, R.-Y., et al. (2013). Direct observations of the evolution of polar cap ionization patches. *Science*, 339(6127):1597–1600.

Zhelavskaya, I., Aseev, N., and Shprits, Y. Y. (2020a). A combined neural network- and physics-based approach for modeling plasmasphere dynamics. *Submitted to Journal of Geophysical Research: Space Physics*.

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

Zhelavskaya, I., Shprits, Y. Y., and Spasojević, M. (2017). Empirical modeling of the plasmasphere dynamics using neural networks. *Journal of Geophysical Research: Space Physics*, 122(11):11,227–11,244.

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

Zhelavskaya, I. S., Shprits, Y. Y., and Spasojevic, M. (2018). Chapter 12 - Reconstruction of plasma electron density from satellite measurements via artificial neural networks. In Camporeale, E., Wing, S., and Johnson, J. R., editors, *Machine Learning Techniques for Space Weather*, pages 301 – 327. Elsevier.

Zhelavskaya, I. S., Vasile, R., Shprits, Y. Y., Stolle, C., and Matzka, J. (2019). Systematic analysis of machine learning and feature selection techniques for prediction of the Kp index. *Space Weather*.