Understanding the Dynamics of the Near-Earth Space Environment Utilizing Long-term Satellite Observations

Artem Smirnov

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

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

24. Oktober 2023

BETREUER:
Hauptbetreuer: Prof. Dr. Yuri Y. Shprits
Zweitbetreuer: Dr. Geoffrey D. Reeves

GUTACHTER*INNEN:
Dr. Viacheslav Merkin
Prof. Dr. Claudia Stolle
Prof. Dr. Jens Wickert
Prof. Dr. Jean Braun
Prof. Dr. Stephan Sobolev
Prof. Dr. Svetlana Santer
Prof. Dr. Fabrice Cotton (Vorsitzender)
PD Dr. Sascha Brune
PD Dr. Elena Kronberg

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-61371
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-613711

Declaration

I, Artem Smirnov, hereby declare that this thesis entitled "Understanding the Dynamics of the Near-Earth Space Environment Utilizing Long-term Satellite Observations" has been composed entirely by myself. The work contained here 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- I have fully acknowledged and referenced the ideas and work of others, whether published or unpublished.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.

24 October 2023

Artem Smirnov

At my tiny flat
There's just my cat,
A bed and a chair.
Still,
I'll stick it till
I'm on a bill
All over Times Square.

— **Stephen Sondheim, *Follies.***

Acknowledgements

I would like to convey my profound gratitude to my advisor Professor Yuri Shprits for invaluable guidance and support throughout my doctoral studies. His expertise and feedback have been immensely helpful for my professional growth. I wish to extend special thanks to my co-supervisor Dr. Geoffrey Reeves for the invaluable discussions on various topics of magnetospheric physics. I am particularly indebted to Professor Hermann Lühr, whose wisdom, advice and encouragement have been paramount to the completion of this thesis.

My PhD journey was, perhaps, not a usual one, at least by pre-COVID standards. Moving to a new town to take the PhD position, little did I know, that in just four months a raging pandemic would change the usual way of life – and work – enormously for the following two years. Nevertheless, the atmosphere in our group was that of mutual support and encouragement. For that, I would like to wholeheartedly thank all members of our Space Physics and Space Weather Section at GFZ Potsdam, past and present. Their expertise, insights, and feedback have facilitated this research and sustained my motivation. I would like to give special thanks to Angelica Castillo, with whom we shared an office and similar PhD struggles, our scientific coordinator Melanie Burns for pushing me to complete this dissertation and proofreading much of my work, and to our IT specialist Alexander Jordan. I am also deeply thankful to my collaborators Dr. Fabricio Prol for his bright ideas and exciting discussions on various ionospheric topics, Dr. Max Berrendorf for agreeing to collaborate with me – then a master student at LMU Munich – on my first machine learning project and for staying in touch and collaborating since then. I am also thankful to my GEM mentor Professor Anthony Chan for our regular discussions about academic and normal life. I am grateful to Dr. Renata Lukianova and Dr. Elena Kronberg, who were my first mentors in the field of space weather, as well as to Professor Stuart Gilder with whom I did my master's studies in geomagnetism that have been a crucial step for delving into magnetospheric research later on. Last but not least, completing this dissertation would really not be possible without coffee, for in order to have a flow of ideas, a flow of coffee must be guaranteed. I, therefore, would like to thank the anonymous coffee donors who brought it in large quantities to our GFZ kitchen.

Finally, I would like to extend my thanks to my family and friends, whose support and encouragement have been so crucial in helping me to complete this PhD. Their belief in me has fueled my determination to achieve my academic goals. I am deeply grateful to all those who have supported me throughout my PhD journey, and I shall carry their lessons, insights, and encouragement with me as I move forward in my professional life.

Artem Smirnov

Potsdam, October 2023

Abstract

The near-Earth space environment is a highly complex system comprised of several regions and particle populations hazardous to satellite operations. The trapped particles in the radiation belts and ring current can cause significant damage to satellites during space weather events, due to deep dielectric and surface charging. Closer to Earth is another important region, the ionosphere, which delays the propagation of radio signals and can adversely affect navigation and positioning. In response to fluctuations in solar and geomagnetic activity, both the inner-magnetospheric and ionospheric populations can undergo drastic and sudden changes within minutes to hours, which creates a challenge for predicting their behavior. Given the increasing reliance of our society on satellite technology, improving our understanding and modeling of these populations is a matter of paramount importance.

In recent years, numerous spacecraft have been launched to study the dynamics of particle populations in the near-Earth space, transforming it into a data-rich environment. To extract valuable insights from the abundance of available observations, it is crucial to employ advanced modeling techniques, and machine learning methods are among the most powerful approaches available. This dissertation employs long-term satellite observations to analyze the processes that drive particle dynamics, and builds interdisciplinary links between space physics and machine learning by developing new state-of-the-art models of the inner-magnetospheric and ionospheric particle dynamics.

The first aim of this thesis is to investigate the behavior of electrons in Earth's radiation belts and ring current. Using ~18 years of electron flux observations from the Global Positioning System (GPS), we developed the first machine learning model of hundreds-of-keV electron flux at Medium Earth Orbit (MEO) that is driven solely by solar wind and geomagnetic indices and does not require auxiliary flux measurements as inputs. We then proceeded to analyze the directional distributions of electrons, and for the first time, used Fourier sine series to fit electron pitch angle distributions (PADs) in Earth's inner magnetosphere. We performed a superposed epoch analysis of 129 geomagnetic storms during the Van Allen Probes era and demonstrated

that electron PADs have a strong energy-dependent response to geomagnetic activity. Additionally, we showed that the solar wind dynamic pressure could be used as a good predictor of the PAD dynamics. Using the observed dependencies, we created the first PAD model with a continuous dependence on L, magnetic local time (MLT) and activity, and developed two techniques to reconstruct near-equatorial electron flux observations from low-PA data using this model.

The second objective of this thesis is to develop a novel model of the topside ionosphere. To achieve this goal, we collected observations from five of the most widely used ionospheric missions and intercalibrated these data sets. This allowed us to use these data jointly for model development, validation, and comparison with other existing empirical models. We demonstrated, for the first time, that ion density observations by Swarm Langmuir Probes exhibit overestimation (up to $\sim 40-50\%$) at low and mid-latitudes on the night side, and suggested that the influence of light ions could be a potential cause of this overestimation. To develop the topside model, we used 19 years of radio occultation (RO) electron density profiles, which were fitted with a Chapman function with a linear dependence of scale height on altitude. This approximation yields 4 parameters, namely the peak density and height of the F2-layer and the slope and intercept of the linear scale height trend, which were modeled using feedforward neural networks (NNs). The model was extensively validated against both RO and in-situ observations and was found to outperform the International Reference Ionosphere (IRI) model by up to an order of magnitude. Our analysis showed that the most substantial deviations of the IRI model from the data occur at altitudes of 100-200 km above the F2-layer peak. The developed NN-based ionospheric model reproduces the effects of various physical mechanisms observed in the topside ionosphere and provides highly accurate electron density predictions.

This dissertation provides an extensive study of geospace dynamics, and the main results of this work contribute to the improvement of models of plasma populations in the near-Earth space environment.

Keywords: Ionosphere; empirical modeling; gradient boosting; machine learning; neural networks; radiation belts; ring current; space physics.

Zusammenfassung

Die erdnahe Weltraumumgebung ist ein hochkomplexes System, das aus mehreren Regionen und Partikelpopulationen besteht, die für den Satellitenbetrieb gefährlich sind. Die in den Strahlungsgürteln und dem Ringstrom gefangenen Teilchen können bei Weltraumwetterereignissen aufgrund der tiefen dielektrischen und oberflächlichen Aufladung erhebliche Schäden an Satelliten verursachen. Näher an der Erde liegt eine weitere wichtige Region, die Ionosphäre, die die Ausbreitung von Funksignalen verzögert und die Navigation und Positionsbestimmung beeinträchtigen kann. Als Reaktion auf Fluktuationen der solaren und geomagnetischen Aktivität können sowohl die Populationen der inneren Magnetosphäre als auch der Ionosphäre innerhalb von Minuten bis Stunden drastische und plötzliche Veränderungen erfahren, was eine Herausforderung für die Vorhersage ihres Verhaltens darstellt. Angesichts der zunehmenden Abhängigkeit unserer Gesellschaft von der Satellitentechnologie ist ein besseres Verständnis und eine bessere Modellierung dieser Populationen von größter Bedeutung.

In den letzten Jahren wurden zahlreiche Raumsonden gestartet, um die Dynamik von Partikelpopulationen im erdnahen Weltraum zu untersuchen, was diesen in eine datenreiche Umgebung verwandelt hat. Um aus der Fülle der verfügbaren Beobachtungen wertvolle Erkenntnisse zu gewinnen, ist der Einsatz fortschrittlicher Modellierungstechniken unabdingbar, und Methoden des maschinellen Lernens gehören zu den leistungsfähigsten verfügbaren Ansätzen. Diese Dissertation nutzt langfristige Satellitenbeobachtungen, um die Prozesse zu analysieren, die die Teilchendynamik antreiben, und schafft interdisziplinäre Verbindungen zwischen Weltraumphysik und maschinellem Lernen, indem sie neue hochmoderne Modelle der innermagnetosphärischen und ionosphärischen Teilchendynamik entwickelt.

Das erste Ziel dieser Arbeit ist es, das Verhalten von Elektronen im Strahlungsgürtel und Ringstrom der Erde zu untersuchen. Unter Verwendung von ~18 Jahren Elektronenflussbeobachtungen des Global Positioning System (GPS) haben wir das erste maschinelle Lernmodell des Elektronenflusses im mittleren Erdorbit (MEO) entwickelt, das ausschließlich durch Sonnenwind und geomagnetische Indizes gesteuert wird und keine zusätzlichen Flussmessungen als

Eingaben benötigt. Anschließend analysierten wir die Richtungsverteilungen der Elektronen und verwendeten zum ersten Mal Fourier-Sinus-Reihen, um die Elektronen-Stellwinkelverteilungen (PADs) in der inneren Magnetosphäre der Erde zu bestimmen. Wir führten eine epochenübergreifende Analyse von 129 geomagnetischen Stürmen während der Van-Allen-Sonden-Ära durch und zeigten, dass die Elektronen-PADs eine starke energieabhängige Reaktion auf die geomagnetische Aktivität haben. Außerdem konnten wir zeigen, dass der dynamische Druck des Sonnenwindes als guter Prädiktor für die PAD-Dynamik verwendet werden kann. Anhand der beobachteten Abhängigkeiten haben wir das erste PAD-Modell mit einer kontinuierlichen Abhängigkeit von L, der magnetischen Ortszeit (MLT) und der Aktivität erstellt und zwei Techniken entwickelt, um die Beobachtungen des äquatornahen Elektronenflusses aus Daten mit niedrigem Luftdruck mit Hilfe dieses Modells zu rekonstruieren.

Das zweite Ziel dieser Arbeit ist die Entwicklung eines neuen Modells der Topside-Ionosphäre. Um dieses Ziel zu erreichen, haben wir Beobachtungen von fünf der meistgenutzten Ionosphärenmissionen gesammelt und diese Datensätze interkalibriert. So konnten wir diese Daten gemeinsam für die Modellentwicklung, die Validierung und den Vergleich mit anderen bestehenden empirischen Modellen nutzen. Wir haben zum ersten Mal gezeigt, dass die Ionendichtebeobachtungen von Swarm-Langmuir-Sonden in niedrigen und mittleren Breiten auf der Nachtseite eine Überschätzung (bis zu $\sim 40-50\%$) aufweisen, und haben vorgeschlagen, dass der Einfluss leichter Ionen eine mögliche Ursache für diese Überschätzung sein könnte. Zur Entwicklung des Oberseitenmodells wurden 19 Jahre lang Elektronendichteprofile aus der Radio-Okkultation (RO) verwendet, die mit einer Chapman-Funktion mit einer linearen Abhängigkeit der Skalenhöhe von der Höhe angepasst wurden. Aus dieser Näherung ergeben sich 4 Parameter, nämlich die Spitzendichte und die Höhe der F2-Schicht sowie die Steigung und der Achsenabschnitt des linearen Trends der Skalenhöhe, die mit Hilfe von neuronalen Feedforward-Netzwerken (NN) modelliert wurden. Das Modell wurde sowohl anhand von RO- als auch von In-situ-Beobachtungen umfassend validiert und übertrifft das Modell der Internationalen Referenz-Ionosphäre (IRI). Unsere Analyse zeigte, dass die größten Abweichungen des IRI-Modells von den Daten in Höhen von 100-200 km über der F2-Schichtspitze auftreten. Das entwickelte NN-basierte Ionosphärenmodell reproduziert die Auswirkungen verschiedener physikalischer Mechanismen, die in der

Topside-Ionosphäre beobachtet werden, und liefert sehr genaue Vorhersagen der Elektronendichte.

Diese Dissertation bietet eine umfassende Untersuchung der Dynamik in der Geosphäre, und die wichtigsten Ergebnisse dieser Arbeit tragen zur Verbesserung der Modelle von Plasmapopulationen in der erdnahen Weltraumumgebung bei.

Schlüsselwörter: Ionosphäre; empirische Modellierung; Gradient Boosting; maschinelles Lernen; neuronale Netze; Strahlungsgürtel; Ringstrom; Weltraumphysik.

Publications

First author peer-reviewed publications

1. **Smirnov, A.**, Shprits, Y., Prol, F., Lühr, H., Berrendorf, M., Zhelavskaya, I. and Xiong, C., (2023): A novel neural network model of Earths topside ionosphere. *Scientific Reports*, 13(1), p.1303.

2. **Smirnov, A.**, Shprits, Y., Allison, H., Aseev, N., Drozdov, A., Kollman, P., Reeves, G.D., Wang, D., and Saikin, A. (2022): An empirical model of the equatorial electron pitch angle distributions in Earth's outer radiation belt, *Space Weather*, e2022SW003053.

3. **Smirnov, A.**, Shprits, Y., Allison, H., Aseev, N., Drozdov, A., Kollman, P., Reeves, G.D., Wang, D., and Saikin, A. (2022): Storm-time evolution of the equatorial electron pitch angle distributions in Earth's outer radiation belt, *Frontiers in Astronomy and Space Sciences*, Vol. 9 (special issue devoted to R.M. Thorne and P. Gary), DOI: 10.3389/fspas.2022.836811.

4. **Smirnov, A.**, Shprits, Y., Zhelavskaya, I., Lühr, H., Xiong, C., Goss, A., Prol, F. S., Schmidt, M., Hoque, M., Pedatella, N., and Szabó-Roberts, M. (2021): Intercalibration of the Plasma Density Measurements in Earth's Topside Ionosphere, *Journal of Geophysical Research: Space Physics*, 126, 10, e2021JA029334, DOI: 10.1029/2021JA029334 – *Top downloaded paper from JGR Space Physics in 2021.*

5. **Smirnov A.G.**, Berrendorf M., Shprits Y.Y., Kronberg E.A., Allison H.J., Aseev N.A., Zhelavskaya I.S., Morley S.K., Reeves G.D., Carver M.R., and Effenberger F. (2020) Medium energy electron flux in Earth's outer radiation belt (MERLIN): a machine learning model, *Space Weather*, 18, DOI: 10.1029/2020SW002532.

6. **Smirnov A.G.**, Kronberg E.A., Daly P.W, Aseev N., Shprits Y., and Kellerman A. (2020) Adiabatic invariants calculations for Cluster mission: a long-term product for radiation belts studies, *Journal of Geophysical Research: Space Physics*, 125, DOI: 10.1029/2019JA027576 – *Top 10% most downloaded papers from JGR Space Physics in 2020.*

Co-authored peer-reviewed publications

1. Shprits, Y.Y., Castillo, A.M., Aseev, N., Cervantes, S., Michaelis, I., Zhelavskaya, I., **Smirnov, A.** and Wang, D., (2023). Data Assimilation in the Near-Earth Electron Radiation Environment, 5, p. 157-172, Chapter in *Applications of Data Assimilation and Inverse Problems in the Earth Sciences*, Cambridge University Press.

2. Prol, F.S., **Smirnov, A.G.**, Hoque, M.M., and Shprits, Y.Y. (2022) Combined model of topside ionosphere and plasmasphere derived from radio-occultation and Van Allen Probes data, *Scientific Reports*, 12(1), 1-11, DOI: 10.1038/s41598-022-13302-1.

3. Landis, D., Saikin, A., Drozdov, A., Zhelavskaya, I., Shprits, Y., Pfitzer, M., and **Smirnov A**. NARX Neural Network Derivations of the Outer Boundary Radiation Belt Electron Flux, *Space Weather*, 20(5), e2021SW002774, DOI: 10.1029/2021SW002774.

4. Xiong, C., Jiang, H., Yan, R., Lühr, H., Stolle, C., Yin, F., **Smirnov, A.**, Piersanti, M., Liu, Y., Diego, P., Zhima, Z., Shen, X., Förster, M., Wuchert, S., and Bilitza, D. Solar flux influence on the in-situ plasma density at topside ionosphere measured by Swarm satellites, *Journal of Geophysical Research: Space Physics*, 127(5), e2022JA030275, DOI: 10.1029/2022JA030275.

5. Kronberg, E.A., Hannan, T., Huthmacher, J., Münzer, M., Peste, F., Zhou, Z., Berrendorf, M., Farman, E., Gastaldello, F., Ghizzardi, S., Escoubet, P., Haaland, S., **Smirnov, A.**, Sivadas, N., Allen, R.C., Tiengo, A., Ilie, R. (2021), Prediction of soft proton intensities in the near-Earth space using machine learning, *The Astrophysical Journal*, 921:76 (14pp), DOI: https://doi.org/10.3847/1538-4357/ac1b30.

6. Kronberg, E. A., Daly, P. W., Grigorenko, E. E., **Smirnov, A. G.**, Klecker, B., and Malykhin, A. Y. (2021). Energetic charged particles in the terrestrial magnetosphere: Cluster/RAPID results, *Journal of Geophysical Research: Space Physics*, 126, e2021JA029273, DOI: https://doi.org/10.1029/2021JA029273 – *Top downloaded paper from JGR Space Physics in 2021.*

7. Kronberg, E.A., Gorman, J., Nykyri, K., **Smirnov, A.G.**, Gjerloev, J.W., Grigorenko, E.E., Kozak, L.V., Ma, X., Trattner, K.J., and M. Friel (2021). Kelvin-Helmholtz Instability associated with reconnection and ultra low frequency waves at the ground : a case study, *Frontiers in Physics*, 700, DOI: 10.3389/fphy.2021.738988.

8. Szabo-Roberts, M., Shprits, Y., Allison, H. J., Vasile, R., **Smirnov, A.**, Aseev, N., Drozdov, A., Miyoshi, Y., Claudepierre, S. G., Kasahara, S., Yokota, S., Mitani, T., Takashima, T., Higashio, N., Hori, T., Keika, K., Imajo, S., Shinohara, I. (2021): Preliminary Statistical Comparisons of Spin-Averaged Electron Data from Arase and Van Allen Probes Instruments, *Journal of Geophysical Research: Space Physics*, 126, 7, e2020JA028929. https://doi.org/10.1029/2020JA028929 – *Top downloaded paper from JGR Space Physics in 2021.*

9. Allison H.J., Shprits Y.Y., Zhelavskaya I.S., Wang D., and **Smirnov A.G.** (2021) Gyroresonant wave-particle interactions with chorus waves during extreme depletions of plasma density in the Van Allen radiation belts, ***Science Advances***, vol. 7, issue 5. DOI: https://doi.org/10.1126/sciadv.abc0380.

10. Kronberg, Elena A., Gastaldello, Fabio, Haaland, Stein, **Smirnov, Artem**, Berrendorf, Max, Ghizzardi, Simona, Kuntz, K. D., Sivadas, Nithin, Allen, Robert C., Tiengo, Andrea, Ilie, Raluca, Huang, Yu, Kistler, Lynn (2021) Prediction of soft proton contamination in XMM-Newton: A machine learning approach, *The Astrophysical Journal*, 903:89, DOI: https://doi.org/10.3847/1538-4357/abbb8f.

White papers

1. Kollmann, P., Allegini, F., Allen, R. C., André, N., Azari, A. R., Bagenal, F., Beddingfield, C. B., Brain, D., Brandt, P., Cao, X., Cao, X., Cartwright, R. J., Clark, G., Cohen, I., Cooper, J. F., Crary, F., Leonard, E. J., Paty, C., Pater, I. d., Desai, R. T., DiBraccio, G. A., Dietrich, W., Dong, C., Ebert, R. W., Felici, M., Filwett, R. J., Fischer, G., Gershman, D. J., Gkioulidou, M., Greathouse, T. K., Griton, L., Gritsevich, M., Hibbitts, K., Hospodarsky,

G., Hue, V., Hunt, G., Huybrighs, H. L. F., Imai, M., Jackman, C. M., Jasinski, J. M., Jia, X., Jun, I., Kotova, A., Kurth, W. S., Lamy, L., Lazio, J., Lejosne, S., Louis, C., Masters, A., Mauk, B., Madanian, H., Mandt, K. E., McNutt, R., Melin, H., Miller, S., Moore, L., Nenon, Q., Neubauer, F. M., Nordheim, T., Palmaerts, B., Paranicas, C., Phipps, P., Regoli, L., Retherford, K., Roth, L., Roussos, E., Runyon, K. D., Rymer, A., Saur, J., Santos-Costa, D., Shprits, Y., Spilker, L. J., Stallard, T., **Smirnov, A.**, Soderlund, K., Stanley, S., Sulaiman, A. H., Szalay, J. R., Turner, D. L., Vines, S., Wang, L., Weiss, B., Wicht, J., Wilson, R. J., Woodfield, E. (2021): Magnetospheric Studies: A requirement for addressing interdisciplinary mysteries in the Ice Giant systems [White Paper], *Bulletin of the American Astronomical Society*, 53(4), DOI: https://doi.org/10.3847/25c2cfeb.d955f654.

Data and model publications

1. **Smirnov, A.**, Shprits, Y., Prol, F., Lühr, H., Berrendorf, M., Zhelavskaya, I., Xiong, C. (2023): Model files for the Neural network-based model of Electron density in the Topside ionosphere (NET). V. 1.0. *GFZ Data Services*, DOI: https://doi.org/10.5880/GFZ.2.7.2023.001.

2. **Smirnov, A.**, Shprits, Y., Allison, H., Aseev, N., Drozdov, A., Kollman, P., Reeves, G.D., Wang, D., and Saikin, A. (2022): An empirical model of electron pitch angle distributions in the outer radiation belt based on Van Allen Probes data, *GFZ Data Services*, DOI: https://doi.org/10.5880/GFZ.2.7.2022.001.

3. **Smirnov, A.G.**, Kronberg, E.A., Daly, P.W., Aseev, N.A., Shprits, Y.Y., and Kellerman, A.C. (2019): Calculations of Adiabatic Invariants and Phase Space Density for Cluster mission, *Zenodo*, DOI: http://doi.org/10.5281/zenodo.3519999.

Selected conference presentations and talks

Invited presentations at international meetings

1. Neural network model of Electron density in the Topside ionosphere (NET), *IAG Comission 4 Symposium*, 4 – 8 September 2022, Potsdam, Germany.

2. Intercalibration of the plasma density measurements in Earth's topside ionosphere, *AT-AP-RASC 2022 (3rd URSI Atlantic Radio Science Meeting)*, 29 May – 3 June 2022, Gran Canarias, Spain.

3. Storm-time evolution and empirical modeling of the equatorial electron pitch angle distributions in Earth's outer radiation belt", *Geospace Environment Modeling (GEM)-2020 workshop*, 20 – 23 July 2020, online due to the pandemic.

Oral presentations at international meetings

1. Machine learning modeling in magnetospheric physics, *ISSI team meeting*, 21 – 26 May 2023, Bern, Switzerland.

2. Neural network model of Electron density in the Topside ionosphere (NET), *11-th Swarm Data Quality Workshop*, 10 – 14 October 2022, Uppsala, Sweden.

3. Calibration of Swarm Plasma densities using neural networks, *11-th Swarm Data Quality Workshop*, 10 – 14 October 2022, Uppsala, Sweden.

4. Neural network model of Electron density in the Topside ionosphere, *D4G (1st workshop on Data science for GNSS remote sensing)*, 13 – 15 June 2022, Potsdam, Germany.

5. Calibration of Swarm Plasma densities using neural networks, *Swarm-EFI data quality meeting*, 17 July 2022, online.

6. Neural network model of Electron density in the Topside ionosphere, *44th COSPAR Scientific Assembly*, 16 – 24 July 2022, Athens, Greece.

7. "Equatorial electron pitch angle distributions in Earth's outer radiation belt: Storm-time evolution and empirical modeling", *EGU22*, 24 – 28 May 2022, Vienna, Austria.

8. "Intercalibration of the plasma density measurements in Earth's topside ionosphere", *AGU-2021*, 13 – 17 December 2021, New Orleans, USA (presented online).

9. "Equatorial electron pitch angle distributions in Earth's radiation belts and ring current: Van Allen Probes observations", *43rd COSPAR Scientific Assembly*, 28 January – 4 February 2021, online due to the pandemic.

10. "Medium energy electron fluxes in Earth's outer radiation belt: a machine learning model", *AGU-2019*, 9 – 13 December 2019, San Francisco, USA.

Poster presentations at international meetings

1. A neural network model of Earth's topside ionosphere, *EGU-2023*, 23 – 28 April 2023, Vienna, Austria. *Received the Outstanding Student and PhD candidate Presentation (OSPP) award.*

2. Storm-time evolution and empirical modelling of the equatorial electron pitch angle distributions in Earth's outer radiation belt, *AGU-2021*, 13 – 17 December 2021, New Orleans, USA. *Highlighted.*

3. Medium energy electron flux in Earth's outer radiation belt (MERLIN): a machine learning model from GPS data, *43rd COSPAR Scientific Assembly*, 28 January – 4 February 2020, online due to the pandemic.

4. Continuous 3D model of ionospheric electron density based on machine learning, *AGU-2020*, 1 – 17 December, online due to the pandemic.

5. Equatorial pitch angle distributions in Earth's radiation belts: an empirical model from Van Allen Probes data, *EGU-2020*, 4 – 8 May 2020, online due to the pandemic.

Awards and recognition

- **Young Scientist Award (YSA)** from the International Union of Radio Science (URSI), for the summary paper "A Neural network model of Electron density in Earth's Topside ionosphere (NET)" submitted to URSI-GASS-2023 conference.
- **Outstanding Student and PhD candidate Presentation (OSPP)** award at EGU-2023 for the poster entitled "A neural network model of Earth's topside ionosphere".
- **Top downloaded paper in JGR Space Physics in 2021 (first author):** Intercalibration of the plasma density measurements in Earth's topside ionosphere.
- **Top downloaded paper in JGR Space Physics in 2021 (co-author):** Preliminary statistical comparisons of spin-averaged electron data from Arase and Van Allen Probes missions.
- **Top downloaded paper in JGR Space Physics in 2021 (co-author):** Energetic charged particles in the terrestrial magnetosphere: Cluster/RAPID results.
- **Top 10% most downloaded papers in JGR Space Physics in 2020 (first author):** Adiabatic invariants calculations for the Cluster mission: a long-term product for radiation belts studies.

Contents

List of Figures

List of Tables

— 1 —

Introduction

1.1 The magnetic field of the Earth

Many of the phenomena that occur in the near-Earth space environment are possible owing to the fact that our planet is magnetized. The geomagnetic field can be observed at any point inside the planet, at the surface and in the near-Earth space, and presents a combination of internal and external sources [e.g., Olsen and Stolle, 2017]. Internal sources are located within the Earth, while external sources originate at altitudes generally above ~100 km. Over 95% of the internal field is attributed to the *core field*, which is believed to arise from a self-exciting dynamo action in the fluid outer core [e.g., Merrill and McElhinny, 1983; Gubbins and Herrero-Bervera, 2007]. The resulting field is mostly dipolar and the dipole component is tilted by around 11° from the rotation axis of the Earth. The second important component of the geomagnetic field (up to ~4% of the total intensity) is the *crustal field*, which arises due to magnetized material in regions where temperatures are below the Curie temperature. It can make a significant contribution to local measurements of the magnetic field at the ground, and has been mapped by a variety of marine, aeromagnetic and ground-based campaigns. The *external* magnetic fields can be induced by electric fields in the Earth's ionosphere and magnetosphere. The geomagnetic field can exhibit changes over a range of timescales [e.g., Constable, 2016], from millions of years for the crustal field (thus it is often assumed to be time-independent) to years for the core field, and several minutes or even seconds for external fields.

Figure 1.1 shows the total intensity of the geomagnetic field at the Earth's surface, calculated using the International Geomagnetic Reference Field (IGRF-13) model [Alken et al., 2021]. A notable region in this map, known as the South Atlantic Anomaly (SAA), is where the total field intensity exhibits a strong depletion (below a designated threshold of 32000 nT [Pavon-Carrasco and De Santis, 2016]). This depletion affects the behavior of charged particles in geospace, causing them to mirror closer to Earth over the SAA longitudes.

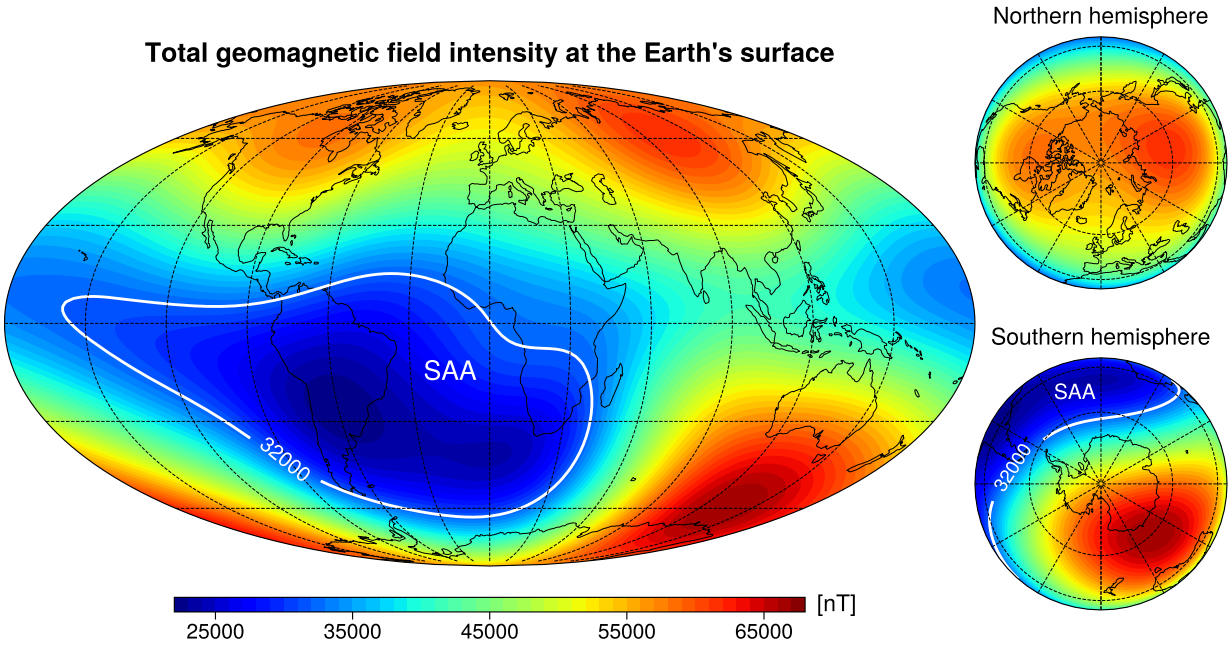

Figure 1.1. Total intensity of the geomagnetic field at the Earth's surface calculated using the IGRF-13 model for November 2022.

The magnetic field of a centered dipole can be expressed in spherical coordinates as:

$$B_r = 2\frac{M}{r^3}\cos\theta, \tag{1.1}$$

$$B_\theta = \frac{M}{r^3}\sin\theta, \tag{1.2}$$

$$B_\phi = 0, \tag{1.3}$$

where θ is the magnetic colatitude, ϕ is the longitude, r is the distance from the Earth's center and M is Earth's magnetic moment. The field's total intensity can be written as:

$$B = \frac{M}{r^3}\left(3\cos\theta + 1\right)^{1/2}. \tag{1.4}$$

For many space physics problems, it is useful to define a special coordinate system that follows the geometry of the Earth's magnetic field. It consists of three components, namely the L-shell, which defines the radial distance to a magnetic field line, magnetic latitude λ (also abbreviated as MLat), which specifies the vertical position along the field line, and magnetic local time (MLT), which represents the azimuthal location. The equation for a dipole magnetic field line can be written down as:

$$r = LR_E\cos^2\lambda, \tag{1.5}$$

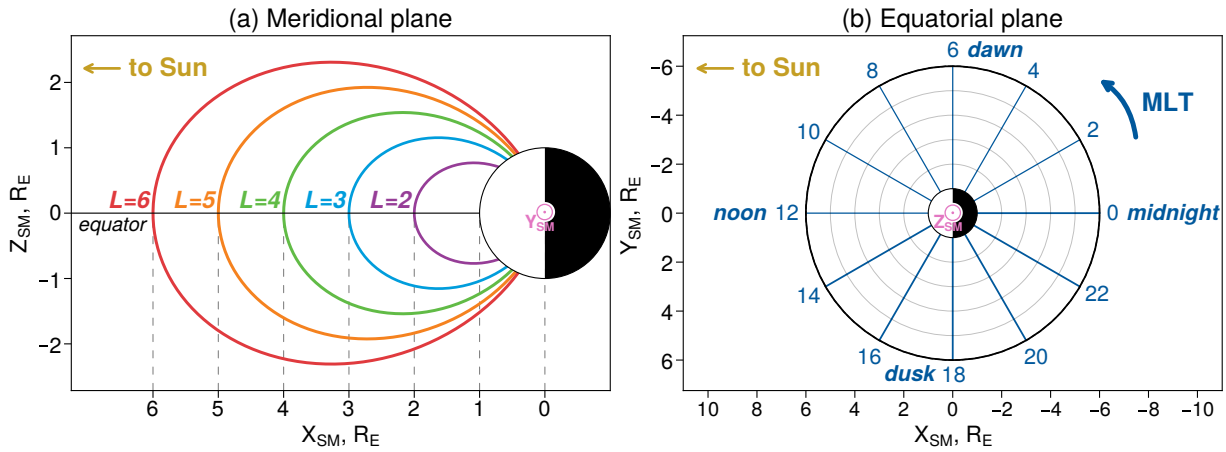

Figure 1.2. A sketch of the dipole magnetic field lines and the corresponding L-parameters (a), and MLT sectors (b) in solar magnetic coordinates.

where $R_E \approx 6371.2$ km is the Earth's radius. In this equation, L is the radius of the equatorial crossing point of a field line, measured in Earth radii (see Figure 1.2) and is referred to as the L-value. MLT, representing the azimuthal component, is measured anti-clockwise in the direction of $-X_{SM}$, where SM denotes the solar magnetic coordinate system. The subsequent Chapters will frequently employ the L-MLat-MLT coordinate system, especially in Chapters 2–5 when analyzing trapped particle populations in the inner magnetosphere.

1.2 The Sun and solar wind

The star at the center of our solar system, the Sun, radiates the light and heat necessary for sustaining life on Earth. The Sun is classified as a yellow dwarf star with an effective black body surface temperature of $\sim$ 5772 K, mass of around $1.99{\cdot}10^{30}$ kg and an average radius of about $6.96{\cdot}10^{5}$ km [Foukal, 2004]. Simply put, the Sun is a sphere of plasma under the gravitational forces that create a pressure high enough to drive a fusion reaction in its core. This reaction, in which four hydrogen atoms combine to form one helium atom, produces energy that travels for millions of years until it reaches the Sun's surface and is then released outward.

The Sun constantly emits ionized plasma that propagates into the interplanetary medium. The existence of such a flow was originally deduced from observations of comets, namely because

one of the tails usually lagged several degrees behind the comets' radial directions. Biermann [1951] showed that this flow is present at all times and has corresponding velocities in the order of hundreds of km/s. Parker [1958] named this outflow as the *solar wind* (SW) and demonstrated that it originates from the supersonic expansion of the solar corona along the solar magnetic field lines. The solar wind consists mainly of protons ($\sim$ 96%) and helium ions ($\sim$ 4%), with a small fraction of heavier ions ($<$ 1% of C, N, O, etc.). The solar wind also has a magnetic field, often called the *interplanetary magnetic field* (IMF), which is embedded into the plasma due to its high electrical conductivity. This is known as the *frozen-in* approximation [e.g., Baumjohann and Treumann, 2012]. When the solar wind reaches the Earth, it has an average flow speed of $\sim$ 400 km/s, a proton number density of around 3-10 cm^{-3}, dynamic pressure exerted onto the Earth's magnetosphere of about 3 nPa, and an average IMF strength of around 5 nT.

Observations of the Sun date back to 800 BC, and the first telescopic observations of *sunspots*, which are regions of reduced temperature at the photosphere caused by concentrations of magnetic flux, were performed in 1610. The continuous record of sunspot observations started in the 19th century. The number of sunspots exhibits variations with a period of $\sim$11 years, which is known as the *solar cycle*. The Sun largely controls our space environment, as structures such as coronal mass ejections (CMEs) and corotating interaction regions (CIRs) can transfer energy to the geospace, causing significant changes in the magnetosphere. Geomagnetic activity is known to be higher during the solar maximum and the declining phase of the solar cycle and lower during the solar minimum [Chapman and Bartels, 1940].

1.3 Regions of the near-Earth space environment

The near-Earth space environment is a region of complex interactions and boundaries [Hargreaves, 1992]. First, there are interactions between the solar wind and the geomagnetic field. The SW propagates at supersonic velocities until it encounters the magnetic field of the Earth. There, a standing shock wave, known as the *bow shock*, is created (see Figure 1.3). The bow shock is responsible for heating the solar wind plasma and slowing it down to subsonic speeds. Most of the solar wind plasma then starts to bypass the Earth through a region known as the *mag-*

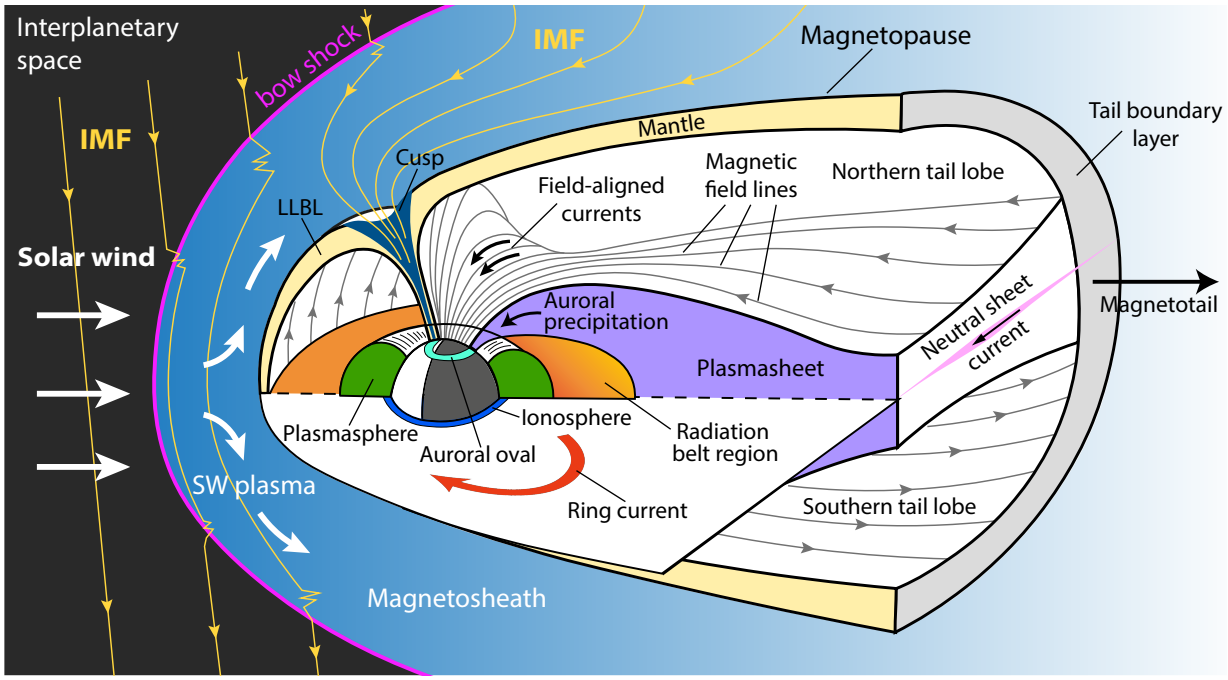

Figure 1.3. Structure of the near-Earth space environment, demonstrating different current systems and plasma regions. This figure is synthesized from Kelley [2009], Hargreaves [1992] and Kivelson and Russell [1995].

netosheath. The region of geospace that is dominated by the terrestrial magnetic field is known as the *magnetosphere* [Kivelson and Russell, 1995]. It is separated from the SW-dominated magnetosheath by a boundary known as the *magnetopause* [Kamide and Chian, 2007]. The magnetopause location is highly variable and is mainly determined by a balance between the pressure exerted by the solar wind and the magnetic pressure of the compressed terrestrial field:

$$\rho_{sw} v_{sw}^2 = \frac{B_{MS}^2}{2\mu_0}, \tag{1.6}$$

where ρ_{sw} is the SW mass density, v_{sw} represents the solar wind velocity, B_{MS} is the terrestrial magnetic field and μ_0 is the permeability of vacuum. The magnetosphere is compressed on the dayside (with stand-off distances of ~9-10 R_E), and this is associated with the magnetopause current. On the nightside, the magnetosphere is stretched up to tens-to-hundreds of R_E, and the magnetotail current exists in this long tail, one part of it flowing from dawn to dusk in the center of the tail, and the other part creating two loops (above and below the center of the magnetotail) that close the central current through the magnetopause [Ganushkina et al., 2018].

The geomagnetic field lines for which both footprints can be traced to the Earth are usually referred to as *closed field lines*, while those with only one footprint connected to the Earth and another one extending into the interplanetary space called *open field lines*. Open field lines correspond to the polar regions and form the *tail lobes*, which are 2 regions with oppositely directed magnetic fields in the tail of the magnetosphere, usually devoid of plasma. In between of the northern and southern lobes lies the *plasmasheet*, which is the main reservoir of plasma in the magnetotail [Kronberg et al., 2014]. The corresponding plasma is mainly of solar wind origin and contains electrons and ions with energies of several keV and number densities of 0.1-1 cm^{-3}. It is worth noting that the plasmasheet is linked to the high-latitude ionosphere along the field lines, and particles precipitating from the plasmasheet hit the neutral atmosphere and produce visible auroras [Chamberlain, 1961]. Plasmasheet particles can also be transported radially towards the Earth, gaining energy and becoming part of the inner-magnetospheric populations.

Although most of the solar wind is deflected around the Earth, some plasma from the solar wind manages to infiltrate the magnetosphere. At high latitudes, there are two funnel-shaped regions with near-zero magnetic field intensities, known as the *cusp regions* (Figure 1.3). They act as gateways that allow heated solar wind plasma to enter the magnetosphere directly [e.g., Reiff et al., 1977]. This entry region spans $\sim 1^{\circ}$ in latitude and 3 hours in local time (LT) and corresponds to the dayside auroral precipitation. Coupling between the solar wind and magnetosphere is furthermore enabled by the process known as the *reconnection*. When the field lines of the IMF and geomagnetic field are directed opposite to one another, they meet at the nose of the magnetosphere, triggering a pattern of magnetic field and plasma motion known as the *Dungey cycle* [Dungey, 1961]. First, the antiparallel field lines of the IMF encounter the geomagnetic field lines, and the so-called *X-line* is formed when they reconnect. This creates two field lines of the new type. Before the reconnection, one field line was closed and could be traced to its 2 footprints on Earth, and the IMF field line was stretched far away from the Earth. The reconnection creates two open field lines that are connected to the planet at one end and extend into the interplanetary space on the other end (marked as 2 and 2' in Figure 1.4). The open ends of these field lines are pushed into the magnetotail, and their footprints are swept across the polar ionosphere from noon to midnight with velocities of several hundred meters per second (marked

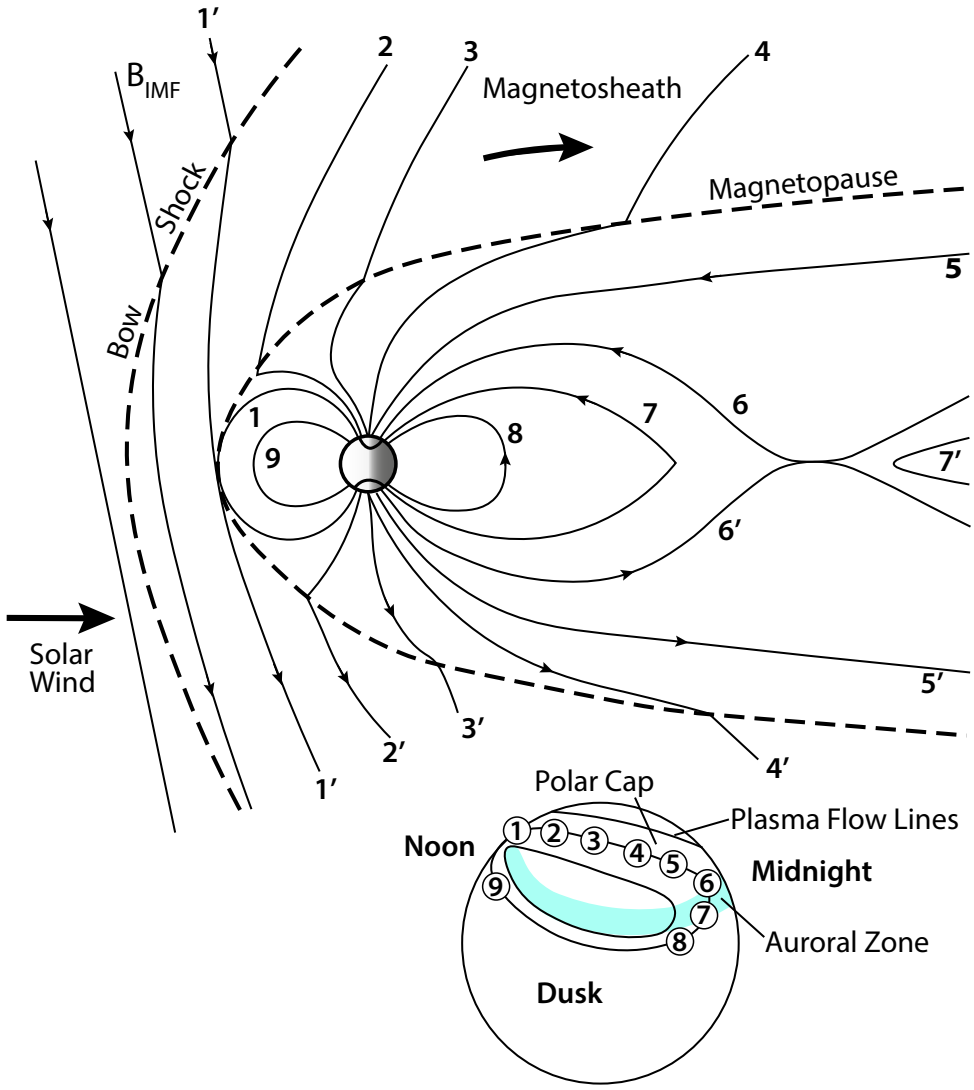

Figure 1.4. An illustration of the Dungey cycle (adapted from Kivelson and Russell [1995]).

as 3-5 in Figure 1.4). This process cannot continue indefinitely, because then all geomagnetic field lines would be open [Kivelson and Russell, 1995]. Instead, due to constant loading of magnetic flux into the magnetotail, the magnetic pressure increases and another reconnection occurs in the tail, where the field lines in the northern and southern lobes are anti-parallel (marked as 6 and 6'). The tail reconnection forms a closed line (marked as 7) and a purely interplanetary line (marked as 7'). The stretched field line (marked by 7) then relaxes and shortens in the Earthward direction (labeled as 8), returning the plasma attached to it towards the Earth [Baumjohann and Treumann, 2012]. As the particles are moved inward to lower L-shells, their ionospheric foot-

prints are dragged through the dusk or dawn side of the Earth and eventually reach the dayside (labeled as 9).

The magnetic reconnection occurring downtail causes convection of ~100 eV – 1 keV plasmasheet particles, and they are energized up to tens-to-hundreds of keV energies by adiabatic compression [DeForest and McIlwain, 1971]. At these energies, together with the inward convective transport, the particles become subject to the gradient and curvature drifts (Section 1.4.3), which depend on particle charge. Therefore, the negatively and positively charged particles drift in opposite longitudinal directions. This differential motion creates an electric current known as the *ring current*. It influences the magnetic field magnitude around the equator: ring current enhancements produce drops in the magnetic field measured at the ground, which are one of the signatures of *geomagnetic storms* [Gonzalez et al., 1994]. Particles with higher energies of hundreds of keV to several MeV form the *radiation belts*, which are toroidal regions of highly energetic but very dilute plasma consisting predominantly of energetic protons and electrons [Van Allen and Frank, 1959; Li and Hudson, 2019]. The protons form a single radiation belt with maximum flux intensities around L~3. Electron radiation belts comprise a two-zone structure, consisting of the inner and outer belts separated by the slot region [Reeves et al., 2016]. It should also be noted that during geomagnetic storms, the third radiation belt can occur as a transient structure lasting several days [Baker et al., 2013]. Around the same radial distances from the Earth, there is a toroidal region populated with a cold (1 eV) and dense (~ 10^3el./cm^3) plasma, known as the *plasmasphere*. The plasmasphere contains particles that originate in the ionosphere and flow out on closed field lines up to radial distances of 4-5 R_E. The plasmaspheric morphology is generally defined by two processes, namely, the corotation with the Earth and convection, and the balance between them gives the plasmasphere its characteristic teardrop shape. The dynamics of the inner-magnetospheric particles are further described in detail in Section 1.5.

Particles in the neutral atmosphere of the Earth can get ionized by solar radiation, mainly of ultraviolet (UV) and extreme ultraviolet (EUV) wavelengths, and this process is known as the *photoionization*. Furthermore, neutrals can experience the so-called impact ionization via interactions with energetic particles from the magnetosphere, the Sun and the galaxy [Kamide and

Chian, 2007]. These particles form the *ionosphere*, which is a partially ionized region of the upper atmosphere extending from ~80 km altitude until a smooth transition into the plasmasphere at ~1000 km. On open magnetic field lines, light ions that have high thermal energies can escape the gravity field and stream out from the polar cap to form the *polar wind*, while the particles outflowing along the closed field lines populate the plasmasphere. The ionosphere is a dispersive medium that, due to high electron concentration, affects the propagation of electromagnetic waves, including signals from the Global Navigation Satellite Systems (GNSS). Even though nowadays dual-frequency observations allow the elimination of the first-order ionospheric effects, the higher-order terms cannot be simply removed [e.g., Hoque and Jakowski, 2007]. This makes ionospheric modeling a crucial task for both space physics and geodetic applications.

The magnetosphere, ionosphere and thermosphere are inextricably tied to one another and can be considered as a joint ionosphere-thermosphere-magnetosphere (ITM) system. It is driven by the solar wind, solar irradiation and a number of feedback mechanisms between different systems within the geospace. In particular, the magnetospheric configuration and convection patterns are determined by the solar wind. The magnetosphere provides particles impacting the ionosphere through precipitation, while the ionosphere has a feedback connection back to the magnetosphere through the polar wind and particle outflows that populate the plasmasphere. The plasmaspheric waves, in turn, affect acceleration and loss of the radiation belts and ring current populations. The ionosphere and thermosphere are also strongly coupled since the thermosphere acts as the source of particles to be ionized and affects the plasma redistribution through neutral winds. It should be noted that the description above only mentions a few of a myriad of processes within the ITM system, while the full picture is much more complex and diverse, and understanding of the ITM system as a whole remains an active area of research.

1.4 Charged particle motion

Plasma can be defined as a state of matter in which a significant fraction of particles are ionized, which results in a mixture of negative and positive charge carriers, and characterized by the collective behavior [e.g., Chen et al., 1984]. Individual components of plasma possess electric

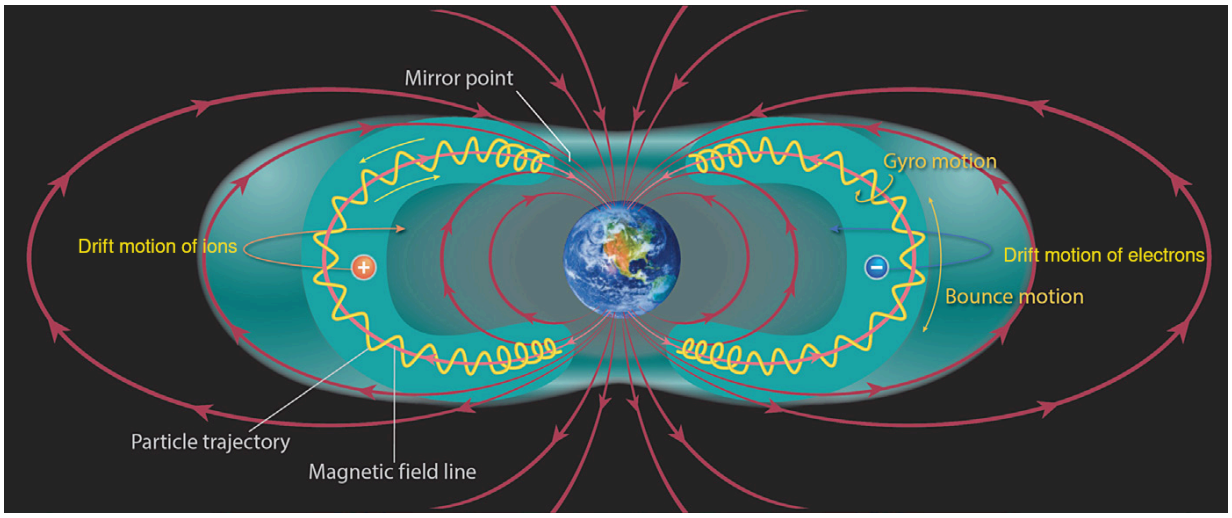

Figure 1.5. Schematics of three types of trapped particle motion in the inner magnetosphere. Figure from Ilie [2020].

charges on a microscale, but on a macroscale, plasma exhibits quasi-neutrality. Regardless of the length scale, the charged particles within plasma react to electromagnetic forces.

Charged particles moving in the magnetic (**B**) and electric (**E**) fields are subject to the Lorentz force, and, assuming that non-electromagnetic forces are negligible, the particle motion can be described as:

$$m\frac{d\mathbf{v}}{dt} = q(\mathbf{v} \times \mathbf{B} + \mathbf{E}), \tag{1.7}$$

where $\mathbf{v}$ is velocity, m is mass, and q is charge of the particle. Equation 1.7 indicates that the magnetic field only accelerates the particles perpendicular to both the magnetic field direction and particle velocity. An electric field can furthermore accelerate the particles parallel to the direction of **E** [Walt, 2005]. Charged particles in the dipole-like geomagnetic field undergo three distinct types of periodic motion with very different timescales (Figure 1.5), namely, the gyration around the magnetic field line, bounce motion between the two mirror points, and drift around the Earth [Schulz and Lanzerotti, 1974].

1.4.1 Cyclotron motion

In a simplified case when no electric field acts on the particles ($\mathbf{E} = 0$), the magnetic Lorentz force determines the particle motion. If one also assumes that the magnetic field is purely vertical

($\mathbf{B} = B\hat{z}$), then the velocity components can be expressed as:

$$v_x = v_\perp \cos \omega_g t, \quad v_y = v_\perp \sin \omega_g t, \quad v_z = \text{const} = v_\parallel, \tag{1.8}$$

where $v_\perp = (v_x^2 + v_y^2)^{1/2}$ is the velocity in the plane orthogonal to $\mathbf{B}$, and ω_g denotes the cyclotron (or, gyro-) frequency, i.e., the rate at which a particle gyrates around the magnetic field line [e.g., Baumjohann and Treumann, 2012]. It depends on the particle charge q that defines the direction of rotation, particle mass m, as well as magnetic field magnitude B:

$$\omega_g = \frac{qB}{\gamma m}, \tag{1.9}$$

where γ is the Lorentz factor. Particles gyrate about the magnetic field line with the cyclotron radius:

$$r_g = \frac{v_\perp}{|\omega_g|} = \frac{m\gamma v_\perp}{|q|B}, \tag{1.10}$$

and the center of the cyclotron motion is referred to as the *guiding center*. When $v_\parallel \neq 0$, particles move along the geomagnetic field line as they gyrate, and the overall particle trajectory takes helical form.

1.4.2 Bounce motion

The terrestrial magnetic field is non-uniform, as the magnetic field lines converge towards the poles. As a result, the field creates a force known as the *magnetic mirror force* [e.g., Baumjohann and Treumann, 2012]:

$$\mathbf{F}_\mathrm{m} = -\frac{p_\perp}{2mB} \nabla_\parallel B(s), \tag{1.11}$$

where s denotes the coordinate along the magnetic field line, and $p_\perp$ is the perpendicular component of the relativistic momentum. It is related to the total momentum through:

$$p_\perp = |p| \sin \alpha, \tag{1.12}$$

where α shows an angle between the particle momentum and magnetic field, referred to as the particle *pitch angle* (PA). When the particle moves into a stronger magnetic field, $\mathbf{F}_\mathrm{m}$ acts in a direction opposite to particle velocity and tries to deflect the particle. The total kinetic energy

is conserved, and therefore, the particle's parallel velocity decreases while the perpendicular velocity increases. Finally, when the parallel velocity is zero, the particle reverses its direction at a point known as the *mirror point*. The particle is then reflected again in the opposite hemisphere, and this back-and-forth motion between the mirror points in the two hemispheres is known as the *bounce motion*. The corresponding bounce period is defined as the time taken for the particle to travel from one mirror point to its conjugate in the opposite hemisphere and back [Lyons and Williams, 2013].

If the magnetic field changes are negligible over the gyroradius, then no work is done on a particle, and the magnetic flux through the particle orbit ($\Phi = B\pi r_g^2$) does not change. Then, for two arbitrary locations (denoted as 1 and 2) on a field line:

$$\frac{\sin^2 \alpha_1}{B_1} = \frac{\sin^2 \alpha_2}{B_2}. \tag{1.13}$$

Equation 1.13 provides an important relation between particles at different locations along the given field line. In particular, it is possible to relate the local pitch angle of a particle (α_{loc}), with the equatorial pitch angle (α_{eq}), although it requires knowledge of the magnetic field values at both points. The equatorial pitch angle can be expressed as:

$$\alpha_{\mathrm{eq}} = \arcsin\left(\sqrt{B_{\mathrm{eq}}/B_{\mathrm{loc}}} \sin \alpha_{\mathrm{loc}} \right). \tag{1.14}$$

It is also possible to express the equatorial pitch angle in terms of the magnetic field values at the equatorial and mirror points:

$$\alpha_{\mathrm{eq}} = \arcsin \sqrt{B_{\mathrm{eq}}/B_{\mathrm{m}}} = \frac{\cos^6 \lambda_{\mathrm{m}}}{(1 + 3\sin^2 \lambda_{\mathrm{m}})^{1/2}}, \tag{1.15}$$

assuming a dipole field configuration [Walt, 2005]. It is of note that the equatorial PA is independent of L-shell and only depends on latitude of the mirror point (λ_{m}). Particles with high equatorial PAs bounce close to the equatorial plane, while particles with small α_{eq} have high parallel velocities and mirror at high latitudes [Baumjohann and Treumann, 2012]. If the mirror points are deep within the atmosphere, the particle will collide with neutrals and be absorbed by the atmosphere before being reflected back. The range of equatorial pitch angles that are too small to be mirrored back before the particles are lost by collisions is known as the *loss cone*.

1.4.3 Drift motion

Electric drifts

In Sections 1.4.1 and 1.4.2, electric field was assumed to be zero. Adding an electric field to the magnetic field results in electric drift motion. For instance, an electric field that only has a parallel component ($\mathbf{E}_\parallel$) yields acceleration along the magnetic field:

$$m\dot{\mathbf{v}}_\parallel = q\mathbf{E}_\parallel. \tag{1.16}$$

It is of note that this acceleration depends on particle charge, and therefore positively and negatively charged particles are accelerated in the opposite directions [Kivelson and Russell, 1995]. This parallel electrostatic electric field is rarely sustained, due to the fact that particles can move quite freely along the field lines and will rearrange to neutralize the field.

When an electric field is perpendicular to the magnetic field ($\mathbf{E} \perp \mathbf{B}$), a velocity increase due to $q\mathbf{E}$ also increases the $q\mathbf{v}_\perp \times \mathbf{B}$ force. The balance between the magnetic and electric Lorentz forces yields a drift of the guiding center of the particles at a velocity $\mathbf{v}_E$:

$$\mathbf{v}_E = \frac{\mathbf{E} \times \mathbf{B}}{B^2}. \tag{1.17}$$

The corresponding net particle motion is referred to as the $\mathbf{E} \times \mathbf{B}$, or *E-cross-B*, drift. It is orthogonal to both electric and magnetic fields and is independent of charge, i.e., positive and negative ions drift in the same direction despite the opposite directions of the gyromotion.

Electric field perpendicular to the geomagnetic field can be produced due to the planet's daily rotation. Both the magnetic field lines and the neutral atmosphere corotate with the Earth. Therefore, plasma that is tied to either of them is also forced into corotation with the planet (this mainly concerns ionospheric flux tubes in the mid-latitude region and their magnetospheric extensions). The ionosphere is not fully ionized, nor is it completely frozen-in due to finite conductivities [Axford and Hines, 1961]. However, the frequent ion and electron-neutral collisions force the ionosphere to also corotate with the Earth. This motion of charged particles can be described in terms of the so-called *corotation electric field*:

$$\mathbf{E}_{cr} = -(\boldsymbol{\Omega}_E \times \mathbf{r}) \times \mathbf{B}, \tag{1.18}$$

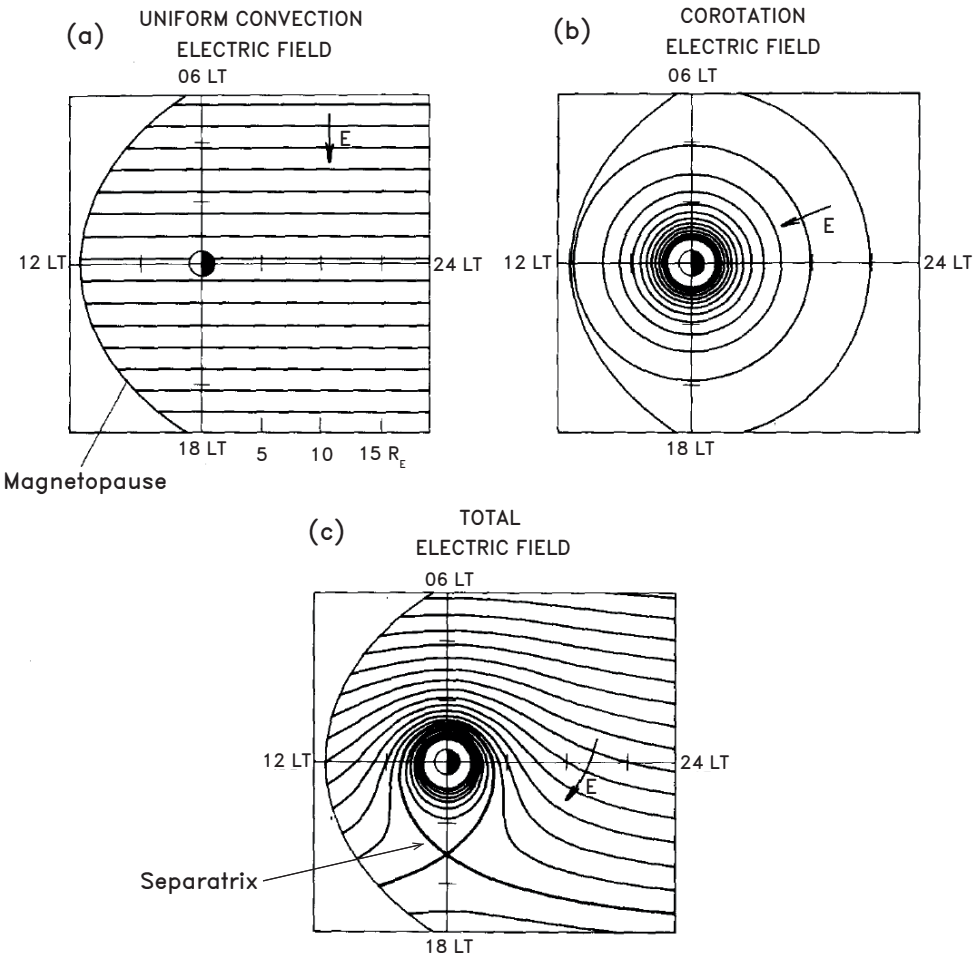

Figure 1.6. Equipotential contours for magnetospheric electric fields in the equatorial plane, adapted from Lyons and Williams [2013]. The convection electric field is assumed to have a strength of 2.5×10^{-4}V/m.

where $\mathbf{r}$ represents the radial position vector, and $\Omega_E = 7.27 \cdot 10^{-5}$ rad/s is the angular velocity of the Earth. Under the simplifying assumption of a centered dipole magnetic field, in the equatorial plane $\mathbf{E}_{cr}$ is directed toward the Earth (Figure 1.6b) and decreases with distance proportional to $1/r^2$. The corotation velocity can be expressed as:

$$\mathbf{v}_{cr} = \boldsymbol{\Omega}_E \times \mathbf{r}, \tag{1.19}$$

which means that the velocity of charged particles is in the same direction as the Earth's rotation.

Another mechanism that produces large-scale electric field is the magnetospheric convection (see also Section 1.3). The flow of plasma across the open field lines in the polar cap leads to

an electric field $\mathbf{E} = -\mathbf{v}_{sw} \times \mathbf{B}$ [Kivelson and Russell, 1995; Baumjohann and Treumann, 2012]. This field is known as the *convection electric field*; it flows from dawn to dusk (see Figure 1.6a) and creates an $\mathbf{E} \times \mathbf{B}$ drift in the anti-sunward direction.

Figure 1.6c shows a combination of the corotation and convection electric fields. If one does not take the magnetic drifts (described below) into account, the charged particles close to the Earth will corotate with the planet and move on closed drift paths. At larger radial distances, the dawn-to-dusk convection electric field dominates the drift motion, and the particles move on open drift paths from the magnetotail in the direction of the Earth, then escaping the magnetosphere through the dayside magnetopause. The main effect of convection is that it constrains radial distances at which corotation is effective and furthermore deforms the drift orbits away from their circular shapes. It should also be noted that only the background component of magnetospheric plasma follows the quasi-potential lines shown in Figure 1.6, while the more energetic populations follow the orbits that deviate from these contours, as they are influenced by the magnetic (gradient and curvature) drifts.

Magnetic drifts

Any linear force added to the Lorentz force can generate a drift motion in the ambient magnetic field. The gravity force can be viewed as one example of an external force, however, its effects are typically negligible in the near-Earth space. Some of the most important forces arise due to the non-uniform structure of the terrestrial magnetic field. For instance, the field strength increases with decreasing radial distance. This gradient in the magnetic field, ∇B, results in a non-uniform Lorentz force throughout the gyromotion (i.e., half of a gyration is within a stronger field than the second half). This creates a drift perpendicular to both the magnetic field and its gradient, referred to as the *gradient drift* with a velocity $\mathbf{v}_G$:

$$\mathbf{v}_G = \left(\frac{m v_\perp^2}{2qB^3} \right) \mathbf{B} \times \nabla B. \tag{1.20}$$

The gradient drift is concerned with the perpendicular velocity of the particle ($v_\perp$). In the terrestrial magnetic field, particles also move parallel to the curved field lines ($v_\parallel$) and therefore experience a centrifugal force $\mathbf{F}_c = m v_\parallel \mathbf{R}_C / R_C^2$. The strength of this force is dependent on the

local radius of curvature R_C. This produces a drift that is perpendicular to both the direction of the background field and centrifugal force, known as the *curvature drift*:

$$\mathbf{v}_C = \left(\frac{mv_\parallel^2}{qR_C^2}\right) \frac{\mathbf{R}_C \times \mathbf{B}}{B^2}. \tag{1.21}$$

From Equations 1.20 and 1.21 it can bee seen that both $\mathbf{v}_G$ and $\mathbf{v}_C$ depend on particle charge, which indicates that species with opposite charges will drift in opposite directions. The terrestrial magnetic field results in both gradient and curvature drifts, and considering that in a near-dipole field $\nabla B = -(B/R_C^2)\mathbf{R}_C$, they can be viewed together as one drift motion known as the *gradient-curvature drift*:

$$\mathbf{v}_B = \mathbf{v}_G + \mathbf{v}_C = \left(v_\parallel^2 + \frac{1}{2}v_\perp^2\right) \frac{\mathbf{B} \times \nabla B}{\omega_g B^2}. \tag{1.22}$$

1.4.4 Adiabatic invariants and phase space density

Motion of trapped particles in the inner magnetosphere can be decomposed into three aforementioned types of periodic motion (Sections 1.4.1, 1.4.2, 1.4.3), which have very different timescales. For instance, MeV electrons in the inner magnetosphere have characteristic times that are in the order of miliseconds for the cyclotron motion, seconds for the bounce motion and minutes for the drift motion. Each type of motion is associated with an adiabatic invariant that remains roughly constant as long as temporal changes acting on the system are slow compared to the periods of the respective motions. In general, an adiabatic invariant can be written down as an integral taken over a period of the corresponding particle trajectory:

$$I = \frac{1}{2\pi} \oint p dr, \tag{1.23}$$

where p and r are the conjugate momentum and position coordinates [Landau and Lifshits, 1960].

The cyclotron motion has a corresponding adiabatic invariant which can be expressed as:

$$\mu = \frac{p_\perp^2}{2m_0 B}, \tag{1.24}$$

where $p_\perp$ is the relativistic momentum perpendicular to the magnetic field direction, and m_0 is the particle rest mass. It is of note that at the mirror point, $p_\perp = p$, and if μ is conserved and no

electric field acts on the particles, the magnetic field magnitude B_m is also conserved. The first adiabatic invariant is typically violated for processes that interact with particle gyration [Schulz and Lanzerotti, 1974].

The second adiabatic invariant, associated with the periodic bounce motion along the magnetic field line, is the integral of the parallel moment $p_{\parallel}$ along magnetic field between the two mirror points (s_m and s'_m) along a given field line:

$$J = 2\int_{s_m}^{s'_m} p_{\parallel} ds = 2\frac{p}{\sqrt{B_m}}\int_{s_m}^{s'_m} \sqrt{B_m - B(s)} ds. \tag{1.25}$$

where B_m and $B(s)$ are magnetic field values at the mirror point and satellite location, respectively. It is also convenient to replace J with another quantity:

$$K = \frac{J}{\sqrt{8m_0\mu}} = \int_{s_m}^{s'_m} \sqrt{B_m - B(s)} ds, \tag{1.26}$$

which is independent of particle energy (unlike the invariant J) and is only a function of geomagnetic field geometry.

The third invariant is the amount of magnetic flux enclosed by the drift shell of a particle [Schulz and Lanzerotti, 1974], and can be expressed as

$$\Phi = \oint \mathbf{B} \cdot d\mathbf{S}, \tag{1.27}$$

where $d\mathbf{S}$ is an area swept over the drift period. Equation 1.27 implies that the total flux enclosed in a particle drift shell remains constant when the perturbations in the magnetic field are slower than the particle drift period. The $L^\star$ parameter, which is defined as:

$$L^\star = \frac{2\pi B_E R_E^2}{\Phi}, \tag{1.28}$$

B_E is the magnetic field value at the equator, is often used instead of Φ. In a perfect dipole field, $L^\star$ is equal to the L-shell parameter given in Equation 1.5.

The quantity often used to describe the trapped particle radiation is the particle flux, measured by the satellite detectors. Let us denote the sensor area as A (in units of cm^{-2}), acceptance solid angle as Ω (in steradians, sr) pointing in the direction θ, and energy as E (in keV). Then, number

of electrons passing through an element of area, element of solid angle, and energy interval $[E, dE]$ in 1 second is expressed as:

$$dN(\mathbf{r}, E, \theta) = j(E, \theta)\, dA\, dE\, d\Omega, \tag{1.29}$$

where $j(E, \theta)$ is the directional differential particle flux in units of $\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}\mathrm{keV}^{-1}$ [Walt, 2005].

The values of electron flux are measured at the satellite location. At the same, it is useful to know how the particle distributions at different locations along the drift or bounce orbits are connected. Liouville's theorem helps to describe the so-called phase space density, denoted as F, which represents a number of particles in a unit volume of a six-dimensional space composed of 3 orthogonal spatial dimensions $\mathbf{r} = (x, y, z)$ and their conjugate momenta $\mathbf{p}$. The corresponding phase space density (PSD) is related to particle flux as:

$$F(\mathbf{r}, \mathbf{p}) = \frac{j(\alpha, E)}{p^2}. \tag{1.30}$$

It is evident from Equation 1.30 that if p^2 remains constant, no changes in flux will be observed along a dynamical path [Walt, 2005]. PSD is a very important quantity in inner magnetospheric research, as it provides a diagnostic tool that allows to separate the non-adiabatic from adiabatic processes and distinguish between different acceleration and loss mechanisms.

1.5 Dynamics of trapped particles in the inner magnetosphere

The inner magnetosphere comprises three main particle populations, namely, the cold ($\sim$ 1 eV) and relatively dense particles in the *plasmasphere*, the more energetic (tens-to-hundreds of keV) particles in the *ring current* and the highly energetic (hundreds-of-keV to multi-MeV) populations in the *radiation belts*. It should be emphasized that the three populations have been historically subdivided in this manner due to the difference in instruments used to observe them [Kamide and Chian, 2007], while they overlap in location and their dynamics are strongly coupled.

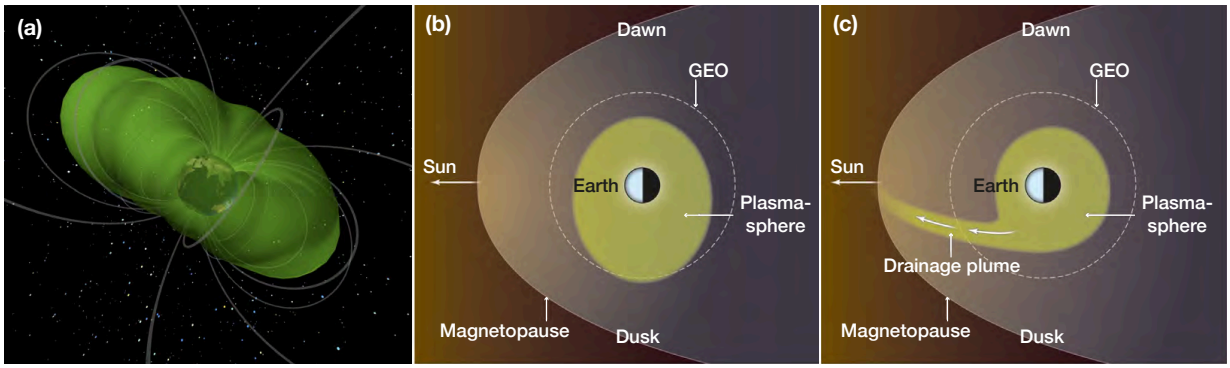

Figure 1.7. (a) Schematic three-dimensional representation of the plasmasphere (image credit: NASA). The two panels to the right show a schematic view of the plasmasphere from above the north pole (b) during quiet times, and (c) following the period of strong convection which shows an eroded plasmasphere with a sunward-convecting plume (adapted from Borovsky [2014]).

The plasmasphere (Figure 1.7) is the innermost region of the magnetosphere [Koskinen and Kilpua, 2022]. It is filled with plasma outflowing from the low- and mid-latitude ionosphere along the closed geomagnetic field lines, and can therefore be considered an upward extension of the ionosphere. Due to the low energy of plasmaspheric particles, the $\mathbf{E} \times \mathbf{B}$ drift dominates over the gradient and curvature drifts [Kamide and Chian, 2007]. The plasmaspheric dynamics are, therefore, mainly controlled by an interplay between the corotation and convection electric fields, and the particles generally follow the contours of constant electrostatic potential shown in Figure 1.6. The overall electron densities in the plasmasphere are of the order of 10^3 el./cm^3, which is larger than in other parts of the magnetosphere. The outer edge of the plasmasphere, where densities drop roughly by a factor of 5 over ~0.5 L, is known as the *plasmapause*.

The morphology of the plasmasphere is strongly influenced by geomagnetic activity. In particular, during geomagnetically quiet times the plasmasphere can extend to radial distances of up to ~6 R_E, while during geomagnetically disturbed conditions the convection electric field causes erosion of the plasmasphere. The plasmapause then moves much closer to Earth, sometimes down to 2 R_E [Baker et al., 2004]. One of the most notable features of the plasmaspheric dynamics is the formation of *plumes*, which are outward-looking teardrop structures typically developed at duskside (see, e.g., Figure 1.7c). Plasmaspheric plumes are very common during active conditions, but can also form at quiet times [Borovsky and Denton, 2008; Moldwin et al., 2016]. The plumes serve as one of the main means of transport of the plasmaspheric material to

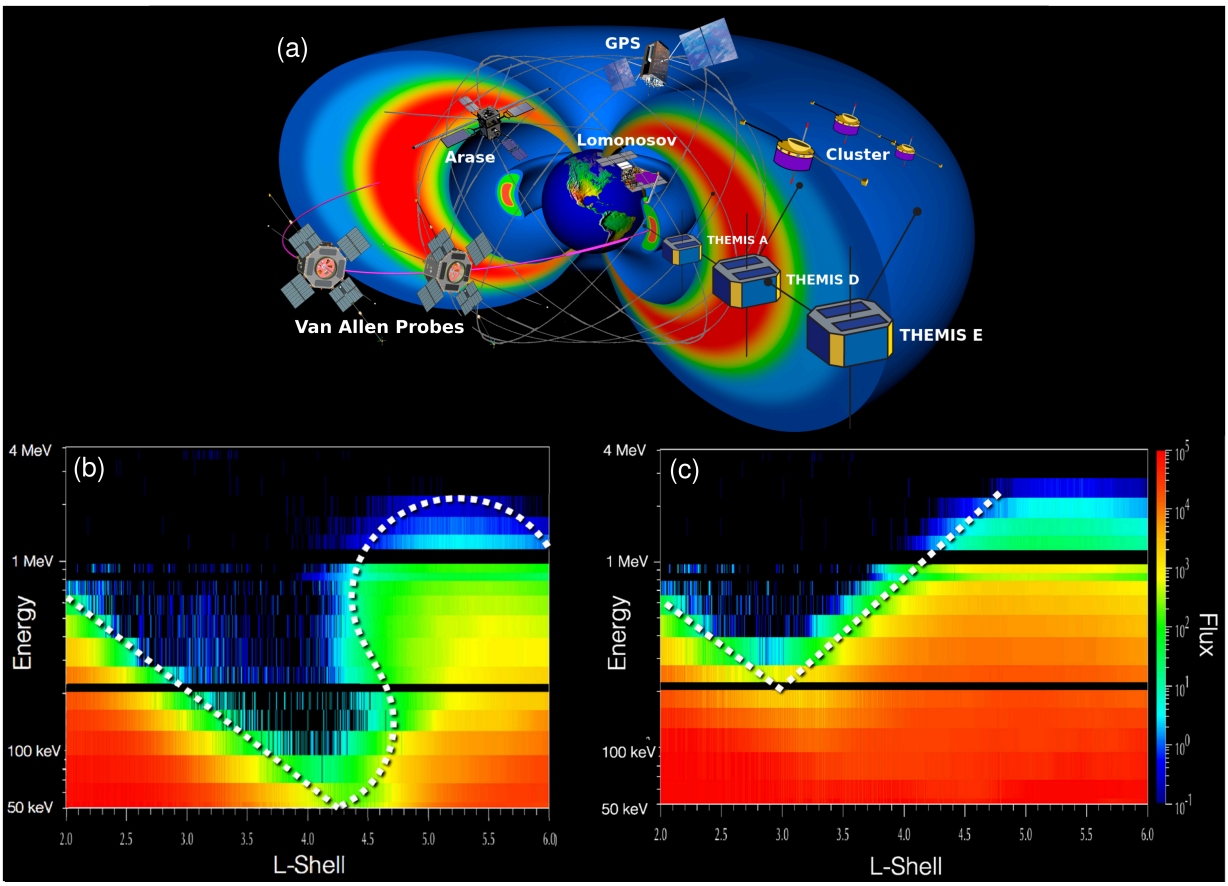

Figure 1.8. (a) A schematic view of Earth's radiation belts and several missions providing energetic particle data (image courtesy of Prof. Shprits; the background image was obtained from Johns Hopkins University Applied Physics Laboratory). Electron fluxes as a function of L-shell and energy are shown in the bottom panels for quiet (b) and active (c) conditions (Figure from Reeves et al. [2016]).

large radial distances during storms, sometimes reaching beyond the geostationary orbit [Goldstein et al., 2019]. The processes acting within the plasmasphere, around the plasmapause and within the plumes are crucial for understanding the wave-particle interactions in the radiation belts and ring current, and significantly affect their dynamics as well [Koskinen and Kilpua, 2022].

Particles that form the ring current and radiation belts have higher energies and are subject to very different kinematics [Kamide and Chian, 2007]. Particles with energies of tens-to-hundreds of keV are subject to both the $\mathbf{E} \times \mathbf{B}$ and gradient-curvature drifts. Since the magnetic drifts depend on particle charge, electrons and ions travel around the Earth in opposite directions. This differential motion gives rise to the ring current at radial distances from ~2 R_E to around 9 R_E.

The ring current particles mainly originate from the plasmasheet and ionosphere [Daglis et al., 1999]. In particular, ions with energies around 1-10 keV in the plasmasheet move toward the Earth and experience an adiabatic increase in energy of up to tens-to-hundreds of keV [Baumjohann and Treumann, 2012]. The particles outflowing from the ionosphere can be affected by a series of acceleration mechanisms that raise their energy from several eV to tens of keV, making them an important ring current constituent [Daglis et al., 1999]. It is of note that all trapped particle populations take part in the ring current formation, but the main contribution to the total current density comes from H^+, He^+ and O^+ ions with energies up to several hundred keV [Kronberg et al., 2014; Ganushkina et al., 2017], while electrons provide up to 10-20% of the total energy. In accordance with the Ampére law, the ring current generates a magnetic field that distorts the geomagnetic field around the equator. This is well reflected in the Disturbance storm-time (Dst) index which represents the hourly averaged perturbations of the horizontal component of the geomagnetic field measured near the equator [Mayaud, 1993] and is routinely used as a proxy of geomagnetic storms in many space weather studies [Borovsky and Shprits, 2017].

The Earth's radiation belts exhibit a two-zone structure. The inner belt lies between ~1.2 and ~3 L and mainly consists of highly energetic protons with energies of up to 10-100 MeV [Selesnick and Albert, 2019], but energetic electrons with energies up to ~800 keV are also present in the inner zone. The energetic protons are produced by the cosmic ray albedo neutron decay (CRAND), while electrons are mainly present due to injections, which are common at energies below ~600 keV but rarely happen for MeV electrons [e.g., Claudepierre et al., 2019]. Observations from the Van Allen Probe mission revealed that electrons with energies higher than ~ 800 keV are generally absent in the inner zone [Fennell et al., 2015]. The outer radiation belt spans from ~3 to ~7-8 L-shell and mainly consists of electrons with energies up to 10 MeV. The outer belt is extremely dynamic [Reeves et al., 2003], and its morphology depends substantially on solar and geomagnetic activity [Meredith et al., 2006; Reeves et al., 2016]. The two belts are separated by the slot region that is usually devoid of electrons and formed by a balance between the pitch angle scattering loss of electrons to the atmosphere and the inward radial diffusion [Lyons and Thorne, 1973]. The slot region can be filled during geomagnetically active times (Figure 1.8c).

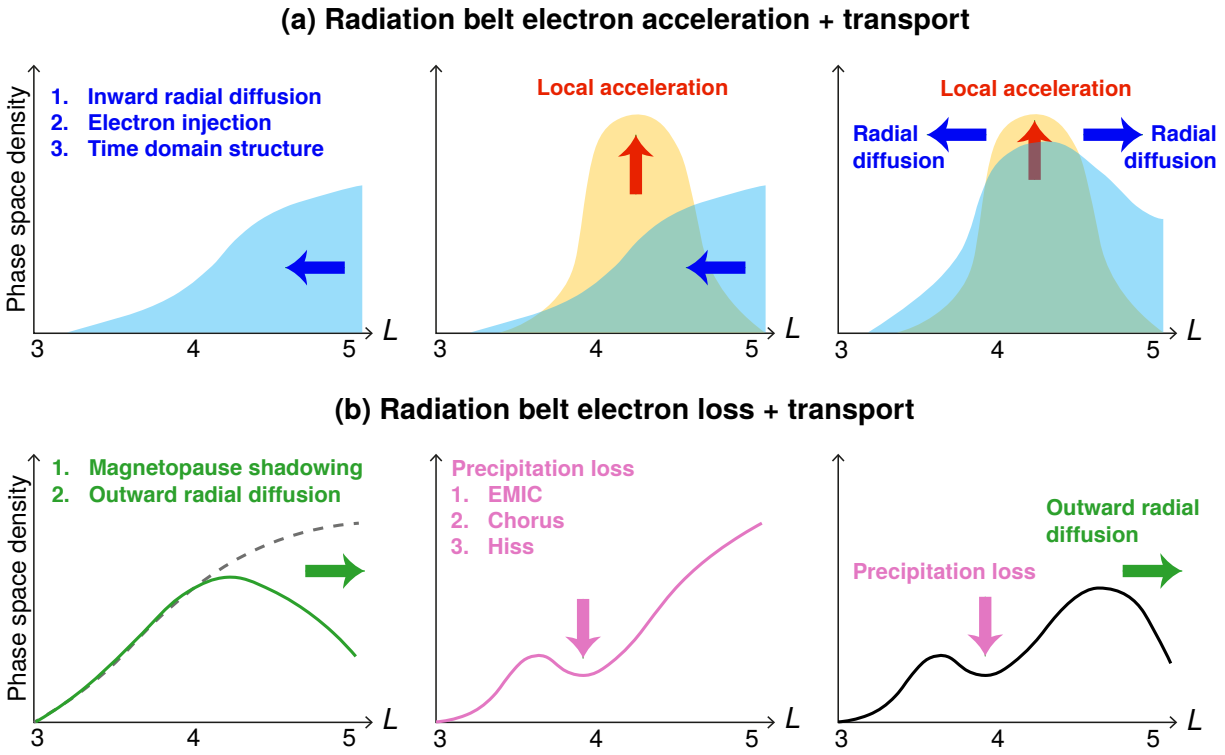

Figure 1.9. Physical processes leading to radiation belt (a) acceleration and transport, and (b) loss and transport. Figure adapted from Li and Hudson [2019].

For the high-energy radiation belt particles, the $\mathbf{E} \times \mathbf{B}$ drift is negligible, and the transport of particles through the magnetic field is determined by a sum of longitudinal gradient and curvature drifts. Electrons drift eastwards around the Earth, while for protons the gradient-curvature drift velocities are directed westward (Fig. 1.5). If no perturbing forces were present, the adiabatic invariants would be conserved and the particles would remain in stable orbits indefinitely. The adiabatic description of trapped particle motion may explain why the radiation belts persist after they are formed [Kamide and Chian, 2007]. The formation and decay of radiation belts and ring current, however, are associated with processes that *violate* adiabatic invariants. For instance, the third invariant is the easiest to violate, and its violation is associated with radial diffusion. Violation of the first and second invariants is typically connected with pitch angle diffusion and local acceleration.

Electrons with energies of hundreds of keV can be considered as a *seed population* for radia-

tion belts [Jaynes et al., 2015]. They are usually provided through substorm injections, enhanced magnetospheric convection and inward radial diffusion. The seed electrons can be further accelerated to MeV energies by local acceleration [e.g., Li and Hudson, 2019; Allison and Shprits, 2020]. One of the most efficient local acceleration mechanisms is through wave-particle interactions with whistler mode chorus waves. This typically creates peaks in the radial phase space density profiles around the heart of the outer belt (L$\sim 4-5$). These newly energized electrons are further redistributed by the radial diffusion due to ultra-low frequency (ULF) waves [Li and Hudson, 2019], which acts to smooth out the PSD peaks. Such a redistribution accelerates particles that are moved inwards and decelerates the ones that move outwards.

Trapped electrons can be lost, for instance, when increased solar wind dynamic pressure compresses the magnetosphere and particles cross the magnetopause before completing their drift orbit. This process, known as *magnetopause shadowing*, creates negative gradients in PSD. The subsequent outward radial diffusion transports the particles to larger distances and depletes the radiation belts [Shprits et al., 2006]. Electrons can also be precipitated locally into the atmosphere through interactions with different types of magnetospheric waves, such electromagnetic ion cyclotron (EMIC), hiss, chorus [see reviews by Shprits et al., 2008a,b; Millan and Thorne, 2007; Friedel et al., 2002]. The combined losses due to pitch angle scattering and magnetopause shadowing can create multiple localized PSD dropouts (Figure 1.9).

Particle populations in the inner magnetosphere are hazardous for satellite operation. In particular, particles with energies of tens-to-hundreds of keV can accumulate at the surface of the satellite and cause electrostatic discharge which may lead to satellite failure [e.g., Boyd et al., 2014]. Particles with higher energies (above 1 MeV) can penetrate into spacecraft components and charge the internal materials of the satellites [Fennell et al., 2000]. This deep dielectric charging can create electric fields that reach breakdown levels, and the following discharge can severely damage the onboard electronics [Baker et al., 1987; Hastings and Garrett, 2004]. Therefore, accurate modeling of these populations at all energy ranges and locations is crucial for mitigating the adverse space weather effects on human infrastructure in space.

1.6 The ionosphere

1.6.1 Properties of the neutral atmosphere

The Earth is blanketed by the atmosphere, which constitutes of neutral gas that is gravitationally bound by the planet. The atmosphere can be described in terms of 4 main quantities: pressure, density, temperature and composition [Hargreaves, 1992]. For a single gas species, the mass density ρ is a product of the average molecular mass m and number density N. The weight of the gas in the cylinder of a unit cross-section and height dh (Figure 1.10) corresponds to the downward force ($-\rho g dh$) that is balanced by the pressure difference between the top and bottom surfaces (dp):

$$dp = -\rho g dh = -Nmgdh. \tag{1.31}$$

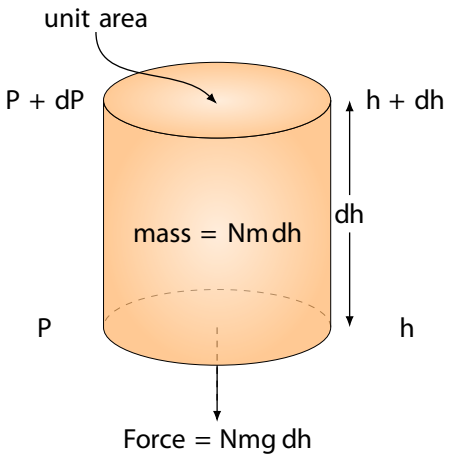

Figure 1.10. Forces acting on an atmospheric column of gas (adapted from Hargreaves [1992]).

At the same time, the neutral atmosphere is subject to the ideal gas law that relates pressure to number density and kinetic gas temperature as follows:

$$p = NkT, \tag{1.32}$$

where k is the Boltzmann's constant and T stands for absolute temperature. Therefore, one can divide the two above-mentioned equations by one another and arrive at a first-order differential equation of the form:

$$\frac{dp}{p} = \frac{-mgdh}{kT} = -\frac{dh}{H}. \tag{1.33}$$

One can define a quantity, referred to as *scale height*, which has dimensions of length and is expressed as:

$$H = \frac{kT}{mg}. \tag{1.34}$$

Integration of Equation 1.34 leads to the *barometric equation*:

$$p = p_0 \exp\left(-\frac{h - h_0}{H}\right), \tag{1.35}$$

where p_0 is the pressure at a reference altitude h_0. A similar relation can be written for number density:

$$N = N_0 \exp\left(-\frac{h - h_0}{H}\right). \tag{1.36}$$

It is also possible to define a dimensionless quantity known as the *reduced height* z [e.g., Rishbeth and Garriott, 1969]:

$$z = \frac{h - h_0}{H}. \tag{1.37}$$

It should be noted that no matter what distribution the atmospheric gas has, its pressure at height h_0 equals to the weight of gas in the column above h_0, and the total number of particles is $N_{total} = N_0 H_0$. If the atmosphere above h_0 was compressed to a uniform pressure p_0, its vertical extent would be the scale height H_0 [Rishbeth and Garriott, 1969], and this holds for any choice of the reference altitude h_0.

Equations 1.35 and 1.36 state that pressure and number density decay with altitude exponentially. Figure 1.11 shows the height distribution of atmospheric pressure, density and temperature given by the Committee On Extension to the Standard Atmosphere (COESA-1976) model [Minzner, 1977]. One can see that temperature only drops uniformly up to around 10 km altitude, in a region known as the *troposphere*. After its boundary, the *tropopause*, there is a region known as the *stratosphere* that corresponds to increasing temperature reaching around 0° C at the *stratopause* (around 50 km in altitude). Such an inverted temperature profile is due to the fact that the molecules such as ozone can strongly absorb the ultraviolet radiation from the Sun [Hargreaves, 1992]. The *mesosphere* is located between ~50 and ~90 km and corresponds to decreasing temperatures, down to −80° around the *mesopause*. The *mesopause* also roughly coincides with the *turbopause*, i.e., a boundary which separates the region of turbulent mixing lying below from the region above the turbopause where the particles stratify according to mass density. At greater heights, there is a region known as the *thermosphere*, which correponds to exponentially increasing temperatures due to solar heating. At altitudes above ~600 km, individual atoms can escape the gravity field, and this region is called the *exosphere*.

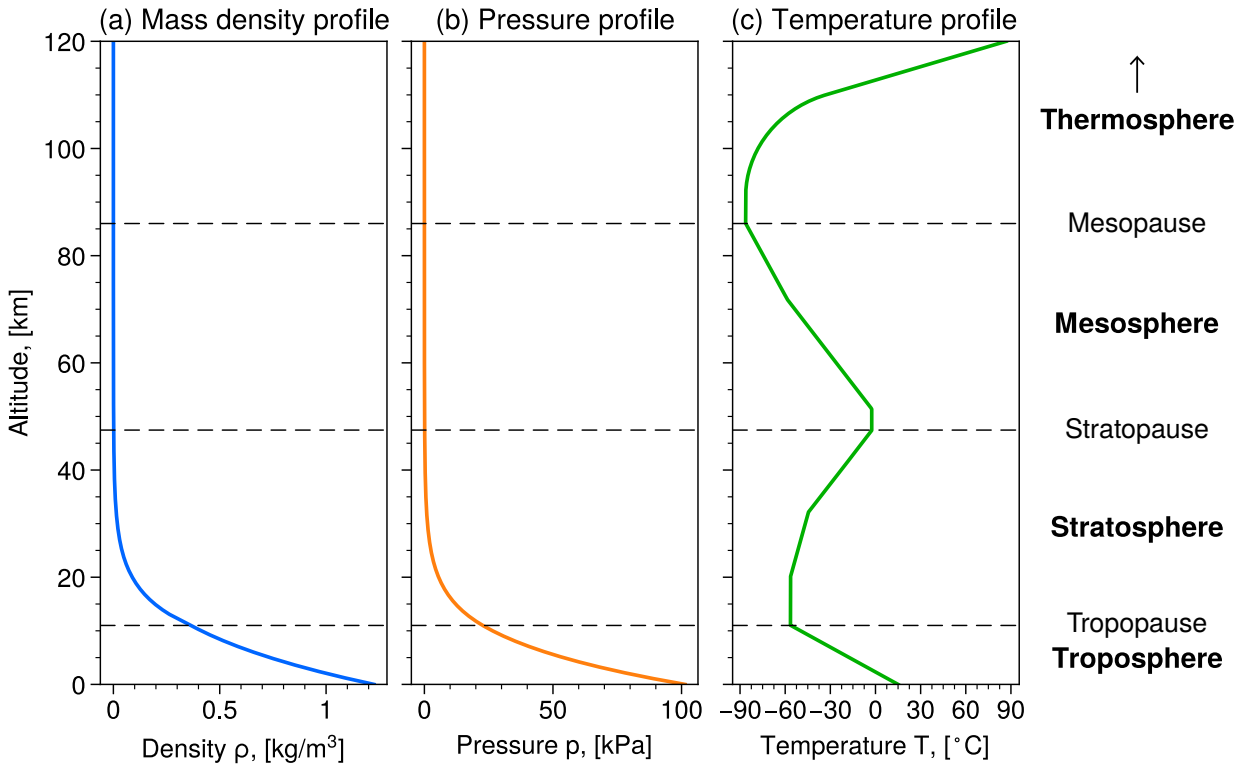

Figure 1.11. Vertical profiles of mass density (a), pressure (b) and temperature (c) as given by the COESA-1976 model.

1.6.2 Balance of ionization

The processes that control the ionosphere can be generally subdivided into three categories: production, desctruction and transport of ionized plasma [Rishbeth and Garriott, 1969]. The relative importance of these categories depends on altitude, as the photochemical processes are very important in the lower layers of the ionosphere, while transport processes play a dominant role in the upper part of the ionosphere. The rate of change of the electron concentration due to the interplay of these processes can be expressed in a continuity equation:

$$\frac{\partial N}{\partial t} = q - l - \nabla \cdot (N\mathbf{v}), \tag{1.38}$$

where q stands for the production rate, l for the loss rate and $\mathbf{v}$ is the net drift velocity of electrons (the change due to transport is the divergence of the flux $N\mathbf{v}$). Below, each of these terms is explained in more detail.

Ionization occurs when the energy of the incoming photons or precipitating particles exceeds the binding energy of the electrons in the neutral gases. There are two most important sources of ionization in the ionosphere, namely the photoionization by the solar EUV and UV radiation, which constitutes the dominant source process on the day side, and the particle precipitation, also called *impact ionization*, which is important on the nightside [Kivelson and Russell, 1995]. The photoionization can be represented in the following equation:

$$A + h\nu \rightarrow A^{+} + e^{*}, \tag{1.39}$$

where e^{*} is a photoelectron, and A is an atom or molecule of neutral gas [Hargreaves, 1992]. Formation of the ionized layer is described below, in Section 1.6.3.

Electrons which are produced in the upper atmosphere can recombine with positive ions; the corresponding process is known as the *ion recombination* [e.g., Rishbeth and Garriott, 1969]. Furthermore, they can get attached to the neutral molecules and form negative ions. The latter loss process is known as the *attachment*, or binding [e.g., Hargreaves, 1992]. The negative ions formed in this way may later be subject to further reations and can get neutralized.

One can express the rate at which the electrons are lost by recombination and attachment as $(-\alpha N_e N_{A^+})$ and $(-bN_e N_A)$, respectively. Here, N_e denotes the electron density, N_{A^+} shows the number density of positive ions and N_A is the neutral density, α is the recombination coefficient, and b - proportionality coefficient [Davies, 1965]. One can assume that only very few negative ions are present and therefore the number density of positive ions is roughly similar to the electron number density ($N_e \approx N_{A^+}$). Moreover, there are much more neutral molecules compared to electrons, and therefore in case of attachment one can neglect changes in the neutral density. The photochemical terms of the continuity equation can be rewritten for the recombination and attachment cases as:

$$\begin{cases} \dfrac{dN_e}{dt} = q - \alpha\{N_e\}^2, \\ \dfrac{dN_e}{dt} = q - \beta N_e, \end{cases} \tag{1.40}$$

where β is the attachment coefficient. Davies [1965] also noted that if the attachment process that produced negative ions is followed by their recombination with positive ions, the overall

continuity equation assumes a form:

$$\frac{dN_e}{dt} = q_{\text{eff}} - \alpha_{\text{eff}}\{N_e\}^2, \tag{1.41}$$

where q_{eff} and α_{eff} are the effective rates of production and loss. Therefore, in such two-step processes the net effect may be represented as a simple recombination.

The last term of the continuity Equation 1.38 is concerned with rate of change of plasma density due to bulk movement of the plasma. Although the transport occurs in both horizontal and vertical directions, for the general morphology of the ionosphere it is important to describe the vertical transport. In particular, in the upper part of the ionosphere one can, for the first approximation, neglect the production and recombination processes, and thus rewrite the continuity equation as:

$$\frac{dN}{dt} = -\frac{\partial(Nw)}{\partial h}, \tag{1.42}$$

where w is the vertical drift velocity at altitude h [Hargreaves, 1992]. Furthermore, if the vertical motion is only due to diffusion, the drift velocity takes the form:

$$w = -\frac{D}{N}\frac{\partial N}{\partial h}, \tag{1.43}$$

where D is the diffusion coefficient (the minority constituent diffuses in the stationary majority gas). The plasma consists of two minority species, namely ions and electrons, and one can define the following quantities:

$$\begin{aligned} D_p &= k(T_e + T_i)/m_i\nu_i, \\ H_p &= k(T_e + T_i)/m_i g, \end{aligned} \tag{1.44}$$

referred to as the plasma diffusion coefficient (D_p) and plasma scale height (H_p). It is possible to show [see, e.g., Hargreaves, 1992] that with increasing altitude, as the photochemistry becomes less important, diffusion becomes exponentially more important and has a significant impact on the ionospheric dynamics.

1.6.3 Chapman layer formation

Solar irradiation, absorbed by the atmosphere, leads to photoionization of electrons. At any given altitude, the production rate of photoelectrons q is proportional to the intensity of ionizing

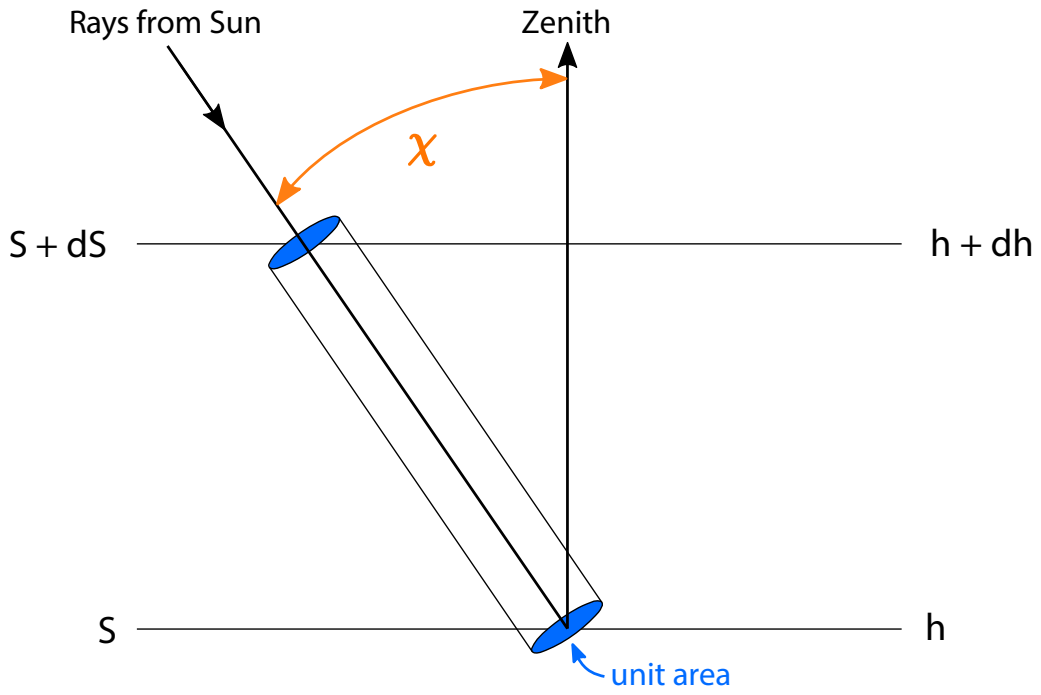

Figure 1.12. Absorption of radiation in a slab of gas (adapted from Davies [1965]).

radiation S_∞, concentration of neutral molecules n_0, and the absorption cross section σ (Figure 1.12). In 1931, Sydney Chapman derived an analytical formula to describe how the production rate of ionization varies with height as a function of the solar zenith angle [Chapman, 1931]. This derivation is based on several assumptions, namely that the atmosphere contains only one species of gas, it is horizontally stratified, it is isothermal and the ionizing radiation is monochromatic and parallel [e.g., Davies, 1965]. Furthermore, it is assumed that the atmosphere does not change with time, and the transport processes are neglected. This Section gives an overview of the Chapman production function, its scaling properties and relation to electron density, while the full derivation is presented in Appendix A. The production rate can be expressed as:

$$\begin{cases} q(\chi, z) = q_0 \exp\left[1 - z - \sec\chi \exp(-z)\right], \\ q_0 = \dfrac{\sigma n_0 S_\infty}{e}, \\ z = \dfrac{h - h_0}{H}, \end{cases} \tag{1.45}$$

where q_0 denotes the production rate at altitude h_0, H is the atmospheric scale height, and z is

the reduced height.

Figure 1.13 shows the dependence of Chapman profiles on solar zenith angle. An interesting scaling rule can be deduced from this figure, namely, that *the shape of the profile remains the same for any zenith angle, and it is only the anchor point of the profile* (q_m, z_m) *that is shifted depending on* χ. In fact, it is possible to show (see Appendix A) that the maximum production rate of a profile q_m and the corresponding reduced height z_m depend on the solar zenith angle as:

$$\begin{cases} q_m(\chi) = q_0 \cos\chi, \\ z_m(\chi) = \ln\sec\chi. \end{cases} \tag{1.46}$$

Therefore, the maximum rate of ionization is scaled by a factor of $\cos\chi$ and the corresponding reduced height is shifted by $\ln\sec\chi$. It is of note that q_0 contains dependence on solar irradiation, which means that the Chapman layer is strongly dependent on solar activity; furthermore, the dependence of the maximum production rate and the respective height on solar zenith angle shows that the Chapman function also contains a significant diurnal variation.

Using Equation 1.46 and derivation presented in Appendix A, one can write an alternative form of the Chapman function:

$$q(\chi, z) = q_m(\chi) \cdot \exp[1 - z - \exp(-z)]. \tag{1.47}$$

In order to determine the distribution of electron density, we assume the photochemical equilibrium with no transport processes, the continuity equation takes the form $q = l$. If one assumes that losses are due to the recombination process ($l = \alpha N_e^2$), then the following expression can be written for electron density:

$$N_e(\chi, z) = N_m(\chi) \cdot \exp\left[\frac{1}{2}(1 - z - \exp(-z))\right], \tag{1.48}$$

with $N_m(\chi) = \sqrt{q_m(\chi)/\alpha}$. This expression defines an alpha-Chapman (or, α-Chapman) layer.

In the case when the loss process is due to attachment of electrons to neutral atoms ($l = \beta N_e$), one can define a beta-Chapman layer as:

$$N_e(\chi, z) = N_m(\chi) \cdot \exp\left[1 - z - \exp(-z)\right]. \tag{1.49}$$

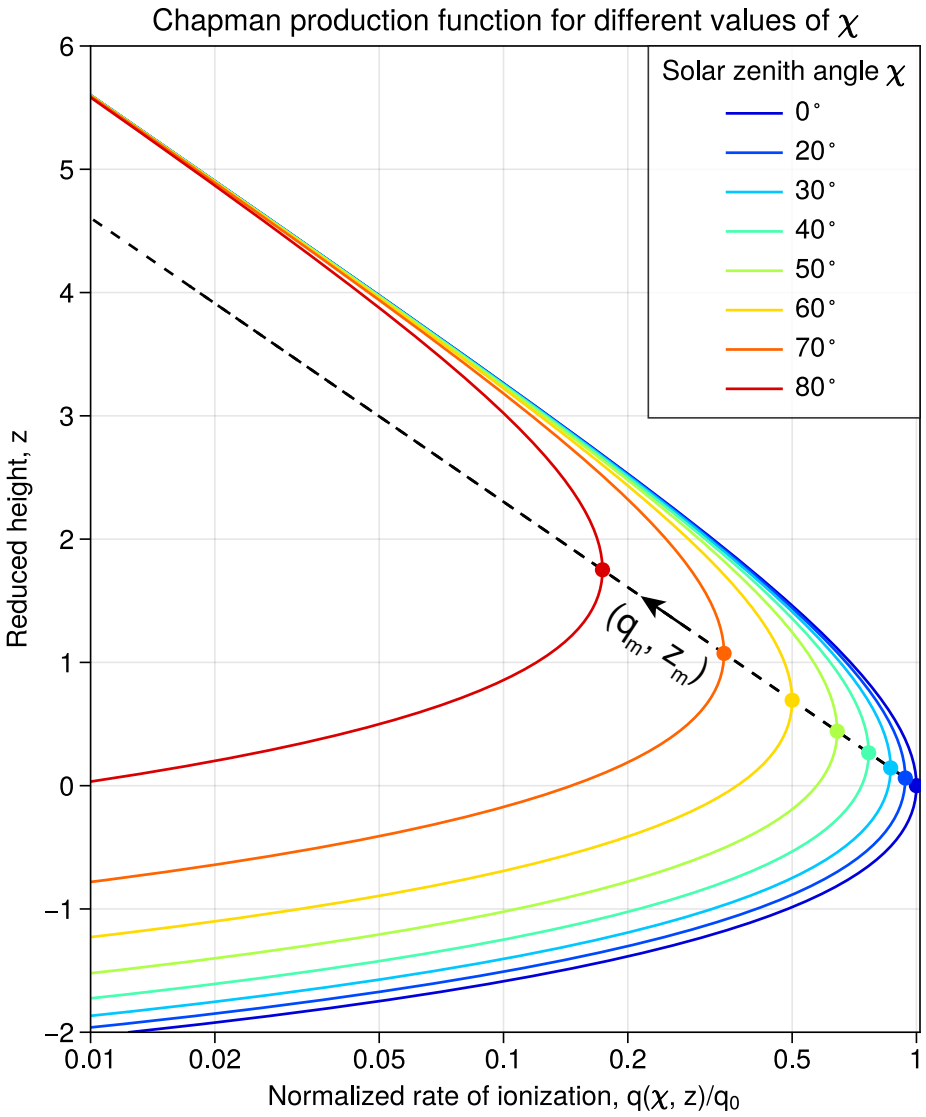

Figure 1.13. Normalized rate of photoionization according to the Chapman's theory for different solar zenith angles (adapted from Hargreaves [1992]). The horizontal axis is in natural logarithmic scale. The dots show points of maximum production rate for each solar zenith angle.

1.6.4 Vertical structure of the ionosphere

The ionosphere can be subdivided into several regions, also called layers, based on the processes that occur within the respective altitude ranges. The main ionospheric layers are denoted as D, E, and F. This naming was used by Sir Edward V. Appleton, who won a Nobel Prize in Physics (1947) for his fundamental work proving the existence of the ionosphere. He initially used the letter *E* to denote the first layer that he had discovered; then, he found another layer above the E-layer, and referred to it as *F* using the next letter in the alphabet. Later on, another layer was found below the E-region. It was therefore denoted with a letter *D*. During the early experiments, it was decided not to start the naming from the letter *A*, due to a possibility that there could be other undiscovered layers below the *D*-region.

The D-layer of the ionosphere spans between ~ 50 and ~ 90 km altitudes and is comprised mainly of NO^+ and O_2^- ions. It is considered to be the most complicated ionospheric region from the chemical perspective [Hargreaves, 1992]. Electron density values in the D-layer are low ($\sim 10^2 - 10^3$ el./cm^3), and the number of electrons is much smaller than the number of neutral molecules. High electron-neutral collision frequencies in the D-region lead to attenuation of radio signals, and the low-frequency signals may not reach the higher layers due to this attenuation. The most important sources of ionization at D-region altitudes include the Lyman-α emission line and X-ray radiation.

The E-layer is located between ~90 and ~140 km altitude. The primary ion constituents are NO^+ and O_2^+, and the secondary ion components are O^+ and N_2^+. The ionization is mainly caused by the X-ray emissions and EUV radiation. Electron densities in the E-layer are much higher than in the D-region, reaching up to $\sim 10^5$ el./cm^3, and the maximum electron density (NmE) is achieved around 100-110 km in altitude. In general, the E-layer can be well-described by the Chapman theory [Hargreaves, 1992], except at nighttime due to the fact that the ionization does not vanish completely but instead the E-region remains weakly ionized.

The F-layer lies at altitudes above ~140 km. A typical ionospheric electron density profile reaches the highest density value within the F-layer. The free electrons in the F-region reflect radio waves in high-frequency communication and broadcasting, making it the main source of

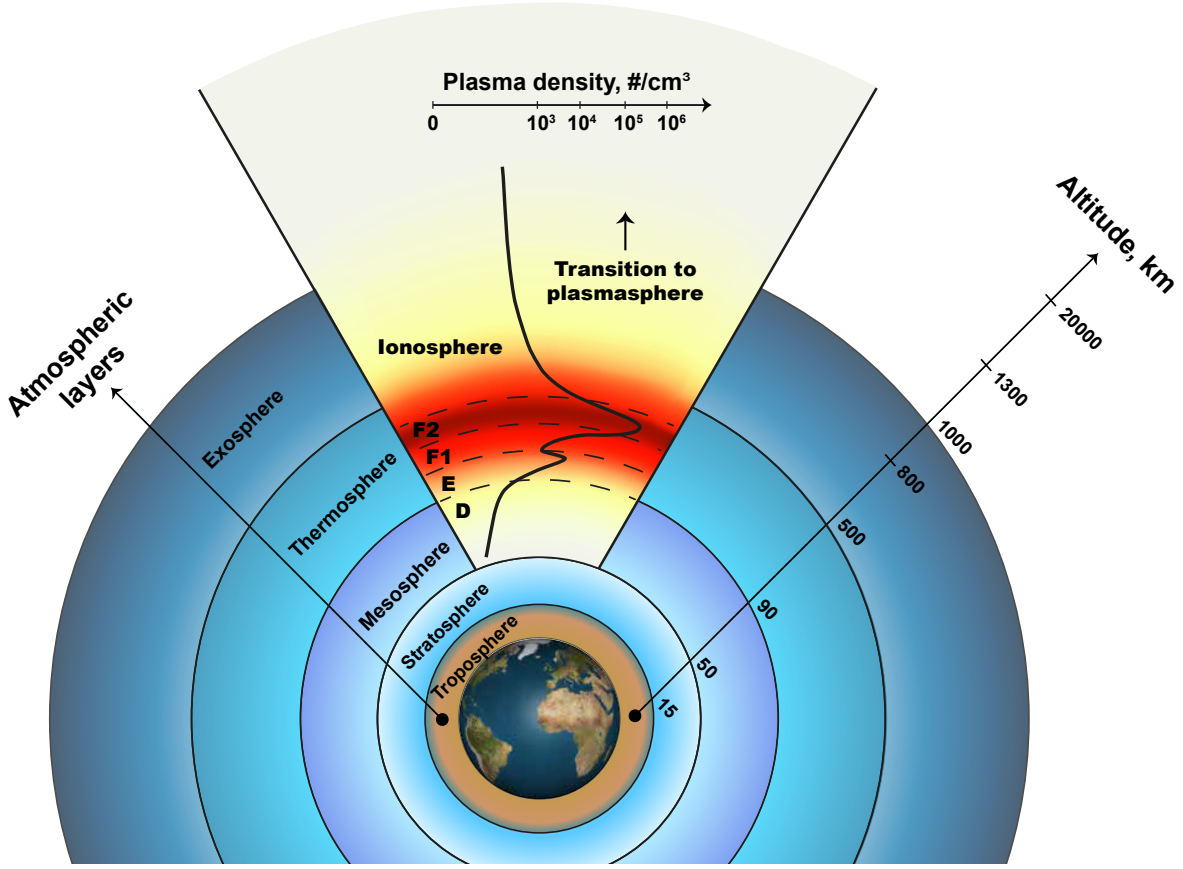

Figure 1.14. A schematics of different layers of the Earth's atmosphere in temperature and ionization domain (adapted from Limberger [2015]).

error for a variety of applications concerned with navigation and positioning [Davies, 1965]. The F-region can be subdivided into the F1 and F2 layers.

The F1-layer is located between ~140 and ~210 km altitudes. The main ion constituents are NO^+ and O_2^+ ions, but O^+ and N^+ are also present as secondary ion components. The main ionization source is the EUV radiation and the typical behavior agrees well with the Chapman theory. The typical electron densities in the F1-region reach $10^5 - 10^6$ el./cm^3. It is worth noting that the F1-layer rarely manifests a distinct peak but rather produces a ledge in the electron density profile [Scotto et al., 1997]. It is more pronounced in summer than in winter, and exhibits a significant diurnal variation, as it vanishes at night and may not be present in winter months even during daytime.

The F2-layer is arguably the most important ionospheric layer due to the fact that it contains

most of the total electron content (TEC) [e.g., Bilitza, 2009]. The F2-layer is maintained both during day and night for any solar activity levels. The peak densities in the F2-layer, also referred to as NmF2, often exceed 10^6 el./cm^3. This layer is highly variable, and includes both long-term trends (e.g., 11-year solar cycle or even longer) and fine variability on timescales of several minutes under geomagnetically disturbed conditions. The main ion constituents are the oxygen (O^+) and light (He^+ and H^+) ions, and the main ionization source is the photoionization by the EUV light. The F2-layer is significantly different from the other photochemically dominated layers, as the transport of plasma produced at lower altitudes plays a major role in the formation and dynamics of the F2-region. The F2-layer peak forms due to the balance between density increase in the lower F2-layer, where the recombination rate falls off quicker than the production rate, and ambipolar diffusion that leads to density decrease with altitude. Therefore, the F2-layer is clearly a non-Chapman layer, and several characteristic anomalies can be found in within the layer, most notably, the equatorial ionization anomaly (EIA). The F2-layer is governed by a variety of competing processes, including the electrodynamic drifts, plasma transport by neutral winds, ambipolar diffusion, changes in the neutral atmosphere, and particle precipitation.

The topside ionosphere lies between the altitude of the F2-layer peak (hmF2) and the upper transition height (UTH), which marks the boundary between the oxygen- and hydrogen-dominated regions. The UTH is defined as the height where the concentration of O^+ equals the concentration of light ions (H^+ and He^+) [Yue et al., 2010a]. The topside ionosphere contains up to ~80% of the TEC, which makes it a region of paramount importance for the GNSS navigation. The dynamics of the topside ionosphere are highly complex and vary significantly with the solar and geomagnetic activity. This part of the ionosphere is located above the critical frequency of the F2-layer, and therefore it cannot be probed by the ionosondes and the available ground-based measurements are mainly performed by the incoherent scatter radars (ISRs), which are only available at a handful of sites around the globe. Satellite measurements present another source of data, but these observations are usually bound to the satellite orbit. The data coverage in the topside ionosphere is, therefore, highly non-uniform, particularly in altitude. This makes the topside modeling a highly complex task, and developing accurate electron density models of the region remains an active area of research.

1.7 Modeling methodology

In recent years, machine learning (ML) algorithms have become an essential component of modern society. They are used in countless applications, such as guiding us to our destinations, providing personalized entertainment recommendations, predicting our shopping preferences, in web applications for spam filtering, email classification and smart replies, for translating text and even detecting fraud in financial transactions. The applications of ML are ubiquitous in our daily lives, and with advancements in technology and increasingly large volumes of data begin collected, the potential for further integration of ML models into various fields of study is immense.

This dissertation delves into the empirical modeling of hazardous particle populations in various regions of geospace, with several publication-based chapters dedicated to this topic. To accomplish this, various modeling techniques have been employed, including machine learning-based approaches. The following subsections provide a general overview of the methods used in the publication-based chapters, while detailed explanations and workflows for each problem are given in the corresponding chapters.

There are several important notes that are worth making before describing the specifics of different ML algorithms. In particular, there are three main "ingredients" for any empirical modeling problem, namely, the data that are used for modeling, the model itself, and the process of learning [Deisenroth et al., 2020]. Therefore, one needs to firstly define what the data are and how they can be represented. Throughout the entirety of this thesis, the input data are assumed to be in tabular form: $\mathbf{X} = \{x_{i,j}\}$ for $i \in [\![1, n]\!]$ and $j \in [\![1, m]\!]$ ($\mathbf{X} \in \mathbb{R}^{n \times m}$)*. In this representation, each of the rows of the matrix, denoted as i, represents a particular measurement instance, and each column, denoted as j, corresponds to the input variable. Throughout this chapter, we will denote matrices with upper-case bold letters, column vectors with lower-case bold letters, and scalars with non-bold symbols. The output data in the following chapters take the form: $\mathbf{y} = \{y_i\}$ for $i \in [\![1, n]\!]$ ($\mathbf{y} \in \mathbb{R}^n$). The next question that arises is what constitutes a machine learning model. Here, ML models are viewed as predictive functions that produce an

*Here, $[\![1, n]\!]$ denotes a set of integers from 1 to n.

output when supplied with input data. Last but not least, the term "learning" refers to improving the model performance on a certain class of tasks with experience.

1.7.1 Types of machine learning problems

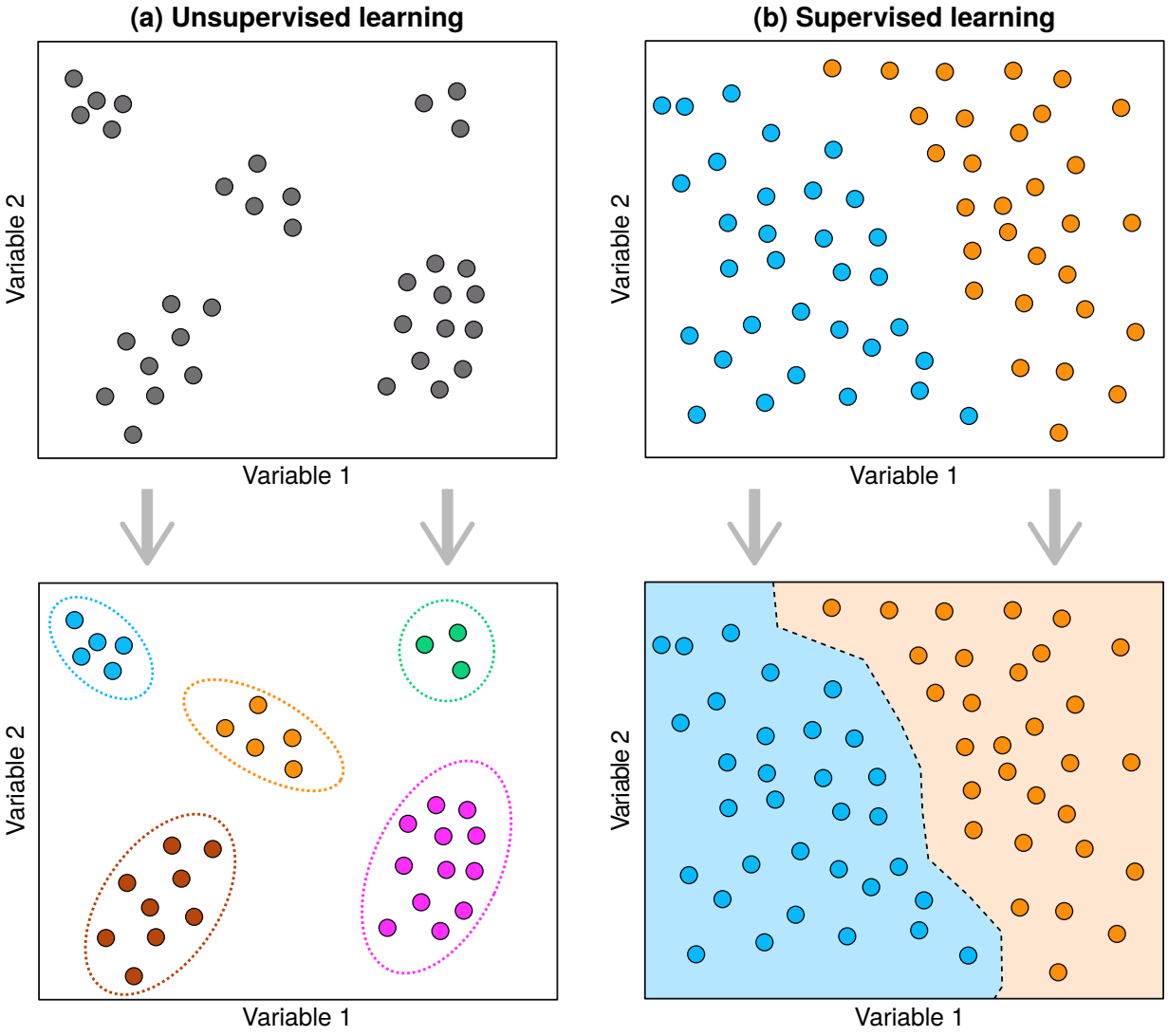

Figure 1.15. An illustration of the unsupervised and supervised learning concepts (adapted from Morimoto and Ponton [2021]). Column (a) shows schematics of the clustering technique, where several collections of points are clustered together without prior knowledge. Column (b) demonstrates a supervised classification problem, where each of the points already has an assigned label (orange versus blue dots), and the classification algorithm draws a boundary between the two classes.

All machine learning algorithms can be roughly subdivided into supervised and unsupervised methods (it is of note, however, that reinforcement learning is sometimes put into a separate category). The schematic difference between the two families of methods is shown in Figure 1.15.

Unsupervised learning methods explore the hidden patters in the data sets without the prior knowledge or inference of these patterns, hence the name "unsupervised" [Goodfellow et al., 2016]. For instance, if several clusters of points are present in the dataset but it is unknown how many clusters there are or which of the features can be used to group these points together, one can employ *clustering* methods. Clustering represents a machine learning technique that allows grouping of the data instances based on their differences and similarities (this is illustrated in Figure 1.15a). Another popular unsupervised approach is the so-called dimensionality reduction (DR). DR is usually employed when the number of input features is too high for a given data set and helps to reduce the number of inputs (or, dimensions) to a more manageable number while maintaining the underlying dependencies. This technique is frequently employed to identify features that can be reduced and are not needed for further model training.

Supervised learning is different from unsupervised methods in that it requires labeled data sets for training. The term "labeled" used here means that for each training example, the value of the output is given. The corresponding methods are designed to "supervise" the model for, e.g., classifying data into several categories (Figure 1.15b) or into the accurate prediction of numerical outcomes based on the training data. If the "true" output values are known, it is possible to evaluate the model accuracy and improve the prediction quality over time. Figure 1.15b shows an example of the supervised classification problem. Unlike in unsupervised clustering described above, each of the dots in Figure 1.15b already corresponds to an existing class of events, and the model finds the boundary between different classes based on the available observations in different input dimensions.

One can subdivide supervised learning methods into classification and regression algorithms. The difference between classification and regression problems lies in the fact that classification operates on a finite number of predefined classes, and therefore the desired output is discrete in nature, while in regression the output can be represented as a continuous numerical variable

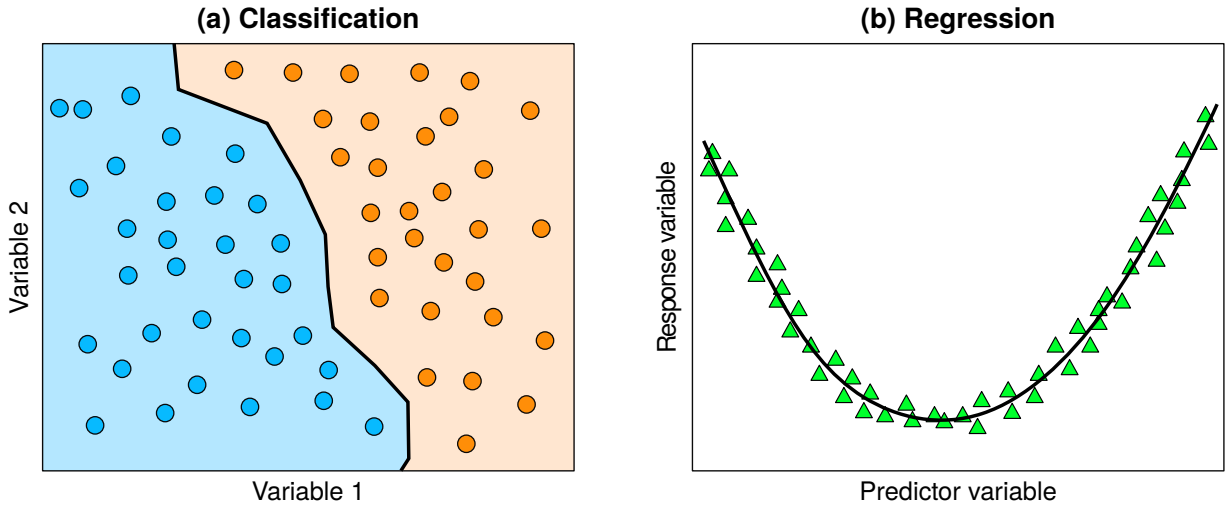

Figure 1.16. An illustration of classification and regression tasks. Panel (a) gives a schematics of a classification problem (similar to Figure 1.15), where each of the data points is assigned to one of several discrete classes (in this case, blue and orange dots represent 2 classes) and the ML algorithm finds a boundary between the classes (black line) based on the input variables. Panel (b) shows an example of a regression problem, where the output can be represented with a continuous variable and the ML algorithm tries to find a suitable predictor function from the input data.

[Hastie et al., 2009]. In particular, let us consider a data set with n data instances and m feature variables $\mathbf{X} = \{x_{i,j}\} \in \mathbb{R}^{n \times m}$ ($i \in [\![1, n]\!]$, $j \in [\![1, m]\!]$). For a classification problem (assuming there are only 2 classes), the task is to find a mapping function that will assign each of the data entries to the corresponding pre-labeled output $y_i \in \{-1, +1\}$ (for $i \in [\![1, n]\!]$), and can be expressed as $\mathbb{R}^m \rightarrow \{-1, +1\}$. In multi-dimensional classification, the output values belong to a set of k discrete labels denoted as $I = \{q_1, ..., q_k\}$, and the task is to find a mapping $\mathbb{R}^m \rightarrow I$. On the other hand, in regression problems, the output can be represented as a continuous numerical variable and therefore the function mapping takes the form $\mathbb{R}^m \rightarrow \mathbb{R}$.

In supervised learning, we would like to find a predictor function defined by a set of model parameters (for instance, weights and biases in the case of neural networks). The main task of the training process is finding the values of these parameters which would produce a "good" estimator [Deisenroth et al., 2020]. Generally, a sensible way to achieve this is to minimize the difference between the predicted and target values using the data in the training set [Ng, 2000]. At the same time, any real-world data sets contain errors of different sources and, therefore, one should not view the learning process as a way to simply memorize the training set, but rather

as a delicate balance between reproducing the training examples well and predicting the unseen events with roughly the same accuracy levels. These concepts are formalized in the following section.

1.7.2 Bias – variance tradeoff

Let a target variable $\mathbf{y}$ and a feature matrix $\mathbf{X}$ be related via $\mathbf{y} = f(\mathbf{X}) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ represents the irreducible error, which is the part of $\mathbf{y}$ that cannot be predicted from $\mathbf{X}$. We are interested in finding an estimator of the mapping f, denoted as $h_\theta(\mathbf{X})$ (the subscript reflects that the estimator is defined by a set of parameters $\boldsymbol{\theta}$) [e.g., Ng, 2000; Goodfellow et al., 2016]. The usual way of training the model $h_\theta(\mathbf{X})$ is by minimizing the difference between the model output (denoted as $\hat{\mathbf{y}} = h_\theta(\mathbf{X})$) and the observations $\mathbf{y}$, which is measured by the cost function J.

Errors of the estimator are usually comprised of three parts, namely, the *bias*, *variance* and *irreducible errors*. The latter describes the fact that predictive capabilities of a trained estimator are always limited by the data noise. Bias and variance, on the other hand, represent reducible errors that should be minimized. The bias of an estimator can be expressed as:

$$\text{Bias}\{h_\theta(\mathbf{X})\} = \langle \hat{\mathbf{y}} - \mathbf{y} \rangle = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i), \tag{1.50}$$

and shows how far on average the estimated values are displaced with respect to the observations [Hastie et al., 2009]. Models with high bias are generally quite rigid and not expressive enough to capture the patterns in the data, whereas low-bias estimators are flexible and can model complicated relationships. The variance of an estimator is given by:

$$\text{Variance}\{h_\theta(\mathbf{X})\} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}_i - \langle \mathbf{y} \rangle\right)^2, \tag{1.51}$$

and shows how the estimate would vary if the data are randomly resampled from the underlying data distribution [Hastie et al., 2009]. High-variance models are too expressive for a given problem and tend to capture not only the useful dependencies but also data noise, while models with low variance are less prone to random fluctuations and have good generalization abilities.

Bias and variance reflect two different error sources in the model estimates. The next logical question is what constitutes a better model – an estimator with low bias but high variance, or

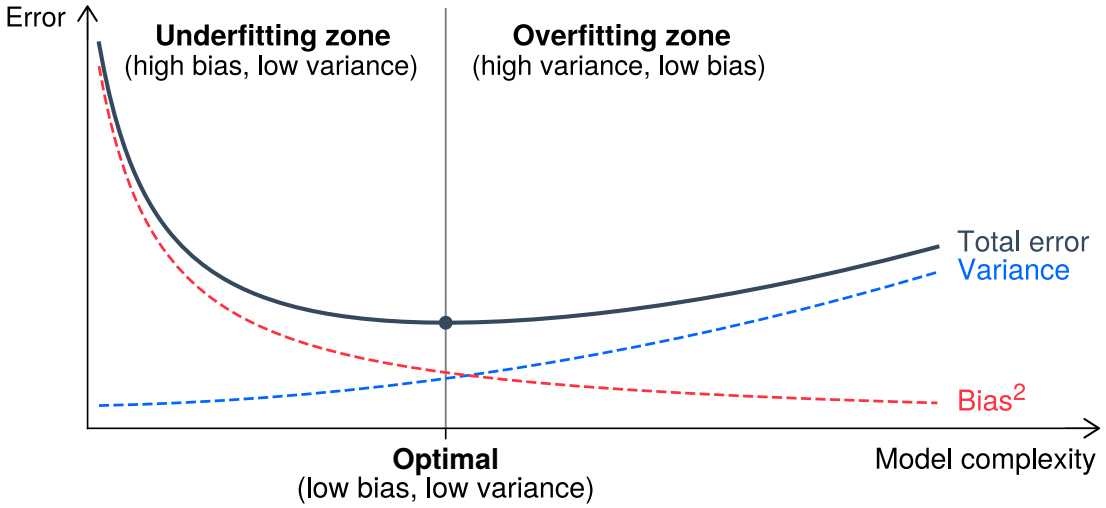

Figure 1.17. An illustration of the bias-variance tradeoff (adapted from Goodfellow et al. [2016]). As the model capacity (on the horizontal axis) increases, bias decreases but variance increases, which yields the U-shaped curve of the generalization error. If we select an "optimal" model complexity, corresponding to the minimum of the generalization error, then the overfitting region corresponds to models that are too expressive for a given data set, and underfitting refers to models that are not complex enough to capture the underlying phenomena.

the one with low variance but high bias? It is obvious that the best case scenario would be to have a model where both error sources are minimized. Finding such a model involves adjusting the model complexity, because the capabilities of an estimator are largely determined by its architecture. Models with few degrees of freedom tend to have high bias. Increasing the model complexity reduces bias but leads to larger variance (Figure 1.17). This phenomenon, known as the *bias-variance tradeoff* [Breiman, 2001], is one of the key concepts of machine learning. Luckily, the following metric, known as the *mean squared error* (MSE), provides a measure of the overall squared deviation between the estimated and true values. It is defined as:

$$\mathrm{MSE}\{h_\theta(\mathbf{X})\} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\,, \tag{1.52}$$

and incorporates both bias and variance. It can be shown [e.g., Hastie et al., 2009] that MSE can be decomposed as:

$$\mathrm{MSE}\{h_\theta(\mathbf{X})\} = \mathrm{Bias}^2\{h_\theta(\mathbf{X})\} + \mathrm{Variance}\{h_\theta(\mathbf{X})\} + \sigma_\epsilon^2. \tag{1.53}$$

Equation 1.53 is known as the *bias-variance decomposition*, and provides a way to negotiate the bias-variance tradeoff through the mean squared error. It is evident that using only the training

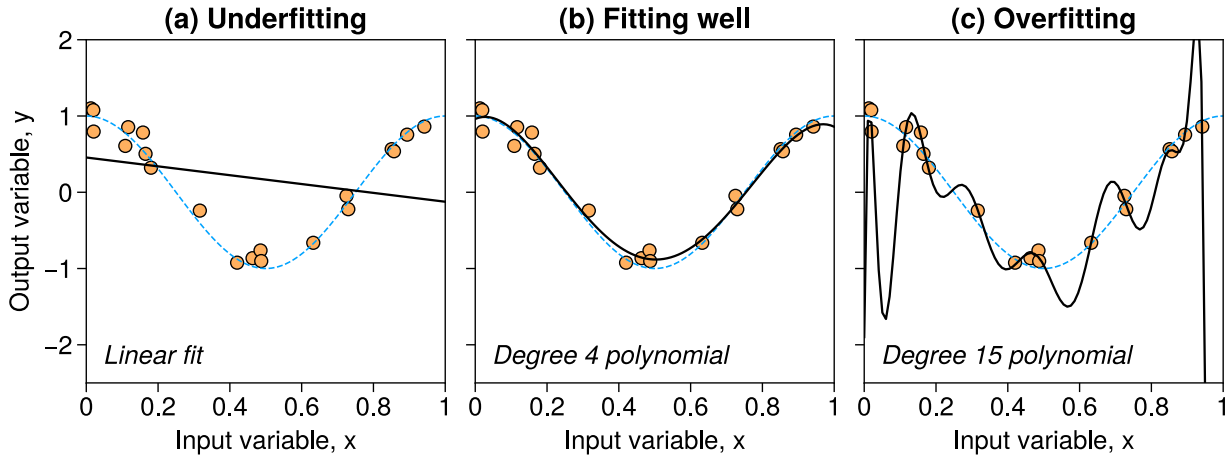

Figure 1.18. An illustration of the underfitting vs overfitting concept. Underfitting (a) refers to a phenomenon when the model is not sufficiently expressive to approximate the synthetic data correctly; panel (b) shows a case when the model can approximate the data well enough and is the output is robust, (c) the model approximates every point but is overly expressive and therefore creates artificial oscillations in between of the data points. The true function from which the synthetic data were drawn is shown with a dashed blue line.

set is not enough to evaluate the model's quality, and that this needs to be done using independent data. Such data sets, which follow approximately the same distribution as the training data but were not seen during the model training, are known as the *validation* sets. During the model training, the validation error is usually tracked alongside the training error, and their difference is referred to as the *generalization error* and can be used to judge the success of the training process. The estimator that achieves the lowest generalization MSE can be regarded as the best estimator, and the set of parameters corresponding to this estimator can be viewed as the optimal combination. Models with architectures that are more complex that this optimal combination would easily *overfit* the training data, while those with less expressiveness would *underfit* the data.

To illustrate these concepts, in Figure 1.18 we randomly draw 20 samples from a synthetic function (shown in blue) and add gaussian noise (with $\sigma = 0.15$) to these synthetic data points. In panel (a), a linear trend line is fitted to these data points. It is evident that the linear function is not expressive enough to approximate the data, and this presents a case of underfitting. On the other hand, in panel (c), a polynomial of degree 15 is used, and the resulting function is too complex for the problem. It fits well the available data points but becomes oscillatory in

between and does not resemble the true function. This is the case of overfitting. In panel (b), we fit the data samples with a polynomial of degree 4, which provides a robust representation of the trends seen in the synthetic data. This simple example is an illustration of one of the most crucial concepts in machine learning, which is to use the independent out-of-sample data samples to make sure that the parameters obtained during the model fitting are adequate and do not lead to overfitting or underfitting.

The subsections above were mainly concerned with the general concepts of empirical modeling. The following subsections give brief formulations and descriptions of different predictor functions used throughout the thesis, while the details of their implementation can be found in the corresponding chapters.

1.7.3 Linear models

Linear regression is a statistical approach to modeling the relationship between one or more predictor variables and a response variable. Using the notations from previous subsections, in linear models one tries to find an affine function h_θ that maps the input $\mathbf{X} \in \mathbb{R}^{n\times m}$ into the corresponding function values $f(\mathbf{X}) \in \mathbb{R}$. The modeled relationship between them is assumed to take the form:

$$h_\theta(\mathbf{X}) = \theta_0 + \theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 + ... + \theta_m\mathbf{x}_m = \theta_0 + \sum_{j=1}^{m} \theta_j\mathbf{x}_j, \tag{1.54}$$

where $\mathbf{x}_j \in \mathbb{R}^n$ are column vectors of the input matrix $\mathbf{X}$, $\boldsymbol{\theta} = \{\theta_1, ..., \theta_m\}$ are the model parameters, referred to as the *model weights*, and θ_0 represents the intercept that allows to make the model more flexible. Assuming $\mathbf{x}_0 = \{x_{0i} = 1\}^\top \in \mathbb{R}^n$ for $i \in [\![1, n]\!]$, it is possible to present the parameter vector and the design matrix in an augmented form as $\boldsymbol{\theta}_A = [\theta_0, \theta_1, ..., \theta_m]^\top \in \mathbb{R}^{m+1}$

and $\mathbf{X}_A = [\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_m]^\top \in \mathbb{R}^{n \times m+1}$. Then, Equation 1.54 can be rewritten as:

$$h_\theta(\mathbf{X}_A) = \underbrace{\begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix}}_{\mathbf{X}_A \in \mathbb{R}^{n \times m+1}} \underbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}}_{\boldsymbol{\theta}_A \in \mathbb{R}^{m+1}} = \underbrace{\begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ \vdots \\ \hat{y_n} \end{bmatrix}}_{\hat{\mathbf{y}} \in \mathbb{R}^n} \tag{1.55}$$

The squared-loss function can be expressed as:

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{1.56}$$

where the goal is to find a set of parameters that minimizes the cost function $\theta_{LSq} = \arg\min_\theta J(\theta)$. They are referred to as the *least-squared estimators*. The minimization of Equation 1.56 can be solved explicitly, using the ordinary least squares (OLS) approach:

$$\boldsymbol{\theta}_{\text{LSq}} = (\mathbf{X}_A^\top \mathbf{X}_A)^{-1} \mathbf{X}_A^\top \mathbf{y}. \tag{1.57}$$

Furthermore, the same problem can be solved using *gradient descent* (GD), which is an iterative optimization algorithm that finds the minimum of a function. Linear regression problems are often put under the umbrella of machine learning, even though in most cases they can be solved efficiently using the OLS method. However, in order to provide a simplified illustration of the gradient descent, we will solve a multiple linear regression problem below.

Let us use gradient descent to minimize the loss function given in Equation 1.56. At the first iteration, one needs to initialize values of the weights, which can be done randomly. An important component of the gradient descent is the *learning rate*, denoted as α, which determines the magnitude of the updates made to the model parameters at each iteration. Furthermore, one needs to calculate the partial derivatives of the loss function with respect to the model parameters. They can be expressed as:

$$\mathcal{D}_j = \frac{\partial J}{\partial \theta_j} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) X_{ij}, \tag{1.58}$$

where θ_j represents the j-th parameter of the model ($j \in [\![1, m]\!]$), and $\mathcal{D}_j$ computes the average gradient across all training samples. At each new iteration, the previous values of the model

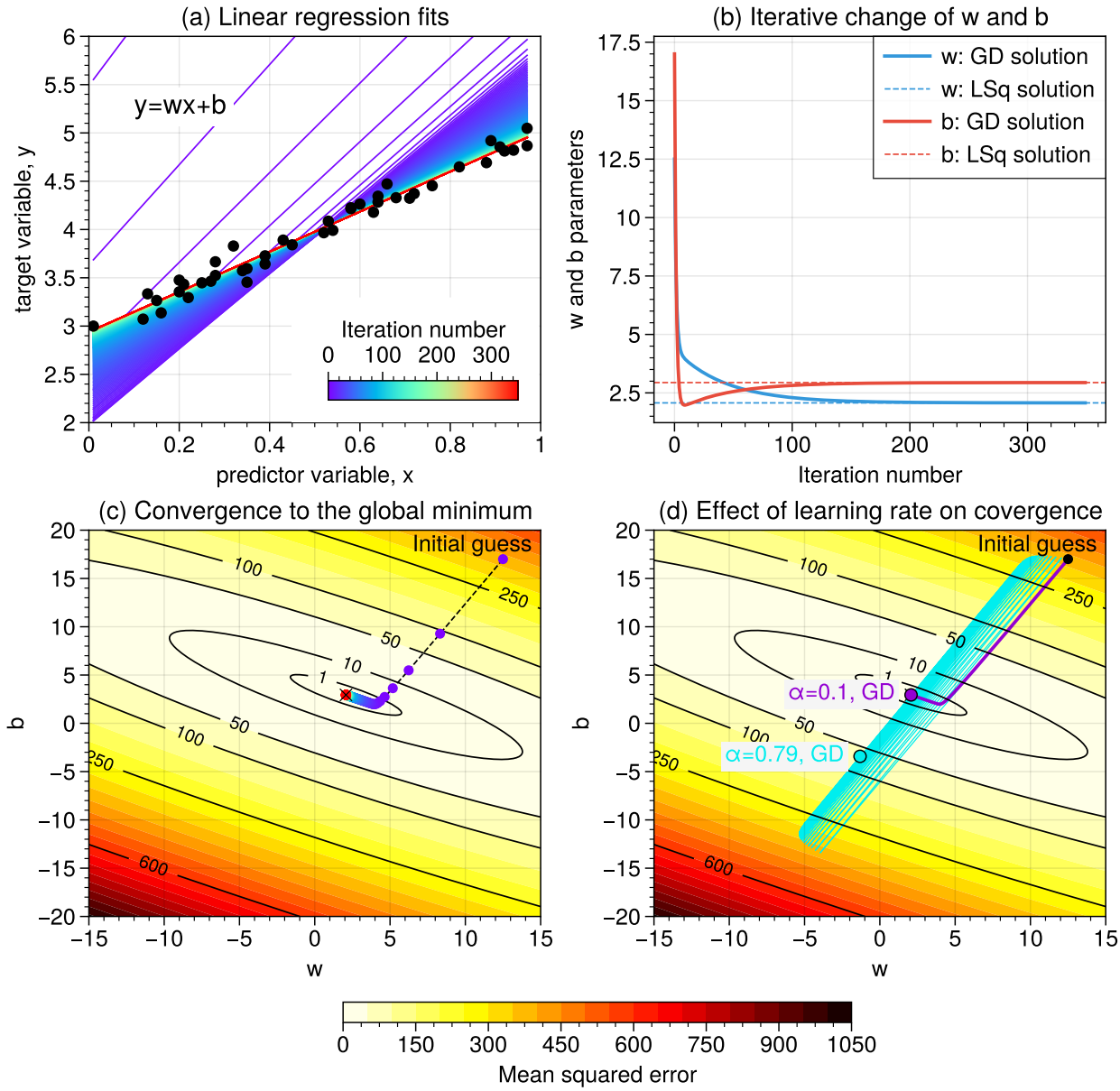

Figure 1.19. An illustration of the gradient descent algorithm. Panel (a) shows synthetic data points (black dots) and linear trend lines colored by the iteration number of the gradient descent. Panel (b) demonstrates the evolution of w and b with the gradient descent iterations (blue and red lines, respectively). After ~200 iterations, the parameters retrieved from GD approach the least square solutions shown as dashed lines of the same color. Panel (c) shows the mean squared error for different combinations of w and b and how the optimal combination of parameters was iteratively obtained. Panel (d) shows an effect of learning rate on the algorithm's convergence. With an appropriate learning rate (0.1, purple line), GD solution approaches the optimal solution (purple dot). For a learning rate that is too high (0.79, cyan line), the iterative update of w and b is much noisier, and for the same number of iterations GD does not converge to the optimal solution (the final iteration is shown as a cyan dot).

parameters θ alongside with input and output values $\mathbf{X}$ and $\mathbf{y}$ are substituted in Equation 1.58, and the parameter values are updated using the following equation*:

$$\theta_j := \theta_j - \alpha \mathcal{D}_j. \tag{1.59}$$

For a clearer understanding, one can invoke the hiking analogy: θ_j can be considered the current position of the person in the j-th dimension, and $\mathcal{D}_j$ is equivalent to the steepness of the slope along that dimension. At each iteration, the updated value of θ_j reflects their new position after taking a step $\alpha \mathcal{D}_j$ in the direction of the negative gradient. Both the learning rate and the partial derivative contribute to the step size, as $\mathcal{D}_j$ gives direction and steepness, and α scales the steps allowing for finer control over the training process. Finally, the hiker arrives at the bottom of the valley which corresponds to the minimum of the loss function.

Figure 1.19 illustrates the gradient descent on a simple univariate linear regression problem. Panel (a) shows synthetic data (black dots) which can be well approximated by an affine function. The straight lines represent the linear trends that are fitted using gradient descent, colorcoded by the number of GD iteration. One can see that the initial guess leads to a line that is very far away from the data points, and that the lines eventually converge to a good approximation of the data. It is of note that the step of gradient descent is proportional to error at each iteration, and therefore one can see larger steps for the first few iterations and smaller updates later on. Panel (b) shows a comparison of the iteratively found slope and intercept values with those obtained using ordinary least squares. After approximately 200 iterations the GD solution converges to the OLS solution. Panel (c) demonstrates the loss function value for different slope and intercept values and demonstrates the trajectory from the initial guess to the global minimum. In panel (d), the effect of learning rate is investigated. In particular, we use two learning rates, 0.1 an 0.79, and perform 300 iterations of the gradient descent. One can see that for $\alpha = 0.1$ there is a smooth convergence towards the global minimum, while for the learning rate that is too high ($\alpha = 0.79$) the update makes large jumps and for the same number of iterations the convergence is not achieved. One should note that this particular optimization problem has only one

*Throughout this Chapter, the notation $a := b$ refers to the replacement of a with b, while equality between a and b is denoted as $a = b$.

global optimum and therefore even for relatively large learning rate the gradient descent would eventually converge to the global minimum. However, for problems with multiple local minima, a learning rate that is too high may prevent gradient descent from converging to the global minimum.

1.7.4 Feature mappings

Natural phenomena often present highly complex data distributions that can be better modeled using non-linear functions, such as polynomial, logarithmic or exponential functions. To address this, an approach known as *feature mapping* can be employed. It allows to transform raw data into a more suitable form and reduces this problem to a multilinear regression task.

In feature mapping, one aims to convert the original input variables into higher-dimensional feature spaces, which helps with modeling more complex relationships. For instance, given a column vector $\mathbf{x}_j$ corresponding to the feature j, one can define a transformation of the form:

$$\phi_k(\mathbf{x}_j) = \mathbf{x}_j^k, \tag{1.60}$$

where $k = [\![0, K-1]\!]$. Such feature mappings convert a one-dimensional input vector $\mathbf{x}_j$ into the K-dimensional feature space that can represent polynomials up to degree $K-1$. This problem can be viewed as a multilinear regression over the new set of feature variables $\phi(\mathbf{x}_j)$. In this case, ϕ is referred to as the *feature map* acting on the original inputs $\mathbf{x}_j$ that are sometimes also called the *function attributes* of the problem [Ng, 2000].

The model function h_θ can be written as

$$h_\theta(x) = \sum_{k=0}^{K-1} \theta_k \phi_k(\mathbf{x}_j) = \boldsymbol{\Phi}\boldsymbol{\theta}, \tag{1.61}$$

where $\boldsymbol{\theta} = [\theta_0, ..., \theta_{K-1}]^\intercal \in \mathbb{R}^K$ are the model parameters, similar to Section 1.7.3, and $\boldsymbol{\Phi}$ denotes the design matrix. The OLS solution can be found using the following equation:

$$\boldsymbol{\theta}_{\text{LSq}} = \left(\boldsymbol{\Phi}^\top \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^\top \mathbf{y}. \tag{1.62}$$

The description above focuses on problems with a single input attribute. However, one can use different forms of feature mappings and transform any number of features while still treating this problem as a multiple linear regression. In space physics problems, several predictor

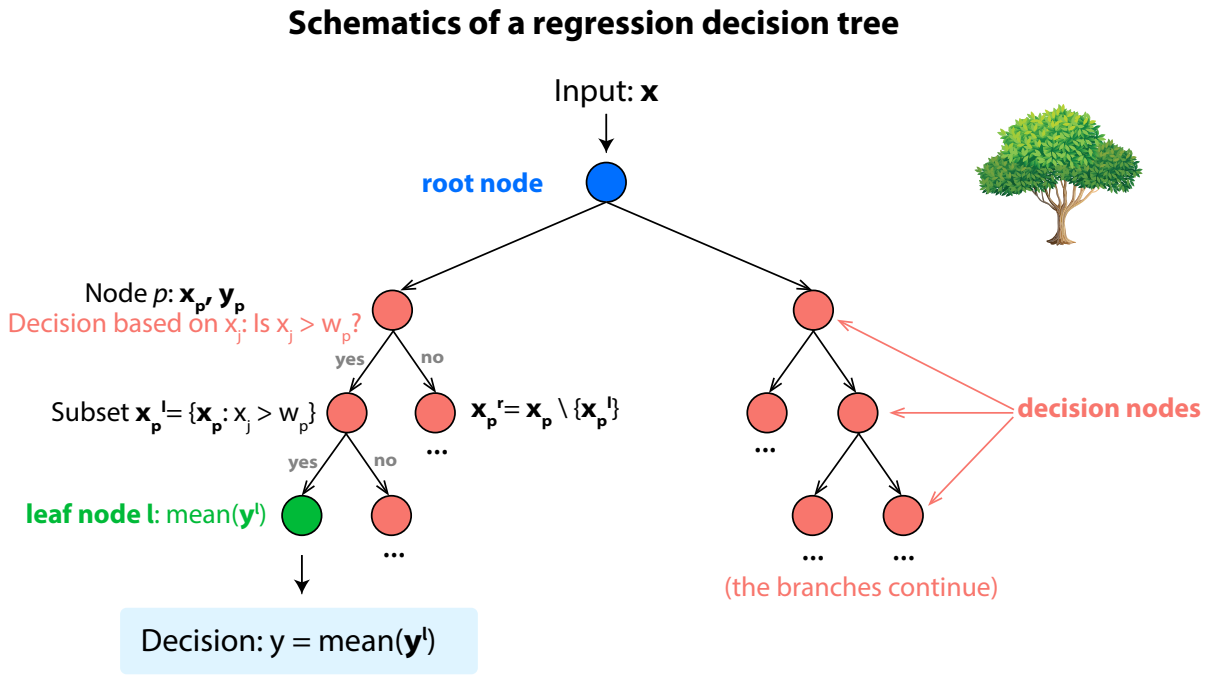

Figure 1.20. Schematics of a regression tree. The root node of the tree, where the decision process starts, is shown in blue, the internal decision nodes are shown in red, together with the decision criteria, and the leaf node is demonstrated in green (the tree image was downloaded from www.freepik.com).

variables that are frequently used for empirical modeling, such as LT and longitude, exhibit discontinuities between their highest and lowest values (e.g., both 0 and 24 hours LT correspond to midnight). Using raw values for such features would result in model artifacts, which can be avoided by using sine/cosine transformations of the values. Such feature mappings have become a common practice in space weather modeling and are utilized, for instance, in Chapters 4 and 6.

1.7.5 Decision trees

Linear methods described in the previous section are *parametric* methods [Alpaydin, 2020]. The parametric methods create a general model over the entire input space and learn the parameters assuming a specific functional form. There also non-parametric techniques that divide the input space into local regions, based on certain measures of distance such as the Euclidian norm, and create local models using the respective subsets of training data. Examples of the non-parametric methods include the K Nearest Neighbors (KNNs) and tree-based methods. There is, however,

an important difference between these two approaches. Specifically, decision trees represent the *model-based learning*, i.e., methods that use the training set to construct the model and fit the model parameters using those data. On the other hand, the KNNs try to find the closest instances to the one being modeled based on a distance in multidimensional space, and therefore employ the whole training set as the model. Such algorithms are called the *instance-based methods.*

Decision trees (DTs) are one of the most popular ML techniques, used both as standalone models and as building blocks for more sophisticated ensemble methods (Section 1.7.6). Trees stratify, or segment, the predictor variable into a number of regions, and return as prediction the average value of the target variable in a region where it is placed based on the input parameter values [e.g., Hastie et al., 2009]. DTs consist of nodes and branches (sub-trees). There are 3 types of nodes, namely, the root node which is the beginning of a tree and contains the entire population before any splitting is done, internal (or, decision) nodes which perform the tests on given attributes, and the leaf (terminal) nodes that contain the output values. DTs are built by a process known as *induction*; it performs recursive splitting of the training data into subsets based on the feature values. Each decision node p implements a certain function $g_p(x)$ that, given an input value, determines which branch should be followed.

The "decision making" process begins in the root node and selects a feature that separates observations into distinct classes of target values in the most optimal way (note that there are multiple ways to approach this and they will be discussed later). Each sub-tree will be split again at the following decision nodes, down to the leaf node that contains the output. In fact, the test function $g_p(x)$ is a simple function that breaks down into a sequence of binary decisions. Each leaf node contains a "label", which is the identifier of a certain class in case of classification and a numeric output in the case of regression.

Let us consider a node with a test function only makes a decision based on a particular input feature j. If the inputs are numeric, there is an empirically chosen threshold w_p such that the decision function

$$g_p(\mathbf{X}_p) : \mathbf{x}_j > w_p \tag{1.63}$$

divides the space of input parameters into two subspaces: $\mathbf{X}_p^r = \{\mathbf{X}_p : \mathbf{x}_j > w_p\}$ and $\mathbf{X}_p^l =$

$\{\mathbf{X}_p \setminus \mathbf{X}_p^r\}$, where $\mathbf{X}_p$ denotes the input space that reached the node p and superscripts l and r refer to the left and right branches generated at node p. Output values are divided using using the same indices. Such divisions are called *binary splits*. They generate sharp boundaries of the input space and produce splits which are orthogonal to each other thus dividing input space into hypperrectangles. After the decision tree has been constructed, it can be used to make predictions given the input, that is, perform *deduction*. At the deduction stage, the model function g which consists of a sequence of binary decisions $g_p(x)$, applies these test conditions to the given data samples and follows the appropriate tree branches, returning the value at the corresponding leaf node. In the case of regression, if the node p is the leaf node, the output value h_p is calculated as the average of training values reaching the node:

$$h_p = \frac{1}{N_p} \sum_{i \in p} y_i, \tag{1.64}$$

where N_p is the number of points at node p.

Decision tree methods can be used both for classification and regression. In both cases, it is necessary to quantify the goodness of split. In the case of regression, splitting criteria are chosen so that they suit regression tasks, for instance, the reduction of variance or MSE are frequently selected. MSE can be reformulated as:

$$J_p = \frac{1}{N_p} \left(\sum_{i \in p^l} \left(y_i - \langle \mathbf{y}_p^l \rangle \right)^2 + \sum_{i \in p^r} \left(y_i - \langle \mathbf{y}_p^r \rangle \right)^2 \right). \tag{1.65}$$

The aim is to find a split that corresponds to the smallest J_p value [James et al., 2013].

Due to the fact that deep DTs are high-variance models, it is common practice to use regularization that could combat overfitting. One popular approach is to constrain the maximum depth of the tree. This is due to the fact that decision trees can, in principle, perform the splitting until each of the samples represents a separate leaf, which would lead to strong overfitting and would take a very long training time. In order to avoid this, it is quite common to put a certain threshold on a depth of a tree.

Another option to regularize DTs is to set an error level that is considered acceptable. If at node p the achieved error J_p is smaller than the defined threshold error, then a leaf node is created

and it contains the h_p value. If the threshold is not yet reached, the data splitting continues further until the criterion is met. For this kind of regularization, the "acceptable" error can be treated as a complexity parameter. When the threshold error is too small, very deep trees that can overfit to training data will be generated, while if the threshold is too large the created trees will be too coarse and will underfit the data.

Another way to constrain the tree depth is by setting the minimum number of samples in a leaf, which can be done, for instance, in terms of the percentage of points of the training set (e.g., each of the leaves should contain no less than 5% of the training set). Constraining the tree depth in this way is called *pruning*. This technique is motivated by the fact that decisions based on too few samples increase variance and thus the generalization error. One can do a *prepruning* where the splitting stops before the tree is fully grown, and *postpruning* where the trees are grown until the training error is pushed to zero, and the splits that do not contribute to the model improvement are identified and removed post-factum based on the data set which was withheld during the fitting process [e.g., Breiman, 2017]. Generally, the postpruning is much faster than the exhaustive search of the best features at every split, and therefore leads to much faster implementation of decision trees.

Decision trees have gained significant popularity in machine learning research. The main reasons are that the DT models are very easy to explain and understand and can be represented graphically, especially if the trees are not very deep. Another reason for this popularity is that decision trees closely mimic the human decision making process. Furthermore, the trees can directly reveal the importance scores for each variable. Last but not least, the tree based methods are model-agnostic and do not require any underlying assumptions about the data. However, DTs also have several drawbacks. For instance, deep trees are prone to data noise and need to be regularized. Furthermore, due to the hierarchical structure of decision trees the errors of the top split can propagate down to all splits below, and therefore the splitting needs to be performed carefully. However, by aggregating many decision trees and creating the so-called modeling ensembles, it is possible to increase the stability of DTs and improve the overall performance of the models.

1.7.6 Ensemble learning: bagging and boosting

Ensemble learning is a machine learning approach where several models with relatively low predictive capabilities are combined in order to achieve better performance [Goodfellow et al., 2016]. The main underlying hypothesis behind all ensemble techniques is that weaker models, also called *base estimators*, can be combined in a way that balances their weaknesses and can thus produce more reliable models. The base estimators usually suffer from either high bias (e.g., have too few degrees of freedom) or high variance (too many degrees of freedom). The ensemble techniques are selected to reduce bias/variance of these weak estimators in order to create a strong learner. For instance, if the base models have high variance (and low bias), the ensemble technique should reduce variance, while if the base estimators have low variance and high bias, then the ensemble method should aim at reducing the bias.

There are two main ways to combine weak learners, known as *bagging* and *boosting*. Bagging (short for bootstrap aggregation) trains the base estimators in parallel and then combines them by taking the average of their predictions aiming to reduce the variance of the ensemble model [James et al., 2013]. Bagging creates multiple (K) learners by creating K subsets of the original data in the training phase. These subsets are created by randomly drawing samples with replacement from the original set, which is referred to as *bootstrapping*. Here, the term replacement refers to the fact that some samples may be drawn multiple times. For the bagging technique to be effective, the subsets need to have a large enough number of samples to capture the underlying phenomena and be representative of the data distribution of the original set. On the other hand, the number of samples in the subsets should be sufficiently smaller than in the original subsets, otherwise the base learners will be trained on the data that are too similar and the ensemble techniques may not bring any improvement.

Decision trees are considered a popular choice as base estimators in ensemble modeling. The combinations of trees are referred to as *forests*. Trees with only a few layers, called *shallow* trees, generally suffer from high bias, whereas trees that contain many layers, called *deep* trees, and are prone to high variance. This makes deep trees a good choice as base models for bagging techniques. A machine learning method that use a bagging approach on deep trees is known

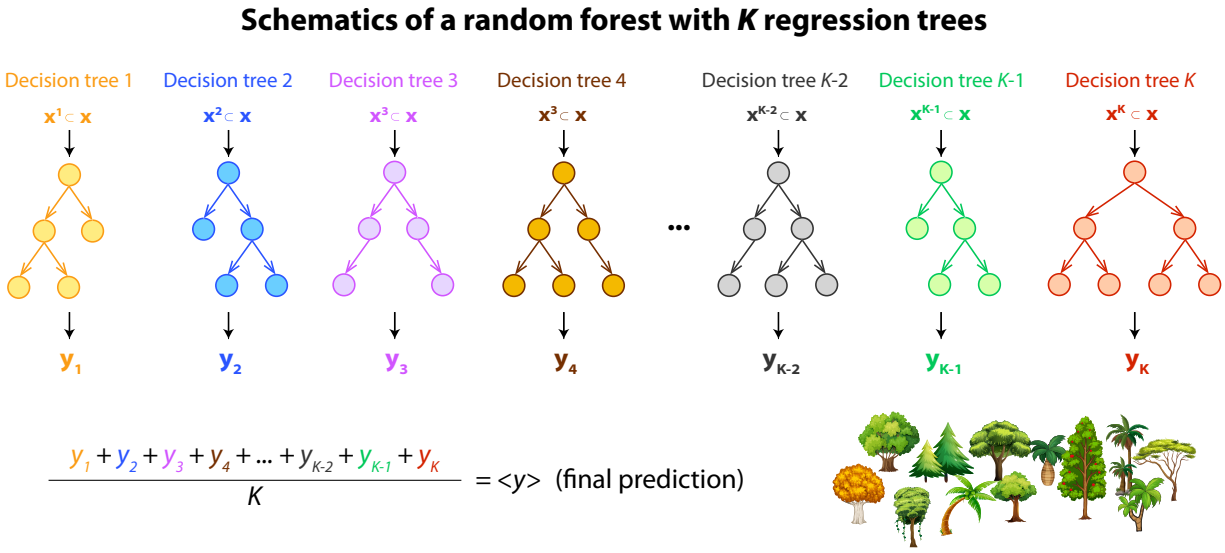

Figure 1.21. Random forest schematics. The RF method trains multiple decision trees in parallel, each of them trained on a subset of the training set drawn with replacement, and then provides a final prediction as an average of predictions by individual trees (tree images were downloaded from www.freepik.com).

as the *random forest* (RF). It has found numerous application in both classification and regression tasks in recent years. In addition to bootstrapping of the data samples, RF tries to further decorrelate the random variables and at each split considers only a fraction of the total number of features. In particular, if one of the variables serves as a very strong predictor, it will likely be selected for splitting in many trees, and to avoid this, a subset of features is selected which makes the individual trees in the ensemble less similar to one another. The main advantages of the RF technique is that the predictions are easily interpretable, and the training phase can be parallelized very well.

While deep trees represent a good choice for bagging methods, shallow trees can be considered a better choice for sequential modeling using boosting. The ensemble "boosted" model learns iteratively to achieve lower bias than the individual weak learners. In particular, in boosting technique, the models on the next iterations give more emphasis to the most difficult data instances that could not be modeled well in previous iterations[James et al., 2013]. It is worth noting that unlike in RF, the conventional boosting methods cannot be parallelized due to the sequential nature of the algorithm. Thus, another motivation behind choosing shallow models as building blocks of the algorithms, as they take much less time to train.

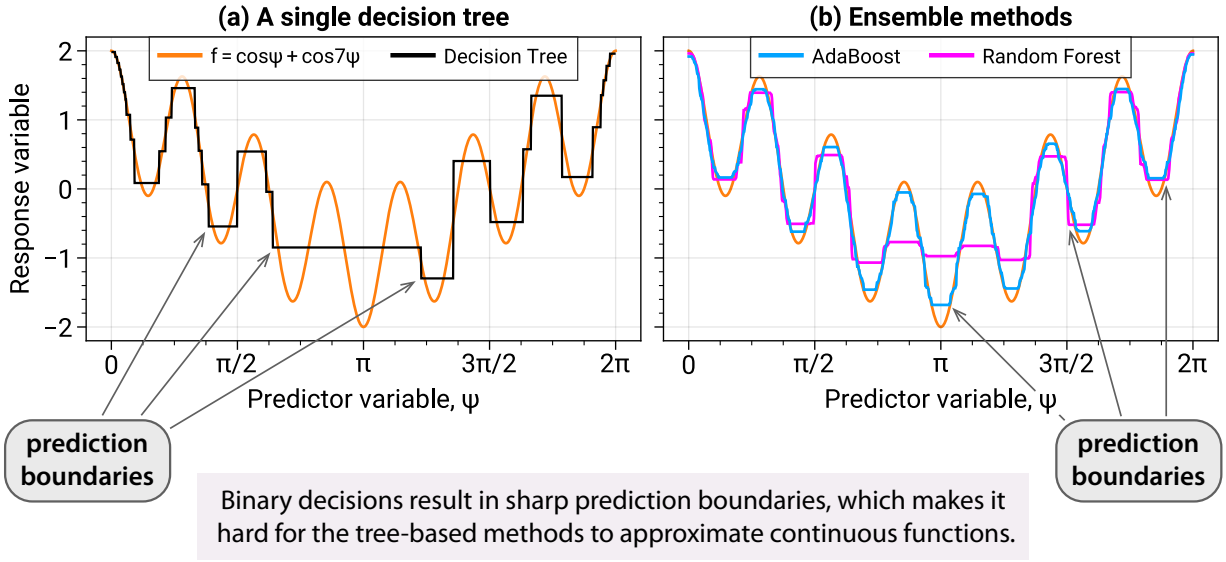

Figure 1.22. An illustration of using the tree-based methods for approximating continuous functions, which shows limitations of these algorithms.

There are two major strategies on how to combine the individual base models in boosting, namely the adaptive boosting (AdaBoost) and gradient boosting. These two approaches refer to updating the weights attached to the individual data points, versus updating values of the data points themselves (e.g., by taking a difference from the previous iteration). In particular, AdaBoost [Freund and Schapire, 1997] attaches weights to each of the training examples, initially chosen as $w_i = 1/n$ for $i \in [\![1, n]\!]$, and updates them at each iteration to give more emphasis to the poorly reproduced data points. Furthermore, each of the weak learners built in sequence also has a corresponding weight proportional to its performance. The final prediction is obtained as a weighted sum of the base estimators.

In gradient boosting [Friedman, 2001], the ensemble model is also built as a weighted sum of weak learners. However, there is a difference in the approach to aggregate the weak learner models, which is treated as a gradient descent problem, hence the name of the algorithm. Each iteration fits a newly created tree on *pseudo-residuals*, which are differences between the known output values of the training set and predictions of the ensemble on the previous iteration. The trees are summarized in the opposite direction to the gradient of the fitting error, and their individual contributions are scaled by the learning rate.

Ensembles of decision trees, including random forests and gradient boosting, have been used in many classification and regression problems with great success. They are known to perform very well on tabular data, sometimes outperforming even deep neural networks. Furthermore, they are easily interpretable and can provide importance scores for different variables. There are also several drawbacks of the tree ensembles. For instance, they are prone to correlated input features and generally require long training times. Moreover, the tree-based methods in general struggle with approximating continuous functions, due to the fact that decision trees perform binary splits which produce sharp prediction boundaries. Ensemble techniques that involve randomization and training many uncorrelated DTs may help with making these boundaries less evident but cannot remove them completely. Figure 1.22 provides an example of approximating a continuous function with a single decision tree consisting of 5 layers (panel a), and ensembles of 150 trees obtained using RF and AdaBoost methods (panel b). It is evident that prediction boundaries are less severe when ensemble modeling is employed but are not completely smoothed out. In several space physics applications, such as ionospheric modeling, it is crucial to have a continuous output that does not exhibit sharp discontinuities. For such problems, Artificial Neural Networks (ANNs), which are universal function approximators [Cybenko, 1989], may be considered a better choice.

1.7.7 Artificial neural networks

Artificial neural networks represent a broad group of non-linear models that consist of series of interconnected nodes. ANNs have been used in a wide variety of applications, such as classification, regression, image recognition, showing state-of-the-art performance. The neural networks are typically comprised of a very large number of parameters (weights and biases); the exact number depends on a problem, but it is not uncommon to have millions of adjustable parameters. These parameters explore the connection between the input and output variables and are capable of finding even the most non-trivial dependencies in the data sets. Determining these parameters is an optimization problem which involves minimization of the objective function and is usually referred to as the model "training".

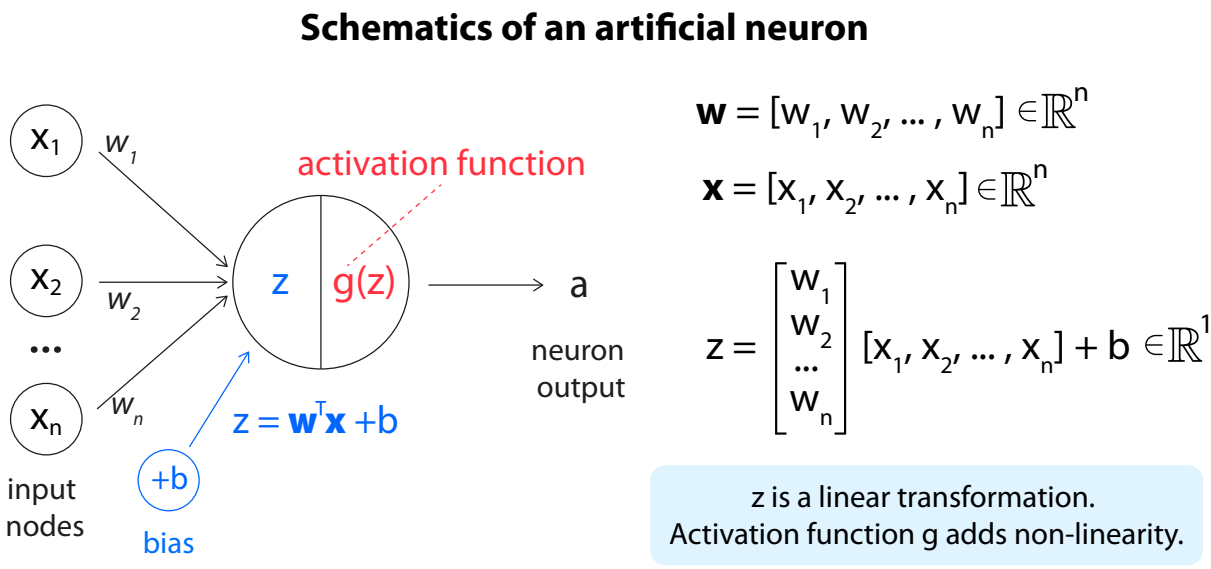

Figure 1.23. Schematics of a single artificial neuron.

The base units of ANNs are referred to as *neurons*. This terminology has been inspired by the biological connection with the human brain and the way that the biological neurons are connected with one another. At the same time, whether ANNs are actually similar to the biological systems remains an open question, as it is unclear if the biological networks can have as many layers as the ANNs (sometimes up to thousands), and it is also uncertain whether the biological networks are updated using back-propagation in the same way as the ANNs [e.g., Ng, 2000].

Let us, firstly, consider the principle of operation of a single neuron (Figure 1.23). Let us assume an n-dimensional input vector $\mathbf{x} \in \mathbb{R}^n$. The output of a neuron can be expressed as:

$$z = \mathbf{w}^\top \mathbf{x} + b, \tag{1.66}$$

where $\mathbf{w}$ is a weight vector ($\mathbf{w} \in \mathbb{R}^n$), and b is the associated bias, introduced to make the model more flexible. The neuron computes a sum of the inputs with an added bias to produce the output z as a weighted sum of the inputs. One can see that a single artificial neuron can be used to solve problems such as linear regression and binary classification, but its capacity is limited, since Equation 1.66 defines an affine transformation. It is also useful to show that stacking many neurons of this type together would not break this linearity, and therefore a modification of this computational system would be necessary. In particular, if we consider a layer with several

neurons $\mathbf{z}^{[1]} = \mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]}$, and pass its output through a second layer $\mathbf{z}^{[2]} = \mathbf{W}^{[2]}\mathbf{z}^{[1]} + \mathbf{b}^{[2]}$, the overall result can be written as:

$$\mathbf{z}^{[2]} = \mathbf{W}^{[2]}\left(\mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]}\right) + \mathbf{b}^{[2]} = \mathbf{W}^{[2]}\mathbf{W}^{[1]}\mathbf{x} + \mathbf{W}^{[2]}\mathbf{b}^{[1]} + \mathbf{b}^{[2]} = \mathbf{W}'\mathbf{x} + \mathbf{b}', \tag{1.67}$$

which is another linear transformation. In order to address more complex problems, it is necessary to overcome this linearity and pass the output of a neuron through some function that could introduce non-linear effects. One example of such function is:

$$g(z) = \mathrm{ReLU}(z) = \max(z, 0), \tag{1.68}$$

which is known as the REctified Linear Unit (ReLU) function [Nair and Hinton, 2010; Glorot et al., 2011]. Such functions are called *activation* (or, transfer) functions, due to their similarity to the neural circuits in neuroscience. The output of a neuron passed through an activation function is:

$$a = g(\mathbf{w}^\top\mathbf{x} + b). \tag{1.69}$$

Different activation functions that are frequently used in neural network training and their first derivatives are shown in Figure 1.24.

Let us now consider a network with an input layer (denoted as the 0-th layer), $r-1$ intermediate layers, which are also called *hidden layers* since the neurons within them are not directly observable, and the output layer (denoted as the r-th layer). This architecture is shown in Figure 1.25. Previously, Equations 1.66 and 1.69 used a column vector of the neuron weights as $\mathbf{w} \in \mathbb{R}^n$ and, therefore, it was necessary to transpose it to take a dot product with the input vector $\mathbf{x} \in \mathbb{R}^n$. When deeper networks with several layers are concerned, we will define a weight matrix for each layer l. The rows of the matrix correspond to neurons within the layer l and contain weights connecting these neurons to the ones in the preceding layer $l-1$; therefore, we will place the transposed weight vectors into the rows of the matrix, and can write:

$$\mathbf{W}^{[l]} = \begin{bmatrix} w^{[l]}_{1,1} & w^{[l]}_{1,2} & \dots & w^{[l]}_{1,N_{l-1}} \\ w^{[l]}_{2,1} & w^{[l]}_{2,2} & \dots & w^{[l]}_{1,N_{l-1}} \\ \vdots & \vdots & & \vdots \\ w^{[l]}_{N_l,1} & w^{[l]}_{N_l,2} & \dots & w^{[l]}_{N_l,N_{l-1}} \end{bmatrix} = \begin{bmatrix} — & \left(\mathbf{w}^{[l]}_1\right)^\top & — \\ — & \left(\mathbf{w}^{[l]}_2\right)^\top & — \\ & \vdots & \\ — & \left(\mathbf{w}^{[l]}_{N_l}\right)^\top & — \end{bmatrix} \in \mathbb{R}^{N_l \times N_{l-1}}, \ \mathbf{b}^{[l]} = \begin{bmatrix} b^{[l]}_1 \\ b^{[l]}_2 \\ \vdots \\ b^{[l]}_{N_l} \end{bmatrix} \in \mathbb{R}^{N_l}, \tag{1.70}$$

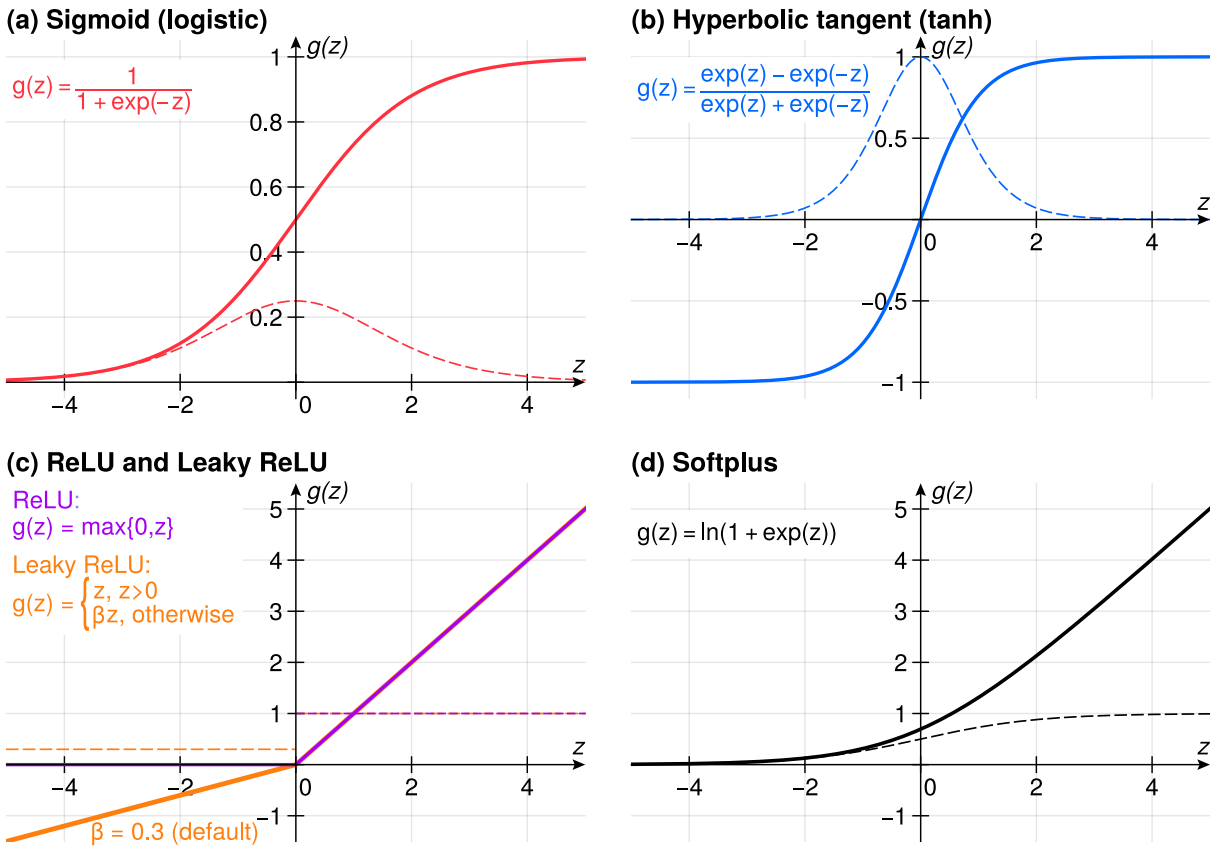

Figure 1.24. Different activation functions (bold solid lines) and their first derivatives (dashed lines).

where N_l is a number of neurons in layer l.

In terms of these notations, let us define how the network computes the output when supplied with the input data. For simplicity, in this chapter we consider a single measurement instance passed through the network, both in the forward and backward directions, while the generalized formulas for minibatch training are given in Appendix B. The *forward pass* of the multilayer perceptron (MLP) can be defined as:

$$\begin{cases} \mathbf{z}^{[l]} &= \mathbf{W}^{[l]}\mathbf{a}^{[l-1]} + \mathbf{b}^{[l]}, \\ \mathbf{a}^{[l]} &= g^{[l]}(\mathbf{z}^{[l]}), \end{cases} \tag{1.71}$$

for $l = 1, ..., r$. This rule also holds for the first layer, considering that $\mathbf{a}^{[0]} = \mathbf{W}^{[1]}\mathbf{x}$, and for the last layer, where the final activation is an identity function and $\hat{y} = a^{[r]} = z^{[r]}$. The network starts by passing the inputs through the first hidden layer, and keeps applying Equation 1.71 until the

output layer; therefore, one can reformulate the forward pass as a composition of functions:

$$\hat{\mathbf{y}} = h_{\boldsymbol{\theta}}(\mathbf{X}) = \underbrace{\left(\mathcal{H}^{[r]} \circ \mathcal{H}^{[r-1]} \circ ... \circ \mathcal{H}^{[1]}\right)_{\boldsymbol{\theta}}}_{\textbf{model}}(\mathbf{X}), \tag{1.72}$$

where $\mathcal{H}^{[l]}$ denotes the transformation given in Equation 1.71 and $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{b}\}$ is the set of model parameters. During each training iteration, the forward pass is used to compute the value of the network output $\hat{y}$ which is then compared to the "true" value y and is used to calculate the loss function value. The loss function quantifies the discrepancy between the model prediction and the target values [e.g., Hastie et al., 2009]. For regression tasks, MSE is frequently used as a loss, but it is worth noting that other functions (e.g., Mean Absolute Error – MAE) are also applied frequently and the choice of the loss function is application-specific. The purpose of the training process is to find such weights that the loss function value is minimal. This minimization is done using gradient descent, which iteratively adjusts the network parameters $\boldsymbol{\theta}$ proportionately to the negative values of the partial derivatives of the loss function with respect to model parameters.

Repeated computations of the partial derivatives are inefficient and computationally expensive, due to the fact that ANNs may have millions of adjustable parameters and these calculations need to be performed many times [James et al., 2013; Goodfellow et al., 2016]. There is a much more elegant approach, known as *backpropagation*. It was proposed by a master student Seppo Linnainmaa [1970] and later on applied to ANN training in a *Nature* paper of Rumelhart et al. [1986], which made this approach widely popular. Backpropagation which is often called *backprop*, tracks changes of the loss backward, from the last NN layer to the first one (hence the name). It exploits the layer-wise network structure to compute the partial derivatives and reuses the quantities that are shared, which helps to significantly reduce the training times.

It is evident that the loss function value depends on the output of the network $\hat{y}$. Since $\hat{y}$ equals to the output of the neuron in the last layer ($z^{[r]}$), one can take a derivative of J with respect to $z^{[r]}$, defined as:

$$\delta^{[r]} = \frac{\partial J}{\partial z^{[r]}} = \frac{\partial J}{\partial \hat{y}} = \hat{y} - y, \tag{1.73}$$

assuming the network has one output neuron. The output of the last layer obviously depends on neurons in previous layers, and therefore it is possible to use the chain rule of calculus and write

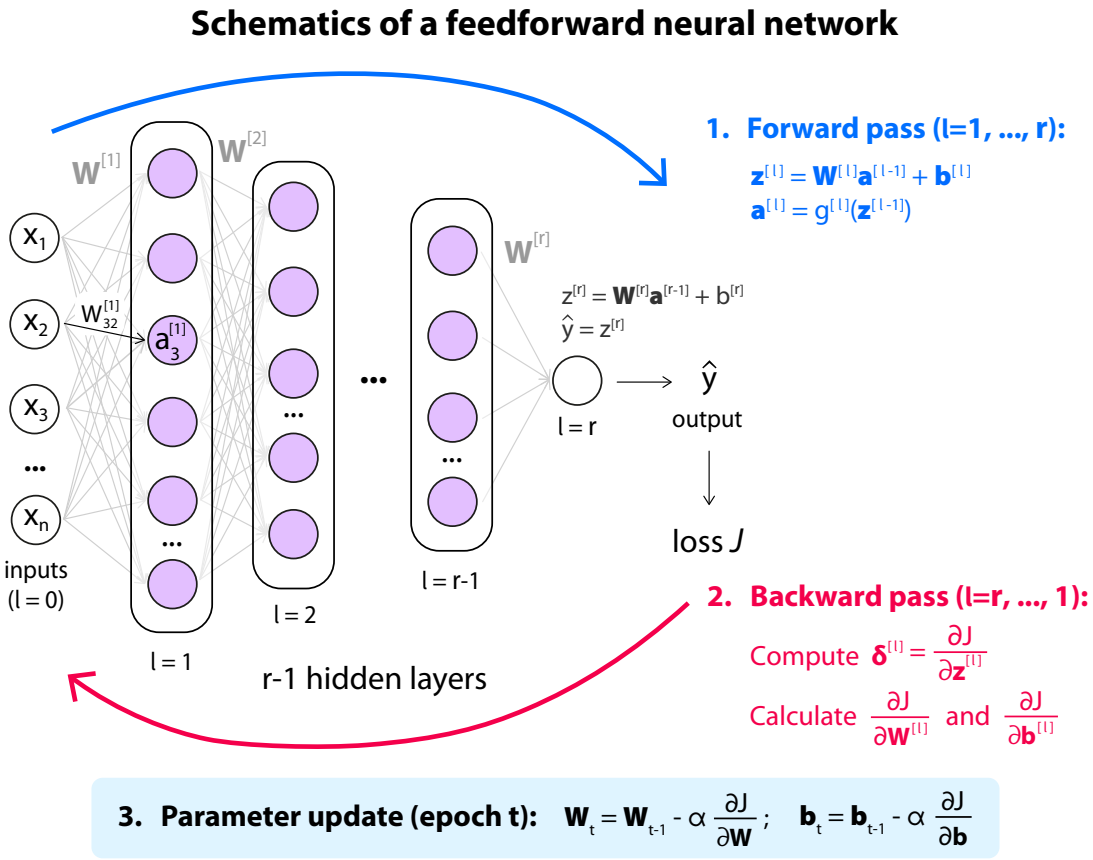

Figure 1.25. Schematics of a feedforward neural network. The forward pass of the network computes the output when supplied with the input. The loss function quantifies the difference between the model output and the target value. It is possible to calculate partial derivatives of the loss function with respect to weights ($\partial J/\partial \mathbf{W}$) and biases ($\partial J/\partial \mathbf{b}$) in different layers (see Equation 1.75), which defines the backward pass.

for layers $l = r - 1, ..., 1$:

$$\boldsymbol{\delta}^{[l]} = \frac{\partial J}{\partial \mathbf{z}^{[l]}} = \left(\left(\mathbf{W}^{[l+1]} \right)^{\top} \boldsymbol{\delta}^{[l+1]} \right) \odot g' \left(\mathbf{z}^{[l]} \right), \tag{1.74}$$

where $\odot$ denotes the Hadamard product. It should be noted that a derivative of a scalar function with respect to a vector has the same dimension as the vector [e.g., Deisenroth et al., 2020], and therefore $\frac{\partial J}{\partial \mathbf{z}^{[l]}} \in \mathbb{R}^{N_l}$. The $\boldsymbol{\delta}^{[l]}$ vectors represent the errors in the activations of layer l, and can be used to express the partial derivatives of the loss function with respect to weights and biases in the layer as:

$$\frac{\partial J}{\partial \mathbf{W}^{[l]}} = \boldsymbol{\delta}^{[l]} \left(\mathbf{a}^{[l-1]} \right)^{\top}, \text{ and } \quad \frac{\partial J}{\partial \mathbf{b}^{[l]}} = \boldsymbol{\delta}^{[l]}. \tag{1.75}$$

These partial derivatives are used to update the weights and biases of each layer as follows:

$$\begin{cases} \mathbf{W}^{[l]} := \mathbf{W}^{[l]} - \alpha \dfrac{\partial J}{\partial \mathbf{W}^{[l]}}, \\ \mathbf{b}^{[l]} := \mathbf{b}^{[l]} - \alpha \dfrac{\partial J}{\partial \mathbf{b}^{[l]}}. \end{cases} \tag{1.76}$$

Equation 1.76 uses a constant learning rate α, but, generally, there can be different learning rates for each of the parameters.

Updating the weights during training is a very crucial step of the model training. The weight updates depend on the derivatives of the activation function $g'\left(\mathbf{z}^{[l]}\right)$, which are evaluated for each neuron. If these gradients have near-zero values, the updates would be negligible and the network would not improve with training. In literature, this is referred to as the *vanishing gradient* problem. In particular, Figure 1.24 demonstrates several activation functions and their first derivatives with respect to z. It can be seen that the sigmoid and hyperbolic tangent activations, which were very popular activation choices in the early days of machine learning research, have derivatives that approaches zero in both left and right directions. Therefore, these functions are known to suffer from the vanishing gradient problem and were later on replaced with the ReLU function and its modifications. ReLU is different from the sigmoid and tanh functions in that its first derivative does not equal to zero most of the time. However, one can notice that for negative input values, the derivative is indeed zero, and therefore another modification of ReLU, known as the *Leaky ReLU*, has been developed. It should also be noted that both ReLU and Leaky ReLU are piece-wise linear functions, and therefore in some cases this can lead to piecewise linear (non-smooth) output of the network. Therefore, in order to produce a smooth output, other modifications of ReLU can be used, such as, for example, the softplus activation function (see Figure 1.24d).

1.7.8 Machine learning pipeline

All of the ML algorithms described above, and many more methods that were not used within this thesis but are frequently employed in various space physics problems share similar steps of model development. Therefore, one can finish the description of the ML techniques with a general "recipe" to building machine learning models.

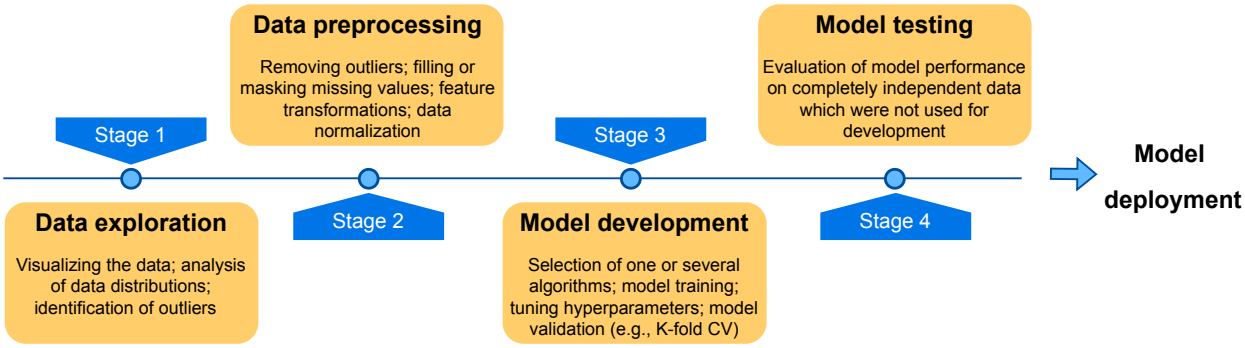

Figure 1.26. Schematics of the general machine learning pipeline.

Step 1 – Data exploration. One of the important initial steps of the machine learning pipeline is to collect and explore the data before proceeding with model development. In particular, the quality of data will, in many ways, determine the model's ability to learn and therefore needs to be checked in advance. In broad terms, a "good" data set for machine learning should contain only very few missing or repeated values and should also be balanced with respect to different classes of events (e.g., have sufficient coverage of different solar/geomagnetic activity levels). Therefore, at this step it is crucial to analyze data distributions, identify the outliers, and explore the general patterns seen in the data set.

Step 2 – Data preprocessing. The next important step involves preparation of the data for training. In particular, one needs to merge observations from different sources and unify them, clean out the outliers, missing values and duplicate instances. Furthermore, after visualizing the data and exploring the preliminary patters and dependencies, one may need to use feature engineering, for instance, by applying the feature mapping technique (Section 1.7.4) to computes different transformations of the initial attributes. This involves, for example, replacing the cyclic features with their sine and cosine values, calculating different powers of the initial input variables, etc. Moreover, another crucial step for modeling the time-series (and many of the space physics problems deal with time-series) is to calculate the time-history of the predictor variables, as the near-Earth space environment generally responds with a time lag to solar wind driving, and this time lag needs to be determined separately for each problem. After the input and output features have been prepared, one can apply data normalization and split the available

observations into the data set which will be used for model development (consisting of training and validation) and test sets.

Step 3 – Model development. To develop a model, one needs to choose one or several algorithms that are relevant to the task. For instance, different methodologies should be applied to the time-series forecasting versus image recognition. Furthermore, it is necessary to make sure that the method is suitable for the numerical or categorical data which are used. In practice, different machine learning algorithms have their own advantages and disadvantages, and it is often a combination of several methods that yields the best result. The models which combine several ML models are known as *superlearners* and are becoming increasingly popular.

The model training is the core of the machine learning pipeline. The selected algorithm is optimized to find the best model parameters within this step. Over time, the model is able to better and better capture the underlying phenomena and to reproduce the training data. However, one needs to note that if left unchecked the model can easily overfit the training data and learn the noise instead of the useful dependencies. Therefore, one needs to regularly check that the model is not only fitting well to the training set, but also to more independent validation data. These checks also need to be performed when tuning the model hyperparameters. There are multiple strategies to split the data into the training and validation data, including the K-Fold cross-validation (CV); they are described in detail in the publication-based chapters. The bottomline for this step is to select an algorithm suitable for the problem, tune the model architecture and minimize overfitting by monitoring the validation error during the model training.

Step 4 – Model testing. After the model has been trained, it is necessary to check its performance on the data that were not seen during the model construction. If the testing is performed on the same data that were used to fit the model, it would not be possible to get a realistic measure of the model performance, and therefore these data need to be withheld during the development, and separated in time from the training and validation sets. The test set can be viewed as the ultimate "judge" of the model performance, and if the model achieves similar accuracy levels on the training and test sets, is has a good generalization ability.

Final steps - Model deployment and maintenance. After the model has been developed and tested on unseen data, it can be deployed for usage by the scientific community, stakeholders,

etc. At this step, it is necessary to save the model in a format that is easy to understand and use, provide scripts to run the model, and upload it to an open access repository. It is also beneficial to include a feedback loop from the community, so that the users could point out where the model can be improved in the future. Maintaining the model and updating it as new data sources become available also constitute a very important part of the modeling process and can lead to gradual improvement of the developed models.

1.8 Contributions of the dissertation

The scope of this dissertation is to develop new data-driven models of different regions of the space environment that pose hazards to satellite infrastructure and communications. Specifically, this work aims at creating modeling tools for the nowcast prediction of the radiation belts, ring current and ionosphere, based on long-term observations collected over the past two decades.

In Chapter 2, we develop a new global model of ring current electrons for the medium Earth orbit (MEO) using a large data set of electron flux measured by the Global Positioning System (GPS) constellation at energies ~120–600 keV. These electrons are driven by both convective and diffusive transport and, therefore, require sophisticated four-dimensional physics-based modeling codes. We demonstrate that their dynamics can be predicted very well using the Light Gradient Boosting Machine (LightGBM) algorithm. The model is driven by the solar wind and geomagnetic indices, including time-history of SW velocity, and does not require previous flux values as inputs. The developed Medium Energy electRon fLux In Earth's outer radiatioN belt (MERLIN) model is validated using 10-fold cross-validation (CV), and is tested on 1.5 years of unseen observations, achieving very good performance with a correlation between the observed and predicted fluxes of ~80% and near-zero bias. The chapter is based on the results from Smirnov et al. [2020a].

Chapter 3 presents a large-scale statistical study of pitch angle distributions (PADs) of electrons in the radiation belts and ring current. PADs are critically important for understanding the dynamics of electron populations and for conversion of electron fluxes to phase space densities in terms of adiabatic invariants. PADs are highly variable with activity, which has been known

since the early studies by Lyons and Williams [1975a]. However, due to the lack of high quality observations prior to the Van Allen Probes era, it was not possible to undertake a large scale statistical study of their behavior. The Van Allen Probes mission produced high fidelity electron flux observations which allowed us to analyze PAD behavior with a very high energy resolution. To understand the impact of geomagnetic storms on the equatorial pitch angle distributions, we perform a superposed epoch analysis study of the PAD morphology during 129 geomagnetic storms throughout the Van Allen Probes era. The chapter is based on the findings from Smirnov et al. [2022a].

In Chapter 4, we use the dependencies reported in Chapter 3 to build an empirical model of the equatorial pitch angle distributions in the outer radiation belt. This is the first PAD model based on the Fourier series expansion, and the first model with a continuous dependence on L, MLT and activity. The chapter proposes two methods to reconstruct full equatorial PADs uni-directional observations, and in particular, it demonstrates a way to reconstruct equatorial electron flux using low-pitch angle data, which can be used to convert the available long-term data sets of uni-and omnidirectional electron fluxes to phase space density. The chapter is based on the results from Smirnov et al. [2022c].

Chapter 5 presents a method of converting electron fluxes measured by the Cluster's Research with Adaptive Particle Imaging Detector (RAPID)/Imaging Electron Spectrometer (IES) detector to phase space density. The Cluster constellation has provided over two solar cycles of electron flux observations in the outer belt, which can be used for the radiation belt science. However, no adiabatic invariants have been calculated for Cluster. The chapter presents a method to convert the RAPID/IES observations to phase space densities and outlines the general characteristics of the radial PSD profiles measured by Cluster. These results were used to create the `LSTAR` product for the European Space Agency's Cluster Science Archive. The chapter is based on the study by Smirnov et al. [2020b].

Chapters 6 and 7 focus on the topside ionosphere, which is the region that contains the largest fraction of the total electron content and is thus critically important for GNSS applications. Over the last two decades, the ionosphere has become a data-rich environment, with several billion observations provided by different observational techniques. While multiple comparisons between

the data sets have been presented over the years, the systematic intercalibration of all available data sources has not been undertaken. In Chapter 6, we present an intercalibration of the plasma density measurements by five most prominent and widely used ionospheric constellations. The intercalibration factors presented in this dissertation eliminate the systematic differences between the most widely used ionospheric data and allow using these data sets together for many applications, including data-driven ionospheric modeling. Furthermore, we showed, for the first time, that the Langmuir Probes (LPs) onboard the Swarm mission overestimate the nighttime electron densities by up to 40-50% during low solar activity. The chapter is based on the results from Smirnov et al. [2021].

In Chapter 7, we develop a novel empirical model of electron density in the topside ionosphere, based on a large three dimensional radio occultation data set covering almost two solar cycles of data. The model predicts 4 parameters of the alpha-Chapman function with a linear decay of scale height with altitude using feedforward neural networks. The developed model is tested on in-situ observations by 3 independent missions and is capable of predicting electron densities with very high fidelity, as the model predictions lie within a factor of 2 from the measurements 90% of the time. The developed model outperforms the current topside options included into the International Reference Ionosphere (IRI) model by up to 80%, especially around 100-200 km above hmF2. The chapter is based on the results from Smirnov et al. [2023].

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

I. The radiation belts and ring current:

   1. Developed the first machine learning-based model of hundreds-of-keV electron fluxes at MEO that does not use flux values as inputs and is driven only by solar wind and geomagnetic indices, based on GPS data.
   2. Selected the input features for the ML model of electron flux, including the time-history of solar wind parameters. Demonstrated that even for hundreds-of-keV electrons, including long-term history of up to 2 weeks significantly improved the model performance.

3. For the first time, applied Fourier sine series to study pitch angle distributions of electrons in Earth's inner magnetosphere. Developed a technique that can fit all PAD types and can be easily integrated to derive the omnidirectional fluxes.
4. Performed, for the first time, a statistical analysis of electron pitch angle distributions during geomagnetic storms using the entire Van Allen Probes data set at energies below ~1.6 MeV using Fourier sine series.
5. Developed the first empirical model of equatorial electron pitch angle distributions that has a continuous dependence on L, MLT and magnetospheric activity.
6. Calculated adiabatic invariants for the Cluster mission and created a product containing the $L^{\star}$ coordinate, which was adopted by the European Space Agency.

II. The ionosphere:

1. Intercalibrated the ionospheric data sets from the Gravity Recovery and Climate Experiment (GRACE), Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC), CHAllenging Minisatellite payload (CHAMP), Swarm, and Communications/Navigation Outage Forecasting System (C/NOFS) missions.
2. For the first time showed that the Swarm Langmuir Probes exhibit nighttime overestimation of electron density by up to ~40–50%.
3. Developed a neural network-based model of electron density in the topside ionosphere that gives highly accurate predictions and outperforms the international standard model, the IRI, by up to an order of magnitude.
4. Validated the developed NN-based electron density model using independent in-situ observations and showed that the model predictions are within a factor of 2 from the data ~90% of the time on all of the employed missions. Furthermore, the model can resolve both the large-scale dynamics of the topside ionosphere and small-scale structures such as the midlatitude ionospheric trough.

Publication-based chapters

— 2 —

Medium energy electron flux in Earth's outer radiation belt (MERLIN): a machine learning model *

Abstract

The radiation belts of the Earth, filled with energetic electrons, comprise complex and dynamic systems that pose a significant threat to satellite systems. While various models of electron flux both for low and relativistic energies have been developed, the behavior of medium energy (120-600 keV) electrons, especially in the MEO region remains poorly quantified. At these energies, electrons are driven by both convective and diffusive transport, and their prediction usually requires sophisticated 4D modeling codes. In this paper we present an alternative approach using the LightGBM machine learning algorithm. The MERLIN model takes as input the satellite position, a combination of geomagnetic indices and solar wind parameters including the time history of velocity, and does not use persistence. MERLIN is trained on >15 years of the GPS electron flux data, and tested on more than 1.5 years of measurements. 10-fold cross validation yields that the model predicts the MEO radiation environment well, both in terms of dynamics and amplitudes of flux. Evaluation on the test set yields high correlation between the predicted and observed electron flux (0.8) and low values of absolute error. The MERLIN model can have wide space weather applications, providing information for the scientific community in the form of radiation belts reconstructions, as well as industry for satellite mission design, nowcast of the MEO environment and surface charging analysis.

*This chapter has been published as Smirnov, A.G., Berrendorf, M., Shprits, Y.Y., Kronberg, E.A., Allison, H.J., Aseev, N.A., Zhelavskaya, I.S., Morley, S.K., Reeves, G.D., Carver, M.R. and Effenberger, F. (2020): Medium energy electron flux in Earth's outer radiation belt (MERLIN): A machine learning model. *Space Weather*, 18(11), e2020SW002532.

Plain Language Summary

The radiation belts of the Earth, which are the zones of charged energetic particles trapped by the geomagnetic field, comprise complex and dynamic systems posing a significant threat to a variety of commercial and military satellite systems. While the inner belt is relatively stable, the outer belt is highly variable and depends substantially on solar activity; therefore, accurate and improved models of electron flux in the outer radiation belt are essential to understand the underlying physical processes. Although many models have been developed for the geostationary orbit and relativistic energies, prediction of electron flux in the 120-600 keV energy range still remains challenging. We present a data-driven model of the medium energies (120-600 keV) differential electron flux in the outer radiation belt based on machine learning. We use 17 years of electron observations by GPS satellites. We set up a 3D model for flux prediction in terms of L-values, MLT and magnetic latitude. The model gives reliable predictions of the radiation environment in the outer radiation belt and has wide space weather applications.

2.1 Introduction

The Van Allen radiation belts, discovered by the array of Explorer satellites [Van Allen and Frank, 1959], are zones of charged energetic particles, mainly electrons and protons, trapped by the magnetic field of the Earth. Energetic electrons (> 100 keV) are mainly confined to two regions - the inner belt, within L from 1.2 to 2.5, and the outer belt located between L from ~ 3 to 7 [Lyons et al., 1972; Summers et al., 2004]. The inner and outer electron radiation belts are separated with a so-called slot region, usually devoid of energetic electrons [Lyons and Thorne, 1973; Kavanagh et al., 2018]. The inner radiation belt is known to exhibit long-term stability, while the outer belt is highly dynamic and depends substantially on solar activity [Meredith et al., 2006].

The dynamics of the outer radiation belt is governed by a complex interplay between acceleration and loss processes [Reeves et al., 2003]. Electrons with energies of tens of keV, called the source population, are injected into the inner magnetosphere during substorms and produce

waves, for instance, the whistler mode chorus [e.g., Boyd et al., 2014, 2016; Jaynes et al., 2015]. Electrons with energies of hundreds of keV, called the seed population electrons, are also injected in the magnetosphere during substorm activity. These electron populations can accumulate at the surface of the spacecraft and lead to satellite loss due to the so-called surface charging effects [e.g., Garrett, 1981; Lanzerotti et al., 1998]. Furthermore, the seed population electrons can be accelerated to relativistic energies by waves. These relativistic (>1 MeV) particles can penetrate through satellite shielding and damage the equipment onboard, also leading to satellite loss [e.g., Fennell et al., 2000].

To date, there are more than 2200 operational satellites in the Earth's orbit and many of them systematically pass through the radiation belts region. Approximately 1400 spacecraft are in Low Earth Orbit (LEO) at altitudes up to 1000 km. The LEO satellites cross the inner radiation belt in the region of the South Atlantic magnetic anomaly and the outer belt at higher latitudes. The second most populated is the geostationary orbit (GEO) with more than 560 satellites flying at altitudes of ~36000 km synchronously with the rotation of the Earth. GEO satellites generally fly close to the outer edge of the outer radiation belt ($L \sim 6.6$). Satellites flying below GEO and above LEO follow the so-called medium Earth orbit. Many GNSS satellites fly at MEO, for instance, the GPS, GLObal NAvigation Satellite System (GLONASS) and Galileo. Furthermore, in order to reach GEO, an increasing number of spacecraft are using the electric orbit raising method and can spend hundreds of days in the MEO region [Horne and Pitchford, 2015; Glauert et al., 2018]. Satellites following MEO systematically pass through the heart of the outer radiation belt and are exposed to the largest values of electron flux. The number of satellites in Earth's orbit will increase significantly in the following years, and in order to ensure the long-term satellite operation stability, it is necessary to have reliable models of electron intensities at different energies (from tens of keV up to several MeV) and locations.

The existing radiation belt models can be divided into three main categories: physics-based, data-driven, and data assimilation models. Several physics-based models of electron flux have been created for the radiation belts and ring current region. Among them, there are the Versatile Electron Radiation Belt (VERB) [e.g., Subbotin and Shprits, 2009], British Antarctic Survey Radiation Belts Model (BAS-RBM) [Glauert et al., 2014] and Dynamic Radiation Environment

Assimilation Model (DREAM) [Reeves et al., 2012; Tu et al., 2013, 2014] codes based on solving the three-dimensional Fokker-Planck equation to reproduce the dynamics and variability of the MeV radiation belts electrons. The physics-based models typically include the radial diffusion, losses due to pitch-angle scattering and magnetopause shadowing [Glauert et al., 2018]. Recently, the VERB-4D code has been developed to extend the VERB code to lower-energy ring current electrons by including advection terms [Shprits et al., 2015; Aseev et al., 2016]. The physics-based Inner Magnetosphere Particle Transport Model (IMPTAM) was developed and shown to give reasonable flux predictions at energies from several eV up to <150 keV [Ganushkina et al., 2019]. The low energy electrons are also modeled by the coupled Fok Ring Current (FRC) [Fok and Moore, 1997] and Comprehensive Inner-Magnetosphere Ionosphere (CIMI) [Fok et al., 2001] models operating online. It should be noted that there is generally a gap in modeling the electron flux at energies 100-600 keV. Physical modeling at these energies is considered difficult due to the fact that electric field effects have to be considered [Ganushkina et al., 2011] and also because the physics governing the dynamics of electrons at medium energies is not entirely understood [Horne et al., 2013].

The data-driven models can be subdivided into static and dynamic ones. AE8 [Vampola, 1997] and AE9 [Ginet et al., 2013] are examples of static models providing the values of electron intensities from 40 keV up to ultra-relativistic energies. AE8 and AE9 models overcome limitations of the individual data sets by combining large scale statistics and are currently used as a reference for engineering purposes [Glauert et al., 2018]. Dynamic data-driven models typically depend on a combination of solar wind parameters and geomagnetic indices. Several data-driven models have been developed for the GEO orbit. Denton et al. [2015] developed an empirical model of electron flux for low energies (10 eV - 40 keV) based on 82 LANL satellites data, driven by the Kp index. Later, the upstream solar wind conditions were incorporated into the model [Denton et al., 2016], and it was also expanded to 6-20 R_E [Denton et al., 2019] using Cluster data. For relativistic energies, Balikhin et al. [2011] employed a Nonlinear AutoRegressive Moving Average with eXogeneous inputs (NARMAX) technique to predict daily flux at GEO orbit at energies 800 keV and 2 MeV, using solar wind parameters and the previous time-series of daily flux from GOES satellites. The NARMAX model was further extended to

a broader energy range, including electron flux at energies of hundreds of keV [Boynton et al., 2016]. However, the model needs satellite time-series as inputs and therefore for now is confined to the geostationary orbit. Several models of the relativistic electron flux are based on the artificial neural networks, e.g., Ling et al. [2010]; Kitamura et al. [2011]; Chen et al. [2019]; de Lima et al. [2020]. Other empirical models of relativistic flux developed for the GEO region include the Relativistic Electron Flux Model (REFM) driven by solar wind velocity [Baker et al., 1990], an empirical function D_0 dependent on several solar wind parameters and Kp [Li, 2004], and linear regression models [e.g., Sakaguchi et al., 2015; Simms et al., 2014, 2016; Osthus et al., 2014] which take as inputs the solar wind parameters and previous values of flux at GEO. Tsutai et al. [1999] used linear filter to predict the values of $>$ 2MeV flux at GEO 1 day ahead using GOES magnetic field data over the preceding 6 days.

Although a variety of models have been developed for the geostationary orbit, few data driven models exist that give reliable electron flux predictions at MEO. One of the recently developed models of electron flux at MEO working with a daily cadence is the SHELLS model [Claudepierre and O'Brien, 2020] which takes as input the LEO electron flux values and Kp index and returns the flux along Van Allen Probes orbit at energies 0.35-1 MeV. The general lack of models at MEO comes from the fact that many GEO satellites provide continuous high quality observations of electron flux, while at MEO the temporal and spatial coverage of observations remains rather sparse [Sakaguchi et al., 2015]. Indeed, only $\sim$ 100 satellites reside in MEO, and only few of them provide measurements of the radiation belt populations. Among other data sets of electron flux measurements in the MEO region, the recently released GPS energetic particle data have notable advantages such as, for instance, the large number of satellites (23) and uniform MLT coverage, as well as availability of 18 years of observations covering almost 2 solar cycles. Furthermore, most of the GPS satellites carry onboard identical Combined X-ray Dosimeter (CXD) detectors measuring electron flux at energies 0.12-10 MeV. The GPS/CXD data have been inter-calibrated with Van Allen Probes electron flux measurements and the two missions were in good agreement at energies below 4 MeV [Morley et al., 2016].

In the current paper we present the data-driven MERLIN electron flux model, based on machine learning. For model training we employ the LightGBM algorithm, which is known for its

high efficiency and accuracy [Ke et al., 2017]. The model takes as input satellite position in an Lshell - MLT - magnetic latitude (MLat) frame, solar wind parameters with history of velocity, and geomagnetic indices. The model returns values of locally omnidirectional electron flux averaged over the viewing angle of the instrument at energies 120-600 keV as outputs. The structure of the paper takes the form of five parts, including this introductory Section. Section 2.2 describes the data set used for model construction. Section 2.3 is concerned with the methodology used for this study. Section 2.4 presents the results. The conclusions are drawn in Section 2.5.

2.2 Data set

2.2.1 GPS electron flux data

The GPS spacecraft are distributed across six orbital planes, nominally inclined at 55°. The satellites follow near-circular medium Earth orbit, with 12h revolution period, at a constant altitude. As of 2020, the constellation consists of 74 spacecraft, of which 31 are operational, 9 reserve, 2 being tested and 32 no longer in use [www.gps.gov]. Due to its fixed altitude of 20,200 km (R$\sim$ 4.2), the GPS constellation travels through a range of L-shells providing the particle measurements in the outer radiation belt. We note, however, that the inclination of the GPS orbit restricts the range of equatorial pitch angles as a function of L-shell. The satellite at off-equatorial magnetic latitudes cannot observe the particles mirroring at lower MLats and therefore samples only a part of the equatorial pitch angle distribution.

Since the early 2000s, newer GPS satellites are equipped with either of the two instrument series: the improved BDD-IIR or the CXD. Most of the satellites currently carry aboard the identical CXD detectors. Their response is well-known and their electron flux data have been used in previous radiation belts studies [e.g., Olifer et al., 2018; Pinto et al., 2020]. The CXD instruments measure the electron flux using two sensors: the low energy particle (LEP) and the High Energy Particle (HEP) sensors, with the typical sampling rate of 240 seconds [Morley et al., 2016]. In the present study we use data of the first 5 evaluated CXD energies, namely 120, 210, 300, 425 and 600 keV. As of 2020, 21 GPS satellites are equipped with the CXD detectors,

providing more than 200 years of satellite data. The CXD measurements were previously cross-calibrated with the Van Allen Probes electron flux and showed good agreement. Morley et al. [2016] analyzed the response functions for different GPS satellites and found a small systematic difference between satellites up to ns-62 and satellites with designators greater than ns-62. To minimize these differences and cross-calibrate against the Van Allen Probes ECT suite in 2012-2014 the GPS instrument response functions were updated. The publicly-released data product uses the updated response functions. The GPS/CXD data showed a good agreement with Van Allen Probes for energies of up to 4 MeV, while at higher energies larger variance was observed due to the unaccounted instrumental backgrounds.

The present study is based on 17 years of GPS/CXD electron flux measurements at energies 120-600 keV. The flux values were first cleaned using the `'dropped_data'` flag, and also outliers in the data were removed by setting the minimum allowed flux values to 1 $cm^{-2}kev^{-1}s^{-1}sr^{-1}$. Furthermore, in order to remove the unrealistic values coming from the forward flux retrieval procedure, we use a new fit quality flag defined as:

$$F = \max \left| \log_{10} \left(\frac{\text{predicted counts}}{\text{observed counts}} \right) \right|, \tag{2.1}$$

where predicted and observed are arrays of the counts for the first 5 energies (`'model_counts_electron_fit'` and `'rate_electron_measured'` products in the data files, respectively). The flag values corresponding to low quality fits were selected to be $F > 0.11$, and such data entries were removed from the data set.

We use data from 20 satellites (ns53–ns59 and ns61–ns73) carrying the CXD detectors [for details see Carver et al., 2018]. It is of note that the data from ns60 were not used due to the intermittent quality issues. We train the model on data with the original cadence of measurements equal 240 seconds. Before fitting the model, we applied the base 10 logarithm to the GPS electron flux, as the data variance can be up to several orders of magnitude.

2.2.2 Solar wind and geomagnetic indices

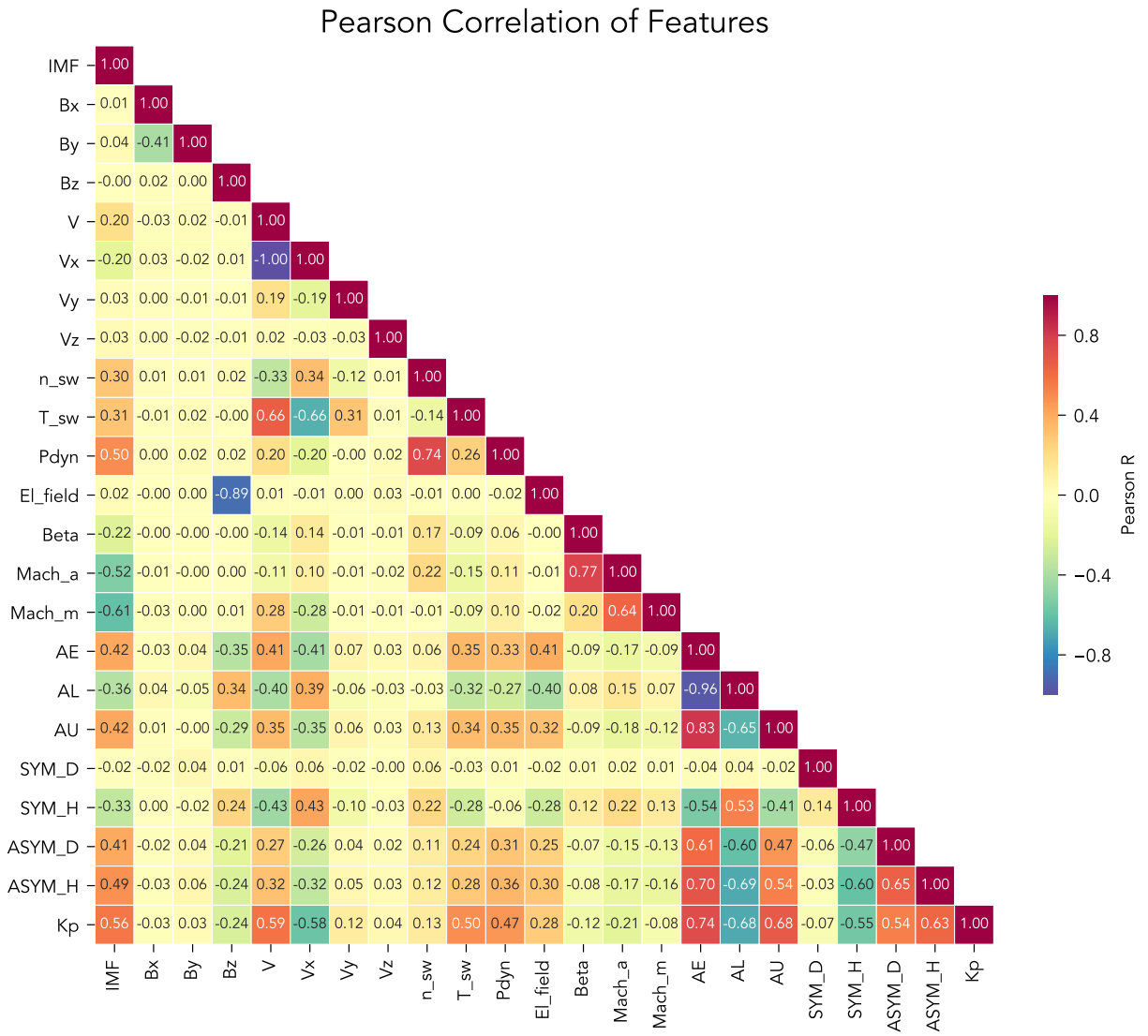

Figure 2.1. Correlation between different solar wind and geomagnetic parameters based on 2001-2016 OMNIWeb data.

The relationship between the electron flux intensities in the outer radiation belt and solar wind parameters has long been recognized [e.g., Paulikas and Blake, 1979; Reeves et al., 2011]. Numerous studies have analyzed contributions of the solar wind parameters to flux enhancements. The independent contributions of solar wind velocity and number density were investigated, for instance, by Balikhin et al. [2011]; Kellerman and Shprits [2012]; Simms et al. [2014]. A com-

bination of velocity and density, pressure and geomagnetic indices, combined with the previous daily flux value, was examined by Sakaguchi et al. [2015]. Long-term relationship between velocities and MeV electron fluxes was discussed in Reeves et al. [2013].

It has been well established that the radiation belts flux enhancements are connected with changes in solar wind speed. Reeves et al. [2011] analyzed the relativistic electron flux at GEO with respect to the solar wind speed and noted that the resulting distribution resembled a triangle. Such a shape was explained as follows. The V_{sw} values rarely fall below 300 km/s, and this leads to a left-hand side of the triangle [see also Wing et al., 2016]. The top side of the triangle forms due to the fact that the flux values seem to have a sharp maximum at higher V_{sw}, for which multiple explanations have been given. One of the most puzzling features of the triangular distribution is that the variability of electron flux at lower V_{sw} is much larger than at higher V_{sw}. Reeves et al. [2011] noted that the electron flux can exhibit large values under any V_{sw} values. The triangular form demonstrates that using the values of the solar wind velocity and density is not enough to fully explain the variability of flux and therefore other parameters have to be taken into account.

We consider the following solar wind parameters and geomagnetic indices obtained at the OMNIWeb database [omniweb.gsfc.nasa.gov]. First, amongst the solar wind drivers, we analyze the solar wind velocity, and its components V_x, V_y, V_z. We analyze the IMF magnitude, B_x, B_y and B_z components, and also solar wind density n_{sw}. We employ the derived solar wind quantities: magnetosonic and alfvenic Mach numbers (Mach_a and Mach_m, respectively), solar wind temperature T_{sw}, electric field (v*B_z), dynamic pressure (Pdyn) and plasma Beta. From geomagnetic indices, we select SYM-H, SYM-D, ASYM-H and ASYM-D indices, planetary Kp index and auroral AL, AU and AE indices. It has been previously established that many of these features are, in fact, correlated with one another. In Figure 2.1, we show the Pearson linear correlations between different solar wind and geomagnetic parameters in order to examine which features can be used for the model setup.

We find that V_x is perfectly anticorrelated with V, with the -1.0 coefficient, which is as expected since V_x consitutes most of the V amplitude. V_y correlates with V with only 0.19 correlation, and V_z shows zero correlation with velocity magnitude. SYM-D does not correlate with

any of the features, except for very weak (0.14) relationship with SYM-H. SYM-H, on the other hand, correlates with several parameters. For instance, it exhibits a moderate positive correlation with the solar wind velocity (0.43), and negative correlation with ASYM-H (-0.6) and ASYM-D (-0.47). Furthermore, it correlates with auroral geomagnetic indices with approximately 0.5 correlation coefficient, and is also weakly anti-correlated with the electric field (v*B_z). ASYM-D index is correlated with ASYM-H with the R-value of 0.65, and also with auroral indices with the absolute value of the linear correlation ~0.6. In turn, ASYM-H index shows weak linear correlation with solar wind velocity (0.32), IMF (0.49), dynamic pressure (0.36), and exhibits higher correlation with the auroral indices with the corresponding R values of up to 0.7. Kp index exhibits moderate positive correlation with the IMF magnitude (R=0.56), solar wind velocity (0.59), temperature and dynamic pressure (R~0.5), stronger positive correlation with AE (R=0.7) and the corresponding AU and AL indices, along with 0.5 correlation with SYM-H. It should be noted that the Kp index has a 3-hour cadence, and therefore shows lower correlation with AE than one would expect. By averaging the AE index to the same 3h cadence, one obtains a correlation of 0.82. IMF magnitude shows weak correlation (0.3) with solar wind density and temperature, moderate (0.5) correlation with dynamic pressure and Mach numbers.

We note that although the gradient boosting regression trees are prone to the multi-collinearity of features [e.g., Maloney et al., 2012; Ding et al., 2016], using highly correlated inputs can pose a disadvantage for machine learning studies. For example, when 2 parameters are correlated we can achieve the same reduction in variance as by using only one of them. Here we remove several correlated and derived quantities leaving the more in-depth analysis of the influence that different parameters have on the electron flux for further studies. First, we exclude directional components of magnetic field and velocity. Furthermore, we exclude all of the derived quantities, because they encompass information of their original constituent variables (for instance, dynamic pressure strongly correlates with density). Magnetosonic and Alfvenic Mach numbers essentially represent the normalized velocity and it is enough to consider the velocity itself. We apply the same reasoning to the geomagnetic indices selection: AE is a product of AL and AU indices, and also correlates with them, which is why we only use AE for model setup. AE and Kp indices are generally strongly correlated. While Kp is a measure of the planetary geomagnetic

activity, substorms are better resolved by the AE index. Furthermore, Smirnov et al. [2019b] reported the long-term positive correlation of electron flux at energies up to 400 keV with AE index along the solar cycles 23 and 24. The same conclusion was drawn in [Smirnov et al., 2020b] by analyzing the long-term PSD variations of low-μ electrons, indicating the importance of the substorm activity for radiation belts transport processes. For these reasons we include both Kp and AE indices as inputs.

The inner edge of the outer belt is highly dynamic and can move inwards during slot-filling events and outwards during the quiet periods. Li et al. [2006] reported a correlation between the 30-day averages of the innermost edge of the outer belt and the plasmapause location (Lpp) using 12 years of SAMPEX data. The flux values in the slot region, located below the Lpp, are lower than those beyond the plasmapause due to loss processes attributed to storm-enhanced EMIC and plasmaspheric hiss waves [Li et al., 2006] . O'Brien and Moldwin [2003] presented a model of the plasmapause location parametrized as a function of the maximum AE value over the preceding 36 hours. Furthermore, the Lpp model based on the AE index was found to perform better than that using the Dst values. In order to account for the dynamics of the inner edge of the outer belt, we include the maximum value of the AE over 36 hours as an input parameter. We do not apply the linear regression coefficients to convert the max(AE) to Lpp, as the Regression Trees are invariant to linear scaling operations [e.g., Druzhkov et al., 2011].

After the enhancement events, the flux of medium energy electrons decay to their pre-storm values gradually over a period of up to 20 days [e.g., Meredith et al., 2006]. Such a slow decay can be explained by the longer hiss lifetimes, which by different estimates vary from several up to tens of days [Orlova et al., 2016]. Hence, it is crucial to include some indication of the previous state of the radiation belts into the model. This is usually done by adding the preceding values of flux as model inputs [e.g., Simms et al., 2016, 2014; Boynton et al., 2016]. Instead, in the MERLIN model we include the history of solar wind velocity as a proxy of the previous activity. We use the averaged v_{sw} values over the preceding 1, 2, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42 hours and 2, 3, 7 and 14 days. It should be noted that the averages over longer periods of time also carry part of the information from shorter scales averages. In this study we only consider the history of solar wind velocity, while adding the history of number density leads

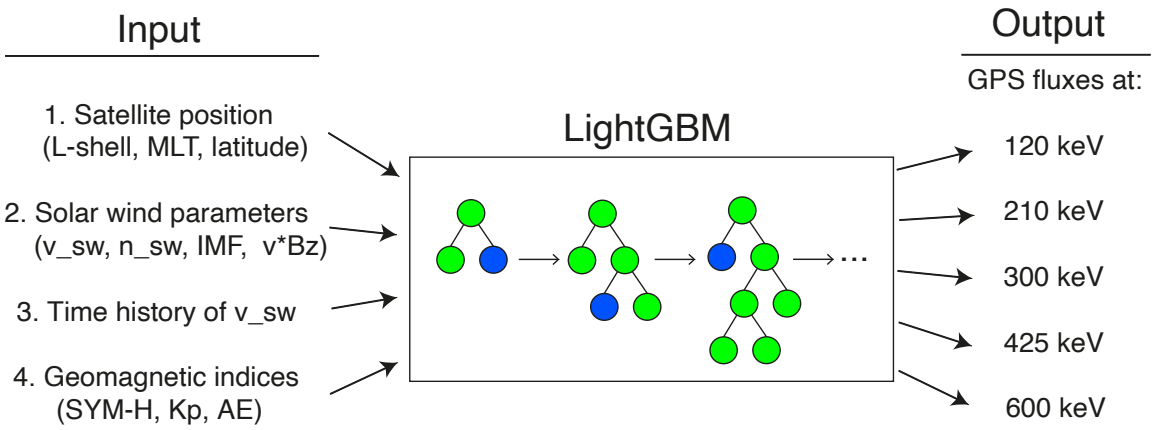

Figure 2.2. Schematic representation of the model workflow. The input parameters include the satellite position in L-MLT-latitude frame, solar wind parameters with history of velocity, and geomagnetic indices. The inputs are supplied to the LightGBM algorithm in order to return the flux values at energies 120-600 keV.

to overfitting, as discussed in Section 4.1. In sum, as input parameters we select the satellite position in the L-MLT-magnetic latitude frame, geomagnetic indices SYM-H, Kp and AE, solar wind parameters - number density, electric field (v*B_z) and velocity with 2 weeks of history (see also Figures 2.2 and 2.4). L-shell values were calculated using T-89 model with the internal field given by IGRF. More features can be incorporated into the model in future, however preserving the methodology, and the more refined feature selection will be performed in a separate study.

2.3 Methodology

Large and growing volumes of data have been provided by the satellite missions in the Earth's radiation belts region. One of the efficient ways to utilize these long-term data sets for modeling is to apply ML techniques. Over the years, machine learning has found numerous applications in the field of space physics research. ML methods have been employed for the geomagnetic indices forecast [e.g., Bala et al., 2009; Shprits et al., 2019], global reconstructions of the plasmaspheric dynamics [Zhelavskaya et al., 2017; Bortnik et al., 2018], solar activity prediction [e.g., Colak and Qahwaji, 2009]. Prediction of the electron flux in the outer radiation belt remains one of the

most challenging tasks in the space weather research [Camporeale, 2019]. In the present study, we use the Light Gradient Boosting approach, described in detail below, to predict the flux of medium energy electrons using the GPS particle data.

2.3.1 Light Gradient Boosting Machine (LightGBM)

One of the predictive approaches, widely used in machine learning, is the so-called Gradient Boosting Decision Tree (GBDT) method. The GBDT algorithms gained popularity for being efficient, highly accurate and interpretable. GBDT is an ensemble model of usually shallow decision trees, also called weak learners, trained in sequence [Friedman, 2001]. Growing each individual tree starts from the source set contained in a root node of the tree (shown in Figure 2.2). When a split is made, the root node is divided into two subsets, and 2 branches are generated [Ren et al., 2019]. The procedure is repeated recursively until either the subset at each node contains all identical values of the target variable, or when the splitting is constrained by the algorithm's hyperparameters (e.g., `max_depth` or `num_leaves` is reached). The GBDTs are grown iteratively and each new tree fits the residuals of the previous iteration to account for the mis-modeled instances [Freund et al., 1999]. GBDTs have been applied successfully to many machine learning problems, performing well for regression and classification tasks alike. Numerous GBDT implementations have been developed, starting from Adaptive boosting (AdaBoost - [Freund et al., 1999]). One of the most popular gradient boosting methods up to date is the Extreme Gradient Boosting Machine (XGBoost) [Chen and Guestrin, 2016], famous for winning machine learning competitions and out-performing even the deep learning neural network models on tabular data. Even though the gradient boosting methods are capable of giving high quality predictions, their main limitation is the unsatisfactorily long training time and poor scalability [e.g., Zhang et al., 2019].

The main cost in GBDT lies in learning the decision trees, and the most time-consuming part in learning each tree is finding the optimal segmentation points. In the conventional algorithms this is usually done using the so-called pre-sorted algorithm. This method enumerates all possible split points on the pre-sorted feature values. While being simple and effective, this method is also

known to be inefficient in training speed and memory consumption [Ke et al., 2017]. Another more recent method is the histogram-based approach. It divides the continuous features into k intervals and selects the split points from those k values [Ju et al., 2019]. Such an approach also has regularization effect and helps prevent overfitting.

One of the GBDT implementations, called the Light Gradient Boosting Machine, or LightGBM, has been recently developed by Microsoft [Ke et al., 2017]. LightGBM addresses the traditional GBDT performance issues by, first, using the histogram approach to find segmentation points, and second by uzilizing a different approach to the tree growth. The conventional gradient boosting, and also other tree-based methods such as random forest, grow the trees level-wise. This means that when it is necessary to make a new split, a new level of leaves will be grown. In contrast, the LightGBM method grows the trees leaf-wise which adds only one more leaf and not level when a split is made. Such an approach leads to much faster and less computationally expensive implementation of the gradient boosting [Ke et al., 2017]. It has been demonstrated that LightGBM can be as much as 20 times faster than XGBoost while reducing more loss.

The objective of this study is to predict values of electron flux at a range of L-shells throughout the outer radiation belt. Since the quantity being modeled can be represented by real numbers, we construct a regression gradient boosting model. The model setup is described in detail in Section 2.3.3.

2.3.2 Test - train splitting of the data

Any supervised machine learning model learns on the so-called training dataset. The training set is seen by the model and usually contains most of the employed data points. Machine learning algorithms are trained iteratively trying to reduce the cost function value at each training iteration [Camporeale, 2019]. At some point, the model learns not only the useful dependencies from the data, but also noise due to reducing the cost function to extremely low values, which results in over-fitting. The performance of the model cannot be adequately evaluated on the data that were used to train it. Therefore, another set is needed to give an unbiased estimate of model

performance and also for tuning the hyperparameters. This second set is called a validation set. Since these data had not been seen during training, the loss function value would decrease only in case when the model captured the underlying general dependencies from the data. However, it has to be noted that since the loss function is being routinely evaluated on the validation set, the model occasionally sees it as well, although never learns from it. Therefore, it is essential that after training the model and checking its performance on the validation set, the model be evaluated one last time on the fraction of data that had never been used before, neither for training nor validation. The data set used for this purpose is called a test set. In many machine learning competitions (e.g., Kaggle), only the training and validation sets are given to the competing teams, while the test set is released after the model submissions and decides the winner.

Multiple ways have been proposed to separate the data set into the training, validation and test parts. In fact, very different strategies can be applied based on the field and objective of the study. For instance, in social sciences the accepted methodology would be to use the random test-train split, that is, when points for training, validation and test sets are selected randomly from the original full dataset. This, however, should not be applied for modeling the time-series and physical processes, due to the fact the validation points (usually 10-20 percent of values) can then be obtained by linearly interpolating the typically larger training set. Such a selection technique might introduce what is referred to as the data leakage [e.g., Camporeale, 2019]. Hence, it is important to validate, and then test the model on the events unseen during training. For that purpose, we should select the consecutive time intervals for validation and testing. One of the ways to achieve this is to use, for example, first 80 percent of data for training, the next 10 percent for validation and then test the model on the last 10 percent of the data. However, in case of magnetospheric phenomena we have to take into account the solar cycle evolution of the processes we model. Indeed, the radiation belts dynamics during the descending and quiet phases of the solar cycle are vastly different. Therefore, the model will be trained on some part of the solar cycle and validated on another, which would not yield an adequate estimate of its performance. In order to take the solar cycle dependence of the radiation belt dynamics into account, we perform a 10-fold cross-validation, described below.

We first reserve >1.5 years of data (March 2016 - January 2018) for testing the model. This is

done due to the fact that on one hand, this interval had numerous solar wind enhancement events and also had enough data points (~1.100.000) for such evaluation. The entire data set consists of ~ 5.5 million data points, and therefore approximately 20% are reserved for test set.

The rest of the dataset is used for the K-fold CV, illustrated in Figure 2.3. The data are divided into K roughly equal parts (in our case, K=10). At every such split the model is fitted on the (K-1) parts and validated on 1 part that was left out. For instance, during the first split we use the first fold for validation and fit the model on the rest of the data. This splitting process is repeated K times, each time selecting a different interval for validation. At each split, the previous model is being discarded and a new one is being trained and validated on the newly selected data. The K-fold allows one to utilize all observations for training and evaluating the model, and each of the data points is used for validation only once. The K-fold CV is used to optimize the hyperparameters and also to retrieve the accuracy of the model, averaged over different phases of the solar cycle. The final model is then trained on the entirety of the training and validation sets.

2.3.3 Model setup

In this study, we use the `LightGBMRegressor` method as implemented in the Python version of `lightgbm` library [Ke et al., 2017].

LightGBM has a variety of algorithm parameters, also called hyperparameters, that can have a non-negligible effect on the model performance. Several of them have to remain fixed, while others need to be optimized to achieve higher accuracy. Some of the parameters that do not change, include the `objective` (in our case, `regression`), the booster method (we use `gbdt`, although we note that other possibilities are available using the novel DART and GOSS methods [Ke et al., 2017]), `metrics` of the loss function values (we select both L1 and L2 metrics, representing the MSE and MAE, respectively), and the `early_stopping_rounds` parameter which is used to stop the model training once overfitting occurs. The `learning_rate` controls how much the model is adjusted on each iteration. In case of the high learning rate, the algorithm makes faster fits that can cause overfitting, while in case of extremely low values the training

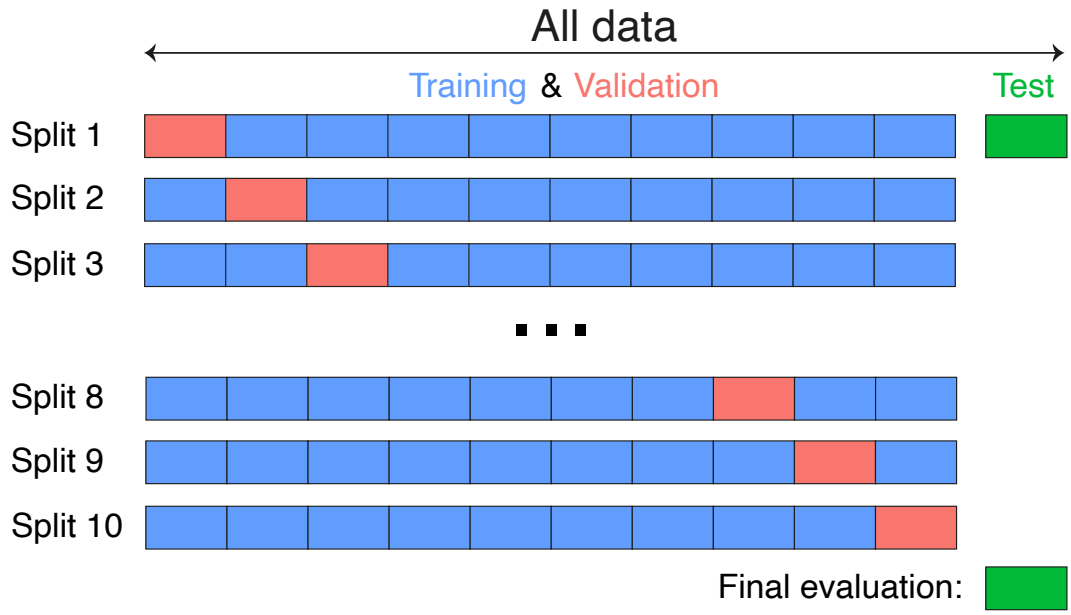

Figure 2.3. Schematics of the K-fold cross-validation CV. The last 1.5 years of data are reserved as the test set, never to be used during training and validation. The rest of the data are then divided into K equal parts (here, K=10) and at every split the model is trained on 9 parts and validated on 1 part. The training process is repeated 10 times, each time withholding a different set for validation. Thus, the model uses all observations for training and validation, and each of the data points is used for validation only once. The final evaluation is performed on the test set.

speed drops sufficiently and more iterations are needed to reach convergence. We select the `learning_rate` of 0.05 and keep it fixed throughout further tuning. Other parameters need to be optimized. The most sensitive hyperparameter is the maximum number of leaves in a tree, or `num_leaves`. Deeper trees have better learning capabilities, but since the gradient boosting model represents an ensemble of weak learners, the `num_leaves` is usually not very high, in our case ranging from 15 to 250. Another important parameter is the minimum number of data points in leaf (`min_data_in_leaf`), which has regularization effect and stops the model from learning the noise. However, the large values can cause decrease in model accuracy. One can also use the subset of the input features for training each individual tree by setting the `colsample_by_tree` value. To optimize the said hyperparameters we use the `hyperopt` Python library [Bergstra et al., 2013] which employs the Tree of Parzen Estimator (TPE) approach. The resulting hyperparameters values, as well as their search domains are given in Table 2.1.

Table 2.1. LightGBM hyperparameters used for model set up. First 5 parameters were kept fixed, while the next ones were optimized using HyperOpt.

Name	Search Range	Value
'objective'	fixed	'regression'
'boosting_type'	fixed	'gbdt'
'metric'	fixed	'L1' & 'L2'
'learning_rate'	fixed	0.05
'early_stopping_rounds'	fixed	50
'num_leaves'	15 – 250	93
'reg_alpha'	0 – 1	0.80
'reg_lambda'	0 – 1	0.09
'min_data_in_leaf'	10 – 1000	79
'colsample_by_tree'	0.7 – 1	0.99

2.4 Results and discussion

2.4.1 Feature importances

One of the advantages of the tree-based machine learning methods, LightGBM included, lies in the fact that the resulting model can be easily interpreted. Every tree comprising the model can be visualized and can give direct information about the individual split gains, internal values in each leaf and the decision making process in general. In practice, it is often not possible to analyze the model in this way due to the large number of trees in an ensemble (in our case, ~300). The insight on the model construction can then be obtained indirectly, for example, by analyzing the importance scores of each variable, also called feature importances. They are computed for each GBDT and then averaged across all of the trees forming the model. There are multiple ways to retrieve the feature importances. LightGBM utilizes the so-called `split` or `gain` methods. The `split` method counts the number of times each variable was used to make a split. The `gain` method summarizes all gains of splits which use each of the features. It has been well established that the two methods can, in fact, give different results, and also that feature importances estimated this way only carry information about how the particular model was constructed, rather than physical meaning. Furthermore, removing one of the features can potentially redistribute its feature importance between several other variables and yield a different result altogether. Other

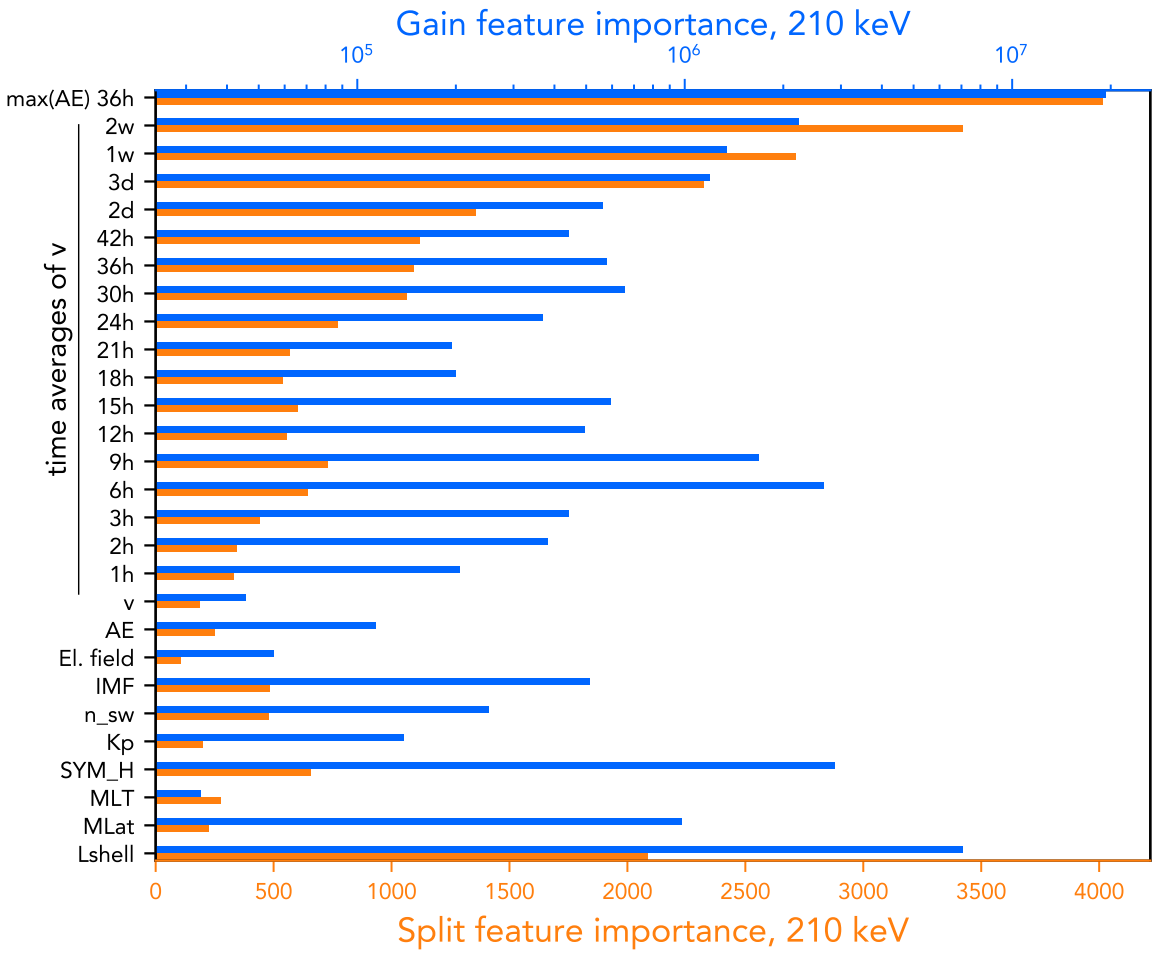

Figure 2.4. Feature importances estimated using the intrinsic LightGBM gain and split methods for 210 keV electron flux.

methods, which are more stable, include the mutual information criterion, permutation method [Auret and Aldrich, 2011] and the recently developed Shapley values technique [Lundberg and Lee, 2017]. These methods require an in-depth analysis which is beyond the scope of the present paper and will be evaluated in future studies. In the present Section we confine to describing of the key attributes of the MERLIN model. Feature importances estimated using the `split` and `gain` methods are shown in Figure 2.4. MERLIN uses the satellite position (Lshell, magnetic latitude, MLT), as well as values of the SYM-H, Kp, and AE indices and solar wind density (n_sw), IMF, electric field (El_field) and solar wind velocity. Plasmapause location is denoted as Lpp. Furthermore, the time history of velocity is incorporated into the model in the form of

averages over the certain time intervals. We use the progressively increasing time steps, from 1 hour (1h) up to 3 days (3d) and also the averages over 1 and 2 weeks (2w).

We find that the most important features by gain are the L-shell and maximum of AE over 36 hours, which is a proxy of the plasmapause location (Figure 2.4). Indeed, the values of electron flux depend on these quantities to a large extent, due to the fact that the electron intensities are higher in the heart of the outer belt, and then decrease to the outer edge of the belt and also in the opposite direction towards the slot region, which is reflected in the L-values. The importance of the plasmapause location has been discussed in [Li et al., 2006], and it was shown that Lpp correlated with the inner edge of the outer belt. The flux values drop significantly below the Lpp, although it is of note that the relationship between the innermost location of the outer belt and Lpp is energy dependent [e.g., Reeves et al., 2016; Ripoll et al., 2016]. Among the instantaneous values of solar wind and geomagnetic indices, the SYM-H index, which is a proxy of geomagnetic storms, shows the most importance in both split and gain methods. We also note that the time history of solar wind velocity plays an important role. For the 210 keV electrons, the average velocity over 6 hours and also over 1 and 2 weeks have the most importance based on impurity reduction (Figure 2.4). The importance of the 2 weeks average of v_{sw} likely comes from the fact that following the flux enhancement events, velocity drops to quiet time values faster than the electron intensities which remain at elevated levels for longer periods of time. Using Combined Release and Radiation Effects Satellite (CRRES) data, [Meredith et al., 2006] demonstrated that the flux values elevated by the substorm or storm processes decay to their pre-storm values gradually, and that the flux can fall by 2 orders of magnitude over a period of approximately 20 days. Therefore, an indicator of the past events is needed in order to correctly reproduce the dynamics of the flux decay. In the MERLIN model we do not use any previous values of flux and hence it is the history of velocity that the model uses as such an indicator. We note that AE and Kp indices exhibit very close feature importance values, because being highly correlated with one another, they produce equal reduction in variance when making splits.

Figure 2.5 shows the influence of the solar wind history on the model performance on the example of 425 keV electrons. The MSE of the model with no solar wind history employed is shown as black dots. We further add the history of solar wind velocity of up to 14 days

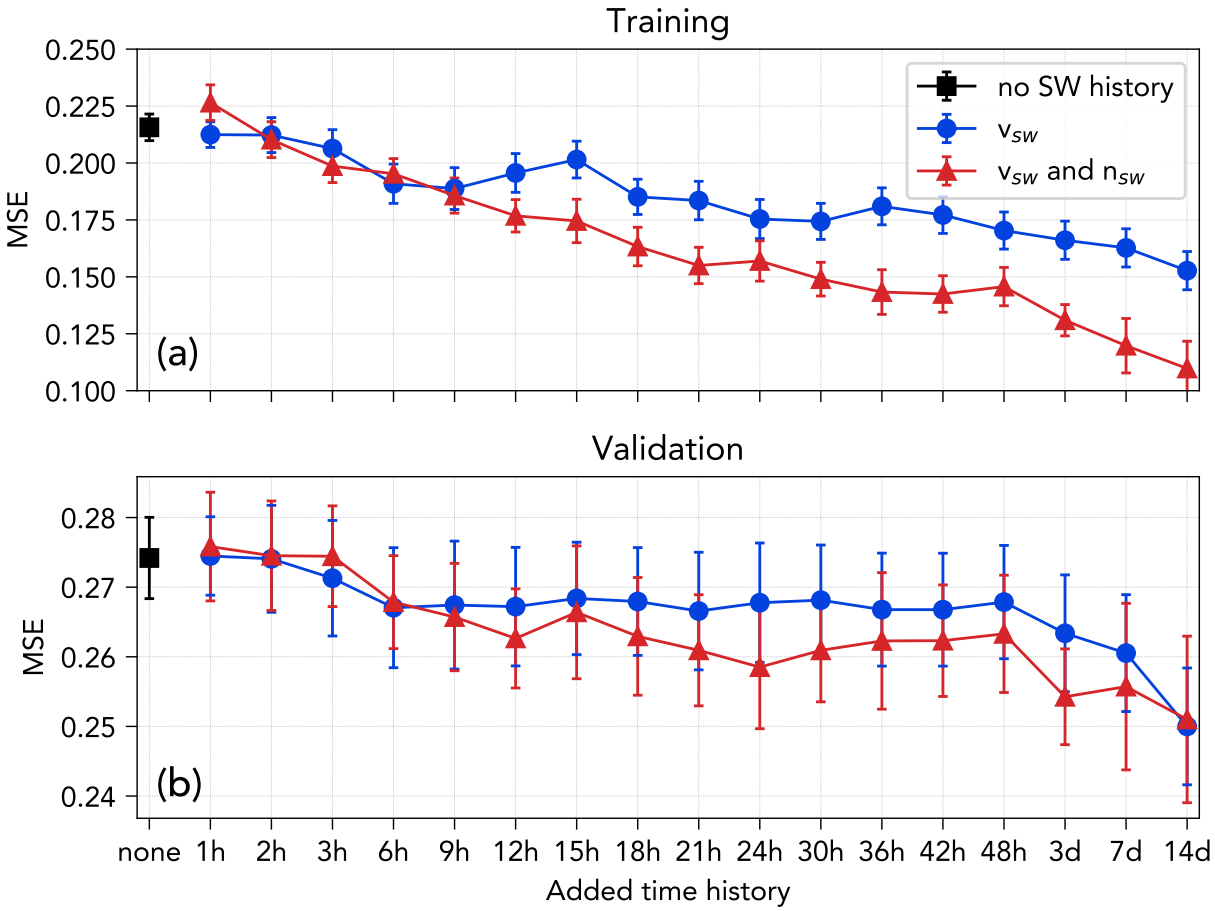

Figure 2.5. MSE on the training (a) and validation (b) data for 425 keV electron flux depending on solar wind history. The black dot represents the model with no solar wind history employed. Averages of solar wind velocity (blue curve) are added as model inputs and reduce both training and validation error. Including also the history number density reduces the training but not the validation error and leads to overfitting. The vertical dashes represent the standard errors of the 10-fold CV.

with gradually increasing time steps. The MSE gradually decreases as more history is included, both on training and validation data. On the other hand, including also the time history of solar wind density decreases the training but not the validation error. In Figure 2.5 it can be seen that while the training MSE is lower at every time step when density is included, the validation MSE changes very slightly as compared to using velocity history alone and remains within the error bar, indicating overfitting. Furthermore, it was observed that adding averages of solar wind velocity of more than 2 weeks also results in overfitting. Therefore, we only use the history of velocity of up to 14 days as inputs.

2.4.2 Results of 10-fold cross validation

Multiple metrics can be used to evaluate the model accuracy [for details see Morley et al., 2018; Liemohn et al., 2018]. The LightGBM library offers a variety of metrics implemented for model analysis. The default metric is the MSE, which is computed at each training iteration. It should be noted that the electron flux can exhibit strong depletions, up to several orders of magnitude, over short times [Ganushkina et al., 2019]. The mean squared error is susceptible to outliers , and therefore we also evaluate the median of the squared error. In Table 2.2 it can be seen that both for training and validation sets, median of the squared error is ~3-4 times lower than the MSE. This means that while some of the rapid depletions/enhancements are not completely reproduced, the value of the median squared error of 0.06 shows that the model predictions are very close to the observed data.

Table 2.2. Metrics evaluated on the training and validation sets during 10-fold CV, and on the test data for 425 keV electrons. The standard error for the 10-fold CV are shown in brackets.

Metrics	Training	Validation	Test
Mean SE	0.12 (±0.010)	0.26 (±0.019)	0.20
Median SE	0.04 (±0.004)	0.06(±0.009)	0.06
MAE	0.19 (±0.010)	0.26 (±0.012)	0.24
NRMSD	0.04 (±0.006)	0.08 (±0.007)	0.07
r^2	0.84 (±0.015)	0.63 (±0.019)	0.59
Spearman ρ	0.89 (±0.005)	0.85 (±0.012)	0.81
Ratio	1	1	1

Another useful metric, implemented in the LightGBM library, is the median absolute error, denoted as MAE. Evaluating the median error allows one to pay less attention to the outliers but gives a good proxy of model performance otherwise. In our case, MAE values are close for the training (0.19) and validation sets (0.26). These values of MAE mean that in general, the predicted values differ from observations by a factor of ~1.5, which is considered acceptable for radiation belt modeling. The scale-dependent metrics can be difficult to interpret, since the model predicts base-10 logarithms of flux and also because the level of flux is generally different for different energies. We evaluate the normalized root mean squared deviation (NRMSD), defined in [Denton et al., 2019, eq. 1]. The zero value of NRMSD corresponds to the perfect prediction,

and the values of < 1 generally indicate a good match between the model and data [Denton et al., 2019]. From Table 2.2 one can see that the values of NRMSD are close to 0 (0.04 for training and 0.08 for the validation set), indicating that the model reproduces the observations very well.

The metrics mentioned above mainly quantify how far the predictions deviate from observations at stationary points. In order to evaluate how well the model reproduces the dynamical behavior of the electron flux, we employ the following metrics. First, we evaluate the correlation between the predictions and observations. The standard Pearson linear correlation coefficient is susceptible to outliers, and therefore we use the Spearman rank correlation coefficient (ρ). From Table 2.2 it can be seen that the correlations are high both for training and validation sets, based on the 10-fold CV. The average correlation on the training set is 0.89 and is slightly higher than that on the validation (0.85), which is as expected since the model was fitted on the training data. The high values of the correlation coefficient show that the model captures the behavior of flux at energies 120-600 keV, which are known to be very dynamic and can exhibit drastic changes on the order of several minutes.

Another popular metric widely used in machine learning is the r^2-score. This indicator is used to quantify the fraction of variance explained by the model [e.g., Morley et al., 2018]. The average r^2 for the training data is 0.85, and 0.63 for the validation set. While these values appear low, we conduct a more detailed analysis of the r^2 on the synthetic data (shown in the Supporting Information of Smirnov et al. [2020a], Figure S1). We first generate a harmonic function - a sinusoid, for 20 periods of 2π each, and then add random noise of maximum magnitude equal 0.2. The electron flux are known to exhibit rapid depletions of up to several orders of magnitude, and to account for this we add 60 dropouts where we subtract 2 units from the synthetic signal. We compute the values of the r^2 for the case with and without outliers. The initial sinusoid signal compared to the data with no outliers yields prediction efficiency of ~0.9, while the data in presence of outliers give a lower r^2 value of ~0.6, which is approximately equal to the value we have for MERLIN on the validation set. This means that the model can adequately reproduce the behaviour of flux but miss some of the dropout magnitudes, which results in the lower prediction efficiency. This will be further discussed while analyzing the performance on the test data.

Figure 2.6 demonstrates the model performance, averaged over the 10-fold CV, for individual

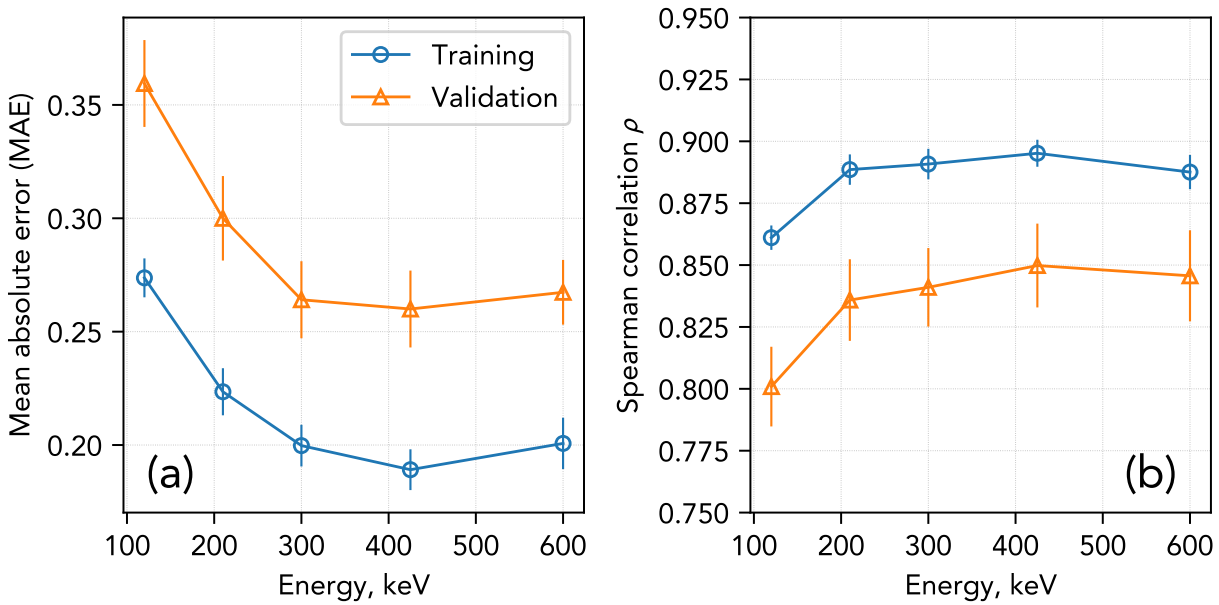

Figure 2.6. Model performance on training and validation sets averaged over the 10-fold cross validation. Median absolute error values are shown in (a), and Spearman correlation coefficients are given in (b).

energy channels. Figure 2.6a shows the median absolute error. It can be seen that the best performance is achieved for 425 keV electrons, although we note that the MAE for other channels is only slightly different. We find that the accuracy improves with higher energies (MAE for 120 keV on the validation set is 0.36 compared with ~0.26 for 425 keV). The accuracy then slightly decreases from 425 keV to 600 keV electrons. Same can be seen in terms of the correlation (Figure 2.6b) - the averaged correlation on the validation sets is ~0.80 for 120 keV and ~0.85 for 300 keV. The ρ value then slightly decreases to 0.84 for 600 keV. The difference between Spearman correlation coefficient is generally very small between the energy channels (on the order of a few percent). It should be noted that the values of MAE and the correlation for each individual channel show that the MERLIN model well predicts both amplitudes and flux dynamics throughout the considered energy range.

2.4.3 Performance on test data

The training set is used to learn the model, and the model is constructed so as to minimize the error on the validation set. On each iteration, the MSE and MAE values should decrease for the

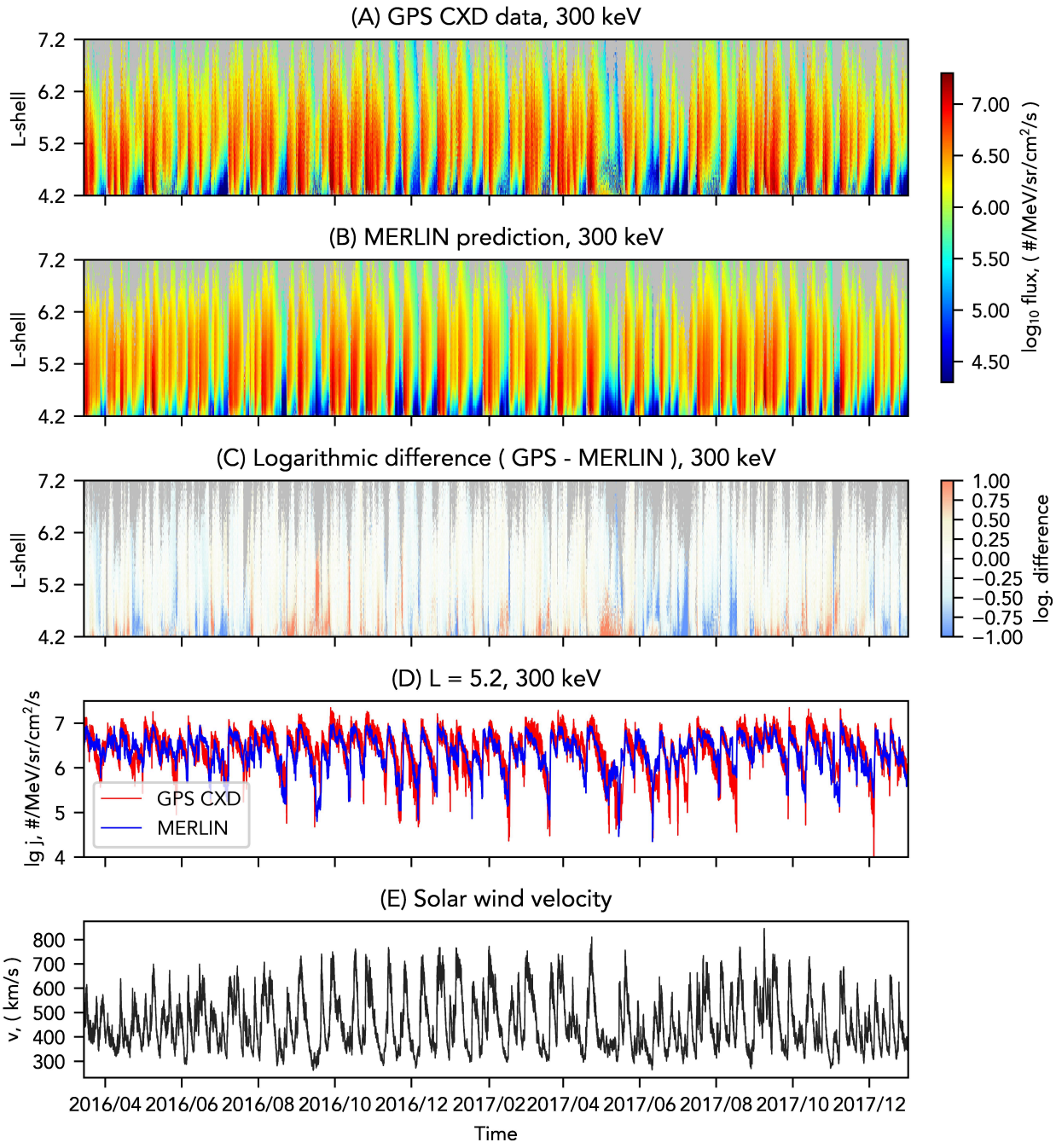

Figure 2.7. Model performance on the test set for 300 keV electron flux. (a) GPS CXD measurements; (b) prediction using the MERLIN model; (c) logarithmic difference between the observed and predicted flux; (d) comparison of observed (red) and predicted (blue) flux at the fixed L-shell of 5.2 with the original satellite cadence; and (e) solar wind velocity over the test time interval.

training set as the gradient boosting model adjusts the residuals to fit the training set better, and then the updated model is evaluated on the validation set. If the validation error reduces, the training continues to a new iteration (i.e., grows a new GBDT). The model never learns from the validation set, however, it is still being used on every training iteration. We therefore need another totally independent set of values to check that the model generalizes well onto the unseen data. It is of note that this test set can only be used once to evaluate the performance, after the model training has been completed, and no further changes to the model can be made. More details on the train-validation-test splitting of the data can be found in Section 2.3 and in Figure 2.3.

The values of different metrics discussed above are given in Table 2.2. We find that the values on the test set are close to those on the average of the validation sets. MAE values are equal to 0.26 and 0.24 on the validation and test data sets, respectively. MERLIN yields NRMSD values of 0.08 for test and 0.07 for validation data, which show that the model performs well on both sets as the NRMSDs are close to zero. The values of the Spearman correlation coefficient are very close for the test data (0.81) and the average validation (0.85). The mean and median squared errors also yield almost identical values for validation and test sets, and their differences are within the error bar. In general, all of the metrics exhibit values which are sufficiently close on test and validation sets and are slightly lower than on training data. This means that the model successfully learnt the underlying relationships between the input parameters and the resulting electron flux on the training data, yields good accuracy on the validation intervals, and generalizes well onto the unseen data.

Figure 2.7a shows the GPS electron flux from all 21 spacecraft for March 2016-December 2017 for 300 keV. Figure 2.7b gives the flux values from the MERLIN model, and the difference between the predicted and observed flux is shown in Figure 2.7c. Figure 2.7e demonstrates the solar wind velocity for the test set. It should be noted that numerous solar wind velocity enhancements happened during the test interval with v_{sw} rising up to >700 km/s. These events generally correspond to increases in electron flux, but it can be seen that after the velocity drops to its quiet time values the flux remains elevated for longer periods of time. For instance, v_{sw} increased up to 700 km/s at the beginning of September 2016 (detailed illustration is in the

supplementary Figure S2 of Smirnov et al. [2020a]) and caused a rapid increase in electron flux by over an order of magnitude. The velocity then started decreasing and within 1 week dropped to 500 km/s while the flux remained at the elevated levels. Within a few days, velocity continued decreasing until it reached <300 km/s but the electron flux had a much longer decay, and even after ~1 week of very low velocities the flux did not fall to the pre-event values.

In general, the model adequately reproduces all of the major flux enhancement events and also reproduces well the flux decay due to consideration of the velocity history. We note that for L-shells lower than 5, the flux enhancements are sometimes followed by periods of the increased data variance (Figure 2.7a). The most likely explanation lies in the fact that the GPS electron flux data are a derived quantity. As such, the fluxes are computed using a forward model combining 3 relativistic Maxwellians (in energy) and a Gaussian (in log of momentum) [see Morley et al., 2016, for details]. The electron energy spectra inside the plasmapause can have local peaks, and intense plasmaspheric hiss can generate a reverse spectrum in the energy range of hundreds of keV [Zhao et al., 2019]. At lower energies, spectra are not always well-captured due to the limitations of the spectral form [see Section 3.2 of Morley et al., 2016], while this effect is not observed at higher energies. As expected, such events are not reproduced by the MERLIN model.

The Spearman correlation between the observed and predicted flux is approximately 0.81. It is, however, important that the model reproduces not only the flux along the GPS orbits but is also capable of giving reasonable prediction at fixed L-values. Figure 2.7d shows the GPS observations and the MERLIN output for L of 5.2. The correlation between the two is 0.76 and the mean squared error is 0.06. One can therefore conclude that the model generalizes well on the unseen data, both at a fixed L and along the orbit. The same conclusion can be drawn for the 600 keV population, demonstrated in Figure 2.8.

Figure 2.9a-e shows the occurrence density plots of the observed versus predicted flux for all 5 energies considered. Solid white lines show the one-to-one relationship between observations and predictions, and the dashed likes represent the flux deviating by a factor of 5. In general, the occurrence maxima follow the trend and most of the points are within the factor of 5 from the trend line. Figure 2.9f provides an example of the histogram of model residuals for the 210 keV population. From the Figure it is evident that the model has very low bias of ~ 0.05, depicted by

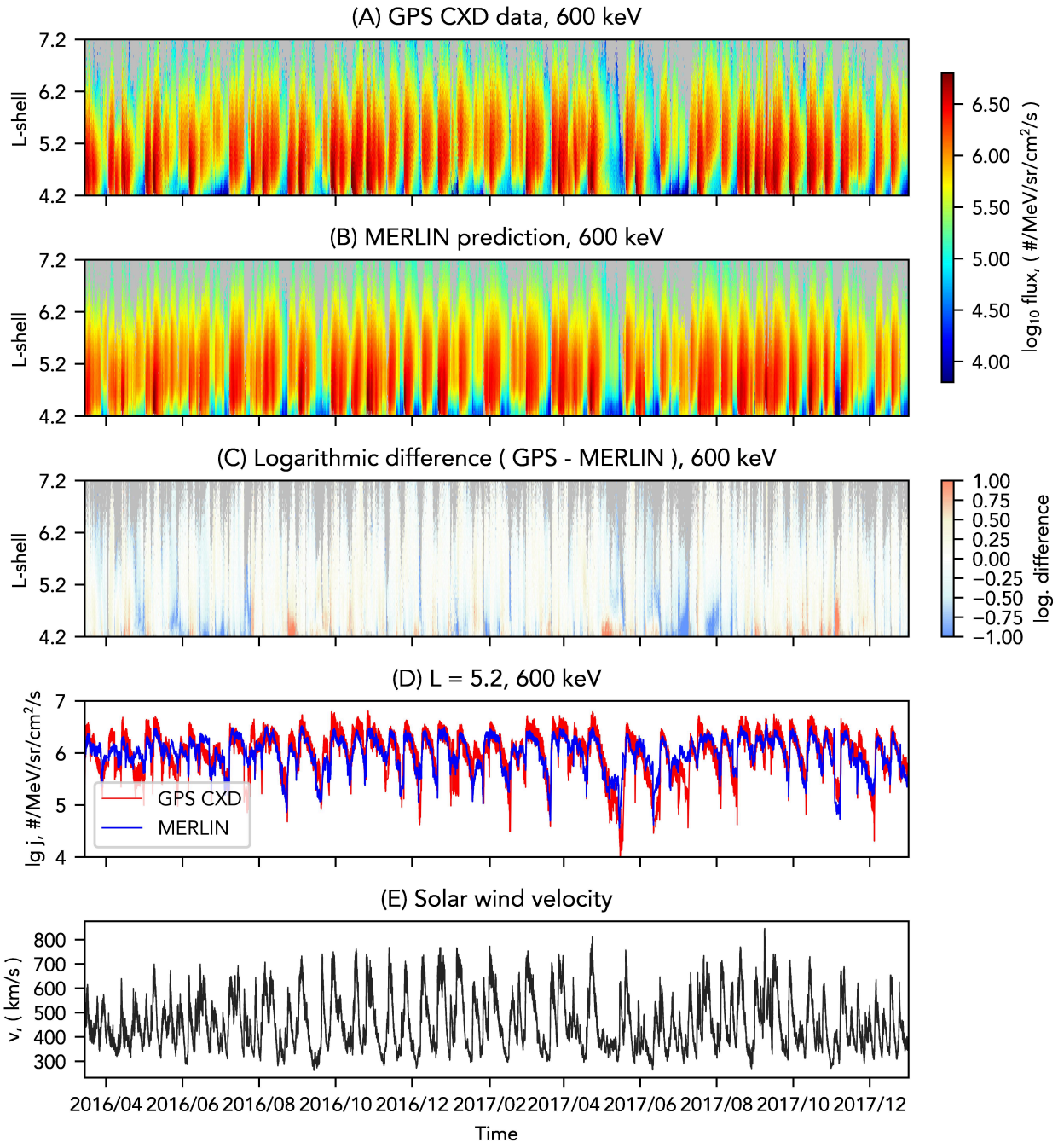

Figure 2.8. Same as Figure 2.7 but for 600 keV electron flux.

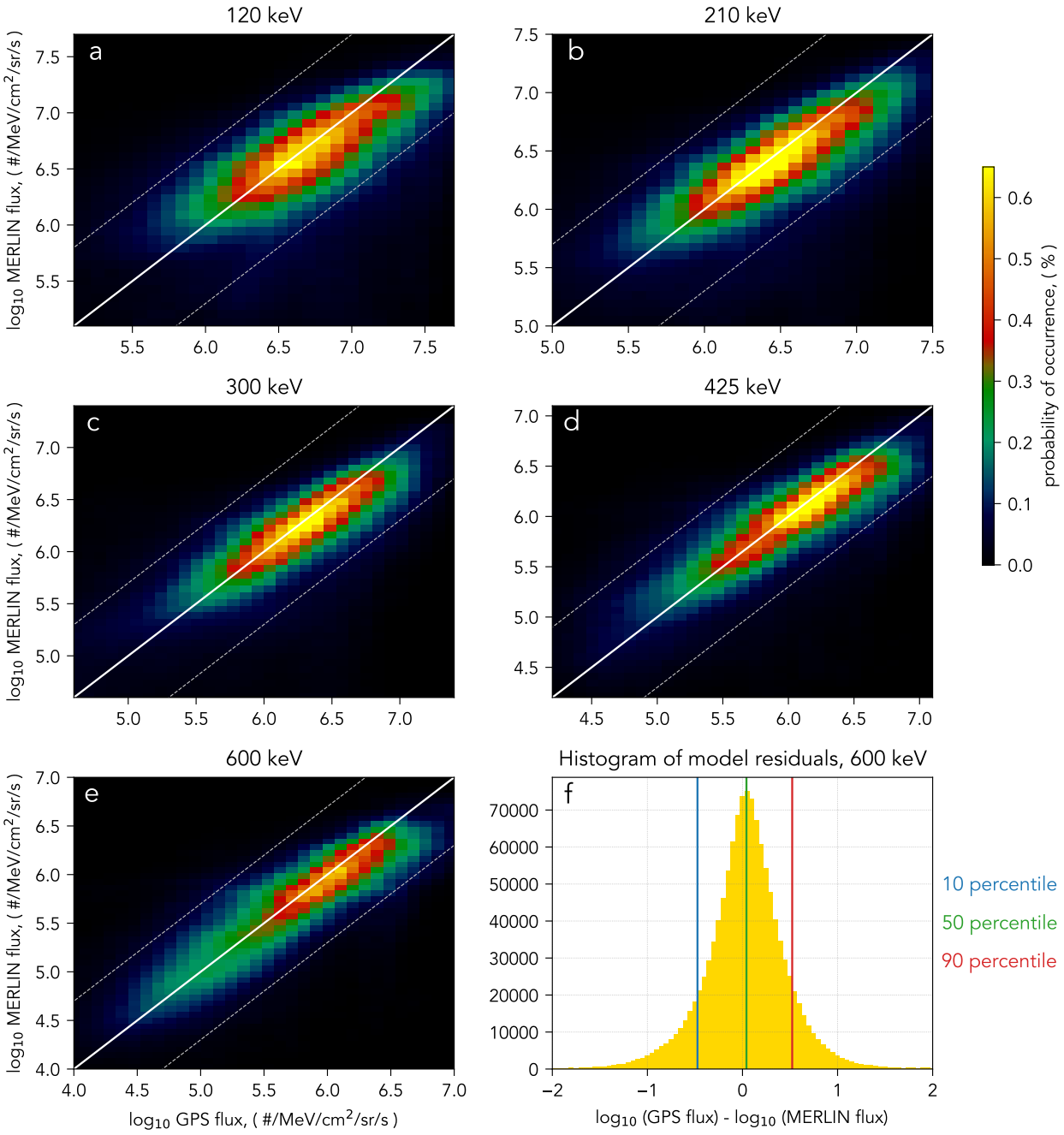

Figure 2.9. Probability of occurrence of the observed (on x-axis) versus predicted (on y-axis) electron flux for (a) 120, (b) 210, (c) 300, (d) 425, (e) 600 keV for test data. The white lines show the one-to-one ratio between the observed and predicted flux. The silver dashed lines give the threshold within a factor of 5. Panel (f) shows an example of the histogram of the model residuals for 600 keV electron flux.

a green line. Furthermore, the errors are normally distributed and the 10-th and 90-th percentiles are ~ ±-0.51 and 0.52, respectively, meaning that 80% of the model residuals are within a range of ±0.5. Therefore, we can conclude that MERLIN predicts the electron flux in 120-600 keV energy range well, has low bias and captures all of the general trends represented in the data.

2.5 Conclusions

A new data-driven model of the electron flux in the outer radiation belt is presented. The model uses satellite position and a combination of geomagnetic indices and solar wind parameters (with time history of velocity) in order to predict the flux values at energies 120-600 keV. The model has been trained and validated on more than 15 years of GPS electron flux data, and tested on >1.5 years of observations. The 10-fold cross validation shows that the MERLIN model predicts the MEO radiation environment well, capturing both the dynamics and amplitudes of electron flux. The results of the 10-fold cross validation agree well with the evaluation on the test data meaning that the model is able to generalize well onto the unseen events. Predicted values of flux exhibit high correlation with the observations (~0.8) and low values of error.

The MERLIN model can have wide Space Weather applications. It can be used by the scientific community to analyze specific events as well as to reconstruct the long-term radiation belt dynamics at the 120-600 keV energy range, for which there is generally a lack of models. Furthermore, it is of use for satellite operators for the nowcast of the MEO environment and can provide information for the surface charging analysis. We note that the model was trained on the data from the GPS constellation, which follows an inclined orbit. Therefore, at higher magnetic latitudes the satellites sample only a part of the equatorial pitch angle distribution. However, using the appropriate pitch angle models it is possible to reconstruct the values of equatorial flux from MERLIN predictions [e.g., using a methodology of Allison et al., 2018]. Further directions for the present study include, first, a more refined analysis of the feature importances using the appropriate permutation and Shapley values methods and the corresponding feature selection. Second, as GPS continues to probe the outer radiation belt, more data can be incorporated into the model.

Acknowledgements

The resulting model files in `.txt` format can are available in open access at: `https://doi.org/10.5281/zenodo.3783757`. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870452 (PAGER). The data of solar wind parameters and geomagnetic indices were taken from OMNIWeb database: `omniweb.gsfc.nasa.gov`. Contributions by S.K. Morley and M.R. Carver were performed under the auspices of the US Department of Energy and supported by the Laboratory Directed Research and Development (LDRD) program, award 20190262ER. We acknowledge the CXD team at Los Alamos National Laboratory (LANL), operating the CXD detectors aboard the GPS constellation. The GPS particle data can be obtained in open access at `https://www.ngdc.noaa.gov/stp/space-weather/satellite-data/satellite-systems/gps/data/`. The analysis in this study was conducted using gradient boosting with regression trees, as implemented in the LightGBM Python library: `lightgbm.readthedocs.io`.

— 3 —

Storm-time evolution of the equatorial electron pitch angle distributions in Earth's outer radiation belt *

Abstract

In this study we analyze the storm-time evolution of equatorial electron pitch angle distributions in the outer radiation belt region using observations from the Magnetic Electron Ion Spectrometer (MagEIS) instrument aboard the Van Allen Probes in 2012-2019. The PADs are approximated using a sum of the first, third and fifth sine harmonics. Different combinations of the respective coefficients refer to the main PAD shapes within the outer radiation belt, namely the pancake, flat-top, butterfly and cap PADs. We conduct a superposed epoch analysis of 129 geomagnetic storms and analyze the PAD evolution for day and night MLT sectors. PAD shapes exhibit a strong energy-dependent response. At energies of tens of keV, the PADs exhibit little variation throughout geomagnetic storms. Cap PADs are mainly observed at energies <300 keV, and their extent in L shrinks with increasing energy. The cap distributions transform into the pancake PADs around the main phase of the storm on the nightside, and then come back to their original shapes during the recovery phase. At higher energies on the dayside, the PADs are mainly pancake during pre-storm conditions and become more anisotropic during the main phase. The quiet-time butterfly PADs can be observed on the nightside at L> 5.6. During the main phase, butterfly PADs have stronger 90°-minima and can be observed at lower L-shells (down to L= 5), then transitioning into flat-top PADs at L$\sim$ 4.5 – 5 and pancake PADs at L< 4.5. The resulting PAD coefficients for different energies, locations and storm epochs can be used to test the wave models and physics-based radiation belt codes in terms of pitch angle distributions.

*This chapter has been published as Smirnov, A., Shprits, Y., Allison, H., Aseev, N., Drozdov, A., Kollmann, P., Wang, D. , and Saikin, A. A. (2022): Storm-time evolution of the equatorial electron pitch angle distributions in Earth's outer radiation belt. *Frontiers in Astronomy and Space Sciences*, 9, 62.

3.1 Introduction

The radiation belts of the Earth contain charged energetic particles, mainly electrons and protons, trapped by the geomagnetic field. The energetic electrons are primarily observed in two regions, namely the inner (L< 2.5) and outer (3.5 <L< 7) belts, separated by the slot region where fluxes typically drop by several orders of magnitude [e.g., Ganushkina et al., 2011]. The radiation belt electrons can be characterized in terms of their flux intensity and angular distributions [e.g., Clark et al., 2014]. These distributions, also named the pitch angle distributions, play a crucial role in understanding the dynamics of the radiation belts, as specific PAD types can reveal the processes governing the particle transport, source and loss mechanisms, and wave activities [e.g., Horne et al., 2003; Gannon et al., 2007; Ni et al., 2015].

There are several common types of pitch angle distributions in the radiation belt region. The so-called pancake, or normal, PADs have a maximum flux at 90° pitch angle (PA) with a smooth decrease in flux towards the loss cone [e.g., West et al., 1973]. The pancake PADs can be formed as a result of the particle PA diffusion, inward radial diffusion, as well as wave-particle interactions (e.g., with hiss and chorus waves in the outer belt [Su et al., 2009; Meredith et al., 2000]), and consititute the most dominant PAD type in the inner magnetosphere on the dayside [Gannon et al., 2007]. The PADs where electron flux at 90° is smaller than at intermediate pitch angles (~30°-75°) are called the butterfly distributions [West et al., 1973]. The butterfly PADs in the outer radiation belt are mainly present at nightside magnetic local times (MLTs) and form due to drift shell splitting [Roederer, 1967; Sibeck et al., 1987], magnetopause shadowing [West et al., 1973] and wave activity [Artemyev et al., 2015; Ni et al., 2020]. The flat-top PADs exhibit a relatively constant flux at a wide range of pitch angles around 90°. The flat-top PADs can be a transition phase between the butterfly and pancake PADs, and they also occur due to strong wave-particle interactions with whistler mode waves in regions of low electron densities [Horne et al., 2003]. The head-and-shoulders, or cap, distributions resemble pancake PADs for non-equatorially mirroring electrons but have an additional bump in flux around 90° PA; they generally result from resonant interactions with the plasmaspheric hiss waves [Lyons et al., 1972]. Wave-particle interactions with ULF waves, particularly in the Pc4-Pc5 range, are known

to also affect the electron pitch angle distributions [Zong et al., 2017]. For instance, during the drift resonance events electron flux oscillations around 90° are observed faster than at lower PAs, which leads to the formation of the so-called boomerang stripes in the pitch angle distributions [e.g., Hao et al., 2017; Zhao et al., 2020].

Pitch angle distributions can be approximated using different trigonometric functions. The standard formulation used in several previous studies includes fitting PADs to the $\sin^n \alpha$, where α is the particle pitch angle and n is the steepness of the distribution [see e.g., Vampola, 1998; Gannon et al., 2007, etc.]. This parametrisation, however, has several limitations. For instance, it fails to capture butterfly distributions which constitute the dominant PAD shape on the nightside at L>5, as well as cap distributions occurring at lower L-shells [e.g., Zhao et al., 2018; Allison et al., 2018]. To mitigate this limitation, Allison et al. [2018] employed a combination of two terms of the said form, which allowed to resolve the cap distributions. Another formulation used in the radiation belts research includes fitting equatorial PADs with Legendre polynomials which comprise a set of spherical functions [see e.g., Chen et al., 2014; Zhao et al., 2018, 2021]. In particular, Zhao et al. [2018] demonstrated that equatorial PADs in the outer zone can be approximated by the first 3 even terms of the Legendre series expansion, while at L<3 it was necessary to include higher harmonics due to the larger loss cones and generally steeper PAD shapes. In this study, we use Fourier sine series expansion to approximate equatorial electron PADs. The Fourier expansion has been used for PAD approximation in the planetary magnetospheres [for Saturn's radiation belts - Clark et al., 2014] but to our knowledge has not been applied to study electrons in the Earth's radiation belts yet. One of the main advantages of using the Fourier sine series expansion is a possibility to integrate Equation 3.1 over the solid angle to derive omnidirectional flux.

Several studies have investigated the morphology of electron pitch angle distributions in the inner magnetosphere, both during quiet and geomagnetically active times. Roederer [1967] analyzed effects of the drift shell splitting on energetic electrons in the model magnetosphere and demonstrated that the drift shell splitting effects could only be observed above L= 5. The spatial structure of 80 keV - 2.8 MeV electron PADs was described by West et al. [1973] using Ogo-5 satellite data. The dayside PADs were found to exhibit mainly pancake shape, while the

nightside distributions at $L >\sim 6$ showed butterfly shapes, which were attributed to a combination of the drift shell splitting in presence of a negative flux gradient in L, and magnetopause shadowing. Selesnick and Blake [2002] computed anisotropies of relativistic electron PADs by tracing drift paths of particles for different pitch angles and levels of Kp and found a good agreement with average flux anisotropies calculated from Polar electron data under quiet geomagnetic conditions. Lyons et al. [1972] was one of the first papers that computed the PA-diffusion of electrons at energies 20 keV - 2 MeV driven by the resonant interactions with whisler mode waves and showed the existence of the cap pitch angle distributions could be attributed to the resonant interactions with the plasmaspheric hiss waves. Furthermore, it was demonstrated that with increasing energy, the bump in flux at 90° PA, characteristic of the cap PADs, decreased in magnitude as the cyclotron resonance branch extended to higher pitch angles at higher energies. Lyons and Williams [1975a] analyzed the quiet-time structure of electron PADs at energies below 560 keV and observed a generally good agreement in PAD shapes with the theoretical predictions by Lyons et al. [1972]. Furthermore, Lyons and Williams [1975b] reported that the storm-time electron PADs were very different from those during quiet times. In particular, the quiet-time cap distributions on the nightside were found to transform into the pancake PADs and then reform to their pre-storm structure several days after the storm. A comprehensive study by Gannon et al. [2007] analyzed electron PADs at energies of hundreds of keV based on data from Medium Electrons A instrument aboard the CRRES satellite. They reported that butterfly pitch angle distributions were the most prevalent type on the nightside at high L-shells, whereas on the dayside the pancake PADs constituted the dominant PAD shape.

In recent years, several statistical studies analyzed the storm-time evolution of electron PADs using data from the Van Allen Probes constellation. Ni et al. [2015] used fifteen months of the Relativistic Electron Proton Telescope (REPT) data to investigate the variability of PADs of electrons with energies $>$ 2 MeV. By fitting the PADs with a $\sin^n \alpha$ function, the authors tracked the spatiotemporal variability of the sine power n and found that pancake PADs became more peaked at 90° PA during the storm times compared to the quiet times. The occurrence rate of the butterfly distributions were investigated in Ni et al. [2016], and it was found that at nightside MLTs at high L-shells, up to 80% of PADs can be of butterfly type, which was in good

agreement with previous studies by West et al. [1973] and Gannon et al. [2007]. Pandya et al. [2020] analyzed the storm-time morphology of PADs of 1.8-6.3 MeV electrons for 27 CME– and 28 CIR–driven storms using the REPT data. The authors reported a strong dependence of PAD shapes on MLT, while the dependence on the storm driver was found to be negligible. Greeley et al. [2021] used REPT data to analyze PAD evolution during enhancements of the ultra relativistic electron fluxes separately for CIR- and CME-driven storms. The study showed that CME-driven storms generally resulted in more anisotropic PADs than CIR-driven storms. Furthermore, it was shown that PADs return to their pre-storm configurations more rapidly during storms driven by CMEs.

The previous studies that used Van Allen Probes data for analyzing the pitch angle distributions have mainly concentrated on relativistic and ultra relativistic energies sampled by the REPT instrument. Observations by the Magnetic Electron Ion Spectrometer (MagEIS) have been used in several recent studies [e.g., Zhao et al., 2018; Shi et al., 2016; Allison et al., 2018] which aimed at creating statistical PAD models but did not investigate PAD evolution for different phases of geomagnetic storms. In this study, we perform, for the first time, a comprehensive statistical analysis of electron PADs sampled by the MagEIS detector at energies 30 keV - 1.6 MeV. Furthermore, our study is the first one to use the Van Allen Probe dataset during the mission's entire lifespan in 2012-2019 for PAD analysis. We identify 129 storms in 2012-2019, and examine the morphology of the normalized PAD shapes for day and night MLTs at different energies.

Storm-time evolution of omnidirectional electron fluxes has been analyzed in detail in Turner et al. [2019] for both MagEIS and REPT energies. In this study, we concentrate on the PAD shapes normalized from 0 to 1. It should be noted that in the inner belt, pitch angle distributions appear relatively independent of activity [e.g., Ni et al., 2016] and exhibit very steep shapes due to larger loss cones which require higher harmonics for modeling [Zhao et al., 2018], therefore, in the present study we analyze PADs for L-values from 3 to 6. The paper consists of 5 parts. Section 4.2 describes the data set and the methodology employed in this study. In Section 4.3, we analyze the storm-time PAD evolution by means of the superposed epoch analysis of 129 storms during the Van Allen Probes era. The results are discussed in Section 4.4, and the conclusions are drawn in Section 4.5.

3.2 Dataset and Methodology

3.2.1 Data

The Van Allen Probes mission, originally known as the Radiation Belt Storm Probes (RBSP), operated in 2012-2019 and consisted of two spacecraft, denoted as RBSP-A and RBSP-B, flying in a near-equatorial orbit with an inclination of 10.2 degrees [Mauk et al., 2012]. The apogee of the probes was at ~5.8 R_E and perigee at an altitude of ~620 km. The orbital period was equal to 9 hours. The full MLT revolution was achieved every ~22 months. The L-shells sampled by Van Allen Probes range from 1.2 to around 6.2 on the nightside and 5.8 on the dayside.

The MagEIS instruments aboard each of the probes measured electron flux over a broad energy range using one low-energy unit (LOW) for energies 20-240 keV, two medium-energy units (M75 and M35) for observing electron flux at energies 80 keV-1.2 MeV, and a high-energy unit (HIGH) to sample data at energies from 0.8 to 4.8 MeV [Blake et al., 2013]. LOW, HIGH and M75 units were mounted at 75 degrees to the spin axis, and the M35 unit was installed at 35 degrees to the spin axis. Such a configuration was selected to provide broader pitch angle coverage for the MagEIS detector. In this study we employ the full data set of the MagEIS pitch angle resolved electron flux (level 3) in 2012-2019 averaged by 5 minutes with an assumed symmetry with respect to 90° PA. Following Zhao et al. [2018], we remove PADs for which the maximum electron flux value is below 100 $cm^{-2}s^{-1}sr^{-1}keV^{-1}$ as those PADs correspond to background levels of the MagEIS detector and are less indicative of the physics.

In order to analyze pitch angle distributions at the geomagnetic equator, it is necessary to propagate the locally measured electron flux values for each pitch angle to the equatorial plane using an appropriate geomagnetic field model. Since in this study we are concerned with variation of pitch angle shapes with increasing geomagnetic activity, the values are propagated to the magnetic equator using the TS04D storm-time model [Tsyganenko and Sitnov, 2005] with the internal field specified by the IGRF model. The TS04D model requires as inputs the By and Bz components of the IMF, solar wind velocity, density and dynamic pressure and the Dst index, as well as special W and G indices defined in [Tsyganenko and Sitnov, 2005]. The values of the

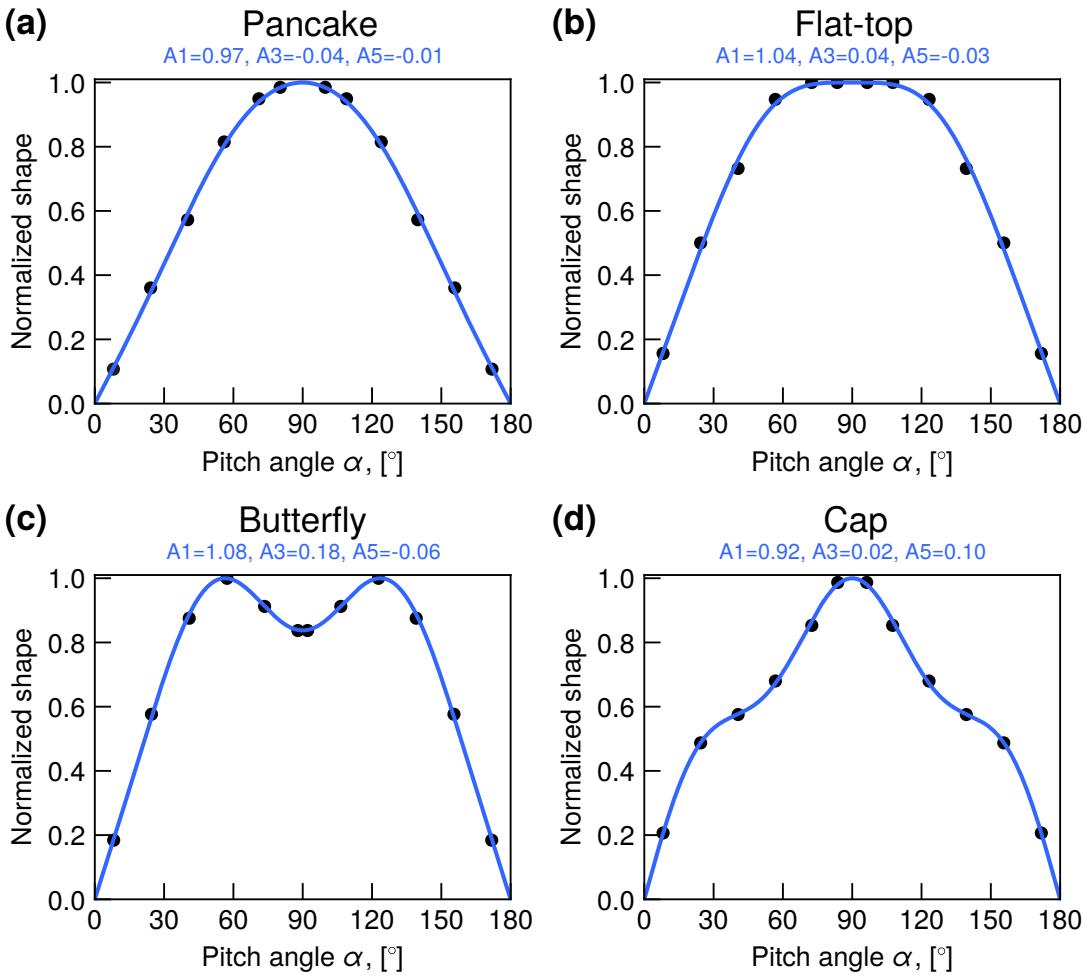

Figure 3.1. Examples of the fitted pancake (a), flat-top (b), butterfly (c) and cap (d) equatorial PADs, normalized using Equation 3.2. The dots show normalized MagEIS observations and the solid lines give fitted shapes.

solar wind parameters superposed for 129 storms used in this study are shown in the Supporting information of Smirnov et al. [2022a] (Figure S2).

3.2.2 PAD approximation using Fourier sine series

In this study we approximate equatorial electron PADs using the Fourier sine series expansion of the form:

$$j(\alpha) = A_0 + A_1 \sin\alpha + A_3 \sin 3\alpha + A_5 \sin 5\alpha, \tag{3.1}$$

where j is electron flux as a function of pitch angle α. It should be noted that the even terms ($\sin 2\alpha$ and $\sin 4\alpha$) represent shapes that are asymmetric around 90°, which is inconsistent with the trapped particle populations considered here, and therefore are omitted. In this study, we use the Fourier expansion up to degree 5, as this combination can effectively fit all the PAD types observed in the outer belt (shown in Figures 3.1 and 3.2 and discussed below). We fit the values of electron flux to equatorial pitch angles in linear scale using least-squares and obtain values of the coefficients A_0, A_1, A_3 and A_5. As can be seen from Equation 3.1, the A_0 coefficient shows the value of electron flux in the edge of the loss cone and represents the minimum flux value for a given PAD. Furthermore, we determine maximum value of electron flux, denoted as j_{max}, within each pitch angle distribution. Carbary et al. [2011] proposed a criterion to remove low quality PAD fits, which has been used in several other studies [e.g., Ni et al., 2016]. This criterion uses the normalized standard deviation of the difference between the observed and fitted electron flux and is defined as $\sigma_N = \sigma/(j_{\max} - A_0)$. The fits with the corresponding σ_N values ≤ 0.2 represent good quality fits [for details, see Carbary et al., 2011], while entries with $\sigma_N > 0.2$ were deemed as bad fits and excluded from the analysis. Examples of high- and low-quality PAD fits with the corresponding σ_N values are shown in the Figure S1 in the Supporting Information of Smirnov et al. [2022a].

In order to normalize PADs to span from 0 to 1, we apply the following equation to the coefficients $A_i, i = \{1, 3, 5\}$:

$$\tilde{A}_i = \frac{A_i}{j_{\max} - A_0}, \tag{3.2}$$

where $\tilde{A}_i$ denotes the normalized value of the respective coefficient A_i. It should be noted that the PAD shapes normalized by Equation 3.2 do not carry information about the flux levels and only reflect the shape of the distributions. In this study we analyze the storm-time evolution of these normalized pitch angle shapes for day and night MLTs at energies 30 keV - ~1.6 MeV. The dependencies observed here are used to create a PAD model in the outer radiation belt, which is presented in a companion paper [Smirnov et al., 2022c].

Figure 3.1 shows examples of the four main types of pitch angle distributions observed in the outer radiation belt region, namely the pancake, flat-top, butterfly and cap PADs, fitted using

Equation 3.1 and normalized by Equation 3.2. Note that while distinguishing these shapes is useful for discussions, we do not assign these discrete categories to our PADs but keep describing them through the continuous A values. Generally, the pancake PAD shape, shown in Figure 3.1a, resembles the first sine harmonic, and therefore has a large value of the corresponding coefficient A1 and low values of coefficients before the third and fifth terms (A3 and A5, respectively). The A3 coefficient corresponding to the $\sin(3\alpha)$ term shows contribution of the butterfly shape. Since $\sin(3\alpha)$ exhibits two peaks at 30° and 150° PA with a minimum at 90°, it can be used together with the first sine harmonic to approximate butterfly PADs (an example is given in Figure 3.1c). The flat-top PAD shape (Figure 3.1b) corresponds to high values of A1 with low values of A3, and small negative values of A5. The cap distributions (Figure 3.1d) can be fitted by a combination of the first and fifth sine harmonics. The $\sin(5\alpha)$ function has 3 peaks (at 30°, 90° and 150°) with two depletions in between of the peaks. In combination with the the general pancake shape given by $\sin(\alpha)$ it can well fit the head-and-shoulder structure.

In Figure 3.2 we show different combinations of the pancake (A1), butterfly (A3) and cap (A5) coefficients, and the resulting PAD shapes. In panel (a), we fix A1=1.1 and A5=0 and start increasing the A3 magnitude from 0 to 0.5. It can be seen that for A3=0 a pancake shape is observed, while under higher A3 values butterfly PADs are created, and the 90° minimum becomes more pronounced. In Figure 3.2b, the A3 coefficient is decreased from 0 to -0.2. As a result, the pancake distribution becomes steeper (i.e., the anisotropy increases). In panel (c) we decrease both A1 and A3 coefficients, which also results in narrower PAD shapes. Such an increase in anisotropy of electron PADs is observed on the dayside during geomagnetic storms, and will be discussed later in Sections 3 and 4. In Figure 3.2d we vary the cap coefficient A5 under fixed A1 and A3. When A5 is zero, a perfect pancake distribution is observed. When A5 becomes negative, there is a transition of pancake PADs into the flat-top distributions. When A5 increases and becomes positive, cap distributions are produced. It should be noted that in case of the butterfly distributions, the A3 coefficient can increase to relatively large values (up to

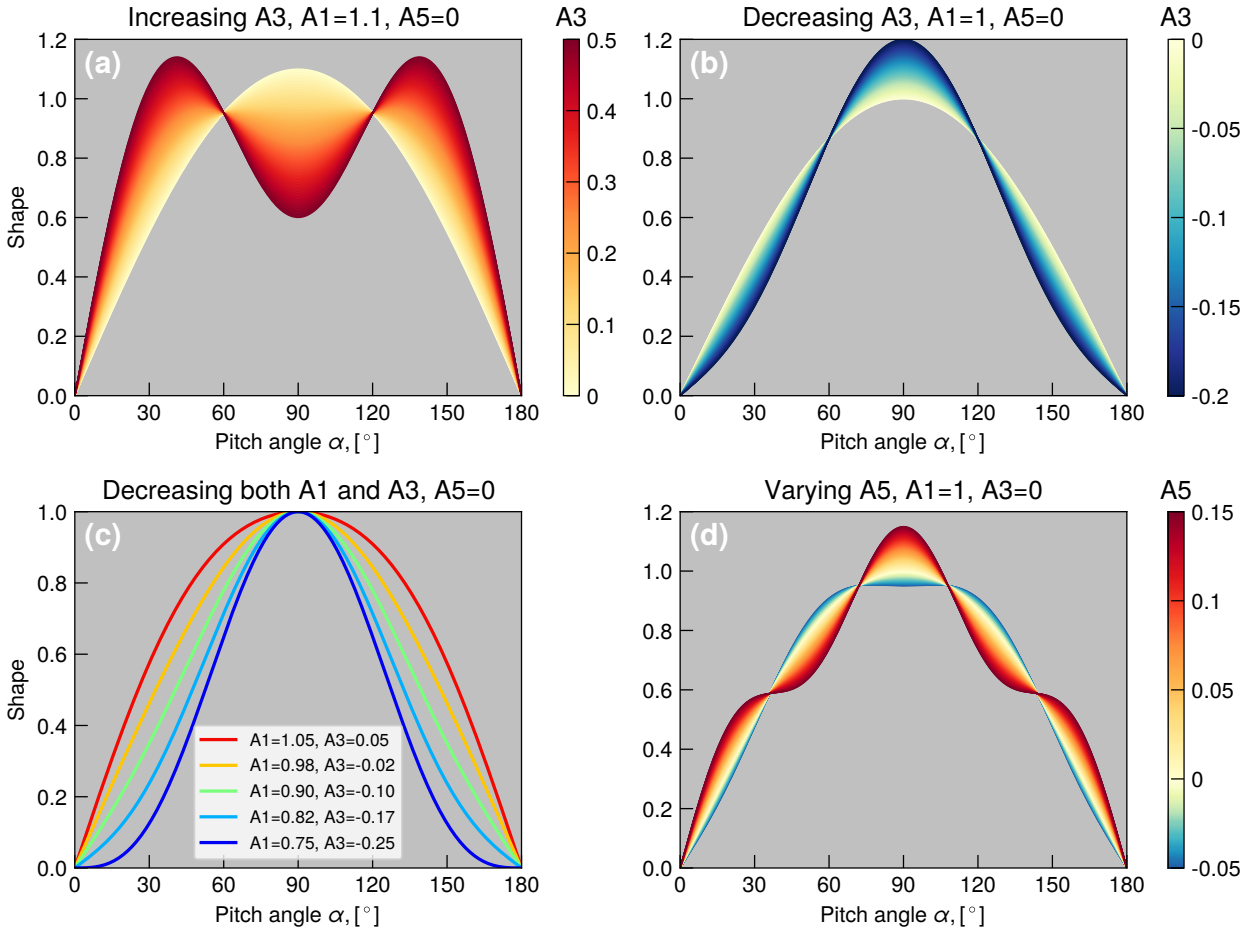

Figure 3.2. Sketches of PAD shapes resulting from different combination of the A1, A3 and A5 coefficients. (a) Increasing the butterfly coefficient A3 under fixed A1 and A5=0 leads to stronger 90° minima; (b) negative values of A3 (also under fixed A1 and A5=0) lead to more anisotropic pancake distributions; (c) decreasing both A1 and A3 also creates steeper pancake PADs; (d) Negative A5 values result in flat-top PADs, while positive A5 values correspond to cap PAD shapes.

0.7), while only small A5 values (0.05-0.15) are needed to resolve the head-and-shoulders PAD shape. As can be seen from Figures 3.1 and 3.2, the Fourier approximation fits well all main types (pancake, butterfly and cap) of equatorial pitch angle distributions. Other PAD shapes can also be resolved by this approximation, for instance, the field-aligned distributions [Clark et al., 2014].

3.3 Superposed epoch analysis of storm-time PAD evolution

3.3.1 L and MLT dependence

In this section, we analyze evolution of the PAD shapes in the outer radiation belt for day and night MLTs during geomagnetic storms in 2012-2019. To select the storm events for this analysis, we follow the methodology of Turner et al. [2019]. As a proxy of the magnetic storm strength, we use the SYM-H index. We select events corresponding to the minimum SYM-H of less than -50 nT, while also requiring that there are no storms 2 days before nor after the event in question, to avoid the repeat events. Using this procedure, we find 129 storms throughout the Van Allen Probes era (the list of storms from October 2012 until October 2017 is given in Turner et al. [2019], their table A1, and the additional events in starting from October 2017 and until the end of the Van Allen Probes mission in late 2019 are listed in Table 3.1). Figure 3.3a shows the SYM-H index, Pdyn and SW electric field (-v·Bz) superposed for the storms used in this study. Following Turner et al. [2019], we select the values of indices, as well as PAD shape coefficients, starting from 84 hours before the SYM-H minimum for each storm and up to 84 hours after the minimum. The data are binned into 3-hour epochs with the zero epoch corresponding to the time of SYM-H minimum. For the spatial binning, 0.2 L was selected as an appropriate step.

In Figure 3.3, the rows (b-e) show the superposed evolution of the PAD shape coefficients A1, A3 and A5 for 58 keV, as well as a higher energy of 735 keV for day (09-15 hr) and night (21-03 hr) MLTs. In each subplot, the x-axis represents the time epochs with respect to min(SYM-H), and the y-axis gives the McIlwain L parameter (Lm), calculated using the TS04D storm-time model. The color-coded values of the pancake (A1), butterfly (A3) and cap (A5) coefficients in these coordinates are shown in the left, middle and right columns, respectively. The corresponding standard deviations and number of points in each bin are shown in the Supporting information of Smirnov et al. [2022a] (Figure S3). In Figure 3.3 we concentrate on two energies, while the energy dependence will be later generalized in Section 3.2.

At lower energies on the dayside, shown in Figure 3.3b, the pancake coefficients A1 generally

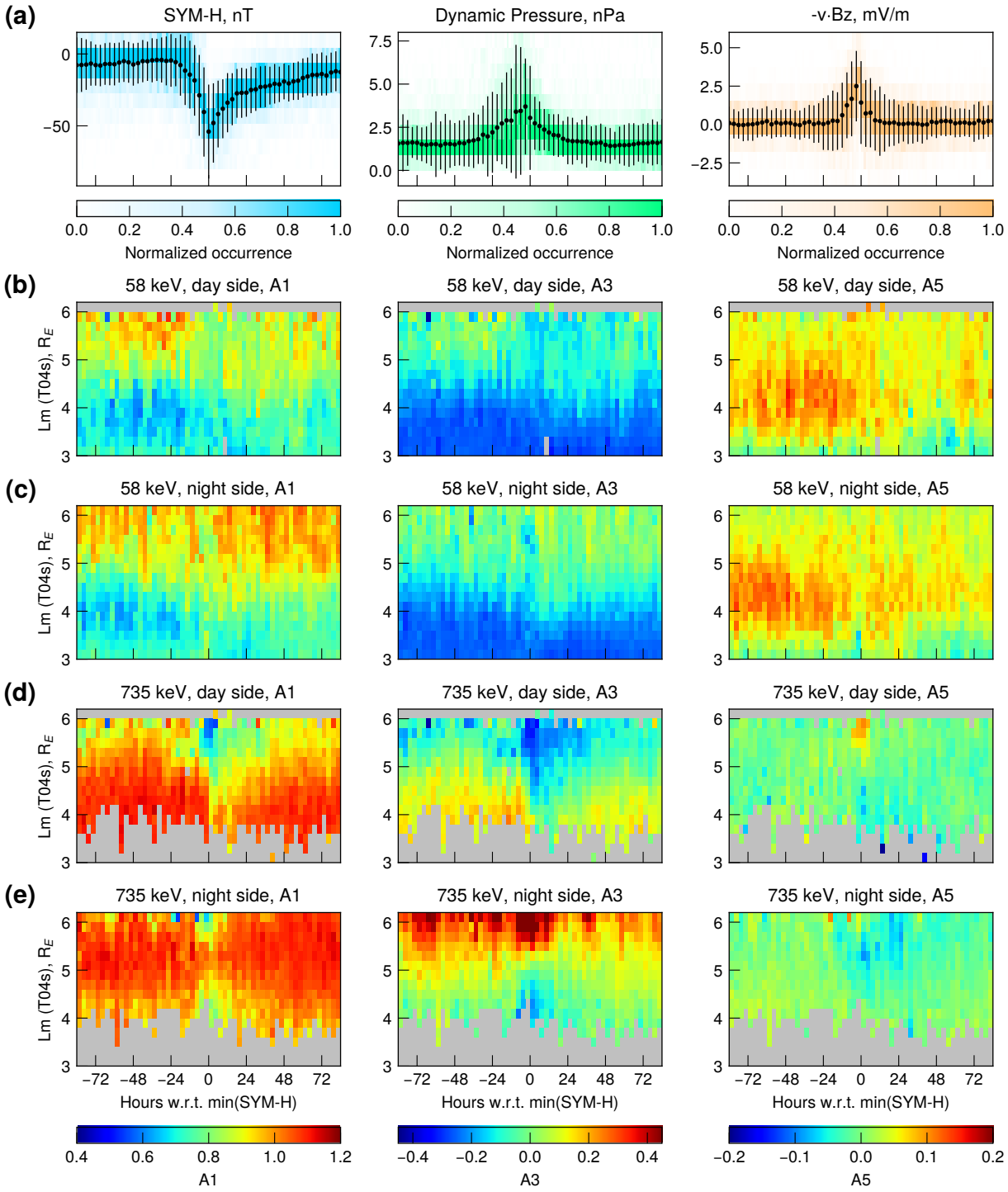

Figure 3.3. (a) SYM-H index, SW dynamic pressure and electric field superposed for 129 storms in 2012-2019, (b) 3-hour averaged dayside values of the A1 (pancake), A3 (butterfly) and A5 (cap) coefficients for 58 keV electrons, binned by the epochs (with respect to the SYM-H index minimum) and Lm. The coefficients in the same format for nightside evolution of 58 keV coefficients is shown in row (c). The storm-time evolution of the 735 keV electron PADs for day- and nightside is given in rows (d) and (e), respectively.

Table 3.1. Additional list of geomagnetic storms in October 2017—September 2019 used in this study (storms before October 2017 are listed in Turner et al. [2019], their Table A1.

Number	Date (yyyy/mm/dd)	time UT of min(SYM-H)
1	2017/10/14	05:35
2	2017/11/08	04:05
3	2017/11/21	06:50
4	2018/02/27	13:00
5	2018/03/10	04:35
6	2018/03/18	21:45
7	2018/04/20	09:30
8	2018/05/06	02:30
9	2018/06/01	07:15
10	2018/08/26	07:10
11	2018/09/11	10:10
12	2018/09/22	07:45
13	2018/10/07	21:50
14	2018/11/05	06:00
15	2019/05/11	03:10
16	2019/05/14	07:50
17	2019/08/05	11:45
18	2019/09/01	06:25
19	2019/05/05	05:20

decrease from 0.9 to 0.7 with decreasing L-values. The same can be seen for the A3 coefficients that turn from around zero at L=6 to negative values of approximately -0.2 at L=3. When both A1 and A3 coefficients decrease, the pancake distributions become narrower (see Figure 3.2c). At the same time, the cap coefficient shows positive values at L=3.5-$\sim$ 5.6 and has a maximum at L$\sim$ 4. This indicates that at a broad range of L-values, the cap distributions will be present (see also Figure 3.5), and that they are most pronounced around $L \sim 4$. In Figure 3.3, the row (c) shows the storm-time evolution of the coefficients for night MLTs at the same energy of 58 keV.

Figure 3.3d shows the evolution of the PAD coefficients for the 735 keV MagEIS energy channel. For this energy, a very strong evolution during the geomagnetic storms can be observed. Based on the row (d), two regions separated by L$\sim$5.2 can be qualitatively defined during pre-storm times. At L<5.2 the A1 values are around 1.1 which indicates broad pancake shapes. At L>$\sim$5-5.2, the A3 coefficient becomes close to zero and turns negative at higher L values.

At the same time, the pancake coefficient A1 decreases, which corresponds to steeper pancake PADs (the cap coefficient at high L-shells remains close to zero). When approaching the storm's main phase, both A1 and A3 decrease which can be observed down to L~4.5. This indicates that during the main phase of the storm, the pancake distributions at higher L-shells become steeper, which is also demonstrated in Figure 3.4d. The coefficients return to their pre-storm configuration in the slow recovery phase.

In contrast, on the nightside, the butterfly coefficients A3 increase at high L-shells when approaching the storm's main phase. In Figure 3.3e, one can also distinguish two distinct regions separated by L=5. At L< 5, the A3 coefficient is small with values around zero, while at L higher than 5, the A3 values are much larger, both for quiet and disturbed times. The A3 values during the pre-storm phase reveal the contribution of quiet-time drift shell splitting, whereas those around the maximum of the dynamic pressure indicate a combination of the drift shell splitting and magnetopause shadowing. It can be seen that at L>5, the values of the pancake coefficient decrease, whereas the butterfly coefficient exhibits a significant increase around the main phase (see also Figure S4 of Smirnov et al. [2022a]). This means that the resulting PADs will have a stronger minimum around 90° PA. After the drop in dynamic pressure, the A3 values return to their original value range. It is worth mentioning that this recovery is faster than on the dayside. Indeed, already at 12-18 hours after the min(SYM-H) the A3 and A1 coefficients are restored to their quiet time range for nightside MLTs, whereas on the dayside it takes 36-48 hours to return to the pre-storm configuration. In Figure 3.5 one can see that for higher energies, the butterfly coefficient becomes even more pronounced. Interestingly, at L between 3.5 and 5, one observes the steepening of the pancake distributions at nightside MLT during the storm-times, which will be discussed in detail below. It is also worth noting that the standard deviations of the *A* coefficients increase around the main phase of the storm (Figure S3 of Smirnov et al. [2022a]). This indicates that although our analysis well depicts the average storm-time behavior of electron PADs, geomagnetic storms correspond to a variety of complex processes [e.g., Reeves et al., 2003] that may not be captured without distinguishing other factors, such as the storm driver, the storm strength, etc.

Figure 3.4 shows the normalized PAD shapes for the four phases of the geomagnetic storms.

In the top row, we show the SYM-H index, solar wind dynamic pressure and y-component of the solar wind electric field superposed for 129 storms analyzed in this study. To demonstrate the PAD shapes corresponding to different phases of the storms, we select 4 epochs, the first one at around 54 hours before the SYM-H minimum (indicating the pre-storm conditions), the second one coinciding with the maximum of the SW dynamic pressure (3 hours before min(SYM-H), indicative of the main phase conditions), the third one at 12 hours past the SYM-H minimum (referred to as 'fast recovery'), and the fourth epoch at 68 hours after min(SYM-H) (i.e., in the 'slow recovery'). These four epochs are marked on the superposed SYM-H, Pdyn and SW electric field plots. In Figure 3.4, row (b) shows normalized PAD shapes for the 58 keV MagEIS channel on the dayside. The pitch angle distributions at low L-values (L< 3.5) exhibit pancake shapes, while at L~ 3.5, the PADs transition into cap shapes (the morphology of cap PADs with respect to energy, L and MLT is described in detail in the following subsection). In the row (c) the PAD shapes are evaluated for 58 keV electrons at nightside MLTs. Generally, they look similar to the dayside shapes, except during the main phase (slice 2) where the nightside PADs exhibit pancake shapes for $L > 5.2$ while on the dayside distinct cap distributions can be observed. These results are in line with previous observations reported by Lyons and Williams [1975a]. They showed that at energies of tens to hundreds of keV there is a loss of head-and-shoulder structure during the main phase of the storm. The cap distributions generally arise due to interactions with the plasmaspheric hiss waves [e.g., Lyons et al., 1972]. Due to erosion of the plasmasphere around the main phase, cap distributions transition into the pancake PADs. During the recovery phase, the plasmapause extends to higher L-values, and interactions with plasmaspheric hiss cause the quiet-time cap distributions to re-form.

In Figure 3.4 (row d), we show the PAD shapes of the 735 keV electrons on the dayside (09< MLT <15). It can be seen that for L >3.5, the distributions during pre-storm conditions exhibit pancake shapes. During the main phase, as noted above and shown in Figure 3.4d, the pancake distributions become narrower and then gradually recover to their original broader shapes. The magnetosphere is compressed around the main phase of the storm, which is reflected in the peak of the SW dynamic pressure (Figure 3.4a). This compression is more pronounced at dayside MLTs and gives rise to a westward electric current which will move ions and electrons inwards

and adiabatically increase their kinetic energy [Walt, 2005]. In this case, the third adiabatic invariant breaks down, whereas the first two invariants are conserved. Equatorially mirroring particles experience the highest adiabatic energy change. When comparing 90° particles with other pitch angles after inward transport, the 90° particles originated from the lowest energies where the phase space density was highest, explaining why the PAD is more peaked at 90° than before. As the system goes into the recovery, the pitch angle scattering and radial diffusion smooth out these highly anisotropic pancake distributions into their broader pre-storm shapes [Walt, 2005], which can be seen in slices 3 and 4 in Figure 3.4d. One interesting feature following the main phase is the appearance of the butterfly distributions in the inner zone. This has been previously attributed to the magnetosonic waves [e.g., Ni et al., 2016] and hiss waves [Albert et al., 2016].

The normalized PAD shapes of the 735 keV electrons at nightside MLTs for the 4 storm phases are demonstrated in Figure 3.4 (row e). Our statistics cover L from 3 to around 6.4 for nightside MLTs, and for the pre-storm conditions butterfly distributions are observed at L= 5.6 − 6.4. At L= 5 − 5.6, we observe the flat-top distribution shapes (as a transition region between butterfly and pancake PADs), and pancake distributions for L<5. The pre-storm butterflies observed on the nightside at high L-shells are indicative of the quiet-time drift shell splitting effects. Roederer [1967] showed that due to the asymmetry of the geomagnetic field particles starting at the same point but with different pitch angles will end up at different radial distances at the opposite side of the magnetosphere. If electrons start from the same point on the nightside, the near equatorially-mirroring particles would drift further from the Earth on the dayside than particles of lower pitch angles. The PA-dependence of the drift paths leads to the formation of the butterfly distributions. While the drift-shell splitting is not energy dependent, it has been well established that a negative radial gradient in phase space density (PSD) is a necessary component to create the butterfly PADs [Roederer, 1967]. At lower energies, the flux gradient is smaller than at high energies [see e.g. Figure 3 in Turner et al., 2019], and therefore the drift shell splitting has less influence on the PADs. This is consistent with the absence of butterflies at low (58 keV) energies (Figure 3.4b,c). In Figures 3.3e and 3.5e we demonstrate that during the main storm phase, the 90°-minimum in the butterfly distributions becomes stronger at

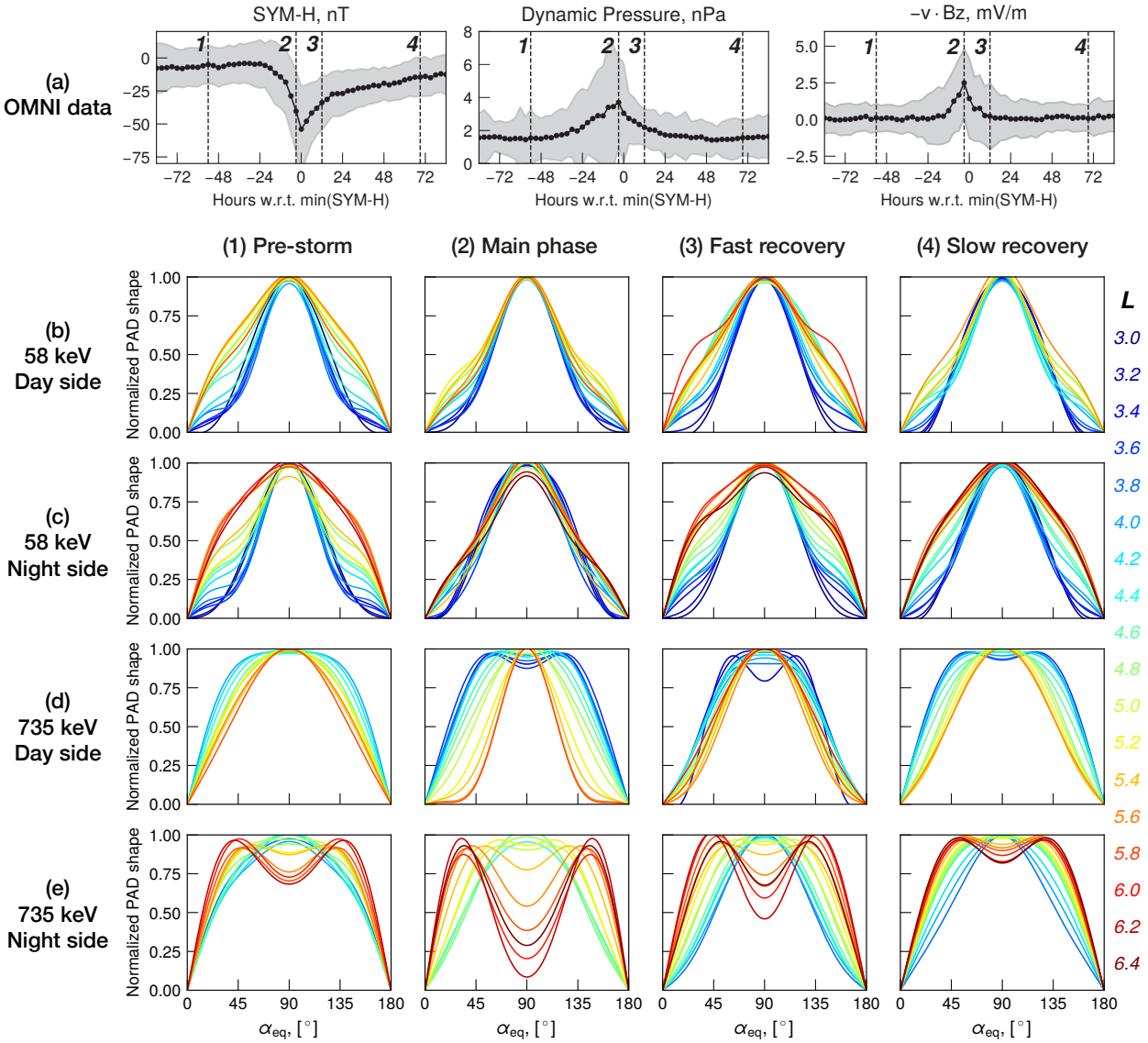

Figure 3.4. (a) Superposed SYM-H index, SW dynamic pressure and electric field. The dashed lines denote epochs at which the PAD shapes are shown below. (b) Normalized PAD shapes of 58 keV electrons for 4 phases of the storm at dayside MLT. The lines are colored by the Lm values shown in a legend below. (c) Same as (b) but for the nighttime MLT. (d) Normalized PAD shapes for 735 keV electrons at dayside, and (e) at nightside.

high L on the nightside. In Figure 3.3e this manifests as an increase in the A3 coefficients around the maximum of the dynamic pressure. In Figure 3.4e, slice 2 shows that the butterfly PADs also extend to lower L-values. With increasing dynamic pressure, drift shell splitting intensifies and thus the butterfly distributions can be observed at lower L-shells. At around L= 4.8 − 5.0, PADs transition into the flat-top/ broad pancake and then to steeper pancake distributions at L<4.5. In Figure 3.4e, PADs at L< 3.6 during the pre-storm phase corresponded to background levels of electron flux and were removed, whereas after the main phase the flux values are increased in the slot region [e.g., Reeves et al., 2016], and the pitch angle distributions exhibit pancake shapes.

3.3.2 Energy dependence

Figure 3.5 demonstrates the energy dependence of the PAD shape coefficients for L= 3 − 6 at day and night MLTs. The resulting PAD shapes plotted as a function of energy under several fixed values of L are shown in Figure 3.6. The values shown here are superposed for 129 storms for the pre-storm and main phase conditions. We first analyze the day-time morphology. It can be seen from Figure 3.5 (a and b) that at energies of <∼ 100 keV, the pancake coefficient A1 appears similar for storm and quiet conditions. The same can be observed for the butterfly coefficient A3 in panels (e and f). At higher energies at L< 5, both A1 and A3 coefficients generally decrease with decreasing energy during quiet times. At the same time, in Figure 3.5i, the cap coefficient is intensified at energies of hundreds of keV and below. It is worth noting that the peak of the cap coefficient across the L-shells is strongly energy-dependent, and moves inward with increasing energy (Figure 3.5i). At higher energies the cap coefficient values are generally around zero (at L> 3.2), which is consistent with the theoretical results by Lyons et al. [1972] and is attributed to the fact that with increasing energy, the dominant first-order cyclotron resonance extends to higher PAs (see Figure 4 in Lyons et al. [1972]) and pitch angle scattering can then affect all equatorial pitch angles.

During the main phase at dayside MLTs, the A1 and A3 coefficients decrease (Figure 3.5, panels b and f), which is especially evident for L> 4.5. At energies below 300 keV, the values of the cap coefficient remain at around 0.05 which means that the cap distributions persist at lower

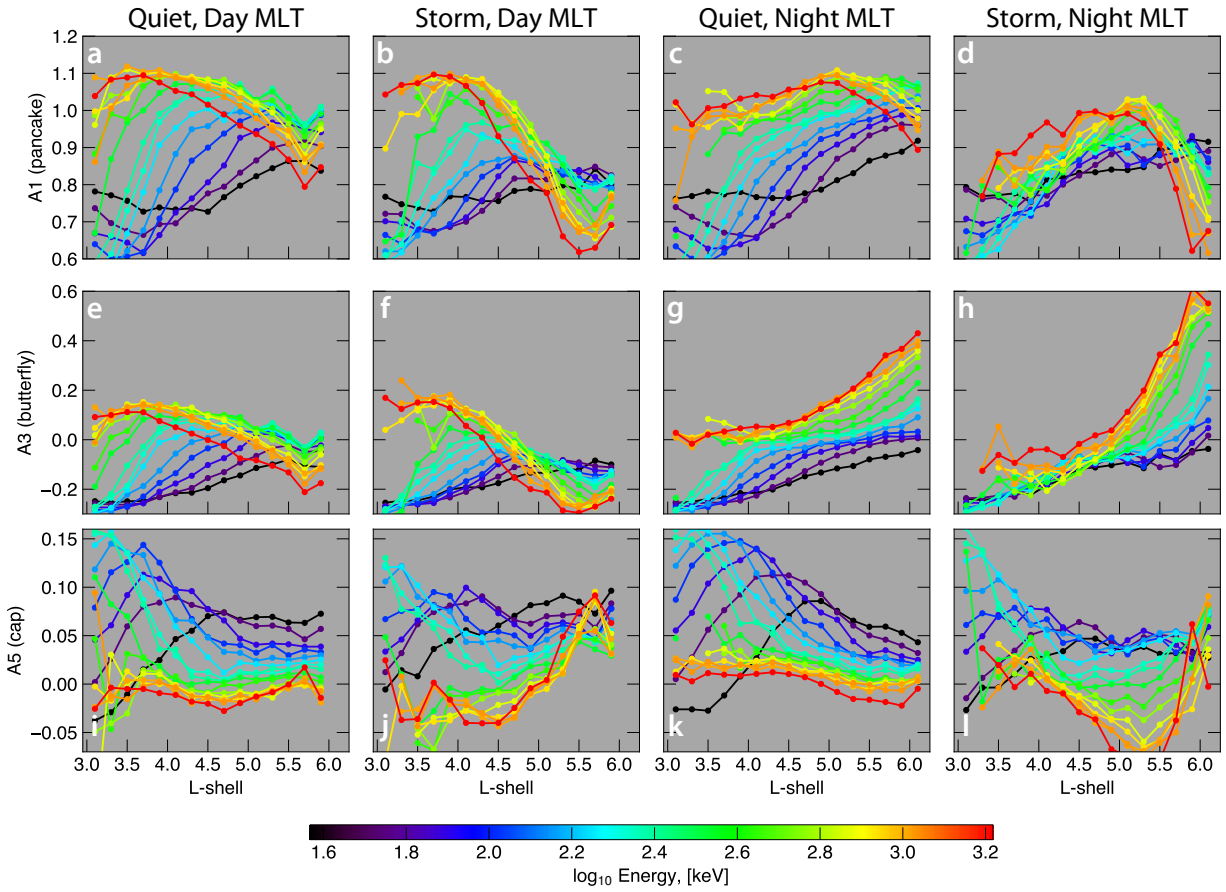

Figure 3.5. Energy dependence of the PAD shape coefficients for pre-storm and main phase conditions at day and night MLT sectors.

energies during the main phase. The cap coefficient decreases at L< 4.5 for energies above 300 keV and turns negative, which results in the flattop distributions (Figure 3.6f). At L> 4.5, the A1 and A3 coefficients decrease, while the A5 coefficient increases. This increase in the A5 coefficient corresponds to strongly anisotropic pancake PADs (see also Figure 3.4d, slice 2) during the main phase on the dayside. Furthermore, the degree of this anisotropy increases with energy (see also Figure 3.6i,j).

During the pre-storm phase the butterfly distributions dominate at L> 5.5 on the nightside (also discussed in Section 3.1). In Figure 3.5d one can see that at L> 5 the pancake coefficient equals approximately 0.9-1.1, while during the main phase the values drop down to 0.6-0.8 across the MagEIS energy range. At the same time, there is a dramatic increase in the butterfly coefficient A3 at large L-shell on the nightside. For instance, at energies of ~ 300 keV the pre-

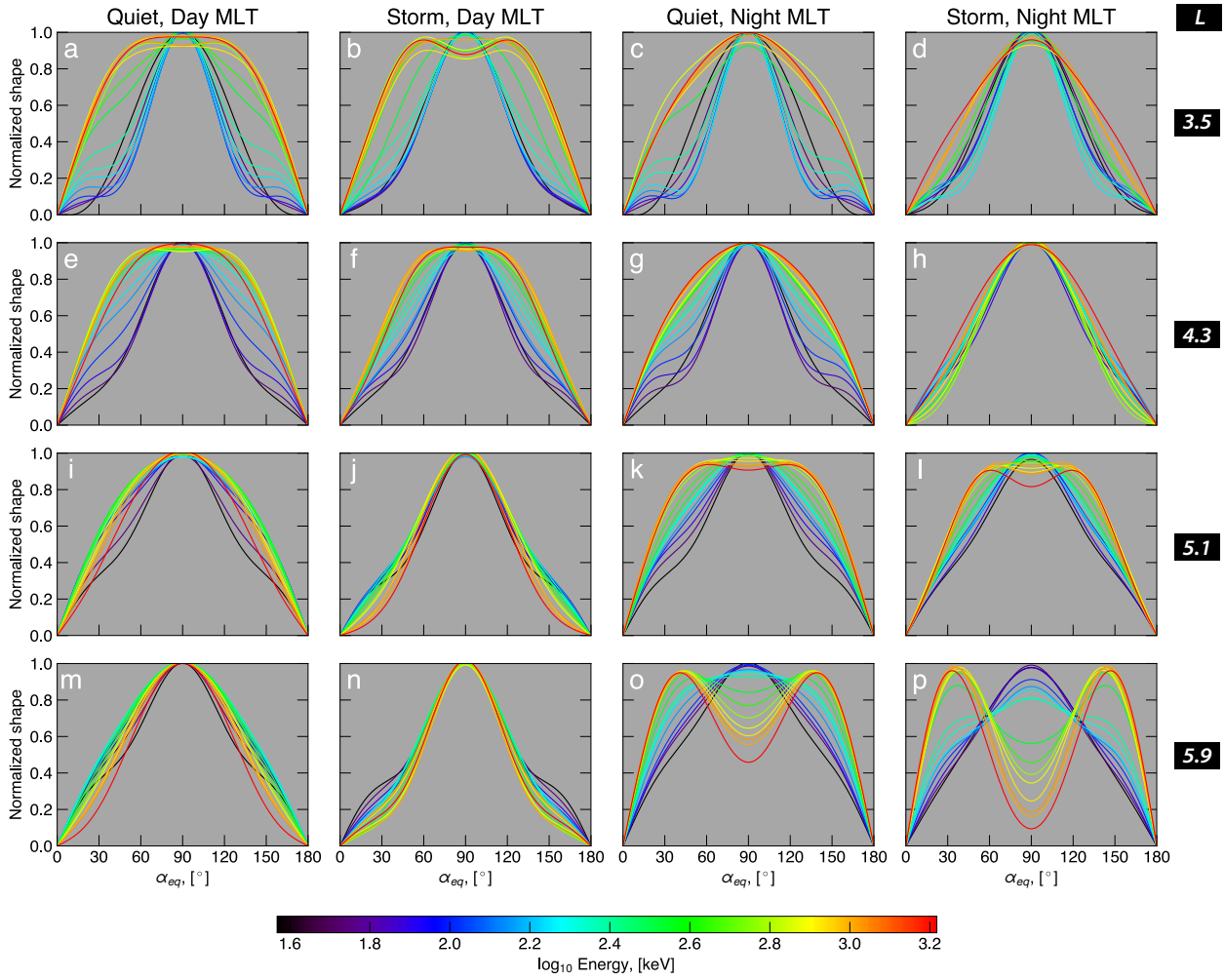

Figure 3.6. Energy dependence of pitch angle shapes at different L-values for quiet and storm conditions on the day and nightside MLT.

storm A3 values were around 0.2, while during the main phase they are magnified by a factor of 2. Furthermore, from Figure 3.6p it is obvious that the 90° minimum gets stronger with increasing energy. At L of ~5, the A5 coefficient becomes negative, and as shown in Figure 3.1 this corresponds to the flat-top PAD shape. At $L < 5$, the butterfly coefficients A3 significantly decrease at all energies (Figure 3.5h) and remains small at lower L-shells, where PADs have pancake shapes.

3.4 Discussion

In this study we employed a Fourier sine series expansion to approximate electron pitch angle distributions. It was shown that a combination of the first, third and fifth sine harmonics can effectively fit all main types of PADs in the outer radiation belt. This approximation has previously been used for analyzing distributions of electrons in planetary magnetospheres, for instance, by Clark et al. [2014]. In case of the terrestrial radiation belts, most studies fitted electron PADs to the $\sin^n(\alpha)$ function, where n shows the steepness of the distribution (larger values of n correspond to more anisotropic distributions). This approximation is very easy to use, and the resulting values can be incorporated into the radiation belts simulations [e.g., Shi et al., 2016]. However, the $\sin^n(\alpha)$ function can only approximate flat-top, pancake and isotropic distributions but is not capable of fitting butterfly and cap shapes. Butterfly distributions can account for up to 80% of PADs on the nightside, and cap distributions dominate lower energies both for day and night-side MLTs during geomagnetically quiet times. Allison et al. [2018] combined two $\sin^n(\alpha)$ terms with different n values which helped two resolve the cap but not butterfly distributions. Due to the fact that different types of PADs can be linked to specific processes acting within the radiation belts, it is crucial to use an approximation that can fit all of the PAD types. We have shown (Figures 3.1 and 3.2) that the Fourier series are capable of resolving all main PAD shapes. Furthermore, the expression used here (Equation 3.1) is easy to integrate over the solid angle and can be used to compute omnidirectional flux values using an analytic expression. We note that the methodology developed in this study can be extended to a range of magnetospheric problems, for instance, to analyzing pitch angle distributions of low-energy electrons and ions.

It is well-known that electrons in the inner radiation belt are stable both in terms of their amplitudes [e.g., Shprits et al., 2013], and pitch angle distributions [e.g., Zhao et al., 2018]. Furthermore, electrons at energies over 1 MeV are generally absent in the inner zone [Fennell et al., 2015]. Therefore, for analyzing the PAD dynamics during geomagnetic storms we concentrated on the outer radiation belt region. While several studies have already used RBSP data for analyzing the storm-time PAD evolution [e.g., Ni et al., 2015, 2016; Drozdov et al., 2019; Ni et al., 2020; Pandya et al., 2020; Greeley et al., 2021], they focused on relativistic and ultra-relativistic

energies sampled by the REPT detector. In the current paper we concentrated on lower energies, from 30 keV to around 1.6 MeV, using observations by the MagEIS detector during the entire lifespan of the Van Allen Probes mission in 2012-2019. Our study has a certain overlap in energies (> 1 MeV) with the previous works by Pandya et al. [2020] and Greeley et al. [2021] and therefore our results can also be compared to those studies. Due to the fact that RBSP orbit was revolving in MLT quite slowly, with a full revolution being completed every 22 months, the statistics for analyzing both MLT and storm driver dependence would be limited, and therefore in the present study we concentrated on the day- and night-time PAD morphology. Furthermore, Pandya et al. [2020] showed that at least for relativistic energies, the dependence of PAD evolution on storm driver was negligible.

It has been shown that the morphology of electron PADs is significantly different during geomagnetically quiet times compared to the active times [Lyons and Williams, 1975b]. During periods of low geomagnetic activity, pitch angle distributions at energies below ~300 keV exhibit cap (or, head-and-shoulder) shapes. This configuration results from cyclotron and Landau resonance with hiss waves, and is strongly energy-dependent [Lyons et al., 1972; Lyons and Williams, 1975a]. In particular, in Figures 3.4 and 3.5 we demonstrate that the cap coefficient A5 exhibits a peak at L~ 5 for 37 keV electrons, and then moves inward in L with increasing energy, which is in line previous results of [e.g., Shi et al., 2016] who reported that peak of PAD anisotropy, corresponding to cap PADs, was moving to smaller L-values under increasing energy. The same conclusion was reported by Allison et al. [2018]. For each energy, Allison et al. [2018] highlighted regions where the combination of two $\sin^n(\alpha)$ terms was performing better than a single sine term. Those regions corresponded to the cap distributions. It was found that the extent of this region diminished with increasing energy, but was practically independent of Kp levels. In the current study we observed loss of the head-and-shoulders structure during the main phase of the storm at nightside MLTs, which is consistent with previous studies by Lyons and Williams [1975b] and Zhao et al. [2018]. Due to the fact with increasing geomagnetic activity the plasmasphere is eroded [e.g., Goldstein et al., 2019], there would be no hiss waves at higher L-shells that could generate cap PADs [see e.g., Lyons and Williams, 1975b]. Furthermore, during active times the low-energy particles are injected from the tail [e.g., Reeves

et al., 1996]. Turner et al. [2015a] demonstrated that injections of electrons with energies below 240 keV can be frequently observed within the geostationary orbit. Furthermore, Motoba et al. [2020] performed a superposed epoch analysis of dispersionless injections using RBSP data and showed that the corresponding pitch angle distributions of tens-of-keV electrons exhibited pancake shapes. Therefore, the transformation of cap PADs into pancakes during the main phase at low energies on the nightside (Figures 3.3c and 3.4c) is likely due to the combination of the plasmasphere erosion and particle injections from the tail.

At energies $>\sim$ 150 keV, pancake distributions are observed on the dayside at $L > 4$ (Figure 3.6). The pancake distributions generally result from the particle pitch angle diffusion [e.g., West et al., 1973]. During geomagnetically quiet times, pancake distributions exhibit relatively broad shapes. During active times, we observe narrowing of the pancake PADs. During the main phase, PADs at high L-shells (L> 5.8) become strongly anisotropic, but distributions at lower L-shells still have broad shapes. The opposite is observed during the fast recovery phase (approximately 12 hours after the SYM-H minimum) - the narrowing moved inwards in L-shell, while at higher L-values the distributions already started to recover to their pre-storm shapes. In the slow recovery phase, the distributions returned to their pre-storm morphology. These results go well with previous findings of Pandya et al. [2020] and Greeley et al. [2021] who also reported narrowing of the pancake PADs during the main phase of the storm. These signatures (narrowing of pancake PADs which progressively moving inward and the subsequent relaxation to pre-storm shapes) are indicative of the inward radial diffusion [Schulz and Lanzerotti, 1974].

On the nightside, the PAD morphology is very different than on the dayside. At energies > 200 keV the quiet-time distributions are mainly of pancake type at L< 5. For L-shells around 5 and energies of around 200 keV, the pancake distributions transform into flat-top PADs. Then, at higher energies there is an emerging minimum at 90° PA which intensifies with increasing energy (see Figure 3.6). Energy dependence of the butterfly distributions can be explained as follows. During quiet conditions at L> 5 the equatorially mirroring particles in the nightside are transported to larger radial distances in the dayside than the lower PA particles, which is known as magnetic drift shell splitting [e.g., Roederer, 1967; Sibeck et al., 1987]. This effect, however, depends only on magnetic field and not on particle energy. At the same time, it has been well

established that a negative radial gradient in PSD is a necessary component for the drift shell splitting to be effective [Roederer, 1967]. Recently, Turner et al. [2019] presented a superposed epoch analysis of omnidirectional electron flux observed by the Van Allen Probes mission. In their Figure 2 one can see that flux gradient in L at lower energies is relatively flat. On the other hand, at higher energies there is a strong negative flux gradient which enables the drift shell splitting to have the full effect on pitch angle distributions and create butterfly PADs. Therefore, with increasing energy the drift shell splitting effects become more evident due to stronger radial gradients in flux.

During the main phase of the storm, the butterfly distributions on the nightside are intensified (Figures 3.4 – 3.6). Such an intensification is most likely due to a combination of the enhanced drift shell splitting and magnetopause shadowing. While the magnetopause is usually located at radial distances of $> 10R_E$, it is well-known that during active times the last closed drift shell can move inward down to L$\sim$ 4. Equatorially mirroring electrons travel to larger distances than low-PA electrons and therefore can get lost to the magnetopause, creating the butterfly distributions. The storm-time butterfly distributions are not observed below L=5 (see Figures 3.3 – 3.5). At L around 5, one can observe a transition of the butterfly into flat-top PADs. Horne et al. [2003] proposed two potential explanations for such a transition at higher energies, namely the inward radial diffusion and wave particle interactions. They concluded that the radial diffusion could be an important factor but did not account for energy dependence of flat-top PADs. By considering the cyclotron resonance with whistler mode chorus waves in presence of low plasma densities, Horne et al. [2003] were able to reproduce a realistic energy dependence of the flat-top PADs.

In this study, we analyzed the normalized electron pitch angle distributions measured by the MagEIS detector onboard the Van Allen Probes mission. The storm-time spin-averaged electron flux intensities were previously investigated by Turner et al. [2019], and it was reported that electrons of different energies exhibited a significantly different response to geomagnetic storms [see also Turner et al., 2015b]. In our study, the storm-time morphology of the PAD shapes was also found to vary greatly with energy. It is worth noting that the results presented here can be combined with averaged picture of the spin-averaged electron flux evolution from Turner et al. [2019], as both the flux intensities and pitch angle distribution shapes are important for

distinguishing between the loss and acceleration processes [e.g., Chen et al., 2005]. Furthermore, due to complex mechanisms acting in the outer belt during the storm times, different particle populations can evolve into one another. For instance, electrons with energies up to tens of keV (the so-called seed population) can be injected during the substorm activity and produce waves that can energize the tens to hundreds of keV electrons to higher energies. In order to investigate these processes in detail, it is beneficial to analyze phase space densities under different values of μ and K, which gives an opportunity to see the time lags between the different populations [see, for instance, Boyd et al., 2016]. We note, however, that in the current study we did not make any assumptions on the processes acting within the outer belt, and therefore the PAD shapes averaged for different storm epoch and energies obtained in this study already include part of the information on the mechanisms mentioned above and provide a good indication of the averaged storm-time behavior of the 30 keV - 1.6 MeV electrons.

3.5 Conclusions

Using the full MagEIS data set of pitch angle resolved electron flux at energies 30 keV - ~1.6 MeV in 2012-2019, we analyze equatorial electron PADs at L=3-6. We use a combination of the first, third and fifth sine harmonics to approximate the pitch angle distributions. The corresponding expression can be analytically integrated, and the values of coefficients before the three terms relate to the main PAD shapes. We perform a superposed epoch analysis of 129 strong geomagnetic storms during the Van Allen Probes era for day and night MLTs. Our findings are as follows.

1. Cap distributions are mainly present at energies < 300 keV, and their spatial extent in L shrinks with increasing energy. During the main phase on the nightside, cap PADs transform into pancakes at L> 4.5, likely due erosion of the plasmasphere and particle injections from the tail. During the recovery phase, the cap distributions are re-formed at high L-shells.

2. At higher energies on the dayside, the distributions are mainly pancake. They exhibit broad

shapes during quiet conditions and become more anisotropic during the main phase of the storm due to the field's compression. The degree of this anisotropy smoothly increases with energy.

3. The butterfly distributions can be observed on the nightside at L> 5.6 during the pre-storm phase. During the main phase, the butterfly PADs can be found at lower L-values (down to L=5), likely due to the combination of drift shell splitting and magnetopause shadowing. Furthermore, the 90° minimum intensifies with increasing energy. This is consistent with stronger negative radial flux gradients at higher energies, which allow the drift shell splitting to create stronger butterfly PADs.

4. On the nightside, there is a transition region between the butterfly and pancake PADs, populated by the flat-top distributions. During quiet conditions, this transition is located at $L \sim 5.5$ and moves inward to $L \sim 5$ during the main phase of the storm.

The PAD shape coefficients for different L-shells, MLTs and phases of geomagnetic storms obtained in this study can be used for the comparisons with the results achieved through the existing wave models when the flux magnitude is taken into consideration, as well as the outputs of the physics-based radiation belt simulations, in terms of PAD shapes. Furthermore, the dependencies reported here can further be used to improve the existing empirical models of the pitch angle distributions in Earth's outer radiation belt.

Acknowledgments

The pitch angle resolved MagEIS data are publicly available at `https://www.rbsp-ect.lanl.gov/rbsp_ect.php`. The solar wind and geomagnetic indices were downloaded from the OMNIWeb database: `https://omniweb.gsfc.nasa.gov`. Artem Smirnov acknowledges support from the International Space Science Institute (ISSI – Bern, Switzerland) through the ISSI team on "Imaging the Invisible: Unveiling the Global Structure of Earth's Dynamic Magnetosphere". This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870452 (PAGER). Hayley Allison was supported by the Alexander von Humboldt foundation.

— 4 —

An empirical model of the equatorial electron pitch angle distributions in Earth's outer radiation belt *

Abstract

In this study we present an empirical model of the equatorial electron PADs in the outer radiation belt based on the full data set collected by the MagEIS instrument onboard the Van Allen Probes in 2012-2019. The PADs are fitted with a combination of the first, third and fifth sine harmonics. The resulting equation resolves all PAD types found in the outer radiation belt (pancake, flat-top, butterfly and cap PADs) and can be analytically integrated to derive omnidirectional flux. We introduce a two-step modeling procedure that ensures a continuous dependence on L, MLT and activity, parametrized by the solar wind dynamic pressure, for the first time. We propose two methods to reconstruct equatorial electron flux using the model. The first approach requires two uni-directional flux observations and is applicable to low-PA data. The second method can be used to reconstruct the full equatorial PADs from a single uni- or omnidirectional measurement at off-equatorial latitudes. The model can be used for converting the long-term data sets of the omnidirectional electron fluxes to phase space density in terms of adiabatic invariants, for physics-based modeling in the form of boundary conditions, and for data assimilation purposes.

*This chapter has been published as Smirnov, A., Shprits, Y. Y., Allison, H., Aseev, N., Drozdov, A., Kollmann, P., Wang, D., and Saikin, A. (2022): An empirical model of the equatorial electron pitch angle distributions in Earth's outer radiation belt. *Space Weather*, 20(9), e2022SW003053.

Plain Language Summary

Pitch angle distributions are critically important for understanding the dynamics of trapped electrons in Earth's radiation belt region. Specific PAD types are linked to processes acting within the radiation belts which relate to the origins and loss mechanisms of the particle populations, as well as wave activity. In this study we present a polynomial model of the equatorial electron PADs at energies 30 keV - 1.6 MeV with a continuous dependence on L-shell, magnetic local time and activity driven by the solar wind dynamic pressure. The model can be used to reconstruct equatorial electron flux from observations at high latitudes and can be applied for converting the long-term electron flux data sets to phase space density, driving the boundary conditions for the physics-based simulations and for data assimilation.

4.1 Introduction

The Van Allen radiation belts contain charged energetic particles, primarily protons and electrons, trapped by the terrestrial magnetic field. Electron radiation belts exhibit a dual-zone structure, consisting of the inner and outer belts [e.g., Van Allen and Frank, 1959]. The two belts are separated by the slot region which is usually devoid of electrons but can be filled during geomagnetically active times [e.g., Reeves et al., 2016]. Charged particles that move much faster than the bulk motion of the plasma organize very well relative to the magnetic field, so that measurements over the full sky are not needed. This is generally the case for electrons. Particle detectors on spinning satellites can sweep through all angles to the magnetic field, measuring intensities for local pitch angle bins. The resulting angular distributions, referred to as the pitch angle distributions, play an important role in understanding trapped electrons' dynamics in the radiation belt region. In particular, specific PAD shapes are linked to distinct processes that relate to the origins of these electrons, their loss mechanisms, and wave activity [e.g., Roederer, 1967; West et al., 1973; Lyons et al., 1972; Walt, 2005; Horne et al., 2003; Ni et al., 2015].

Several PAD types have been identified in the radiation belt region [e.g., West et al., 1973].

Some of the most common types include pancake, butterfly, flat-top, and cap PADs. The pancake distribution, also known as the normal PAD, exhibits a maximum flux value around 90° PA, which smoothly decreases toward lower PAs. It is considered to be the most frequently observed PAD type in the inner magnetosphere, especially on the dayside [e.g., Gannon et al., 2007], and generally results from the particle pitch angle diffusion and loss of the field-aligned electrons to the atmosphere. The pitch angle distributions that exhibit a maximum flux at around 30°-60° PA with a relative deficit of the 90° electrons are called butterfly distributions [e.g., West et al., 1973]. The quiet-time butterflies in the outer belt are prevalent at high L-shells on the night side and mainly formed as a result of drift shell splitting [e.g., Selesnick and Blake, 2002]. They can also be caused by wave activity in the outer radiation belt [e.g., Artemyev et al., 2015; Kamiya et al., 2018] and in the inner zone [e.g., Albert et al., 2016]. Another common PAD type, a flat-top PAD, has a relatively isotropic flux around 90° PA. Horne et al. [2003] showed it to be a transition shape between the butterfly and pancake PADs. The cap (or head-and-shoulders) PAD shape manifests as a pancake with an additional peak in flux close to 90° PA. It has been shown to form due to interactions with the plasmaspheric hiss waves which scatter the near equatorially mirroring electrons slower than those at lower pitch angles [e.g., Lyons et al., 1972].

Particles trapped in the terrestrial magnetosphere experience a bounce motion between the two mirror points [e.g., Roederer and Zhang, 2016]. At the geomagnetic equator, every bouncing particle can be observed, whereas at an off-equatorial magnetic latitude (MLat) the satellite cannot sample the below-mirroring particles and resolves only a part of the equatorial PAD. Therefore, the most complete information about electron PADs can be obtained through measurements in close proximity to the geomagnetic equator [e.g., Chen et al., 2014]. However, many satellites operate at non-equatorial orbits, for instance the LEO, where the range of observed equatorial pitch angles in the outer belt is limited to small PAs (less than several degrees). In recent years, there has been an increasing interest in using LEO data for radiation belt monitoring [e.g., Allison et al., 2018; Claudepierre and O'Brien, 2020; Green et al., 2021; Reidy et al., 2021]. In particular, the Polar Operational Environmental Satellites (POES) and Meteorological Operational Satellites (MetOP) missions provide long-term data sets that can be used for radiation belt

modeling in the form of boundary conditions and for data assimilation. Both constellations provide observations of only a narrow range of pitch angles by two detectors pointing in vertical and horizontal directions, and in order to reconstruct the full equatorial PADs it is necessary to employ appropriate pitch angle models. In addition, several long-term missions, including GPS, Cluster, etc. measure omnidirectional electron flux and do not provide pitch angle-dependent distributions. In order to convert these long-term data sets to phase space density in terms of adiabatic invariants, it is necessary to have a reliable model of the pitch angle distributions.

In a companion paper, Smirnov et al. [2022a] analyzed the storm-time evolution of equatorial electron PADs at energies 30 keV–1.6 MeV. The authors performed a superposed epoch analysis of 129 strong geomagnetic storms in 2012-2019 and examined the morphology of electron PADs in the outer radiation belt for day and night MLTs. It was shown that for energies >150 keV, the PADs undergo significant systematic changes during geomagnetic storms. In this paper, we aim to generalize this dependence on activity by creating an empirical polynomial model of the equatorial electron PADs. Previous PAD models binned the data into several activity intervals based on the Dst [Chen et al., 2014], AE [Zhao et al., 2018] and Kp [Allison et al., 2018; Zhao et al., 2021] indices, as well as into spatial bins in L-MLT coordinates. In this study, we use the solar wind dynamic pressure as a driving parameter and employ a 2-step modeling procedure (described in Section 4.3) which provides a continuous dependence on L, MLT and activity for the first time. Furthermore, our model can be applied to reconstruct the full equatorial PADs from both uni- and omnidirectional observations at off-equatorial latitudes by using the selected parameterization technique.

This Chapter is divided into six parts. Section 4.2 describes the data and methodology used in this study. Section 4.3 is devoted to model construction. In Section 4.4 we introduce two methods to reconstruct equatorial flux from observations at off-equatorial latitudes and show the model validation. The results are discussed in Section 4.5, and the conclusions are presented in the Section 4.6.

4.2 Dataset and Methodology

4.2.1 Dataset

The Van Allen Probes mission, initially called the Radiation Belt Storm Probes (RBSP), operated in 2012-2019 and consisted of two identical spacecraft (probes A and B) following a highly elliptical orbit with an apogee of $\sim$ 620 km, a perigee of around 30400 km and an orbital period of 9 hours [e.g., Mauk et al., 2012]. The probes were spinning with a period of around 11s [e.g., Claudepierre et al., 2021]. Van Allen Probes covered L-shells from 1.2 to $\sim$ 5.8 on the dayside and $\sim$ 6.2 on the nightside and revolved in MLT covering all magnetic local times every $\sim$22 months [e.g., Mauk et al., 2012]. Onboard, the MagEIS instruments measured intensities of trapped electrons and ions [e.g., Blake et al., 2013]. Each Probe had four MagEIS units providing excellent coverage in energies and pitch angles [e.g., Claudepierre et al., 2021]. MagEIS maintained a quasi-constant magnetic field inside the instrument chamber, which allowed to separate the incoming electrons and protons [for details, see Blake et al., 2013]. Furthermore, using the constant magnetic field created a separation between electrons of different energies which hit different detector pixels numbered according to energy [see, e.g., Figure 1 in Claudepierre et al., 2021].

This study employs MagEIS pitch angle resolved electron flux data (level 3) at energies 30 keV – $\sim$1.65 MeV during Van Allen Probes' entire lifespan in 2012-2019. For each moment in time and each local pitch angle, we computed the corresponding equatorial PA using the relation $\alpha_{eq} = \arcsin\left(\sqrt{\frac{B_{eq}}{B_{loc}}}\sin(\alpha_{loc})\right)$, where B_{loc} and B_{eq} represent the values of the local and equatorial magnetic field, respectively, corresponding to the magnetic field line at the satellite location. The magnetic field magnitude and the McIlwain L parameter (Lm) were calculated using the TS04D storm-time model [Tsyganenko and Sitnov, 2005]. The data from October 2012 until October 2018 are used for model construction, and the subsequent 7 months of data (towards the end of the mission) are reserved for validation. The data are averaged by 5 minutes with an assumed symmetry with respect to 90° PA, and the PADs which had no data points at high ($> 75°$) or low ($< 25°$) equatorial pitch angles were removed in order to exclude unphysical shapes. Electron

flux values in the slot region can often be at background values and the corresponding pitch angle distributions are less indicative of the physics. In order to remove the background, we exclude PADs where maximum value of electron flux is lower than 20 $keV^{-1}cm^{-2}s^{-1}sr^{-1}$. This value was empirically selected to filter out PAD shapes corresponding to background for energies up to 1.65 MeV. Using higher threshold values led to removal of realistic PAD shapes and limited the statistics for model fitting at low L-shells.

4.2.2 PAD approximation using sine series

In order to create a PAD model, it is crucial to approximate PADs with a trigonometric function that can fit all PAD types. In previous studies, pitch angle distributions have been approximated by the $\sin^n \alpha$ function, where the power n shows the steepness of the distributions [see e.g., Vampola, 1998; Gannon et al., 2007, etc.]. This parametrization resolves the pancake, flat-top, isotropic and field-aligned PADs. However, it cannot capture the head-and-shoulders distributions which are frequently observed at low energies, as well as the butterfly PADs which account for up to 80% of all PADs at high L-shells on the nightside [e.g., Ni et al., 2016]. Allison et al. [2018] used a combination of the two $\sin^n \alpha$ terms which helped to resolve cap but not butterfly PADs. Several studies have employed Legendre polynomials to approximate electron PADs in the Earth's radiation belts [e.g., Chen et al., 2014; Zhao et al., 2018, 2021]. In particular, Chen et al. [2014] created the first PAD model in the outer belt, and Zhao et al. [2018] extended this methodology to both inner and outer zones and showed that PADs in the outer belt could be approximated by the first 3 even terms of the Legendre series expansion, while it was necessary to include higher harmonics at $L < 3$ due to the larger loss cones and generally steeper PAD shapes.

In order to approximate the equatorial electron PADs, we use a sine series expansion of the form:

$$j(\alpha) = A_0 + A_1 \sin\alpha + A_3 \sin 3\alpha + A_5 \sin 5\alpha, \tag{4.1}$$

where j is electron flux as a function of PA α. Having electron flux observations at several pitch angles, we use least-squares fitting to obtain the coefficients A_0, A_1, A_3, and A_5. The Fourier

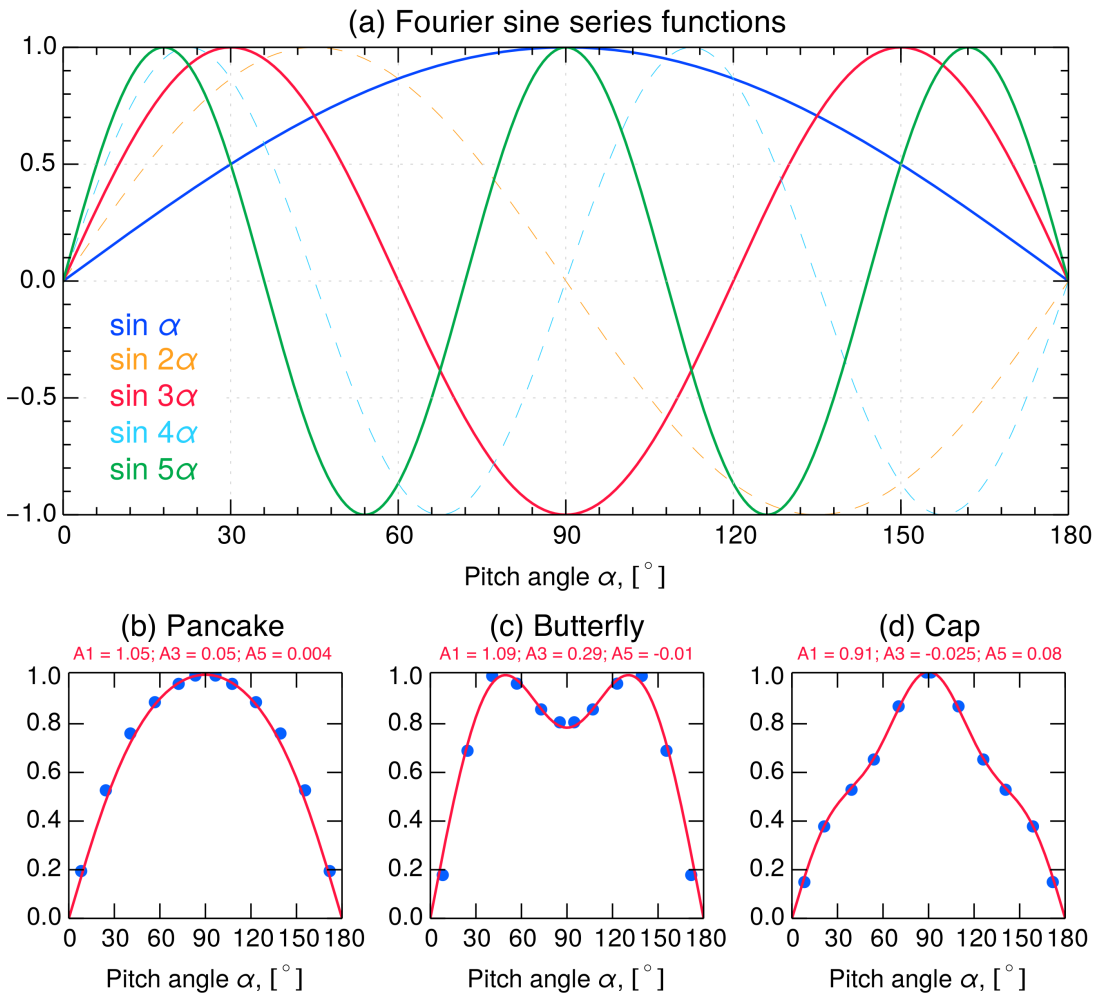

Figure 4.1. (a) Fourier sine functions up to degree 5, (b, c, d) examples of the fitted pancake, butterfly and cap equatorial PADs, respectively, normalized using Equation 4.2. Blue dots show the normalized MagEIS observations and the red lines give the fitted shapes.

sine series expansion has been used by Clark et al. [2014] to analyze pitch angle distributions in Saturn's radiation belts. Smirnov et al. [2022a] used this approximation for the first time to analyze the storm-time PAD dynamics in the Earth's outer radiation belt. The Fourier expansion has notable advantages, for instance, Equation 4.1 can be analytically integrated over the solid angle and the fitted PADs can be converted to omnidirectional flux.

Figure 4.1a demonstrates the shape functions for the first five sine harmonics. The even harmonics ($\sin 2\alpha$ and $\sin 4\alpha$) are asymmetric with respect to 90° PA and therefore are omitted and only odd terms are used further on. From Figure 4.1a it can be seen that generally, the

first Fourier component ($\sin\alpha$) resembles the pancake PAD shape (shown in Figure 4.1b), corresponding to large values of the A1 coefficient and small values of A3 and A5 coefficients. The coefficient before $\sin 3\alpha$ demonstrates the contribution of the butterfly shape. Indeed, the $\sin(3\alpha)$ function exhibits a double peak at around 30° and 150° PA and has a minimum at 90°. Combined with the $\sin\alpha$ term, it can well approximate the butterfly distributions (Figure 4.1c). An example of the cap distribution is demonstrated in Figure 4.1d. In order to fit this shape, a combination of the first and fifth Fourier harmonics is needed, since the first term, reflected in the A1 value, gives the general pancake shape and the fifth term (A5) introduces the head-and-shoulders structure due to the two depletions in the $\sin 5\alpha$ shape around 45° and 135° PA and an additional bump around 90° PA. It is worth noting that in butterfly distributions the A3 coefficient can increase to relatively large values (up to 0.7), while only small positive A5 values (0.05-0.15) are sufficient to fit the head-and-shoulders PAD shape. As can be seen from Figure 4.1, the Fourier sine series expansion well resolves all main types (pancake, butterfly and cap) of equatorial pitch angle distributions found in the outer belt. The PAD shapes resulting from various combinations of the A1, A3 and A5 coefficients are demonstrated in the companion study (Smirnov et al. [2022a], Figure 2). We note that it is possible to extend this methodology to lower L-shells and resolve the steeper PAD shapes observed there by adding the higher order terms to Equation 4.1, similar to the models based on Legendre polynomials [Zhao et al., 2018, 2021].

In Equation 4.1, the A_0 coefficient represents the value of electron flux in the edge of the loss cone, which corresponds to the minimum flux value for the respective PAD. For each of the PADs, one can also find the maximum flux value defined as $j_{\max} = \max_{\alpha_{eq}\in[0,90]} j(\alpha_{eq})$. In this study we are most interested in the PAD shapes. In Section 4.3 we therefore normalize the pitch angle distributions and develop a model for the normalized shapes, and later present a method to reconstruct absolute values of electron flux with the help of observations. Before proceeding with analyzing the pitch angle shapes, every PAD is normalized as follows. At first, we fit the said PAD to Equation 4.1 and obtain the $A_{0,1,3,5}$ coefficients. After that, the PAD $j(\alpha)$ is evaluated with a very small pitch angle step of 0.01 degrees, and the maximum flux value $j_{\max}$ is empirically found. The odd coefficients $A_{1,3,5}$ are normalized to the minimum (A0) and

maximum ($j_{\max}$) values so that each individual PAD spans from 0 to 1 using the equation:

$$\tilde{A}_i = \frac{A_i}{j_{\max} - A_0}, i = \{1, 3, 5\}, \tag{4.2}$$

where $\tilde{A}_i$ denotes the normalized value of the respective coefficient A_i. In order to remove low-quality PAD fits, we use a criterion proposed by Carbary et al. [2011] which uses the standard deviation of the difference between the measured and fitted flux normalized to the minimum and maximum values ($\sigma_N = \sigma/(j_{\max} - A_0)$). The good quality fits correspond to σ_N values below 0.2, while PADs with $\sigma_N > 0.2$ represent low quality fits and were not used in the model construction. The details on this quality flag and examples of high- and low-quality fits are given in Smirnov et al. [2022a] (e.g., their Figure S1 in the Supporting information).

It should be noted that the PADs normalized by Equation 4.2 do not carry information about the flux levels and should only reflect shapes of the distributions. In the companion paper, Smirnov et al. [2022a] analyzed the storm-time evolution of electron PADs in terms of the normalized shapes. In the present study, we, first, create a model of the normalized pitch angle distributions driven by the solar wind dynamic pressure (Section 4.3). Furthermore, in Section 4.4 we give two methods of reconstructing electron flux from the model by using off-equatorial observations.

4.3 Model Construction

Pitch angle distributions in the outer radiation belt are known to be very dynamic [e.g., Horne et al., 2003; Gu et al., 2011]. It is therefore crucial to select a suitable driving parameter for the model. In a companion study, Smirnov et al. [2022a] analyzed the storm-time evolution of the PAD shapes normalized using the methodology described above. It was found that at energies > 150 keV, PADs undergo a systematic evolution during geomagnetic storms. Furthermore, the main processes governing the PAD dynamics in the outer belt can be attributed to changes in the solar wind dynamic pressure [for details, see Smirnov et al., 2022a]. In Figure 4.2, we demonstrate averaged PAD shapes of 871 keV electrons observed by MagEIS at L~5.4 for the day- and nighttime MLTs, colored by the values of the solar wind dynamic pressure. It can be seen

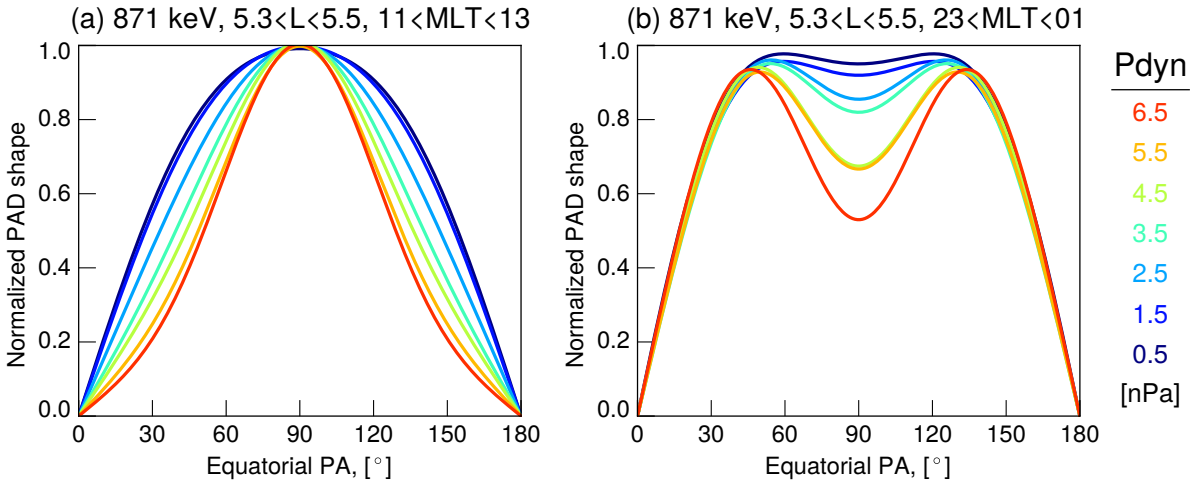

Figure 4.2. Averaged PAD shapes of 871 keV electrons measured by the MagEIS instrument at L~ 5.4 on the dayside (a) and on the nightside (b) for different values of the solar wind dynamic pressure (±0.25 nPa).

that on the dayside, the distributions are of pancake shape and that their 90°-anisotropy increases with Pdyn, likely due to the magnetic field compression. On the nightside, the distributions have butterfly shapes, and the 90°-minima become more pronounced with increasing SW pressure, due to the more intense drift shell splitting (for details, see Smirnov et al. [2022a]). From Figure 4.2 one can see that the PAD shapes exhibit a strong dependence on dynamic pressure, which is analyzed in more detail below. Furthermore, Smirnov et al. [2022a] showed that the morphology and response to activity of electron PADs is strongly dependent on energy. Therefore, the model presented here is constructed for each energy channel.

We first examine the normalized occurrence plots of the PAD shape coefficients $A_{1,3,5}$ within a chosen range of L and MLT with respect to solar wind dynamic pressure, which are constructed as follows. At first, the predictor variable (on the x-axis) is divided into 20 bins. Then, for each interval of the x-axis, the corresponding A_i coefficients are also split into the same number of bins. For each bin, the number of occurrences is calculated and then divided by the total number of points in the respective x-interval. Such a representation effectively demonstrates how the PAD coefficients are distributed for each bin of the predictor variable and essentially represents a normalized 2D probability distribution function [e.g., Kellerman and Shprits, 2012].

In particular, it is well-established that throughout the Van Allen Probes era most of the points were attributed to the quiet conditions, and storm-time events, especially the strong ones, were rare. By normalizing the coefficient values to their respective activity bin, a more balanced estimation of the distributions can be obtained.

In Figure 4.3, we show the normalized occurrence histograms for the day- and night MLTs at high L-values (from 5.7 to 6) of 871 keV electrons. It can be seen in Figure 4.3a that on the dayside, the pancake coefficient A1 decreases with increasing Pdyn, from ~ 1 at low pressure to ~ 0.75 under dynamic pressure of 6 nPa. Furthermore, the butterfly coefficient A3 also decreases, from 0 at low Pdyn values to < -0.2 at higher Pdyn (Figure 4.3c). The cap coefficient A5, while increasing slightly for higher Pdyn, remains very small and is generally close to 0. These results agree very well with the superposed epoch analysis presented in [Smirnov et al., 2022a]. The simultaneous decrease in A1 and A3 coefficients with Pdyn at L>5.7 (see e.g., Figure 3d in Smirnov et al. [2022a]) corresponds to more anisotropic PADs during the storm's main phase as a result of the geomagnetic field compression. On the night side, the pancake coefficient decreases from >1.1 to around 0.9 with increasing Pdyn (Figure 4.3b), while the contribution of the butterfly coefficient A3 roughly doubles with higher Pdyn, which also agrees with findings of Smirnov et al. [2022a]. As the stronger dynamic pressure compresses the Earth's magnetosphere, the effect of drift shell splitting becomes more evident. One notable feature in Figure 4.3b is that the spread of A1 values increases with Pdyn. This indicates that while there is a clear dependence of the PAD coefficients on dynamic pressure, the storm times correspond to complex processes in the outer radiation belt [Reeves et al., 2003], which might not be captured by a simple empirical relation to Pdyn. It is of note that the Pdyn values generally lie from 0 to 6 nPa but can also experience spikes of up to 50 nPa. Figure S1 of Smirnov et al. [2022c] shows that over 97% of values are below 5.5 nPa and therefore the model statistics for the spikes would be limited. Thus, all values corresponding to Pdyn> 5.5 nPa are combined into a separate model bin which corresponds to intervals of very high activity.

The black dots in Figure 4.3 indicate the median coefficient values for each interval of the dynamic pressure. It can be seen that the general dependence can be well approximated using

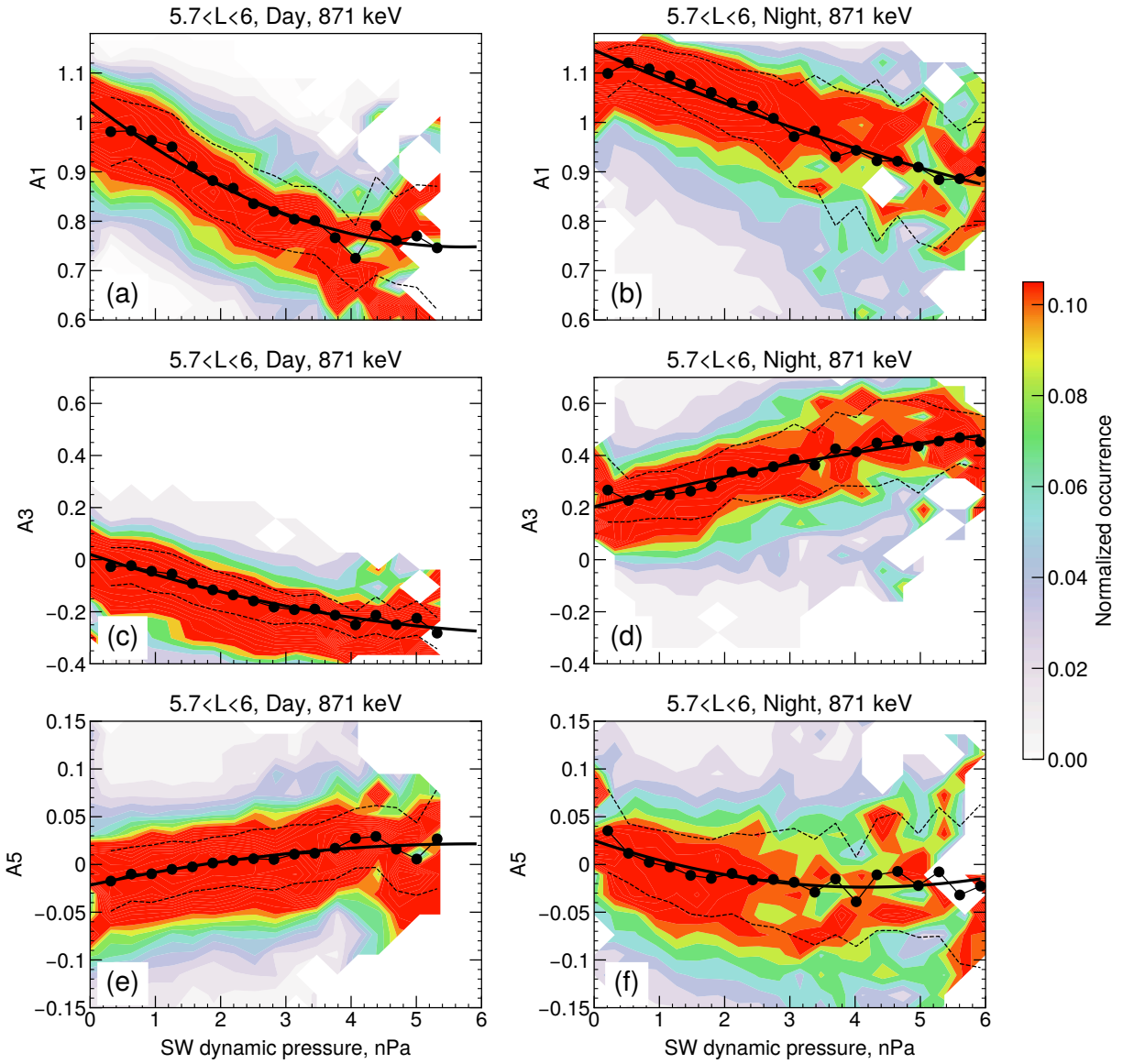

Figure 4.3. Normalized occurrence plots of the A_1, A_3 and A_5 coefficients as a function of solar wind dynamic pressure for 1.04 MeV electrons at day side MLT (a,c,e) and night side MLT (b,d,f), respectively. The median values of the Fourier coefficients in each bin of dynamic pressure are shown as black dots, and the dashed lines give the interquartile range. The bold black lines show the quadratic trends fitted to the medians using Equation 4.3. The contour function was applied to the plot. PADs for the median values of parameters shown here can be found in Figure 4.2.

the quadratic trend line of the form:

$$A_{1,3,5}(\text{Pdyn}) = c_0 \cdot \text{Pdyn}^2 + c_1 \cdot \text{Pdyn} + c_2, \tag{4.3}$$

where Pdyn denotes the solar wind dynamic pressure. We divide data into spatial bins using a step of 1hr MLT and 0.2 L, and in order to achieve better statistics we add data points from within ±3h MLT and ±0.15 L-shell to each of the bins, which is equivalent to smoothing in L and MLT dimensions. At L>5.8, the step in L was increased to ±0.25 L due to the limited Van Allen Probes statistics at high L-shells. By fitting the medians of each of the Fourier coefficients to the dynamic pressure using Equation 4.3, we obtain the matrices of the quadratic trend coefficients $c_{0,1,2}$ as functions of L and MLT. It is now necessary to introduce a continuous dependence on L and MLT into the model.

In order to avoid a discontinuity at the 24-00 MLT boundary, it is commmon practice to replace MLT with its sine and cosine values, denoted as $sMLT = \sin(2\pi \cdot \text{MLT}/24)$, and $cMLT = \cos(2\pi \cdot \text{MLT}/24)$ [see, e.g., Bortnik et al., 2016b; Katsavrias et al., 2021]. Then, the matrix of the trend coefficients c_i is fitted to L, $sMLT$ and $cMLT$ using the equation:

$$c_{0,1,2}(L, sMLT, cMLT) = \sum_{i=0}^{N} \sum_{j=0}^{N} \sum_{\substack{k=0 \\ i+j+k \leq N}}^{N} (p_{ijk} \cdot L^i \cdot sMLT^j \cdot cMLT^k), \tag{4.4}$$

where N is the degree of the polynomial. In this study we use an empirically selected value of N=7. Including higher orders made the results oscillatory, while the polynomials of order less than 7 were missing features seen in the data. The values of p_{ijk} are obtained using Theil-Sen regression which is a modification of the multivariate linear regression robust to outliers [Theil, 1950; Sen, 1968]. The final model is a collection of the p_{ijk} coefficients, and from these the initial Fourier coefficients A_i can be obtained through the script available from [Smirnov et al., 2022b].

Figure 4.4 shows a comparison between the values of PAD coefficients predicted by the polynomial model and those averaged for the MagEIS data under dynamic pressure values of ∼3 nPa (from 2.8 to 3.2 nPa). It can be seen that for all three coefficients, the model values are in very good agreement with data averages. Several regions can be qualitatively defined in Figure

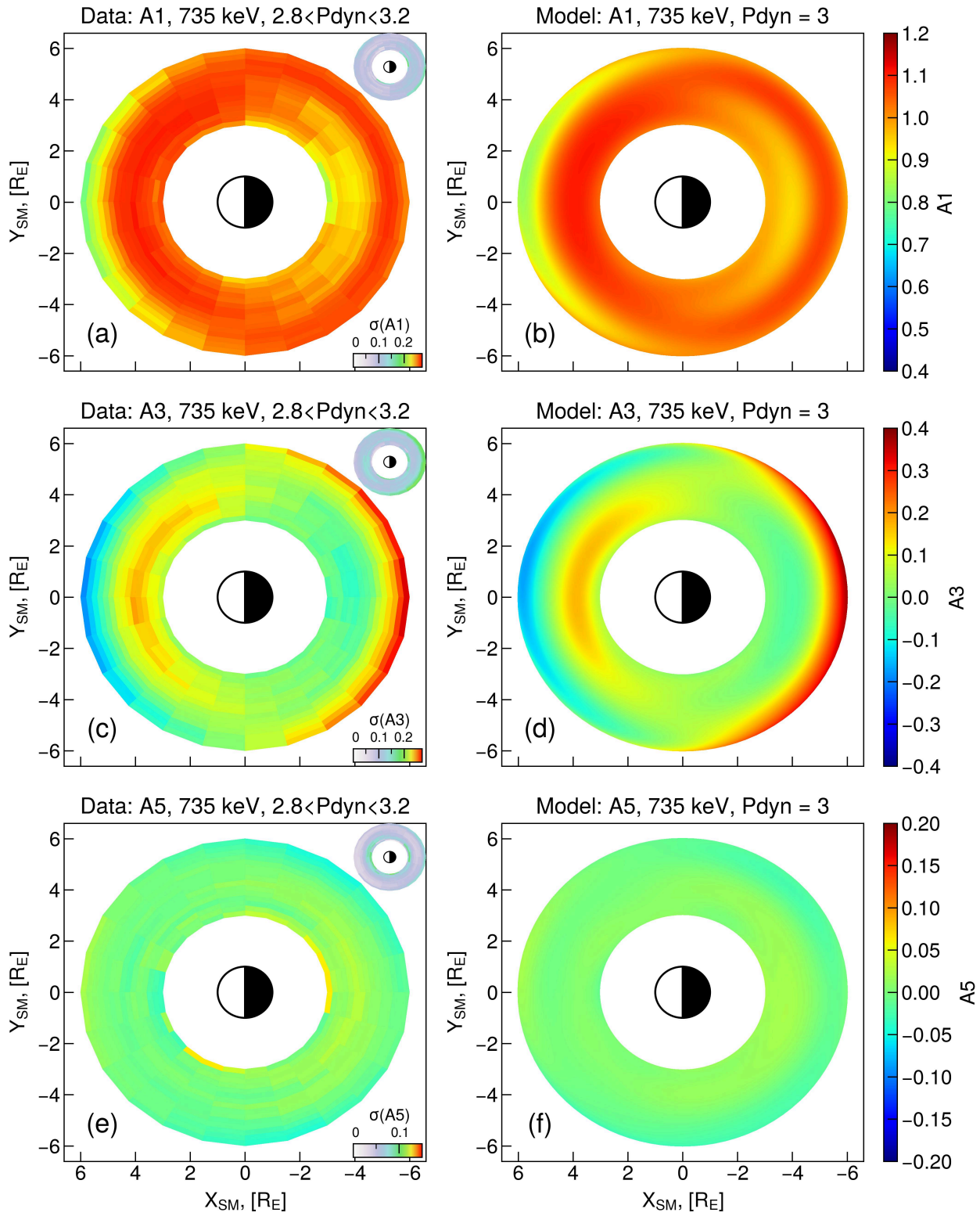

Figure 4.4. A comparison between the median pitch angle coefficients A1, A3 and A5 from MagEIS data (panels a, c, e, respectively) and their model values (panels b, d, f). The inset plots in panels (a, c, e) show the standard deviations of the respective coefficients. The values shown here are for 735 keV electrons and Pdyn~ 3 nPa.

4.4. For instance, one can observe lowered values of A1 and A3 coefficients on the dayside at L> 5, which corresponds to steep pancake distributions. At high L-values on the nightside, on the other hand, there is an increase in the A3 coefficient, which corresponds to the butterfly distributions resulting from the drift shell splitting and magnetopause shadowing. At $L \sim 5$, the butterfly coefficients A3 decrease to zero, and the corresponding PADs are of pancake type. This is in line with theoretical predictions by Roederer [1967], who showed that drift shell splitting is only effective at $L > 5$. One interesting feature in Figure 4.4c is the increase in the butterfly coefficient A3 on the dayside at low L-values, which manifests as a band around 12:00 MLT but is slightly shifted towards the dawn. These L-shells correspond to the slot region for the 735 keV electrons, and the existence of these butterfly distributions has previously been attributed to the wave activity [e.g., Albert et al., 2016]. Smirnov et al. [2022c] demonstrate the same comparison for Pdyn values of 0.5 nPa, corresponding to the quiet times, and Pdyn$\sim$ 5 nPa corresponding to the storm-times (their Figure S2 in the Supporting information). The values of the PAD coefficients are in good agreement with MagEIS observations both for quiet and active conditions, and the observed PAD morphology and activity dependence is in line with previous results based on the superposed epoch analysis presented in Smirnov et al. [2022a]. Furthermore, In Figure S4 in the Supporting Information of Smirnov et al. [2022c] it is demonstrated that observed and modeled PAD shapes are in good agreement for several satellite orbits during different phases of geomagnetic storms. Moreover, Figure S5 Smirnov et al. [2022c] shows the modeled PAD coefficients under low and high (0.6 nPa and 6 nPa, respectively) values of dynamic pressure both for day and night MLTs as a function of energy. It can be seen that the values of the PAD coefficients in Figure S5 agree well with the previous study by Smirnov et al. [2022a] which investigated the storm-time PAD evolution and therefore the model depicts the main feature of the PAD dynamics across different energies.

The model constructed here describes the pitch angle distribution shapes for L=3-6 at energies from $\sim$ 30 keV up to $\sim$1.65 MeV. The model is continuous with respect to L and MLT. Furthermore, the developed model is the first PAD model to have a continuous dependence on activity, in this study parameterized by the solar wind dynamic pressure. It has been established that the model captures the main effects seen throughout the storm-time evolution of pitch an-

gle distributions. The model presented in the current Section gives the normalized PAD shapes. In order to apply the model to the data one needs to be able to connect the PAD shapes to the flux intensities, and the two methods to recover this information are described in the following section.

4.4 Model Validation

4.4.1 Reconstructing Equatorial PADs Using Two Flux Measurements At Low Pitch Angles

In this Section we describe the reconstruction of the full equatorial PADs using the developed model and 2 directional observations of electron flux at low PAs. From Equations 4.1 and 4.2 it is evident that after normalizing the PAD shapes, one needs to retrieve 2 parameters to relate the PAD shapes back to flux magnitudes, namely the A0 coefficient and the scaling factor equal to maximum flux (j_{max}) minus A0. In this case, having two observations allows for an estimation of these parameters using the following procedure.

At each of the two known equatorial pitch angles α_1 and α_2, we can re-write Equation 4.1 in the form:

$$j_{\text{data}}(\alpha_i) = A_0 + A_1 \sin\alpha_i + A_3 \sin 3\alpha_i + A_5 \sin 5\alpha_i, \tag{4.5}$$

with $i = \{1, 2\}$, and for the same equatorial pitch angles, the normalized PAD shapes calculated using the developed model can be expressed as follows:

$$j_{\text{model}}(\alpha_i) = \tilde{A_1} \sin\alpha_i + \tilde{A_3} \sin 3\alpha_i + \tilde{A_5} \sin 5\alpha_i. \tag{4.6}$$

Therefore, one can write:

$$j_{\text{data}}(\alpha_i) = s \cdot j_{\text{model}}(\alpha_i) + A_0, \tag{4.7}$$

where $s = j_{\text{max}} - A_0$ is the scaling factor. It can be retrieved by dividing the difference between the observed fluxes at two pitch angles to that calculated from the model. Then, the A0 value

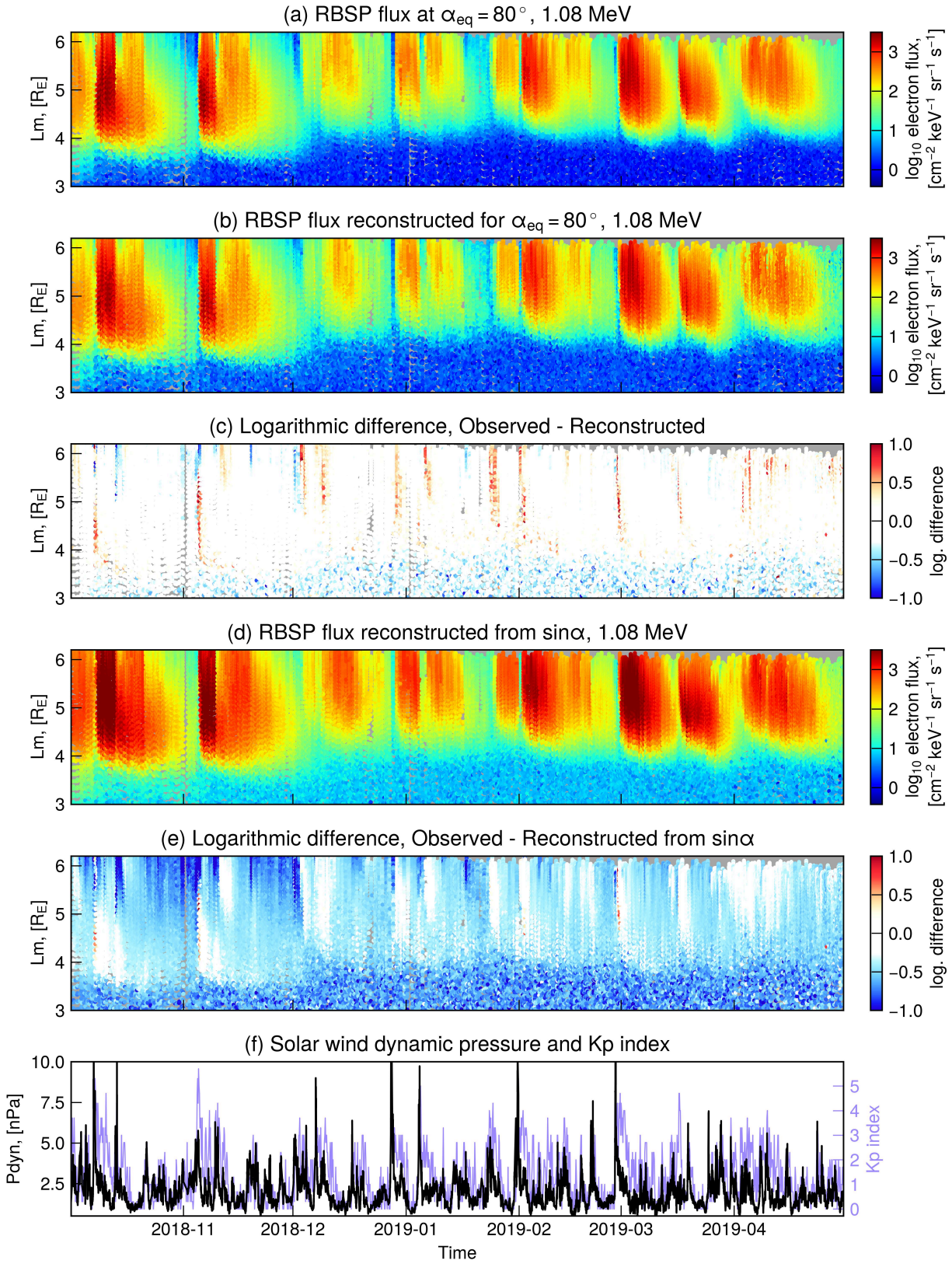

Figure 4.5. (a) RBSP equatorial electron flux for 1.08 MeV, (b) RBSP equatorial flux reconstructed from two lowest MagEIS pitch angles using the developed PAD model, (c) logarithmic difference between observed and propagated flux; (d) equatorial electron flux propagated using the sin(α) approximation, (e) logarithmic difference between observed flux and the propagated one, (f) solar wind dynamic pressure.

can be obtained by substituting values into Equation 4.7 for either of the data points. After both of the parameters have been calculated, the full equatorial PAD can be reconstructed.

Figure 4.5 shows the validation of the modeled electron flux by the RBSP observations for 1.08 MeV MagEIS channel. In Figure 4.5a, electron flux evaluated from RBSP observations at a high equatorial PA of 80° is given. Each of the observed MagEIS PADs was propagated to 80 degrees by using the fit to Equation 4.1. In panel (b) we show electron flux reconstructed using the method described above. We first compute the "observed" flux at 8° and 17° PA (i.e., the two lowest MagEIS pitch angles). Then, these values are supplied as inputs and using the method described above the flux at 80° equatorial PA is retrieved. It can be seen that the observed and reconstructed fluxes are highly consistent, which shows the capacity of the model to reproduce equatorial PADs from low pitch angle data. The logarithmic difference between the observed and reconstructed flux is shown in panel (c). It is the logarithmic difference mostly shows values within ±0.25, although at a relatively small number of points there can be an underestimation on the order of 1. These outliers mainly occur following the spikes of the dynamic pressure, in particular, when the Pdyn values decrease but the Kp index remains at elevated levels. This indicates that the model could benefit from including time-history of solar wind parameters and geomagnetic indices, which can be a topic of a future study. Furthermore, there is slight overestimation of electron flux in the slot region (Figure 4.5c), which appears to be relatively independent of activity. This is likely due to the fact that the flux in the slot region is often at background levels and therefore there are less data points for the model fitting. Nevertheless, the mean logarithmic difference at $L < 4$ equals -0.13, which is considered acceptable for radiation belts modeling.

It is also crucial to establish whether the model presented in this paper performs better than a simple approximation of the form $j(\alpha) = A_n \sin(\alpha)$. To retrieve electron flux at 80° equatorial PA using this approximation, we use MagEIS electron flux evaluated at a low pitch angle of 12.5 degrees and obtain the A_n value. In Figure 4.5d we show the flux propagated using the $\sin(\alpha)$ approach. It is apparent that electron flux is over-estimated compared to observations shown in panel (a). The logarithmic difference shown in Figure 4.5e is much larger than that computed for the Fourier-coefficients-based model. Indeed, while in panel (c) the logarithmic difference was close to 0 throughout the test interval, in panel (e) the values are in general of the order of 0.5. In

order to quantify the improvement in performance by the Fourier expansion compared to $\sin(\alpha)$, we calculated the skill score, given by the formula:

$$SS = 1 - \frac{\sum_{i=1}^{N}(m_i - o_i)^2}{\sum_{i=1}^{N}(b_i - o_i)^2}, \tag{4.8}$$

where m denotes the model values, o stands for observations, and b represents the baseline model output. This metric quantifies the improvement over a baseline model, and is also sometimes referred to as the prediction efficiency (PE) [for details see e.g., Morley et al., 2018; Glauert et al., 2018]. In our case, the skill score equals 93%, meaning that the proposed approximation and modeling of the pitch angle distributions is strongly beneficial compared to the standard simplified approach.

4.4.2 Reconstructing Equatorial PADs From a Single Flux Measurement

In order to reconstruct the equatorial electron pitch angle distributions from a single measurement, be that a uni- or omni-directional one, it is necessary to create a model of the A0 coefficient. Here, we parametrize the A0 in each L-MLT bin (the binning procedure remains the same as before) as a function of the solar wind velocity. The A0 coefficient essentially represents a value of flux in the edge of the loss cone. It should be noted, however, that due to the fact that RBSP did not have a high resolution in the loss cone, the A0 value is not a physical parameter but rather an artificial value from the details of the measurement that averages the zero-intensity loss cone together with the finite intensities at small pitch angles. It has been well established that changes in electron flux are related to the solar wind velocity which is why we use this quantity here instead of the otherwise used pressure [Reeves et al., 2011]. It has also been noted that solar wind velocity is one of the most meaningful predictor variables for empirical flux modeling [e.g., Li et al., 2005; Wing et al., 2016; Sillanpää et al., 2017; Smirnov et al., 2020a; Katsavrias et al., 2022]. For relativistic electrons, Reeves et al. [2011] reported that the distribution of flux versus SW velocity resembled a triangle. Kellerman and Shprits [2012] analyzed the normalized occurrence histograms of electron flux observations at GEO with respect to the solar wind

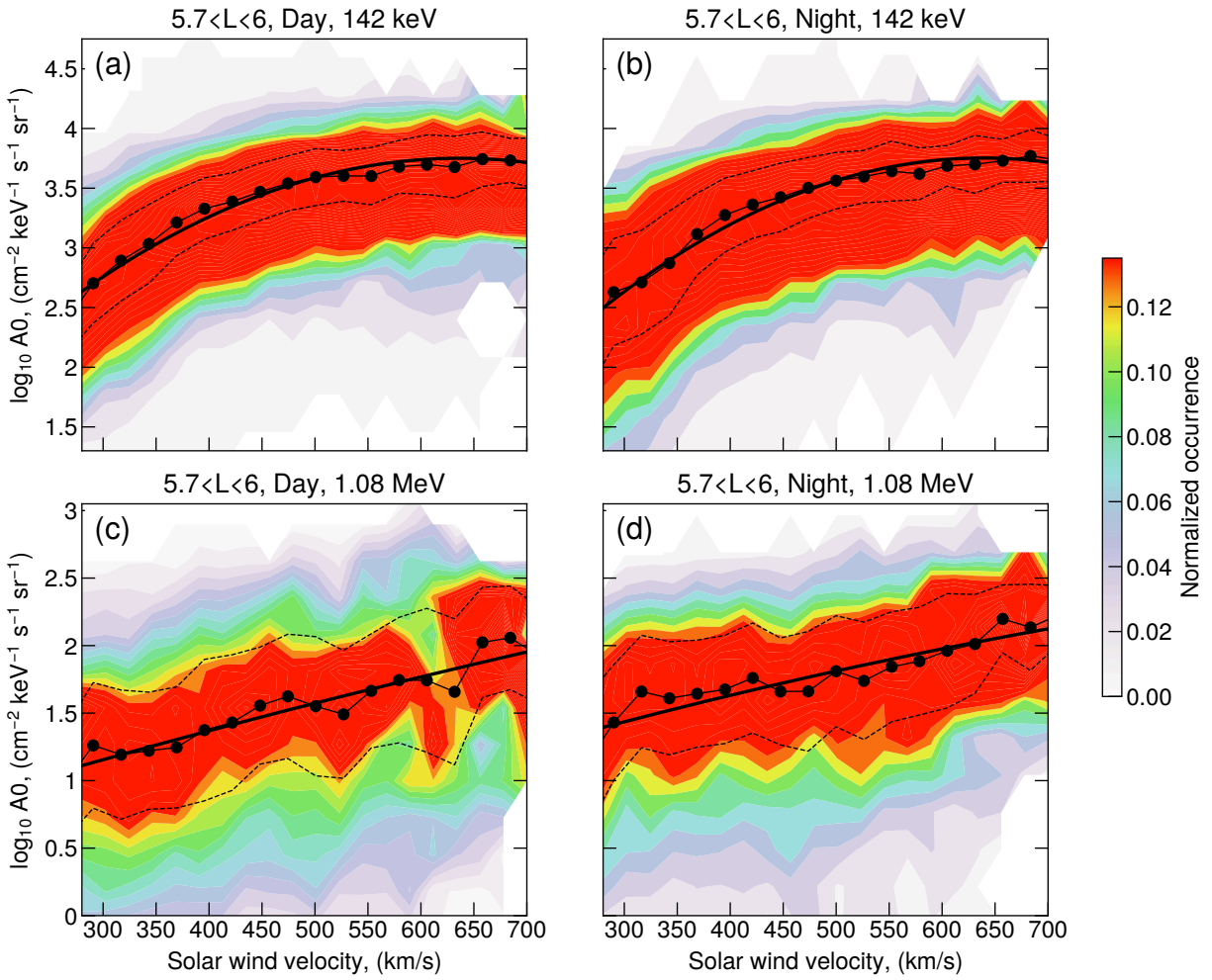

Figure 4.6. Normalized occurrence plots of the A_0 coefficient as a function of solar wind velocity for 142 and 1.08 MeV electrons at day side MLT (a, c) and night side MLT (b, d), respectively. The median A0 values in each velocity bin are shown as black dots, and the dashed lines give the interquartile range. The bold black line shows the quadratic trends fitted to the medians.

velocity and reported a well-observed non-linear trend between them. In the present study, a similar normalization procedure in used, namely, for each bin of solar wind velocity we compute the normalized distribution of electron flux. In Figure 4.6, we show the normalized occurrence histograms of the A0 in logarithmic scale on solar wind velocity for 2 energies (142 keV and 1.08 MeV) and day and night MLT sectors with L> 5.7. One can observe a strong dependence of A0 on Vsw. While for the lower energies this dependence has a quadratic shape, for 1 MeV electrons the trend appears linear.

In case of a single uni-directional measurement, Equations 4.5 and 4.7 can be applied in com-

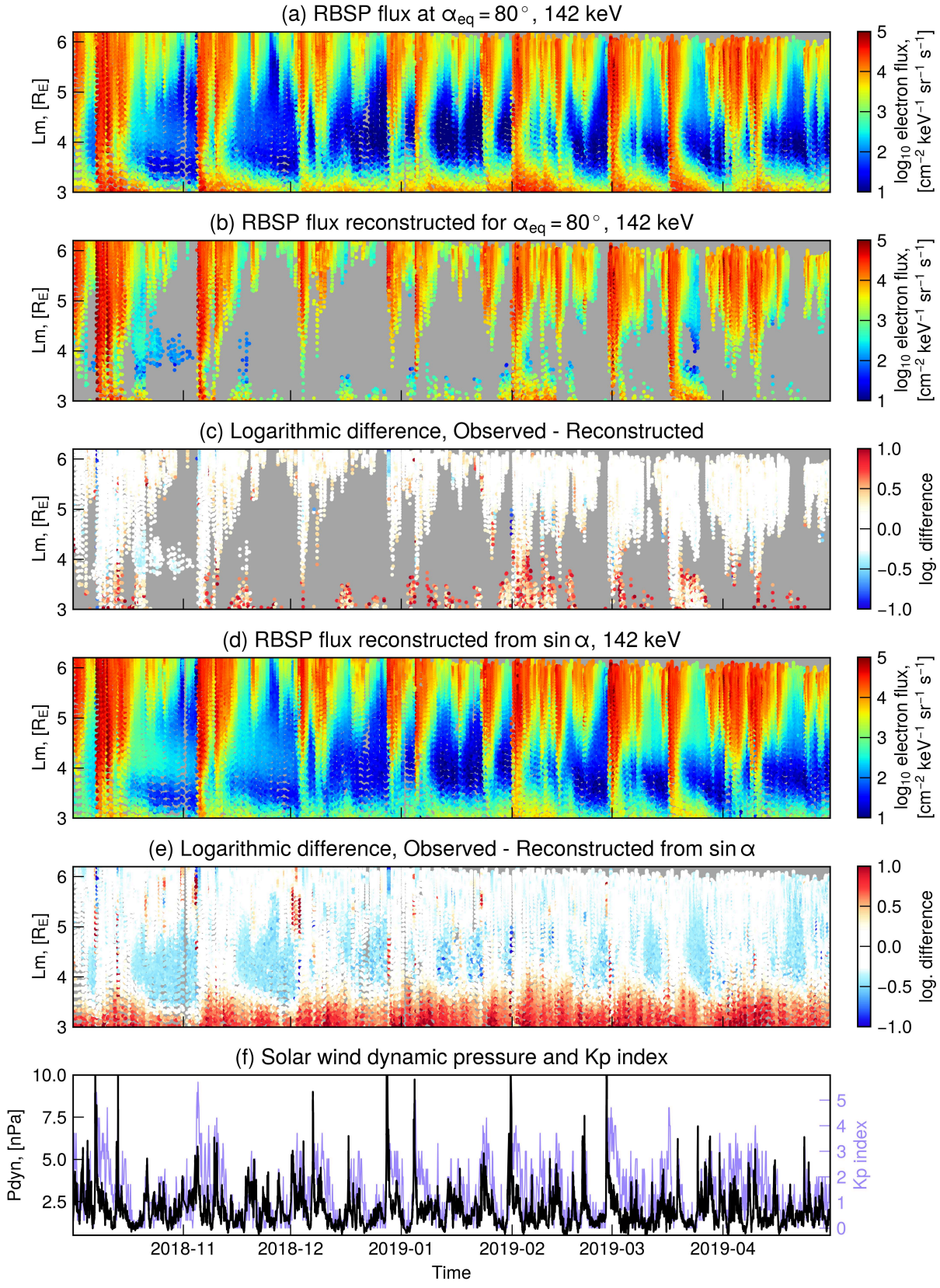

Figure 4.7. Same as Figure 5, but for a single uni-directional observation and the A0 model used for flux reconstruction.

bination with an appropriate model of the A0 coefficient. Furthermore, if the flux measurement is locally omnidirectional, the model PAD can be integrated from 0° up to the maximum equatorial PA, and the scaling factor can be retrieved through the same relation. It should be noted that the A0 model used here is very simple, as it is well known that the solar wind velocity cannot fully explain the variability of flux in the radiation belt region [e.g., Reeves et al., 2011]. For instance, during quiet times the flux intensities can be lower than the averaged A0 value predicted using this simple model. This would cause unphysical PAD shapes. To avoid this, the value of the observed flux at α_0 needs to be higher than A0. The entries where this does not hold are removed from analysis. The reconstructions shown here use the PA value of 20°. The main interest in such formulation lies in retrieving PADs from the omni-directional data for MEO constellations such as GPS. Within the range of L-shells covered by our model (3 to 6), GPS pitch angles vary from 90° to approximately 35-40°, and and the assumption on the selected PA holds most of the time.

Figure 4.7 shows an example of the model reconstruction for the 142 keV electron flux. Similar to Figure 4.5, in Figure 4.7a we show the RBSP flux during the test interval. In pannel (b) the flux reconstructed using the A0 model and the developed pitch angle model is demonstrated. Their logarithmic difference is shown in Figure 4.7c. It is evident that the model performs well in the outer zone, as the difference between model predictions and observations is close to zero. At the same time, a significant portion of the data in the inner zone and slot region is removed during the flux reconstruction. This is likely due to the simplicity of the A0 model we used here. The slot region corresponds to highly complex dynamics, including slot filling events, that are not expected to be captured by a flux model which depends only on solar wind velocity and does not use time-history. It should be noted that at L<4, there is an overestimation of the inner zone fluxes by the model, however it is smaller than for the reconstructions based on $\sin(\alpha)$ function (-0.33 compared to -0.5 in logarithmic scale). Therefore, even though the flux reconstructions based on the A0 model remove most of the points in the slot region, the technique described in this subsection gives good accuracy in the outer belt with an overall skill score of 39% over the standard $\sin(\alpha)$ approximation.

4.5 Discussion

Electron pitch angle distributions are of paramount importance for understanding the physical processes acting in the Earth's radiation belts and ring current. Observations in the vicinity of the geomagnetic equator yield the most complete information about the electron PADs, and the equatorial PADs are also indicative of the processes in other parts of the inner magnetosphere. For example, at an off-equatorial latitude a PAD sampled by a satellite can be retrieved by taking a part of the corresponding equatorial PAD from 0° to the maximum equatorial PA corresponding to the satellite location. The inverse, however, does not hold - it is physically impossible to reconstruct the full equatorial PAD only having observations at high latitudes. However, that information can still be filled in through independent sources as the empirical model that we provide here.

In this study we analyzed electron PADs at energies 30 keV- 1.6 MeV in the outer belt. It has been demonstrated that the PADs can be well approximated by a combination of the first, third and fifth sine harmonics. We create an empirical model of the PAD shapes in terms of the respective coefficients, driven by the solar wind dynamic pressure. We have also examined the probability distributions of the PAD shape coefficients (A1, A3 and A5) with respect to the AE and SYM-H indices, as well as several solar wind parameters (velocity, By and Bz components of the interplanetary magnetic field). It was found that of the considered proxies the dynamic pressure was the best parameter to capture the PAD dynamics with increasing activity, while other parameters exhibited much weaker or irregular trends (see Figure S3 in the Supporting information of Smirnov et al. [2022c]). Furthermore, asymmetries in the terrestrial magnetosphere can to a large extent be attributed to changes in SW pressure, and therefore we used it as a predictor variable for the model. At the same time, the A0 coefficient exhibited a strong dependence on solar wind velocity. This coefficient represents electron flux in the loss cone, and the relation of flux to SW velocity has long been established [see e.g. Reeves et al., 2011]. We note, however, that the quadratic relation of A_0 to v_{sw} introduced here is very simple, and the appropriate time lag between the two, which for energies of > 1 MeV can be in the order of several days, was not investigated.

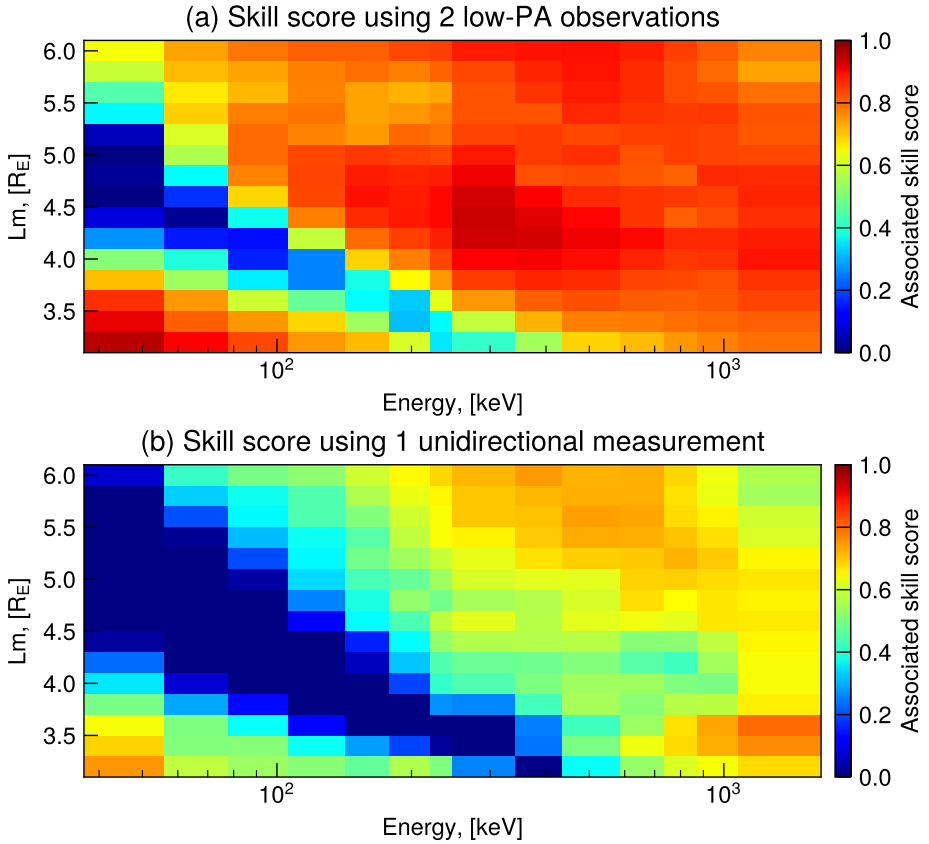

Figure 4.8. Associated skill score (SS) of the 80° equatorial PA electron flux reconstructions using the proposed PAD model over the simple sin(α) approximation using two methods proposed in Sections 4.4.1 and 4.4.2. (a) SS of the reconstructions using 2 observations at low pitch angles; (b) SS of the reconstructions from a single unidirectional flux measurement.

Our PAD model was constructed for the normalized PAD shapes, and in order to reconstruct the equatorial electron flux values it is necessary to employ electron flux observations. We describe two techniques for flux reconstructions. The first method (Section 4.4.1) involves two flux observations at low equatorial pitch angles. Another method, described in Section 4.4.2, can be used when a single flux measurement is available and works for both uni- and omni-directional observations. It was established that the developed model outperforms the standard sine approximation. In Figure 4.8, we demonstrate the skill score of our PAD model over the simple propagation of flux to high PAs using the sin(α) function for both methods, binned by 0.2 L for each of the MagEIS energy channels. The skill score for the first method (involving two low-PA data points) is generally very high (over 90%) at all energies and L-shells, except for the slot

region where the SS values are $\sim$ 40%. Furthermore, as shown in Figure 4.3, the reconstructed flux values are very close to the MagEIS observations. Therefore, the developed model can be a useful tool for highly accurate reconstructions of the equatorial flux from observations at low PAs.

In Figure 4.8b, we demonstrate the skill score for the second method which involves a single flux measurement. It can be seen that at energies $>\sim$ 300 keV the model outperforms the standard approach by > 50% in the outer belt. Furthermore, at lower energies the flux reconstructed using the $\sin(\alpha)$ is underestimated at low L-shells (as shown in Figure 4.7d), and therefore the skill score corresponding to the outer edge of the inner zone is generally above 50%. However, in the slot region, a significant number of points are removed during the flux reconstructions (see also Figure 4.7b) and the corresponding SS values are around zero and sometimes negative. This means that the second method does not lead to any improvement over $\sin(\alpha)$ extrapolation in the slot region. This is due to the fact that this approach relies on the A0 model, which as described above is rather simple and does not carry the full information about the flux variability, even more so in the slot region where the dynamics are very complex. It should be noted, however, that the statistics in the slot region are based on a much smaller number of points, and furthermore, electron flux in the slot region is often at background levels and the performance metrics may not be appropriate in this case. The second approach presented here can be applied to reconstruct the pitch angle distribution of flux in the outer belt at energies $>\sim$ 300 keV, but in order to extend this method to lower energies and L-shells a more precise A0 model needs to be developed in the future. Introducing such a model will likely lead to higher SS values matching those in Figure 4.8a.

4.6 Conclusions

Using the full MagEIS data set of pitch angle resolved electron flux at energies 30 keV - $\sim$1.6 MeV in 2012-2019, we analyze equatorial electron PADs at L=3-6. To fit the pitch angle distributions with a combination of the first, third and fifth sine harmonics. This approximation can fit all PAD types observed in the outer radiation belt, and the coefficients before the three sine

terms relate to the main PAD shapes. Furthermore, the resulting expression can be analytically integrated and the fitted PADs converted to omnidirectional flux.

We present a PAD model with a continuous dependence on L, MLT and activity parametrized by the solar wind dynamic pressure. We demonstrated that the model can be applied to reconstruct the full equatorial PADs from observations at low pitch angles, and therefore can be applied to LEO data. In particular, it can help with using the LEO observations for data assimilation [e.g., Castillo Tibocha et al., 2021], for driving the boundary conditions of the physics-based simulations, and for converting these measurements to phase space density in terms of adiabatic invariants for which the correct PAD shapes are critically important [Chen et al., 2005]. Moreover, the model can be applied to reconstruct the equatorial PAD from a single uni- or omnidirectional measurement. This is of particular interest for the long-term MEO constellations, for instance, the GPS constellation [Morley et al., 2018] and the Cluster mission [e.g., Smirnov et al., 2019b, 2020b; Kronberg et al., 2021], and would add to the availability of the PA-resolved data in the outer radiation belt region. The model can further be extended to higher L-values by using the data from the currently operating Arase constellation, already cross-calibrated with the RBSP [Szabó-Roberts et al., 2021].

Acknowledgements

The model coefficients and the Python script to read them are publicly available [Smirnov et al., 2022b]. The pitch angle resolved MagEIS data are publicly available at `https://rbsp-ect.newmexicoconsortium.org/rbsp_ect.php`. The solar wind and geomagnetic indices were downloaded from the OMNIWeb database (`https://omniweb.gsfc.nasa.gov`). The AE and SYM-H indices used in this paper were provided by the WDC for Geomagnetism, Kyoto (`http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html`). The planetary Kp index of geomagnetic activity was obtained from GFZ Potsdam (`https://www.gfz-potsdam.de/en/kp-index/`). This research is supported by a grant from Deutsche Forschungsgemeinschaft, Germany (SFB 1294, project No. 318763901). Hayley Allison was supported by the Alexander von Humboldt foundation.

— 5 —

Adiabatic invariants calculations for Cluster mission: a long-term product for radiation belts studies *

Abstract

The Cluster mission has produced a large dataset of electron flux measurements in the Earth's magnetosphere since its launch in late 2000. Electron fluxes are measured using RAPID/IES detector as a function of energy, pitch-angle, spacecraft position and time. However, no adiabatic invariants have been calculated for Cluster so far. In this paper we present a step-by-step guide to calculations of adiabatic invariants and conversion of the electron flux to PSD in these coordinates. The electron flux is measured in two RAPID/IES energy channels providing pitch-angle distribution at energies 39.2-50.5 keV and 68.1-94.5 keV in nominal mode since 2004. A fitting method allows to expand the conversion of the differential fluxes to the range from 40 to 150 keV. Best data coverage for PSD in adiabatic invariant coordinates can be obtained for values of second adiabatic invariant, $K \sim 10^2$ $\mathrm{nT}^{1/2}\mathrm{R}_E$, and values of the first adiabatic invariant μ in the range ≈5-20 MeV/G. Furthermore, we describe the production of a new data product `'LSTAR'`, equivalent to the third adiabatic invariant, available through the Cluster Science Archive for years 2001-2018 with 1-minute resolution. The produced dataset adds to the availability of observations in Earth's radiation belts region and can be used for long-term statistical purposes.

*This chapter has been published as Smirnov, A.G., Kronberg, E.A., Daly, P.W., Aseev, N.A., Shprits, Y.Y. and Kellerman, A.C. (2020): Adiabatic invariants calculations for Cluster mission: A long-term product for radiation belts studies. *Journal of Geophysical Research: Space Physics*, 125(2), p.e.2019JA027576.

5.1 Introduction

In 1958, US spacecraft Explorer-1 [e.g., Williams, 1960] was launched into the orbit, carrying onboard the experiment originally intended for measuring the cosmic rays. The measurements were susceptible to the energetic electrons in the regions that later became known as Van Allen radiation belts. There are proton and electron radiation belts. Electron radiation belts exhibit a two-zone structure: the inner belt lies from L=1 to $\sim$ 2.5, the outer belt is located between L-values from $\sim$ 3 to 7 with maximum electron fluxes at L$\sim 4-5$ [Lyons et al., 1972]. Two electron radiation belts are separated with a so-called slot region situated within L-values from $\sim$ 2 to 3 [Baker et al., 2018]. The slot region is formed by a balance between pitch angle scattering loss to the atmosphere and inward radial diffusion [Lyons and Thorne, 1973; Kavanagh et al., 2018] and is usually devoid of energetic electrons. Protons form a single radiation belt with maximum flux intensities between L-values 3 to 4 [Ganushkina et al., 2011].

The radiation belts can pose significant hazard to satellites. Relativistic particles with energies >1 MeV can penetrate through satellite shielding and burn the equipment onboard down [e.g., Koller et al., 2008]. Moreover, particles at smaller energies of $\approx$50-100 keV (a so-called seed population) can accumulate at the surface of the satellite thus generating internal currents that can also lead to satellite loss [e.g., Fennell et al., 2000]. Although most of the spacecraft flying in the radiation belts are affected by the energetic particles, relatively few satellites provide measurements of the radiation belts population [Koller et al., 2008]. Amongst them, one of the longest ongoing missions is the Cluster mission launched in 2000 [Escoubet et al., 1997]. The mission consists of the four identical spacecraft, each carrying onboard the RAPID/IES detector measuring electron fluxes at energies 30-400 keV. It should be noted that the IES detector was not specifically designed for radiation belts studies and as such is subject to background contamination by the high-energy electrons (>400 keV) and inner-zone protons. However, the recently developed background correction techniques proved to efficiently remove the contamination thus making an enormous dataset available for the inner-magnetospheric science [Kronberg et al., 2016; Smirnov et al., 2019b].

The IES detector provides measurements of electron fluxes as a function of energy, pitch

angle, position, and time. However, for many scientific applications, especially those concerning the dynamics of electron radiation belts, it is necessary to investigate the phase space density as a function of the phase space coordinates (also called adiabatic invariants) [Chen et al., 2005; Hartley and Denton, 2014]. In particular, one of the most essential parameters currently used in inner-magnetospheric research is the third adiabatic invariant $L^\star$. The two major uses of the $L^\star$ parameter are the real-time satellite tracking in the $L^\star$ coordinate and mapping between the invariant and phase spaces [Min et al., 2013]. Amongst other applications, it is used in radial diffusion simulations, enables to reconstruct the particles drift and determine the energy exchange in the magnetosphere [Yu et al., 2012]. It is also crucial for the PSD gradient analysis used for studying acceleration and losses of the outer belts electrons [e.g., Green and Kivelson, 2004].

In this paper we present a step-by-step guide of the calculation of the adiabatic invariants and the conversion of electron fluxes to phase space density for Cluster electron data. Among these we describe the details of the new data product developed for the Cluster Science Archive - the `LSTAR`, available with one minute resolution for more than 18 years of the Cluster measurements. The structure of the paper takes the form of three parts, including this introductory section. Section 2 is concerned with data set and the methodology of adiabatic invariants calculations and conversion of electron fluxes to PSD. Section 3 gives the conclusions.

5.2 Methodology

5.2.1 Adiabatic invariants

As defined in classical mechanics, under the changing conditions of motion one can define the quantities, called *adiabatic invariants*, which remain constant when changes are slow [Landau and Lifshits, 1960]. Since the timescales of charged particles motion in the Earth's magnetosphere are much smaller than those of the geomagnetic field change, adiabatic invariants can be introduced in order to understand a variety of physical processes. Charged particles in the inner magnetosphere undergo three types of periodic motion and each type is associated with an

adiabatic invariant. The first type, cyclotron motion around the magnetic field line, conserves the first invariant μ which is expressed as follows:

$$\mu = \frac{p_{\perp}^2}{2m_0 B}, \tag{5.1}$$

where $p_{\perp}^2$ is the electron momentum orthogonal to the magnetic field direction, and m_0 is the electron's rest mass and B is the magnitude of the magnetic field.

This invariant is associated with the magnetic flux through the surface covering the particle's Larmor orbit. In addition, converging of the geomagnetic field toward the poles makes the particles oscillate between two mirror points. This second type of motion is associated with an invariant J. In the general case it includes both parallel particle momentum and magnetic field geometry through the following equation:

$$J = 2\sqrt{2m_0\mu}\int_{s_m}^{s'_m}\sqrt{B_m - B(s)}ds, \tag{5.2}$$

where the integral is taken along the field line between 2 mirror points s_m and s'_m, $B(s)$ is the magnetic field at point s and B_m is the magnetic field at the mirror point. In Equation 5.2, the term outside of the integral is constant. Therefore, one can write another form of the second invariant, depending only on magnetic field and not the particle characteristics, as follows [Chen et al., 2006]:

$$K = \int_{s_m}^{s'_m}\sqrt{B_m - B(s)}ds. \tag{5.3}$$

Furthermore, an increase in distance between the mirror points leads to a decrease in particle momentum parallel to the magnetic field and particle energy, and vice versa, when magnetic mirrors converge the particles are accelerated to higher energies. The third invariant Φ is related to the azimuthal drift around the Earth and represents the conservation of magnetic flux through the closed drift shell of a particle. The third invariant is conserved on timescales of hours in case of weakly changing magnetic field [Chen et al., 2005].

Adiabatic invariants are approximate integrals of motion [Somov, 2012]. In practice it is often more convenient to use some equivalent proxies to estimate the values of the invariants. For instance, when no parallel forces act on a particle, the momentum is conserved along a bounce

path [Green and Kivelson, 2004]. One can therefore define an integral invariant coordinate I, which is a proxy of the second invariant [Roederer and Zhang, 2016]:

$$I = \int_{s_m}^{s'_m} \sqrt{1 - \frac{B(s)}{B_m}} ds, \tag{5.4}$$

I has the units of distance (km or R_E) [Konstantinidis and Sarris, 2015].

The Roederer L parameter, or $L^\star$, is a dimensionless proxy of the third adiabatic invariant and is defined as:

$$L^\star = -\frac{2\pi M}{\Phi R_E}, \tag{5.5}$$

where M is the Earth's dipole magnetic moment. The $L^\star$ parameter shows the radial distance to the equatorial plane where the electron would remain if all external magnetic fields were removed [Green and Kivelson, 2004].

PSD f as a function of adiabatic invariants can be obtained from satellite measurements of particle flux j through the relation:

$$f = \frac{j}{p^2}. \tag{5.6}$$

In order to analyze PSD at fixed values of the adiabatic invariants a number of steps involving interpolation should be performed. The technique used for such conversion applicable to the RAPID/IES electron observations is described in detail below.

5.2.2 Data set

Launched in 2000, the Cluster mission with RAPID/IES detectors onboard has provided a large data set of electron flux observations in Earth's radiation belts and ring current. The Cluster constellation follows the polar elliptical orbit, systematically passing through radiation belts region [Escoubet et al., 1997]. It should be noted that Cluster orbit has been evolving over time due to the lunisolar perturbations. Starting with the initial perigee at around 4 R_E and apogee at 19.6 R_E, the perigee was gradually declining up to 1.2 R_E in 2010 and then rising again up to ≈6 R_E in 2019. Therefore, Cluster has mainly been passing through the outer belt but traveled through

Table 5.1. Lower thresholds of electrons energy channels for RAPID for the omnidirectional flux ('ESPCT6' product) and PA-resolved ('L3DD' product).

Channel	ESPCT6	L3DD
1	39.2	39.2
2	50.5	—
3	68.1	68.1
4	94.5	—
5	127.5	—
6	244.1	—
Upper	406.5	94.5

the inner belt as well in 2007-2013 (for further details on the spatial coverage refer to [Smirnov et al., 2019b]).

Electrons with energies 30-400 keV are measured by the RAPID/IES instrument aboard each of the four satellites. IES is a solid-state silicon detector consisting of three pin-hole systems. Three of these detectors are combined in a configuration providing electron flux observations over a 180° range [Wilken et al., 1997]. The energy channels of the IES detector are given in Table 5.1. Electron distributions are delivered in the nominal telemetry mode (NM, active most of the operation time) and burst mode (BM, allows 4 times more data but only for short periods of several hours per week) [Kronberg et al., 2015]. The omnidirectional electron fluxes are provided in the nominal mode for all 6 energy channels (`'ESPCT6'` product) since the start of measurements, whereas the pitch angle distribution was provided at the beginning only in burst mode (`'E3DD'` product). Starting in April 2004 an additional electron product became available in the nominal mode. This product, called `'PAD_L3DD'`, was added after reprogramming the onboard DPU and provides the electron pitch angle distribution for channels 1 and 3 for nine pitch angles, see details in [Kronberg et al., 2015]. We note that for our purposes the `'E3DD'` product does not have enough temporal coverage, and thus for conversion to PSD we will furtherafter use the `'L3DD'` product.

5.2.3 IRBEM library

A variety of tools has been developed for the adiabatic invariants calculations. Amongst them is the physics-based LANLGeoMag providing $L^\star$ values for the Van Allen Probes constellation in real time [Mauk et al., 2012]. In addition, the LANLGeoMag computes several other magnetospheric parameters by obtaining a numerical solution to the equation of motion with high accuracy [Min et al., 2013]. Koller et al. [2008] used a neural network to predict $L^\star$ values based on the statistical training of the network. In the present paper we use the International Radiation Belt Environment Modeling (IRBEM/ ONERA) library [Boscher et al., 2012], relying on the physics-based principles. The library is fully open access and allows computation of magnetic coordinates and fields as well as geographic and heliospheric coordinate transformations and atmospheric models runs. The IRBEM-lib is written in Fortran but can be called using the IDL and MATLAB interfaces. Recently, the Python interface has been developed allowing to employ the IRBEM-lib functionality through the open access high-level language [Shumko et al., 2018].

Several of the IRBEM functions can be used to compute the adiabatic invariants. Amongst others, there is a `'make_lstar'` function. It allows to calculate magnetic coordinates at any satellite position [Boscher et al., 2012]. In Python interface, the function takes as an input: time in decimal format; specifications of the magnetic field model, output parameters that should be returned and coordinate system; satellite position and the magnetic field input depending on the selected model. We note that since the `'make_lstar'` function does not require pitch angles as input it neglects the slight dependency of $L^\star$ on local pitch angle. In order to fully account for the latter, a `'make_lstar_shell_splitting'` function can be employed. It requires the same input as `'make_lstar'` with addition of local pitch-angles. The outputs include the L and $L^\star$ values, magnetic field strength at the mirror point (B_{mir}) and at position of satellite ($B(s)$), MLT and the I value (see Equation 5.4). Therefore, by calling this function one can obtain the values of the third invariant $L^\star$ as a direct output and also the values of the second adiabatic invariant K through the relation

$$K = I\sqrt{B_m}. \tag{5.7}$$

5.2.4 Conversion of electron fluxes to phase space density

PSD analysis under fixed values of adiabatic invariants has notable advantages over the conventional electron flux studies. In particular, it helps to separate non-adiabatic from adiabatic effects and allows to employ observations from different magnetospheric regions [Chen et al., 2006]. Conversion of the measured electron differential flux $j(E, \alpha, \mathbf{r}, t)$ as a function of energy, pitch angle, position, and time to phase space density $f(\mu, K, L^\star, t)$ as a function of three adiabatic invariants and time is done in the following steps described below. The described calculations were implemented as a set of Python and MATLAB routines and are available in open access at [Smirnov et al., 2019a]. Figure 5.1 gives the schematics of the PSD calculations. We first perform the $L^\star$ calculations, as this parameter is necessary for the background contamination removal.

Preliminary steps

(i) Calculations of $L^\star$ parameter. $L^\star$ coordinate is equivalent to the third adiabatic invariant. The values of $L^\star$ with 1 minute resolution can be obtained through the Cluster Science Archive (`https://csa.esac.esa.int/csa-web/`, `'LSTAR'` product).

For the `'LSTAR'` product, calculations were performed using the IRBEM library (`'make_lstar'` function). As mentioned above, for calculations of L* it is necessary to specify the satellite position, magnetic field model (since the fluxgate magnetometer onboard gives the magnetic field value at only one point) and geomagnetic/magnetospheric conditions. Tsyganenko-89 magnetospheric model, T89, [Tsyganenko, 1989] with Kp-index as an input was employed.

(ii) Background correction. Since the RAPID/IES electron flux measurements are subject to contamination in the radiation belts region, it is necessary to perform a background data correction to avoid erroneous values during the next steps. In the present paper we use the empirical contamination coefficients described in detail in [Kronberg et al., 2016] and [Smirnov et al., 2019b]. Kronberg et al. [2016] derived empirical percentages of contamination for different $L^\star$ values on all IES energy channels. The contamination coefficients for channels 1 and 3

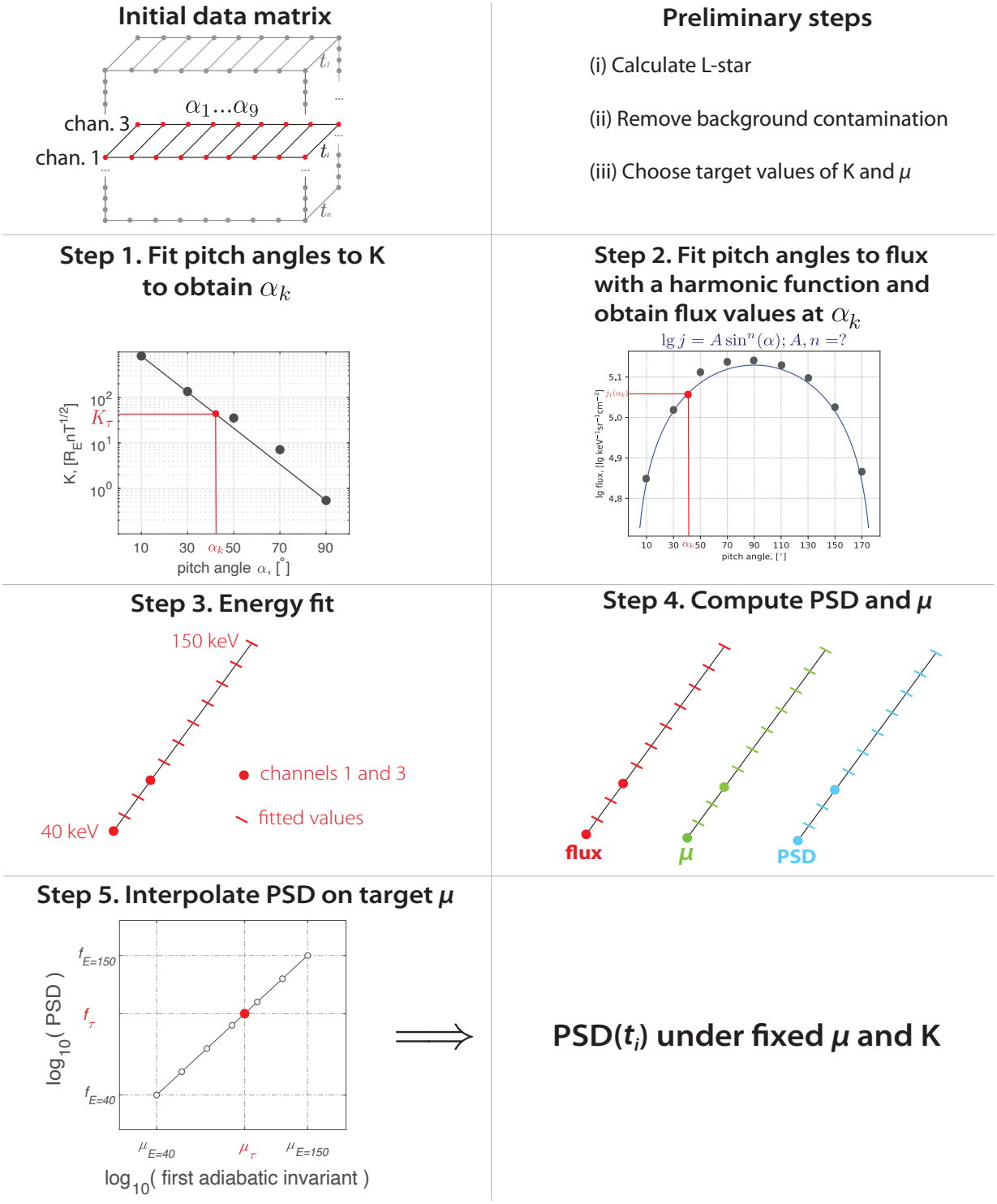

Figure 5.1. Schematics of the electron differential flux conversion to PSD. Initial data matrix contains at each time the flux values at 2 energy channels and 9 pitch angles. After steps 1-5, one value of PSD for the fixed adiabatic invariant values is obtained for each time.

Table 5.2. Total contamination (%) of the IES Detector in the Earth's Radiation Belts [after Kronberg et al., 2016].

Channel	$L^\star = 1$	$L^\star = 2$	$L^\star = 3$	$L^\star = 4$	$L^\star = 5$	$L^\star = 6$	$L^\star = 7$	$L^\star = 8$	$L^\star = 9$
1	2.63	2.87	19.59	31.41	9.27	1.33	0.44	0.26	0.17
3	12.29	12.04	43.18	48.7	20.2	6.15	2.68	1.61	1.06

are shown in Table 5.2. In order to filter the IES electron fluxes, we subtract the part of intensity attributed to the contamination using the following equation:

$$I_{electrons} = I_{measured} * (1 - \frac{p}{100}), \tag{5.8}$$

where $I_{measured}$ is the total measured electron flux intensity, p - percent of contamination from the Table 5.2, $I_{electrons}$ - the filtered electron intensity.

(iii) Calculations of K. K is one of the forms of the second adiabatic invariant. It depends on the configuration of the magnetic field and is independent of the particle's energy, charge and momentum (as can be seen from Equation 5.3). Cluster data resolves 9 pitch angles from 10° to 170° with a 20° step. Since K values are symmetric with respect to 90° pitch angle, we compute K for 5 pitch angles from 10° to 90° via the `'make_lstar_lsplitting'` subroutine implemented in the IRBEM/ONERA library. As input parameters we use local pitch angles at which to compute K, satellite position, and Kp-index as a magnetic input parameter required by the T89 model. The output of the `'make_lstar_lsplitting'` subroutine includes the I values (see Equation 5.4) and magnetic field value at the mirror point. We then obtain K through the Equation 5.7. We note that there are different units commonly used during K calculations, such as $G^{1/2}$km and $nT^{1/2}R_E$ [Hartley and Denton, 2014]. In this paper we use the latter units.

(iv) Choose target values of K and μ. Different values of the first and second adiabatic invariants are related to different physical mechanisms. For instance, for higher K target values the corresponding pitch angle will be smaller and the particle motion will be closer to parallel to the magnetic field. Therefore, one needs to set up the specific values of K and μ before the following steps.

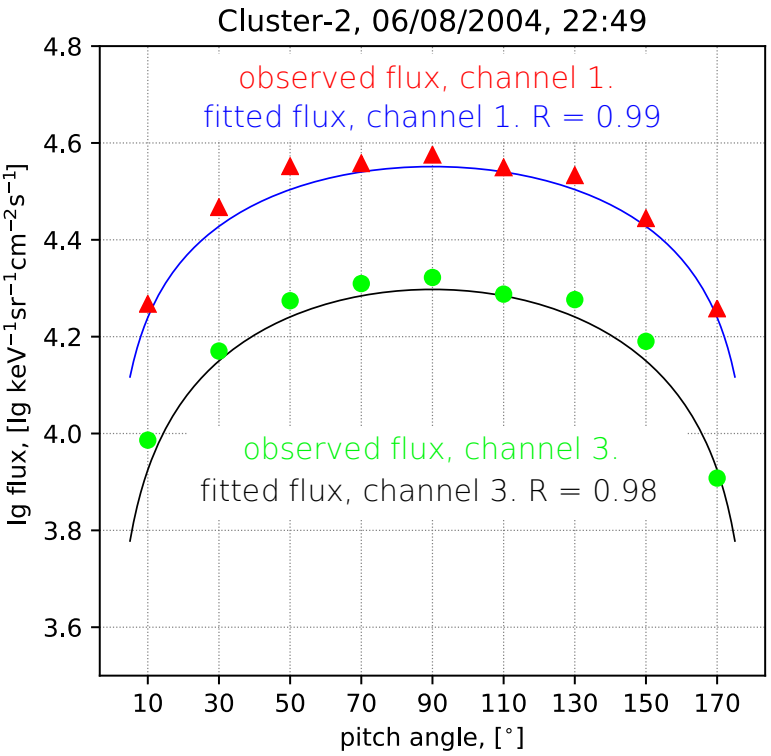

Figure 5.2. Example of pitch angle distribution fit for channels 1 and 3. The fitting is done using equation 5.9 and allows the flux estimation at α_k. The red triangles show the observed flux at different pitch angles and the blue like represents the fitted flux for IES channel 1; the green dots show the observed fluxes at channel 3 and the black line represents the fitted flux for IES channel 3.

Step 1. Fit pitch angles to K to obtain α_k. The second adiabatic invariant K is independent of the momentum and depends only on the satellite position, local pitch angles and magnetic field geometry. To each pitch angle there exists a corresponding value of K. As discussed above, the K values are symmetric with respect to 90° pitch angle and for the fit one can use the pitch angles from 10° to 90°. We perform a linear fit using the `interp` function implemented in the NumPy library [Oliphant et al., 2006]. The illustration of such linear fit is given in Figure 5.1.

Step 2. Fit pitch angles to flux to obtain full pitch angle distribution. In order to retrieve flux values at a specific α_k it is necessary to compute the full pitch angle distribution. This is done by fitting a functional form, relating pitch angles to the observed flux. The standard approach uses the relation:

$$\lg j = A \cdot \sin^n \alpha \tag{5.9}$$

and seeks optimal values of A and n by the least-squares fit. An example of such a fit is given in

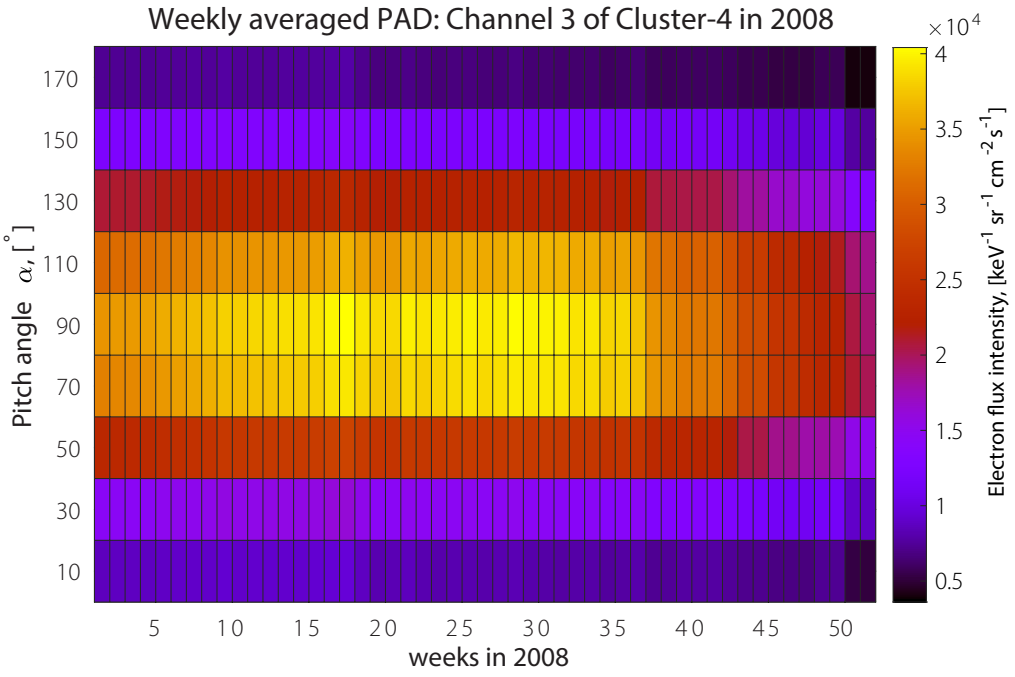

Figure 5.3. Weekly averaged pitch angle distribution (PAD) for channel 3 of Cluster-4 in 2008. The pitch angle distribution has predominantly a pancake shape.

Figure 5.2. It should be mentioned that this method can be used efficiently to fit pancake, field-aligned and flattop distributions but it cannot adequately reproduce butterfly PAD's. However, the electron PAD's observed by the Cluster/IES instrument in radiation belts region predominantly have the pancake shape. An example of the weekly averaged PAD for IES channel 3 is shown in Figure 5.3.

Step 3. Perform the energy fit of differential fluxes using the relation $\hat{j}(E) = a\exp(bE)$. IES measurements are available for channels 1 and 3, therefore we only have two points for such a fit. However, for the IES electron fluxes we adopt the same distributions as in AE9 model [Ginet et al., 2013], where the a and b coefficients remain quai-constant for all $L^\star$ values in energy range ≈40-150 keV. In logarithmic scale it is equivalent to fitting a straight line to the data, which can be done in a unique way using two data points:

$$\log_{10}\hat{j}(E_i,\alpha_k) = \log_{10} j_1(\alpha_k) + \log_{10}\left(\frac{E_i}{E_1}\right)\cdot\frac{\log_{10} j_3(\alpha_k) - \log_{10} j_1(\alpha_k)}{\log_{10} E_3 - \log_{10} E_1}, \tag{5.10}$$

where $\hat{j}$ is the fitted value of flux, E_i is the energy at which we want to estimate the flux, E_1 and E_3 are the effective energies of channels 1 and 3, respectively (all terms are in logarithmic scale)

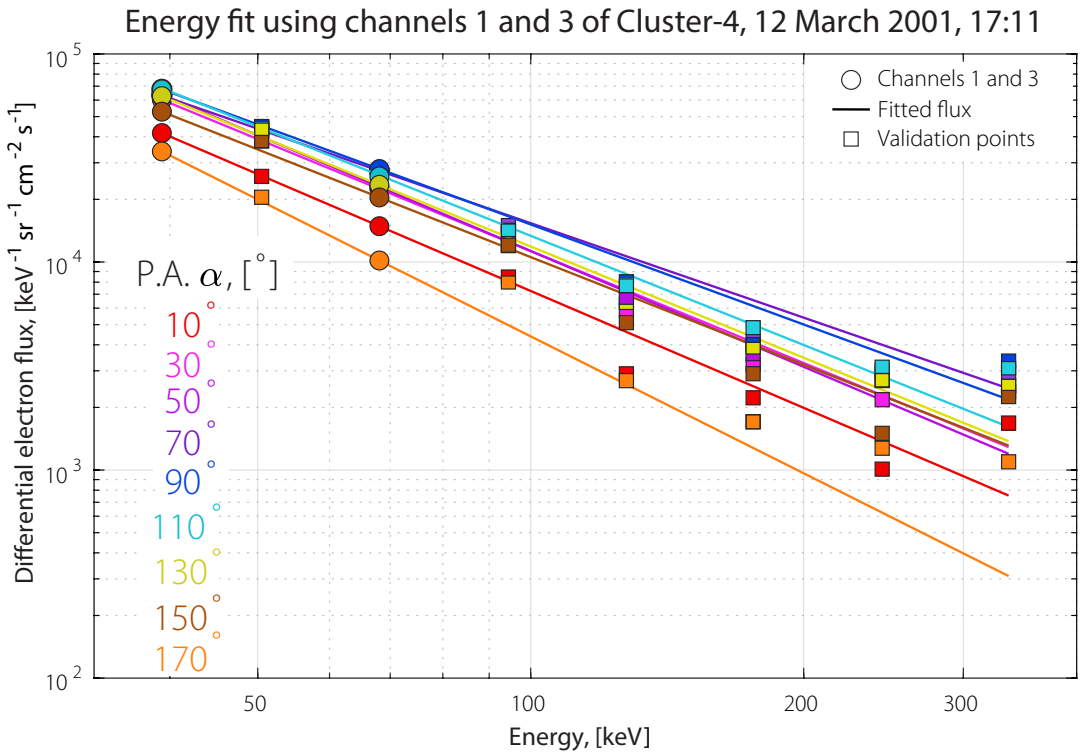

Figure 5.4. Example of the fitting the electron differential flux by energy in double logarithmic scale. Circles show the data points on channels 1 and 3 used for the fit, the solid lines represent the fitted fluxes, and the squares give the IES observations taken from the 'E3DD' product for channels 2 and 4-8. The extrapolation works well up to 5th energy channel (150 keV).

With a step of 10 keV, we obtain the differential flux values using Equation 5.10. The example of such fit is shown in Figure 5.4 based on the full 3D distribution taken from `'E3DD'` product in 2003. From Figure 5.4 it is evident that the proposed fit gives realistic estimates of flux for energies 40-150 keV.

Step 4. Calculate the first invariant μ and PSD values for all energies $E_i = \{40, 50, ..., 150\}$ keV using the following formulas:

$$\mu(E_i, \alpha_k) = \frac{E_i(E_i + 2m_0c^2) \cdot \sin^2 \alpha_k}{B \cdot 2m_0c^2} \cdot 10^5 \left[\frac{MeV}{G}\right], \tag{5.11}$$

where E_i is energy in MeV, m_0c^2 - rest energy of the particle (for electrons equal to 0.511 MeV), B - magnetic field measured by the satellite in nT and α is the pitch angle, in our case equal 90°;

$$f(E_i) = \frac{\hat{j}(E_i)}{E(E + 2m_0c^2)} \cdot 3.32 \times 10^{-8}, \tag{5.12}$$

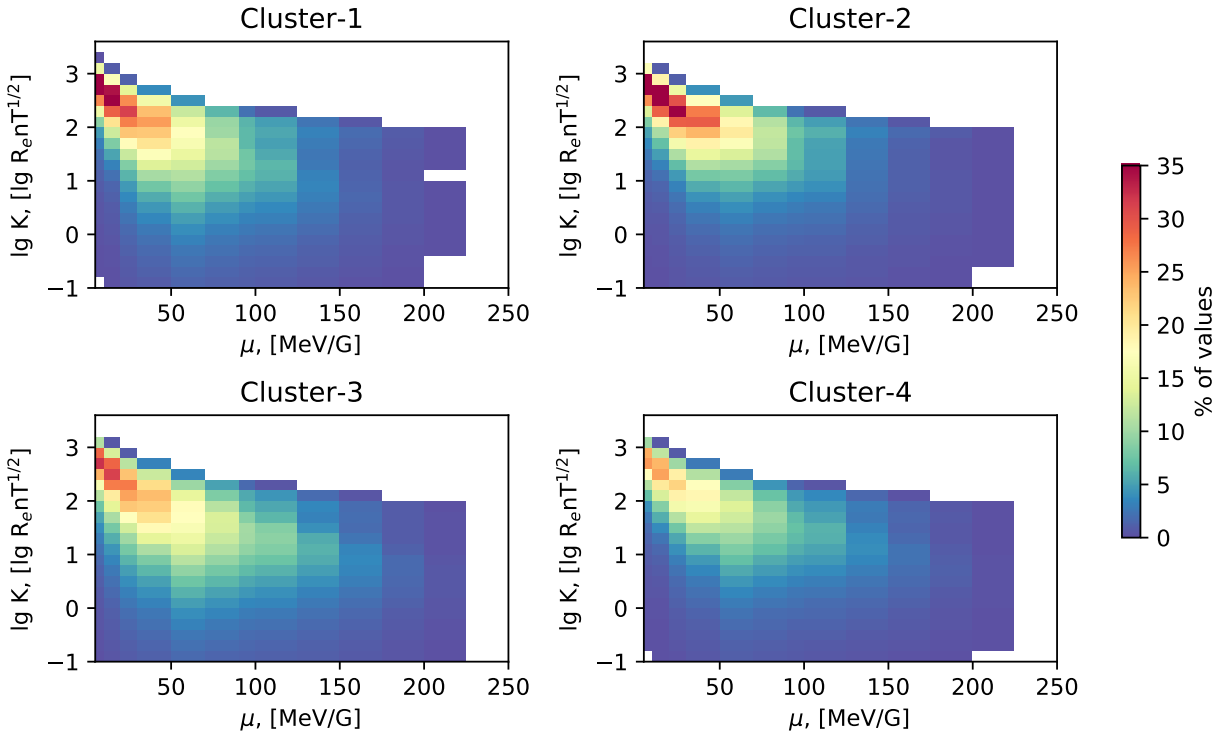

Figure 5.5. Percent of the data points available for different target K and μ values. The most data points are available under $\mu_\tau \approx 5 - 20$ MeV/G and $K_\tau \approx 100 - 1000$ nT$^{1/2}$R$_E$.

where $f(E_i)$ is phase space density in GEM (Geospace Environment Modeling) units (c/MeV/cm)^3 for the corresponding energy, $\hat{j}(E_i)$ is electron differential flux from Equation 5.10 in units cm^{-2}sr^{-1}s^{-1}keV^{-1}. We note that since the Cluster mission has a polar orbit it can travel through the dielectric cusp region, where the physical processes are totally different from those in radiation belts. The L* approximation is based on the T89 model which is computationally efficient but may not account for local field irregularities. In order to avoid data from the cusp region, we put a limit on the magnetic field: it should be higher than 80 nT.

Step 5. Interpolate the PSD values onto the target value of μ. Such interpolation can be performed, for example, using the Python function `numpy.interp` in the following way:

```
PSD_target = numpy.interp(mu_target, mu, PSD),
```

where lengths of `mu` and `PSD` arrays equal the number of E_i. It is a common practice to use logarithmic values of PSD and μ during interpolation. A visual scheme of such interpolation is

given in Figure 5.1. Fitting the differential flux versus energy should be performed in the energy range from 40 to 150 keV.

Thus, one gets the value of PSD $f(\mu, K, L^*, t)$ as a function of adiabatic invariants and time. The percent of data values for different μ_τ in range 5-250 MeV/G and K_τ in range 0.1-1258 $\mathrm{nT}^{1/2}\mathrm{R}_E$ for all Cluster satellites is shown in Figure 5.5. For all four Cluster spacecraft the most data values can be obtained when μ_τ lies in range ≈5-20 MeV/G and for higher values of K (∼ 100-1000 $\mathrm{nT}^{1/2}\mathrm{R}_E$). Examples of PSD merged over all four Cluster spacecraft for the most probable values of K and μ are given in Figure 5.6.

5.3 Discussion

The Cluster mission has been measuring the pitch-angle resolved electron flux in the nominal mode since 2004, thus covering approximately 1.5 solar cycles: the declining phase of solar cycle 23 (2004-2008) and solar cycle 24 (2008-2019). Therefore, the RAPID/IES electron measurements converted to the adiabatic invariants coordinates can be used for long-term statistical analysis of Earth's radiation belts and ring current dynamics. Figure 5.6a,b,c shows the phase space density evolution along the declining phase of solar cycle 23, the historical minimum of solar activity in 2008-2009 and during the following solar cycle 24 for several values of the first and second adiabatic invariants.Figure 5.6d gives the monthly smoothed AE and F10.7 indices.

The F10.7 index is a proxy of the solar emissions from chromosphere and corona, thermal gyroresonance over sunspots and nonthermal emissions [Tapping, 2013]. The F10.7 index follows the general pattern of the solar cycle variation exhibiting peaks during solar cycle maxima. The AE index is calculated as a difference between the upper and lower envelopes of the magnetic variations in H component from 12 geomagnetic observatories located at $61^\circ - 70^\circ$ latitude in the Northern Hemisphere [Kamide and Akasofu, 1983]. The AE index is a proxy of the substorm activity and has peaks during the declining phases of the solar cycles 23 and 24. We use the AE index due to the fact that the Cluster PSD coverage shown in Figure 5.6 is within $L^\star$ from ∼4 to 9, which when mapped to the ground coincide with the latitudes at which the AE index

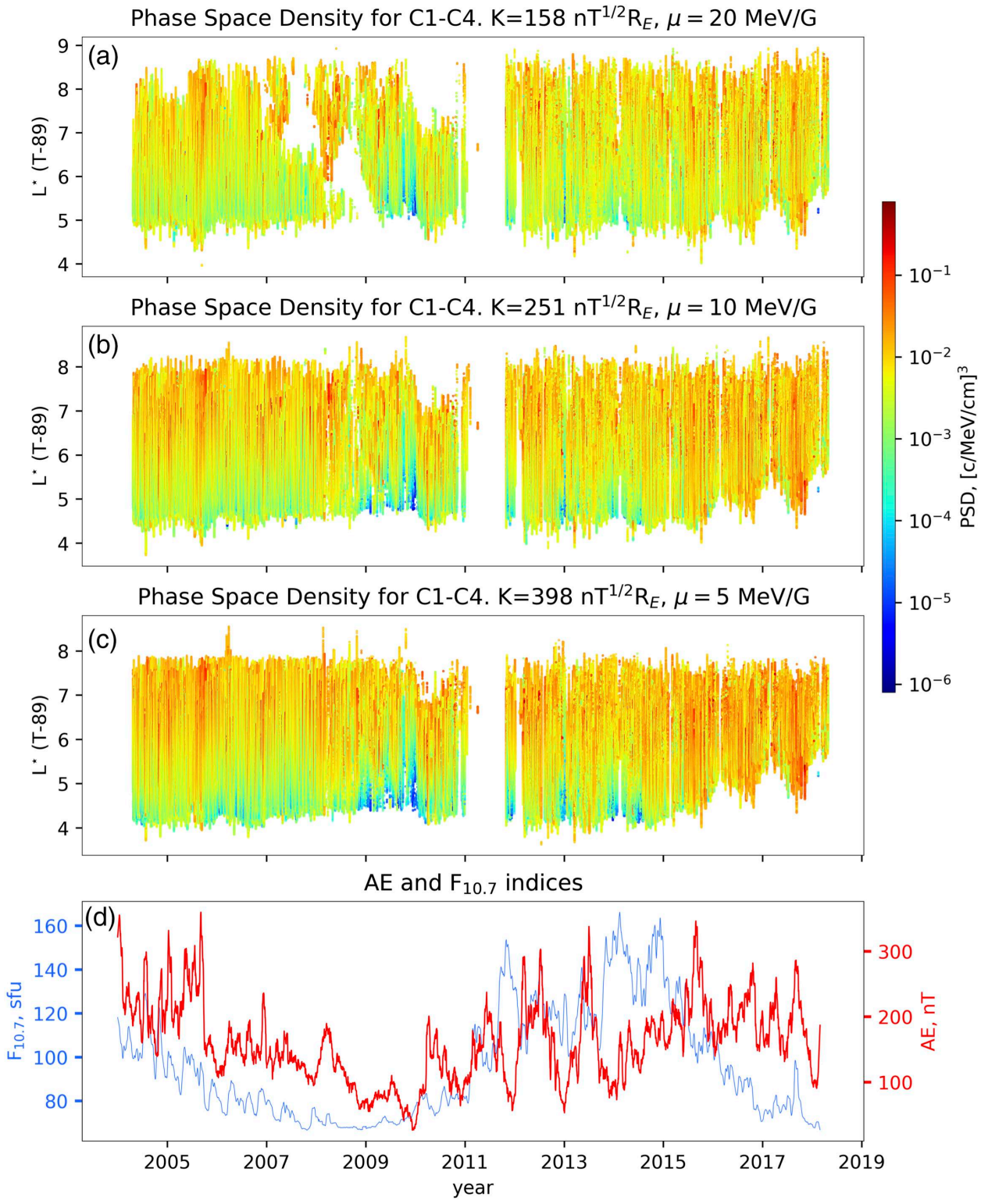

Figure 5.6. Examples of PSD for Cluster-1 - Cluster-4 under the target values of adiabatic invariants (a) $K = 158$ nT$^{1/2}$R$_E$ and $\mu = 20$ MeV/G, (b) $K = 251$ nT$^{1/2}$R$_E$ and $\mu = 10$ MeV/G, (c) $K = 398$ nT$^{1/2}$R$_E$ and $\mu = 5$ MeV/G. Monthly smoothed AE (red) and F10.7 (blue) indices in 2004-2019 are shown in panel (d).

is measured. Furthermore, Smirnov et al. [2019b] showed the correlation between the electron fluxes measured by the IES detector at energies $\sim 40-400$ keV and the AE index and solar wind dynamic pressure, while no significant correlation was found between electron fluxes and the F10.7. This is due to the fact that the F10.7 follows the sunspot cycle while the other parameters follow the occurrence of coronal holes and high speed streams which occur more often during the declining phase causing the radiation belts enhancements.

It has long been established that the substorm activity plays an important role in radiation belts electron transport and acceleration [e.g., Jaynes et al., 2015; Boyd et al., 2016; Zhao et al., 2017]. The source population electrons with energies of tens of keV are injected into the inner magnetosphere during substorms and produce waves (for instance, the whistler mode chorus) [Boyd et al., 2016]. The so-called seed population electrons with energies of up to a few hundred keV are also injected during the substorms and can be accelerated to relativistic energies by wave activity [e.g., Baker et al., 1998]. Zhao et al. [2017] showed that the AL index had good correlation with PSD enhancements for energetic electrons, suggesting the direct transport of low-μ electrons from substorm injections. From Figure 5.6 it is evident that at $L^\star \sim 4-6$ the PSD follows the pattern close to that of the AE index. The PSD values are generally higher during the descending phase of the solar cycle 23 in 2004-2005 and exhibit minimum values during the time of the so-called 'radiation belts desert' in 2009-2010, which also corresponds to the intense dip in the AE index. The largest PSD values at these $L^\star$ can be observed during the descending phase of the solar cycle 24 in 2015-2017. Li et al. [2010] analyzed the global distribution of PSD for low-μ electrons and showed that higher PSD values correspond to the higher AE values (e.g., their figure 9a), which goes well with our findings shown in Figure 5.6. Furthermore, Li et al. [2010] reported that the radial PSD gradient was positive for μ in range 0.05-2 MeV/G suggesting the radial diffusion caused by the magnetospheric activity. Our results are in good agreement with findings of Li et al. [2010], since the PSD values in Figure 5.6 generally increase towards higher $L^\star$.

5.4 Conclusions

Using the step-by-step guide to the adiabatic invariants calculations presented in this paper, the large Cluster dataset covering almost 2 solar cycles can be used for long-term statistical studies in terms of electron phase space density in the corresponding coordinates. Furthermore, since the Cluster constellation consists of 4 spacecraft the data can be useful for PSD gradients analysis.

Using the IRBEM library, the $L^\star$ coordinate was calculated with 1 minute resolution for the Cluster mission in 2001-2018. The created product called `'LSTAR'` is available through the Cluster Science Archive [`https://csa.esac.esa.int/csa-web/`] as one of the "Auxiliary, MAARBLE and ECLAT support data" files. The developed product has numerous scientific applications in the field of radiation belts research, such as, for instance, the more precise satellite tracking in terms of $L^\star$ coordinate compared to the simplified L-shell parameter. Cluster data converted to adiabatic invariants coordinates can be used for physical models validation with data, model set up in the form of initial or boundary conditions (e.g., for the now- and forecasting models), and for data assimilative models. Furthermore, the Cluster dataset can be used to study injections in terms of phase space density. The developed product may be complementary to other data sets and may help more accurately reconstruct the radial profiles of PSD for different energies and locations.

Acknowledgements

When using the `'LSTAR'` product refer to the current paper and Smirnov et al. [2019a]. The authors are grateful to Dr. David Hartley (University of Iowa) and Mykhaylo Shumko (Montana State University) for valuable comments. The Cluster data can be found at CSA Archive: `http://www.cosmos.esa.int/web/csa/`. The data of AE and F10.7 indices were taken from OMNIWeb database [`http://omniweb.gsfc.nasa.gov`].

— 6 —

Intercalibration of the plasma density measurements in Earth's topside ionosphere *

Abstract

Over the last 20 years, a large number of instruments have provided plasma density measurements in Earth's topside ionosphere. In order to utilize all of the collected observations for empirical modeling, it is necessary to ensure that they do not exhibit systematic differences and are adjusted to the same reference frame. In this study, we compare satellite plasma density observations from GRACE, COSMIC, CHAMP, Swarm and C/NOFS missions. Electron densities retrieved from GRACE K-Band Ranging (KBR) system, previously shown to be in excellent agreement with incoherent scatter radar measurements, are used as a reference. We find that COSMIC radio occultation (RO) densities are highly consistent with GRACE-KBR observations showing a mean relative difference of $< 2\%$, and therefore no calibration factors between them are necessary. We utilize the outstanding three-dimensional coverage of the topside ionosphere by the COSMIC mission to perform conjunction analysis with in situ density observations from CHAMP, C/NOFS and Swarm missions. CHAMP measurements are lower than COSMIC by $\sim 11\%$. Swarm densities are generally lower at daytime and higher at nighttime compared to COSMIC. C/NOFS ion densities agree well with COSMIC, with a relative bias of $\sim 7\%$. The resulting cross-calibration factors, derived from the probability distribution functions, help to eliminate the systematic leveling differences between the data sets, and allow using these data jointly in a large number of ionospheric applications.

*This chapter has been published as Smirnov, A., Shprits, Y., Zhelavskaya, I., Luhr, H., Xiong, C., Goss, A., Prol, F.S., Schmidt, M., Hoque, M., Pedatella, N. and Szabo-Roberts, M. (2021): Intercalibration of the plasma density measurements in Earth's topside ionosphere. *JGR Space Physics*, 126(10), p.e2021JA029334.

6.1 Introduction

The ionosphere is an ionized part of the upper atmosphere, spanning from 60 to around 1000 km in altitude [Hargreaves, 1992]. It arises mainly due to the photoionization effects from the solar EUV radiation and charged energetic-particle precipitation [Kivelson and Russell, 1995]. Generally, the ionosphere is strongly coupled with the thermosphere [Astafyeva, 2019]. The latter supplies the neutral particles that can be ionized, and plays a crucial role in the interplay between the production (source) and recombination (loss) processes. The ionosphere affects the propagation of the GNSS signals by introducing frequency-dependent delays. Unlike the neutral atmosphere, which can cause errors in navigation and positioning in the order of several meters, ionospheric effects can yield uncertainties of up to ~100 m [e.g., Petit and Luzum, 2010; Hernández-Pajares et al., 2011]. Ionospheric delays are inversely related to the square of carrier frequency, and directly proportional to electron density integrated along the ray path [e.g., Hobiger and Jakowski, 2017; Goss et al., 2019].

Electron density distribution in the ionosphere strongly depends on altitude and can be divided into several layers, originally identified from ionograms: the D-layer (60-90 km altitude), E-layer (90-130 km), and F-layer (above 130 km), which can be subdivided into F1 and F2 layers [e.g., Astafyeva, 2019]. The dominant contribution to electron density profiles comes from the peak of the F2 layer, generally located between ~250 and 400 km in altitude. The part of the ionosphere located above the F2-peak altitude is referred to as the topside ionosphere; it has a smooth transition into the plasmasphere at approximately 800-1000 km in altitude. Previous studies have indicated that the topside ionosphere and the plasmasphere are a major constituent of the vertical TEC, accounting for up to >60% of the TEC magnitudes [e.g., Yizengaw et al., 2008; Cherniak et al., 2012; Lee et al., 2013; Klimenko et al., 2015]. The F2-peak density (NmF2) and the peak height (hmF2), due to their importance for radio operation, have received a lot of attention in terms of data quality. The existing empirical models, including the IRI [e.g., Bilitza, 2018; Bilitza and Xiong, 2020] and NeQuick [e.g., Nava et al., 2008], give an accurate representation of these parameters. In the topside ionosphere however, the altitude coverage of observations is highly non-uniform. This results in the notable discrepancies between obser-

vations and model predictions above hmF2 [e.g., Kashcheyev and Nava, 2019; Cherniak and Zakharenkova, 2019]. In order to develop high-resolution empirical models of electron density, it is crucial to have dense coverage of inter-calibrated observations in the topside ionosphere.

Ionospheric density profiles have been traditionally monitored through a network of ground-based ionosondes which provide electron density values from ~60km up to the height of the F2 density peak. Observations of the topside densities were at first provided by the ISRs and vertical topside sounders on board several missions (e.g., Alouette) in the 1960-1970s [Benson and Bilitza, 2009]. These observations were, however, very sparse, both temporally and spatially [e.g., Prol et al., 2019]. The traditional ground-based observational techniques have been providing accurate and reliable measurements of ionospheric density and temperature for several decades. Yet, the ionosondes and especially ISRs are only available at a limited number of sites around the globe, and therefore these instruments alone cannot satisfy the increasing demand for high-resolution electron density data, even more so in the topside ionosphere. Over the last 50 years, in situ satellite observations at low Earth orbit, mainly by LPs and retarding potential analyzers (RPAs), have become an important data source. However, these observations are bound to the orbit of their specific spacecraft, and therefore also cannot provide the global three-dimensional coverage of the ionosphere. The GNSS radio occultation represents the only active observational technique to date that allows profiling through the entire F-layer of the ionosphere with global coverage [e.g., Cherniak and Zakharenkova, 2014]. It has been estimated that the precision of the RO observations is $\sim 10^3 \mathrm{cm}^{-3}$ [Schreiner et al., 2007], although the RO profile geometry and assumptions introduced during the density retrieval can lead to an underestimation of the Equatorial Ionisation Anomaly (EIA) crests on the order of $\sim 10^4 - 10^5 \mathrm{cm}^{-3}$ [e.g., Yue et al., 2010b; Liu et al., 2010].

A large number of studies have analyzed the agreement between various plasma density data sets in the Earth's ionosphere. In particular, the RO data from the COSMIC mission, comprising a fleet of 6 satellites, have been validated extensively by ground-based observations. Lei et al. [2007] compared preliminary COSMIC observations of electron density to the ISR data and reported that the two were largely consistent in the topside, although the number of points for the comparison was limited. COSMIC NmF2 observations were found to correspond well

with Arecibo ISR measurements [Kelley et al., 2009]. Similar conclusions were obtained by Cherniak and Zakharenkova [2014] for Kharkov ISR. Chuo et al. [2011] compared the F2-peak parameter observations by the Jicamarca digisonde with those by COSMIC-RO based on data from 2006 to 2008. RO and digisonde NmF2 observations were found to agree well, with a correlation of 94% and near-zero bias. Habarulema and Carelse [2016] performed a comparison of NmF2 and hmF2 between COSMIC and ionosonde data specifically during the geomagnetic-storm times (with Dst< -50 nT). It was found that the precision of COSMIC-RO data did not degrade during geomagnetically active times, compared to the quiet times. Furthermore, several studies performed comparisons of the electron density profiles (EDPs) from RO-devoted missions. For instance, Forsythe et al. [2020] used RO data retrieved from the Spire constellation, comprising 84 satellites, to compare with COSMIC EDPs and reported a close agreement between them. The RO observations by the China Seismo-Electromagnetic Satellite (CSES) were recently found to also be in very good agreement with COSMIC [Wang et al., 2019].

Several studies have performed comparisons between the in situ satellite and ground-based observations. McNamara et al. [2007] compared plasma frequency observations from the CHAllenging Minisatellite Payload (CHAMP) mission and Jicamarca ionosonde. Although generally the two data sets were in good agreement, CHAMP LP densities were on average lower by $\sim 4.6\%$. Recently, Lomidze et al. [2018] compared Swarm Langmuir probe observations to Jicamarca, Arecibo and Millstone Hill ISRs, based on the overhead passes from December 2013 to June 2016. Plasma frequencies measured by Swarm-LP were lower by $\sim 10\%$, which corresponds to a 21% underestimation of electron density. The study showed that application of ISR-derived corrections to Swarm Langmuir probe data improved the agreement with the COSMIC RO measurements. A number of other studies have presented comparisons between radio occultation and Langmuir probe data. Pedatella et al. [2015] compared COSMIC radio occultation densities with in situ measurements by C/NOFS and CHAMP, however, no inter-calibration factors were introduced in that study. Lai et al. [2013] analyzed conjunctions between C/NOFS and COSMIC data and reported a close agreement between the two, although the study was based on a small number of conjunctions occurring on 2 consecutive days.

The aim of the present study is to compare electron density data from several missions collected over the past 20 years and adjust them to the same reference frame. This is a necessary step in order to aid future empirical modeling and data assimilation efforts. Most of the studies mentioned above utilized only a single pair of instruments or observational techniques for comparisons, while in this paper we use data from 5 satellite missions that operate on different observational principles. The ground-based ISR observations, which comprise the golden standard plasma density data set in the topside ionosphere, cannot provide enough spatial and altitude coverage to perform the intercalibration. In this study, we use electron densities retrieved from the GRACE-KBR system as a reference. The KBR densities have been calibrated by incoherent scatter radars at Arecibo, Millstone Hill, Jicamarca and EISCAT and shown to be in excellent agreement with ISR observations [Xiong et al., 2015]. Furthermore, due to their retrieval procedure, KBR densities do not experience quality degradation throughout the mission's lifespan. We compare the COSMIC radio occultation electron densities to the GRACE-KBR data at a large number of conjunctions (> 7100) and find that the two data sets are in very good agreement with a mean relative difference of $\sim$1.6%. Missions providing in situ density observations (e.g., CHAMP, C/NOFS and Swarm) operate at different altitudes, and require a three-dimensional data set to be calibrated against. After establishing that COSMIC concurs with GRACE-KBR, the in situ measurements of plasma density by the CHAMP, C/NOFS and Swarm missions are compared to COSMIC and the cross-calibration factors between them are introduced. This Chapter consists of five parts, including this introductory section. In Section 6.2, we describe the data used in this study. Section 6.3 is concerned with the methodology. Results are presented in Section 6.4. Section 6.5 draws the conclusions and discusses potential applications of the intercalibrated observations.

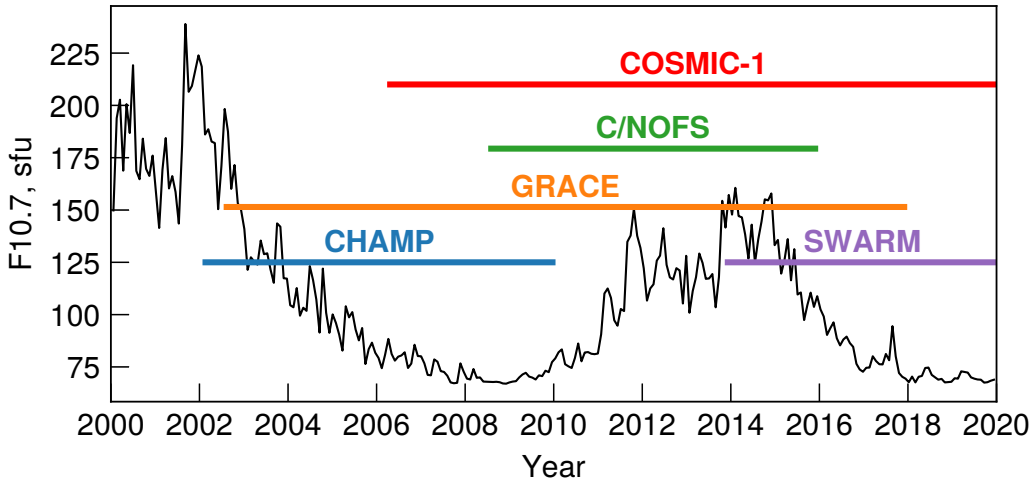

Figure 6.1. Temporal coverage of the missions used in this study.

6.2 Data set

In this Section, we describe the data sources used in the present study. The temporal coverage of the missions with respect to solar cycles 23 and 24 is shown in Figure 6.1, and their altitude range of measurements and horizontal spatial resolutions are specified in Table 6.1. We first describe the GRACE-KBR electron densities, used as a reference for the intercalibration. Then, the radio occultation electron densities from the COSMIC constellation, and in situ observations by CHAMP, Swarm and C/NOFS are described.

6.2.1 GRACE-KBR electron densities

The GRACE mission was launched in March 2002 into a near-circular polar orbit with an inclination of 89° and initial altitude of ~490 km. The mission consisted of two identical spacecraft, GRACE-A and GRACE-B, following each other at a distance of ~200 km. The local time of the mission precessed by 4.5 minutes per day, thus providing coverage of all local times every 160.5 days [e.g., Xiong et al., 2010]. While the primary purpose of the GRACE mission was to construct global high-resolution and time-dependent models of the Earth's gravity field, the satellites were equipped with the KBR system which also allowed the derivation of electron density. The

Table 6.1. Horizontal spatial resolution and altitude range of measurements used in this study.

Mission	Altitude range	Horiz. spatial resolution, km
GRACE-KBR	400-550	~200
COSMIC-RO	150-800	~300
CHAMP/PLP	300-460	~115
CNOFS/CINDI	400-850	3.8
Swarm A/LP	460-490	3.8
Swarm B/LP	490-510	3.8
Swarm C/LP	460-490	3.8

brief description of the electron density reconstruction is given below and the full explanation is provided in Xiong et al. [2010, 2015].

The KBR is one of the core instruments of the GRACE mission; it measures the dual one-way range changes between the two satellites with a precision of 1μm. The level 1B KBR data include an ionospheric correction which can be used for deriving the horizontal total electron content between the two satellites. Furthermore, the position of the two GRACE spacecraft is provided in the GPS Navigation (GNV) data. By dividing the horizontal TEC by distance, the average electron density between the two spacecraft can be retrieved with a resolution of approximately 170-220 km along the ground track [Xiong et al., 2010]. However, it is of note that there is an unknown bias in the ionospheric correction term, which also remains in the derived electron density, but this bias is constant for continuous intervals of GRACE measurements and can be eliminated by using the reference data set. Xiong et al. [2015] used several incoherent scatter radars, namely EISCAT at Tromsø and Svalbard locations, Arecibo, and Millstone Hill to validate the GRACE density measurements. The retrieved KBR electron densities were in excellent agreement with ISR observations, having a correlation of more than 97% and a very low bias of $3 \cdot 10^4$ el./cm^3.

Incoherent scatter radars are one of the oldest traditional instruments for studying the topside ionosphere and have been in use for decades. Due to the near-zero bias of GRACE-KBR den-

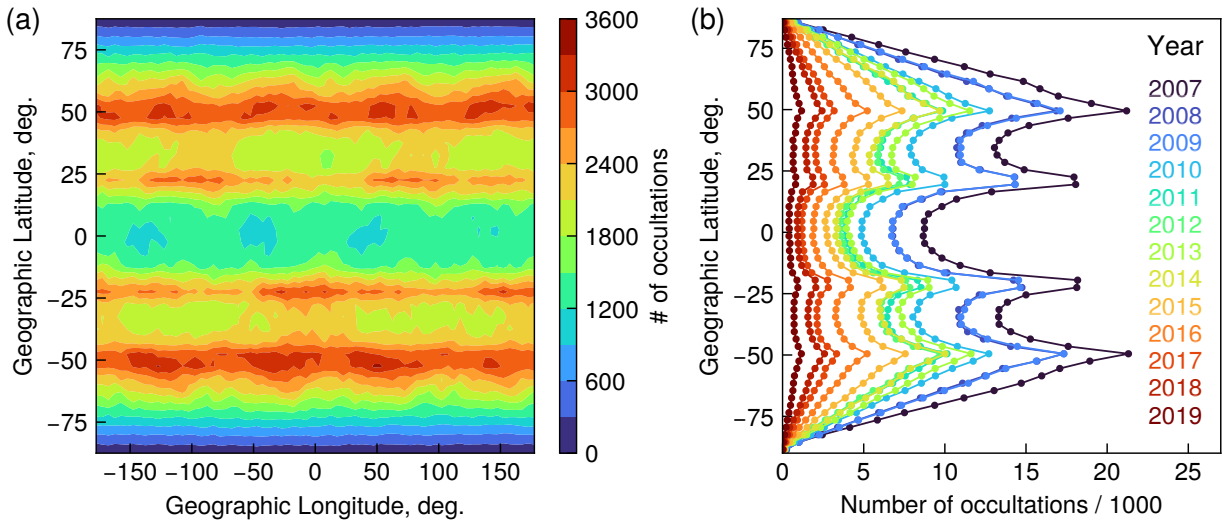

Figure 6.2. (a) Global distribution of the number of occultations observed by the COSMIC-1 mission per $5° \times 5°$ bins in terms of geographic latitude (GLat) and longitude (GLon) for the mission's entire lifespan, with a contour function applied to the plot. (b) Number of COSMIC-1 radio occultations as a function of latitude in 2007-2019.

sity data with respect to the ISR observations, GRACE electron densities comprise a practically calibration-free data set at altitudes 400-500 km, covering ~1.5 solar cycles (2002-2017). Therefore, in the present study GRACE-KBR measurements are used as a reference for intercalibration with other data sources. Using the ISR data alone would not provide enough spatial coverage for the intercalibration, as only a very limited number of the overhead passes can be found for each of the missions. We use GRACE-KBR observations from 2006 to 2015 to compare with COSMIC radio occultation data.

6.2.2 COSMIC radio occultation measurements

The GNSS RO measurements represent a remote sensing technique allowing retrieval of high-resolution vertical profiles of the atmosphere and ionosphere [e.g., Schreiner et al., 2007; Melbourne et al., 1994]. By means of the GPS/MET satellite it was experimentally shown that RO can be used for deriving the vertical electron density profiles [Hajj et al., 1996]. In context with the CHAMP satellite launched a few years later, the RO analysis technique was fully developed for both ionospheric and atmospheric profiling, and the first software package for routine operational evaluation of GPS/RO data was created [Jakowski et al., 2002]. Since then, electron

density profiles retrieved from radio occultations have become a major source of observations used in various ionospheric applications [Pedatella et al., 2015]. The RO missions provide ionospheric profiles from the D-layer up to satellite height and thus give a representation of the three dimensional structure of the ionosphere. Electron density profiles are retrieved using the Abel inversion with several underlying assumptions, including the proportionality between refractivity and electron density, straight-line signal propagation, and spherical symmetry [e.g., Schreiner et al., 2007]. Although the latter assumption can cause systematic errors in the retrieved densities [e.g., Yue et al., 2010b], RO electron densities have been thoroughly validated by ground-based instruments [e.g., Lei et al., 2007; Schreiner et al., 2007; Cherniak and Zakharenkova, 2014].

Multiple spacecraft have supplied the EDP data by means of the RO technique. The largest data source up to date, both in terms of the number of occultations and temporal coverage, is provided by the COSMIC mission. RO-devoted constellations that preceded COSMIC provided much fewer data points, and were estimated to comprise only up to several percent of the COSMIC data set. The COSMIC mission consisted of six microsatellites in 72° inclination orbits. The satellites were launched at the beginning of 2006 with an initial altitude of approximately 500 km, which was increased up to 800 km throughout the following 1.5 years. This created a spatial separation of $\sim$ 30 degrees between the orbital planes [Lei et al., 2007]. Each of the COSMIC satellites carried on board the GPS occultation experiment (GOX) receiver that enables probing the Earth's atmosphere using the RO technique.

One of the notable features of the COSMIC mission is the open loop mode of tracking both the rising and setting occultations, which approximately doubles the number of profiles and thus provides a denser coverage of the ionosphere [Schreiner et al., 2007]. At the beginning of the mission's lifespan, the COSMIC constellation was providing >2500 EDPs per day, while the number gradually reduced to 200-300 profiles by the end of the mission in 2019 [e.g., Wang et al., 2019]. The total number of the COSMIC profiles used in this study exceeds 4.5 million. It is of note that most of the COSMIC occultation events occurred at mid-latitudes, while the equatorial region generally has fewer data points (Figure 6.2, see also Arras et al. [2010]). The follow-up mission COSMIC-2 has been launched in 2019 and mainly focuses on probing the lower equatorial latitudes, providing ~5000 radio occultations per day. The preliminary data

quality analysis [Cherniak et al., 2021; Schreiner et al., 2020] has already demonstrated promising results for ionospheric monitoring, and using these data for the intercalibration can be a topic for further studies.

As described in Section 6.1, the traditional ground-based observational techniques suffer from limited data coverage, both in terms of their location on the globe and in altitude. The electron density profiles retrieved from radio occultations are deprived of these limitations and provide a global three-dimensional data set of electron densities. Furthermore, in order to set up empirical models based on the large scale statistics, it is essential to include the RO measurements for the topside ionosphere, where they would constitute the main and largest data source. In our study we first compare COSMIC electron densities with the selected reference data set (GRACE-KBR). In Section 6.4 it is shown that the two data sets agree very well, and therefore other data sources are compared to COSMIC.

The COSMIC RO data were obtained from University Corporation for Atmospheric Research (UCAR) through the COSMIC Data Analysis and Archival Center (CDAAC). In this study we use the level 2 EDPs, provided through the "IonPrf" product (`https://cdaac-www.cosmic.ucar.edu/`).

6.2.3 In situ plasma density measurements

In this study we consider in situ observations by Langmuir probes and retarding potential analyzers. In general, a Langmuir probe is an electrode, either of planar, cylindrical, or spherical shape, which is extended into the plasma [Hargreaves, 1992]. By applying a variable voltage (V) to the probe, the corresponding current (I) between the probe and the spacecraft is measured. Plasma parameters are retrieved from the current-voltage, or I-V characteristics [Knudsen et al., 2017]. It should be noted that the spacecraft are subject to the sheath effect, which alters the properties of the plasma within the Debye length, and therefore the mounting posts should provide enough spatial separation between the probe and the spacecraft. The RPA is a modification of the Langmuir probe, in which one or more grids, biased at different potentials, are mounted before the

collecting electrode in order to exclude electrons of certain energies from reaching the collectors [Hargreaves, 1992].

The CHAMP mission was launched in July 2000 into near-polar orbit with an inclination of 87.25° [e.g., Reigber et al., 2004; Rother et al., 2010]. The initial orbit altitude was around ~460 km, and slowly decayed to <300 km at the end of the mission's lifespan in 2010. The orbital plane precessed by 1 hour of local time in approximately 11 days, thus covering all local times in roughly 130 days when combining the ascending and descending orbital arcs [e.g., Rother et al., 2010]. Among the instrumental payload aboard the CHAMP satellite, there was a Digital Ion Drift Meter (DIDM) suite, consisting of an ion drift meter (DM) and a Planar Langmuir Probe (PLP). While the DM experienced damage during the satellite ascent and also the subsequent degradation, the PLP instrument provided measurements of electron/total ion density, electron temperature and spacecraft potential throughout the mission's entire lifespan [e.g., McNamara et al., 2007]. The PLP was mounted on the lower front panel of the spacecraft pointing in the ram direction [Reigber et al., 2004]. The CHAMP-PLP represented a golden rectangular plate with a 106 × 156 mm sensing area. The instrument was operating in a voltage sweep mode, taking measurements every 15s, of which the spacecraft potential was tracked for 14s and then the voltage was swept for another second to obtain the electron density and temperature [McNamara et al., 2007]. CHAMP's orbital configuration and velocity of ~7.6 km/s correspond to the horizontal density resolution of approximately 115 km [Rother et al., 2010] (see also Table 6.1). The PLP design and operation are described in more detail in McNamara et al. [2007] and Rother et al. [2010].

The Swarm constellation [Friis-Christensen et al., 2006], launched on 22 November 2013, consists of three identical spacecraft following a near-polar circular orbit. During the commissioning phase, the three satellites Alpha, Bravo and Charlie (also referred to as A, B, and C, respectively), were flying in a configuration following one another at a similar altitude of ~490 km. By April 2014, Swarm A and C satellites were lowered to 460 km, while probe B was raised up to 510 km in altitude. The primary objective of the Swarm mission is to provide highly accurate measurements of the geomagnetic field. The three satellites are also equipped with Langmuir probes in order to take the magnetic field perturbations arising from the diamagnetic effect into

account [Friis-Christensen et al., 2006]. Swarm LP represent spheres of 4 mm radius mounted on 8 cm posts at the bottom front side of the satellites. Each of the satellites is equipped with 2 probes - a high-gain nitrated-titanium probe used for electron density estimation, and a low-gain golden probe providing measurements of the spacecraft potential [Knudsen et al., 2017]. It has been noted that the Langmuir probes configuration on Swarm is rather unconventional, both in their design and usage of the so-called "harmonic" mode, where the voltage changes harmonically at a nominal frequency of 128 Hz. In particular, the length of the LP posts is only a few centimeters, and therefore the Langmuir probes might remain within the spacecraft's Debye sheath, and therefore will not give an accurate representation of density under certain conditions. The ion density data used in this study (version '0502' of the `EFIx_LP_1B` product) are measured with the 2 Hz sampling rate, which corresponds to the spatial resolution of ~3.8 km (Table 6.1).

The C/NOFS satellite was launched into orbit in early 2008. The spacecraft followed the elliptical low Earth orbit, with inclination of approximately 13° [de La Beaujardière et al., 2004]. The satellite altitude was between 400 and 850 km, and its orbit covered all local times. Among the scientific payload, the Coupled Ion Neutral Dynamics Investigation (CINDI) suite, operated by NASA, contained the RPA and an ion drift meter and measured the electron temperature, drift velocities, plasma composition and the ion number density in the topside ionosphere [Heelis et al., 2009]. The satellite velocity was ~ 7.5 km/s [Costa et al., 2014], and the sampling rate of the CINDI/RPA instrument was 2 Hz [Coley et al., 2010], which is equivalent to the spatial resolution of around 3.8 km (Table 6.1).

6.3 Methodology

The intercalibration presented in this study is conducted in two stages. First, we select the GRACE-KBR electron density data as a reference and evaluate whether the COSMIC radio occultation observations agree with GRACE at a number of conjunctions along the GRACE orbit. In Section 6.4, we will demonstrate that the two data sets agree well, with a bias on the order of <2% and a correlation of 96%. At the second stage, it is necessary to compare other

measurements against the reference. Since COSMIC is the only data set capable of providing the three dimensional coverage of the topside ionosphere and also presents a significant interest for empirical modeling, the available in situ observations are compared with COSMIC.

In order to perform the above-mentioned comparisons, it is necessary to impose spatial and temporal conjunction criteria. Several studies have performed comparisons of electron density from radio occultation and in situ observations using different coincidence criteria. Lei et al. [2007] used measurements within 6° geographic latitude (GLat) and 6° geographic longitude (GLon) for comparison of COSMIC densities with ISRs at Millstone Hill, and within $3° \times 9°$ for comparisons with Jicamarca ISR. Pedatella et al. [2015] employed observations within ±2° GLat and GLon and 15 minutes universal time (UT). Shim et al. [2008] estimated the meridional and zonal correlations lengths. It was found that the correlation lengths in latitude were on the order of 2° to 5°, while in longitude the values were much larger, up to $10° - 23°$, based on local time and latitude [Shim et al., 2008; Wang et al., 2019]. The appropriate conjunctions should be within the correlation distances [Wang et al., 2019]. In the present study, we select the following conjunction criteria. We select observations coming from $\pm 1.25° \times 2.5°$ GLat and Glon and ±7.5 minutes universal time, to ensure that the observations in question are close in location and time, within distances not affected by significant horizontal ionospheric gradients. We further remove points coming from geomagnetically active times (i.e., Kp>3) to avoid the storm-time disturbances, although it is of note that COSMIC electron density profiles have been shown to provide high-quality observations during active times as well as quiet times [Habarulema and Carelse, 2016]. The unified conjunction criteria are used for all of the comparisons.

Each of the selected data sources provides daily files with electron density measurements, except for COSMIC which provides one file per individual occultation event (i.e., ~2000 files per day at the beginning of the mission and ~200 files in 2019, see Figure 6.2b). Therefore, the data analysis procedure is as follows. The orbit height corresponding to the in situ measurements is interpolated onto the times of RO events. After that, we linearly interpolate the COSMIC density, as well as position in geographic latitude and longitude, onto the derived altitude. The geographic latitude and longitude corresponding to COSMIC events are then compared to orbital

tracks of another satellite in order to check whether a conjunction occurs. If an event meets the conjunction criteria, it is added to the resulting data frame.

The scatter plots of electron density, shown in the Figures 6.3a, 6.5-6.7a, give information about the approximate data distribution and individual conjunctions. However, it is also important to evaluate how the data sets are distributed with respect to each other by means of the probability distribution functions. Therefore, in Figures 6.3b, 6.5-6.7b we show the normalized occurrence plots. We divide the x-axis into a number of intervals, and for each column the resulting conjunctions are also divided into the same number of bins in y-direction. Then, the number of occurrences in each bin is counted and divided by the total number of points in the corresponding interval on the x-axis. Therefore, the probabilities along each bin in x-direction sum up to 1. We then introduce linear fits to the maxima of these normalized occurrences, which allows for the more correct trend estimation in the presence of outliers. Since for the linear regression it is assumed that any error present in the data set lies exclusively in the y-values, when using regression for calibration purposes the fits are preformed in reverse (i.e., the reference data set is on the x-axis and the data being calibrated - on the y-axis) [e.g., Moosavi and Ghassabian, 2018]. When the linear relationship between the variables is formulated as $y = ax + b$, the final expression for the calibrated data takes the form:

$$y_{\text{calibrated}} = cy + d, \tag{6.1}$$

where $c = 1/a$ and $d = -b/a$. Table 6.3 gives c and d values for each of the missions allowing to calibrate the data to the reference using Equation 6.1.

6.4 Results

6.4.1 Comparison of GRACE-KBR and COSMIC-RO electron densities

Figure 6.3 shows a comparison between the COSMIC-RO and GRACE-KBR electron densities. Using the conjunction criteria described in Section 6.3, we find more than 7100 quiet-time con-

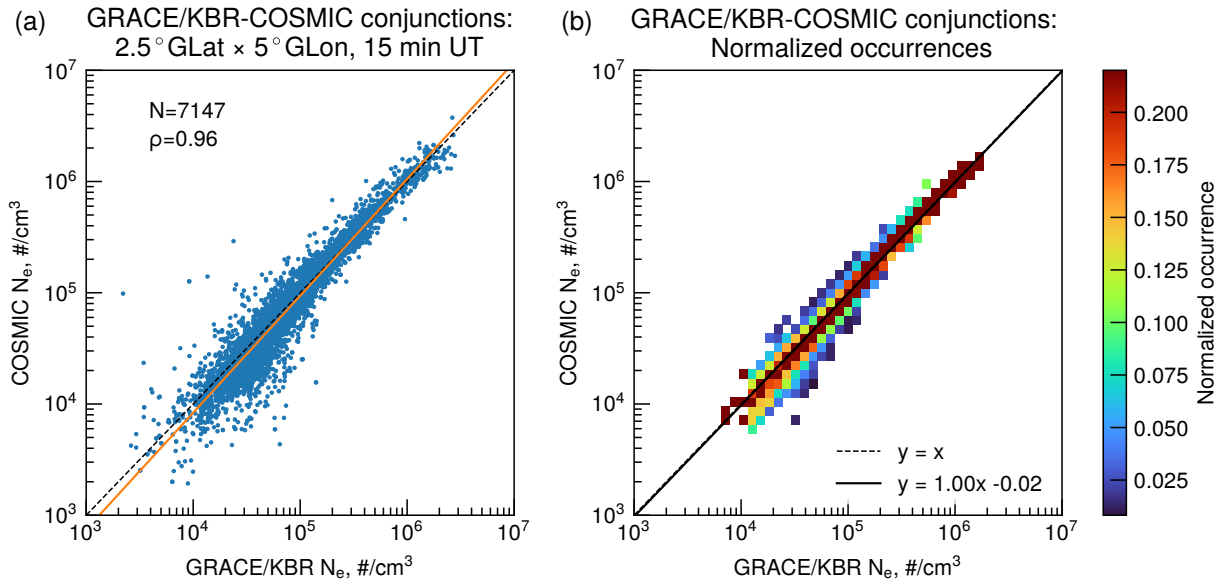

Figure 6.3. Comparison between COSMIC-RO and GRACE-KBR electron densities. (a) Scatter plot of GRACE-KBR versus COSMIC-RO electron densities. The black dashed line shows the one-to-one correspondence between the two data sets, and the orange-colored solid line represents the linear trend, fitted to the scatter plot. (b) Normalized occurrence plot of GRACE-KBR versus COSMIC-RO densities. In each bin in the abscissa, the maximum of the probability distribution function was selected and the linear equation was fitted to these maxima. The resulting trend is shown as a black solid line, and the equation is shown at the bottom right of the panel, with x and y representing the logarithms of GRACE and COSMIC densities, respectively. The black dashed line shows the one-to-one correspondence between two data sets.

junctions between the GRACE orbital tracks and COSMIC tangential lines. From Figure 6.3a it is apparent that the two data sets are in very close agreement. The dashed black line shows the one-to-one correspondence between the data sets, and the majority of points are clustered along the line. The orange-colored line gives the linear fit, performed in double-logarithmic scale. The fitted trend is generally close to the one-to-one line, although at low densities ($< 3 \cdot 10^4$el./cm^3) several outliers are present, presumably due to the plasma bubbles occurring at night side and small-scale ionospheric irregularities not resolved by COSMIC-RO. In order to remove those effects from the linear trend, in Figure 6.3b we show the normalized occurrence histogram. In double-logarithmic scale, the data are divided into 50 bins in x- and y-directions, and bins with less than 5 points are removed. Then, the number of points in each bin is divided by the total number of occurrences in the corresponding interval along the x-axis. By looking at the probability in each bin of GRACE-KBR electron density, one can examine how the corresponding

Table 6.2. Metrics evaluated for the conjunctions between satellites.

Comparison	Mean bias, el./cm^3	Mean bias, %	Median bias, el./cm^3	Median bias, %	St. deviation, el./cm^3	Standard deviation, %	Spearman correlation, %
GRACE/KBR - COSMIC/RO	-1813	-2	2703	3	63495	43	96
COSMIC/RO - CHAMP/PLP	-17754	-10	-14055	-12	51151	29	96
COSMIC/RO - CNOFS/CINDI	-16021	-6	-5458	-7	127468	48	97
COSMIC/RO - Swarm A/LP	-29587	-14	-11915	-11	82727	45	93
COSMIC/RO - Swarm B/LP	-11303	-8	-4046	-5	82762	55	86
COSMIC/RO - Swarm C/LP	-32305	-15	-14142	-13	92376	43	93

COSMIC-RO observations are distributed. Therefore, Figure 6.3b essentially represents a 2D probability distribution function of the conjunctions. Such a representation also helps to avoid overplotting when the number of conjunction is sufficiently large. It can be seen that the largest normalized occurrence values are clustered along the one-to-one correspondence line. The linear trend equation is fitted to the normalized occurrence histogram as follows. For each interval on the x-axis (GRACE-KBR densities), we find the maximum of the probability distribution; then, the orthogonal-distance least-squares fit is performed based on the selected probability distribution maxima. The resulting trend is shown as a solid black line. It can be seen that generally, the trends based on scatter plot and 2D probability distribution are quite similar, although the latter method is more stable due to the outlier removal.

The agreement between the two data sets needs to be evaluated by several metrics. For all comparisons in this study, we use the mean and median bias, both in units of density and normalized, standard deviation and Spearman rank correlation (ρ). The values of these metrics for all comparisons are given in Table 6.2, and the metrics definitions are given in the Supporting information of Smirnov et al. [2021]. In case of COSMIC-GRACE comparison, we find that the mean value of difference between the two data sets (i.e., the mean bias) equals -1813 el./cm^3, and a median bias of 2704 el./cm^3. The mean and median biases normalized are 1.6% and 3%, respectively. The two data sets are in very good agreement and their relative differences are within precision of observations, which for COSMIC-RO was estimated to be in the order of 10^3 el./cm^3 [e.g., Schreiner et al., 2007]. Another metric evaluated for the conjunction analysis is the correlation between the data sets. It has been demonstrated that the linear Pearson correlation

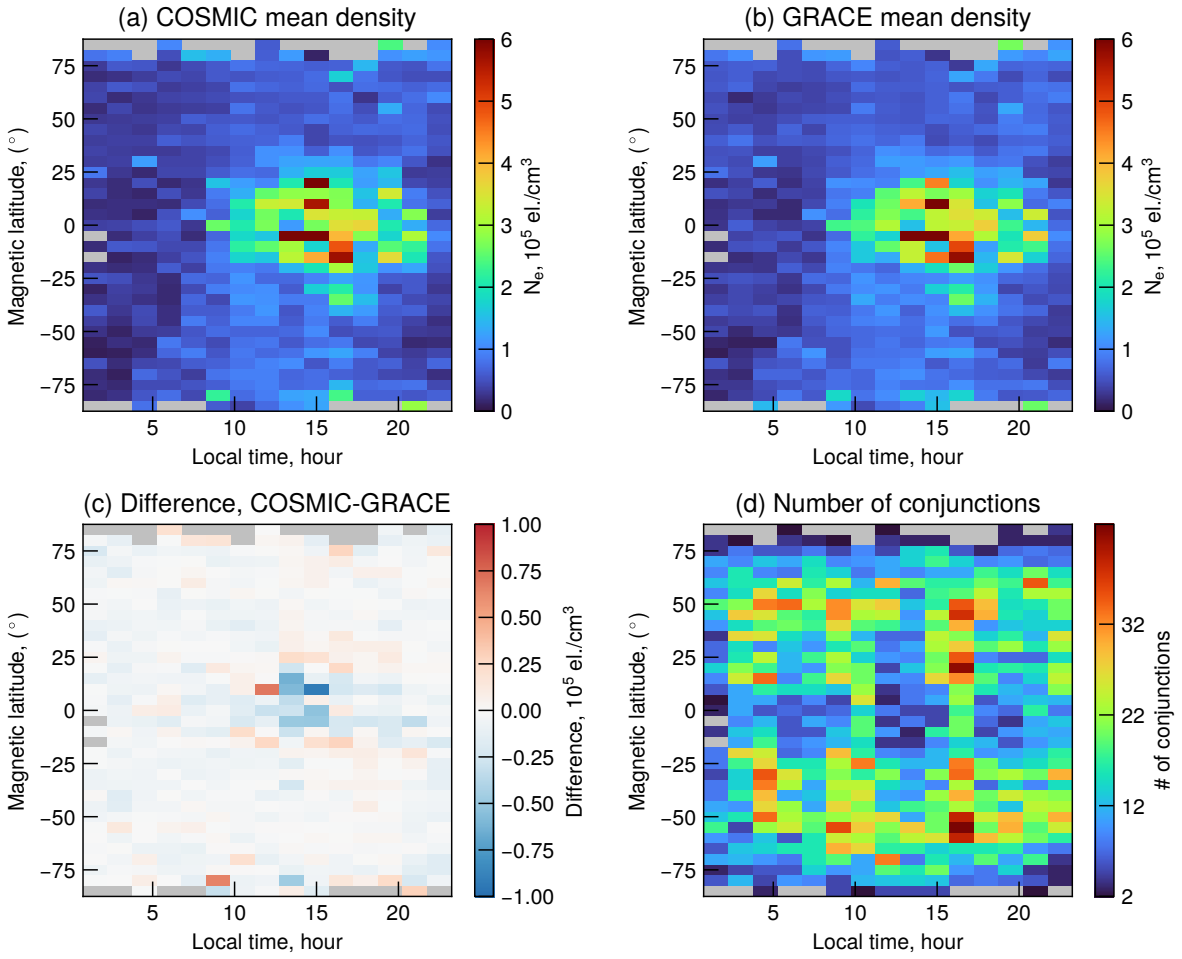

Figure 6.4. Comparison of COSMIC-RO and GRACE-KBR electron densities at conjunction points, binned by magnetic latitude and local time. (a) Mean electron density observed by COSMIC-RO. (b) Mean electron density observed by GRACE-KBR. (c) Mean difference between COSMIC-RO and GRACE-KBR, and (d) number of conjunctions per Mlat-LT bin. In all subplots, the grey background color corresponds to bins with < 2 conjunctions.

can be affected by data noise, whereas the Spearman rank correlation is a more robust metric in the presence of outliers [e.g., Smirnov et al., 2020a]. The value of the Spearman correlation for GRACE-COSMIC comparison is high (0.96), also illustrating that the two data sets closely agree with each other.

Figure 6.4 demonstrates the COSMIC-RO and GRACE-KBR electron densities at conjunctions and their difference, binned by magnetic latitude and local time. In order to obtain a sufficiently high number of collocations in each MLat-LT interval, we bin the data by 5°Mlat and

1.5hr LT. Figure 6.4 shows the corresponding distribution of conjunctions. In general, conjunctions were more frequent at middle latitudes, and fewer conjunctions were observed around the geomagnetic equator. This effect comes from the spatial distribution of the COSMIC measurements, illustrated in Figure 6.2. The bins with less than 2 conjunctions were removed from the analysis, and the average number of occurrences in a bin across all magnetic latitudes and local times equals 12.

From Figure 6.4a,b it is evident that GRACE-KBR and COSMIC-RO measurements at conjunctions are largely consistent. As noted above, the mean bias between GRACE and COSMIC measurements across all latitudes and local times is on the order of $< 2\%$. It is also important to analyze the bias distribution in the MLat-LT frame, shown in Figure 6.4c. It is evident that the difference between the two data sets is close to zero at middle latitudes, while in the equatorial region at ~12 to 18 hours local time, COSMIC underestimates the crests of the equatorial ionisation anomaly, and slightly overestimates the regions poleward from the crests. It should be noted, however, that the conjunction number at the equatorial latitudes is rather small, and the fountain effect is not well-resolved. The bias distribution in Figure 6.4c concurs with the previous study by Yue et al. [2010b], which found that due to the profile geometry and assumptions introduced during the Abel inversion, the radio occultation technique can underestimate electron densities around the EIA crests. We note, however, that in case of the GRACE-COSMIC comparison, the magnitudes of errors are relatively small, with an average value of ~0.3-0.4$\cdot 10^5$el./cm^3, which corresponds to $\sim$ 1.7 MHz when converted to the plasma frequency. Furthermore, EIA is the region where the largest density values ($> 4 \cdot 10^5$el./cm^3) in the F2 layer are manifested, and the mean bias around the EIA crests in Figure 6.4c remains on the order of less than 10%. Therefore, although the Abel inversion introduces a slight underestimation of the EIA crests in COSMIC data, in general the difference between COSMIC and GRACE is very small and is close to zero.

In the present study we use the GRACE-KBR data as reference. Our results are in line with those from previous studies, e.g., from Habarulema and Carelse [2016]. The authors have compared the COSMIC densities to the ionosonde observations during disturbed geomagnetic conditions and concluded that the mean deviation was on the order of 2-3%, which matches our comparisons in the present paper. Since electron densities obtained from COSMIC radio occul-

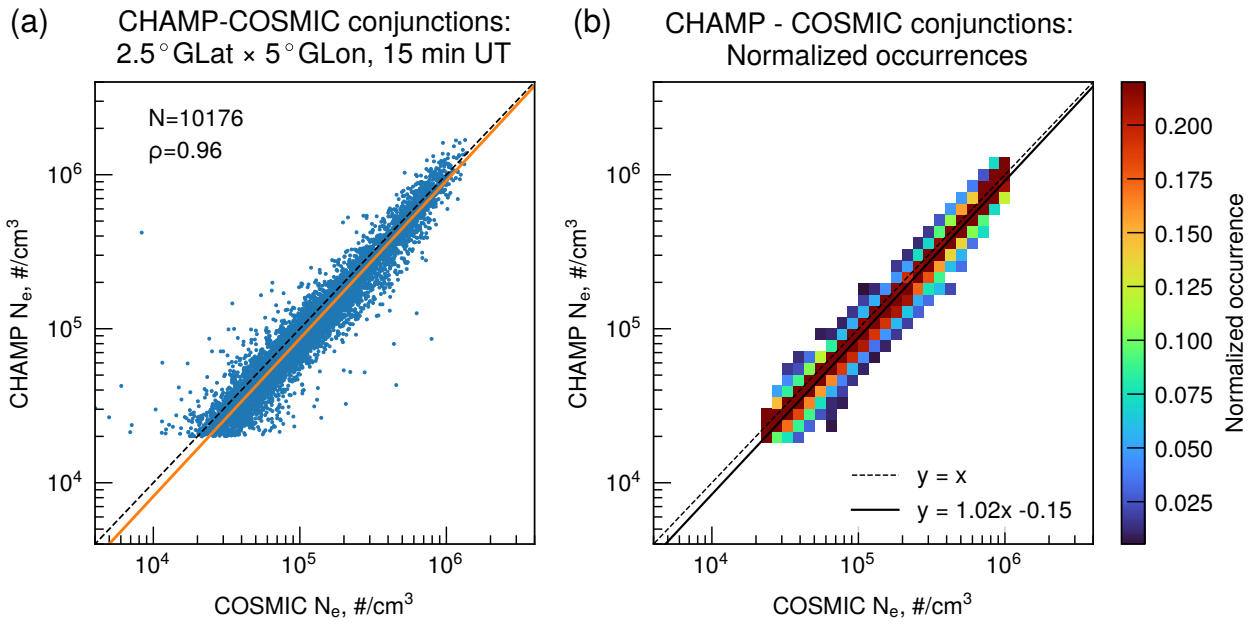

Figure 6.5. Comparison of CHAMP-PLP and COSMIC-RO densities. Notations are identical to those in Fig. 6.3.

tations agree well with GRACE, and they are the only technique that allows evaluating conjunctions on the global scale as well as the biggest data source for empirical modeling collected to date, the other data sources will be compared to COSMIC in the following sections.

6.4.2 Comparison of CHAMP-PLP and COSMIC-RO electron densities

Figure 6.5 demonstrates a comparison between CHAMP planar Langmuir probe and COSMIC radio occultation electron densities. As before, the scatter plot for all conjunctions is given in Figure 6.5a, and the normalized occurrence histogram is shown in Figure 6.5b. We find more than 10100 conjunctions during geomagnetically-quiet times (Kp<3) across all latitudes and longitudes. The collocations corresponding to low values in CHAMP-PLP data ($N_e < 2 \cdot 10^4$el./cm^3) were removed, to exclude the potential negative bias of CHAMP-PLP at very low densities [see, e.g., McNamara et al., 2007; Pedatella et al., 2015].

From Figure 6.5a it can be seen that in general, the scatter points follow a linear trend, although the fitted trend is different from the one-to-one line. The same feature can be observed

Table 6.3. Calibration coefficients used to adjust the data sets to the same reference frame (GRACE-KBR) using Equation 6.1.

Data set	c	d
COSMIC/RO	1	0
CHAMP/PLP	0.980	0.147
CNOFS/CINDI	0.968	0.214
Swarm-A/LP	1.087	-0.380
Swarm-B/LP	1.042	-0.167
Swarm-C/LP	1.087	-0.370

for the trend fitted to the probability distribution maxima. CHAMP-PLP densities are lower than those observed by COSMIC, with a mean bias of -17754 el./cm^3, which corresponds to $\sim -10\%$ relative difference. The median bias exhibits roughly similar values of -14054 el./cm^3, and -12%, respectively (see also Table 6.2). The Spearman correlation between CHAMP and COSMIC electron densities is high (0.96), indicating that while there is a leveling difference, the behavior of two data sets is sufficiently similar.

Our results agree well with previous findings. McNamara et al. [2007] compared the CHAMP-PLP plasma frequencies to ionosonde measurements when CHAMP's altitude was below the F2-layer peak. It was found that the CHAMP plasma frequencies were lower than those by the ionosonde, with the majority of the scatter points being higher than the one-to-one line and a bias of around 5%. It should be noted that the bias value of 5% in plasma frequency corresponds to $\sim$ 10% difference in electron densities, which matches the bias observed in the present study. Ionosondes provide highly accurate observations of ionospheric densities, and the difference of CHAMP-PLP data with respect to ionosondes justifies the leveling correction presented here.

Pedatella et al. [2015] performed a comparison between the COSMIC and CHAMP electron densities, based on data from 2007-2009, although no inter-calibration factors were introduced. In the present study we use the data from the start of the COSMIC mission in late 2006 until the end of CHAMP operation in 2010. Pedatella et al. [2015] reported that CHAMP electron densities were lower by 14.9%, and the correlation between the two data sets was 0.93, which

fits well with our findings. In the present study, our aim is to cross-calibrate data from several missions to further utilize the combined data set for empirical modeling. Therefore, we introduce a linear trend in double-logarithmic scale, and adjust CHAMP-PLP and COSMIC-RO densities using Equation 6.1 (the coefficients are given in Table 6.3). Table 6.4 gives the values of mean and normalized mean bias before and after adjustment. The introduced trend equation helps to eliminate the differences between the data sets, and the relative difference of the adjusted CHAMP-PLP data compared to COSMIC is -3%.

6.4.3 Comparison of C/NOFS-CINDI and COSMIC-RO plasma densities

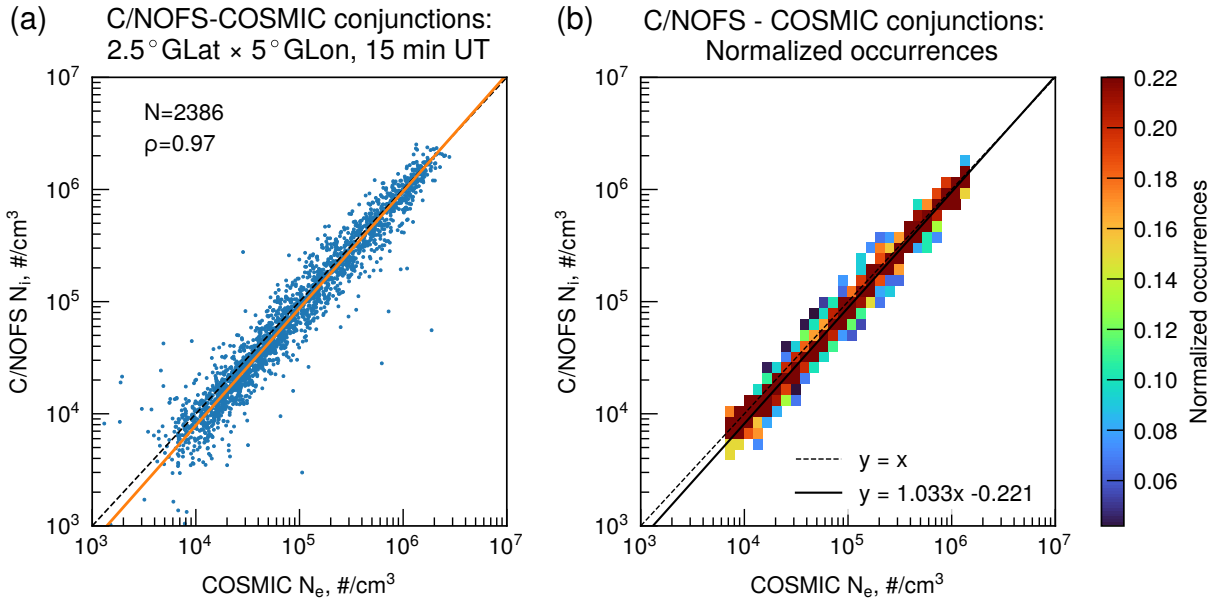

Figure 6.6. Comparison of C/NOFS-CINDI and COSMIC-RO densities. Notations are the same as in Figure 6.3.

Figure 6.6 shows conjunctions between the C/NOFS and COSMIC missions. Figure 6.6a gives a scatter plot of electron density observed by COSMIC and full ion density measured by the CINDI/RPA instrument aboard C/NOFS. It is apparent that the two data sets are highly consistent with each other, based on the high value of the Spearman correlation (0.97) and a relatively low bias ($\sim$ 6%). In general, C/NOFS values of electron density are lower than COSMIC ob-

servations. Figure 6.6b shows that, within the range of measurements, the trend fitted to the normalized occurrence histogram generally lies lower than one-to-one line.

Our results are consistent with previous findings. Lai et al. [2013] compared C/NOFS-RPA densities to COSMIC-RO observations during 2 consecutive days and reported substantial agreement between them, with a correlation of 0.83 and a slope of the regression line close to 1. While the primary objective of that study was to establish the tentative agreement between the two missions in order to use both data sets for the analysis of the ionospheric storm that occurred in March 2013, the study showed that the two data sets can be used together in a variety of applications. Pedatella et al. [2015] compared C/NOFS and COSMIC-inferred densities using larger scale statistics for measurements from 2009 to 2013. The two missions were found to be in good agreement, with a correlation of 95 % and a relative bias of $\sim$ 5.6%. In the present study, we employ a longer data set, comprising collocations between the two missions from 2009 up to the end of C/NOFS operation in 2015, while also using stronger conjunction criteria to decrease the influence of ionospheric gradients. Our results match those of Pedatella et al. [2015]. The C/NOFS mission provides a valuable data set of plasma density observations in the topside ionosphere, covering altitudes from 450 to $\sim$ 800 km, and can be used together with RO data for empirical topside modeling. Furthermore, several recently launched constellations, for instance the Ionospheric Connections Explorer (ICON) mission, are equipped with RPAs. The cross-calibration procedure presented here can be employed in future cross-calibration studies and can be used for combined ionospheric monitoring by ICON and the active RO missions, e.g., CSES, Spire, and the follow-up constellation COSMIC-2.

6.4.4 Comparison of Swarm-LP and COSMIC-RO plasma densities

Figure 6.7 shows a comparison between COSMIC electron densities and Swarm ion densities. In the present study, we use observations from 2013-2020 and find >3800 conjunctions during quiet geomagnetic conditions (Kp<3) for each of the three satellites, distributed across all latitudes and longitudes. Figure 6.7a shows a scatter plot of electron density measurements from Swarm-A versus COSMIC-RO observations, with data points coming from 06 to 18 LT plotted

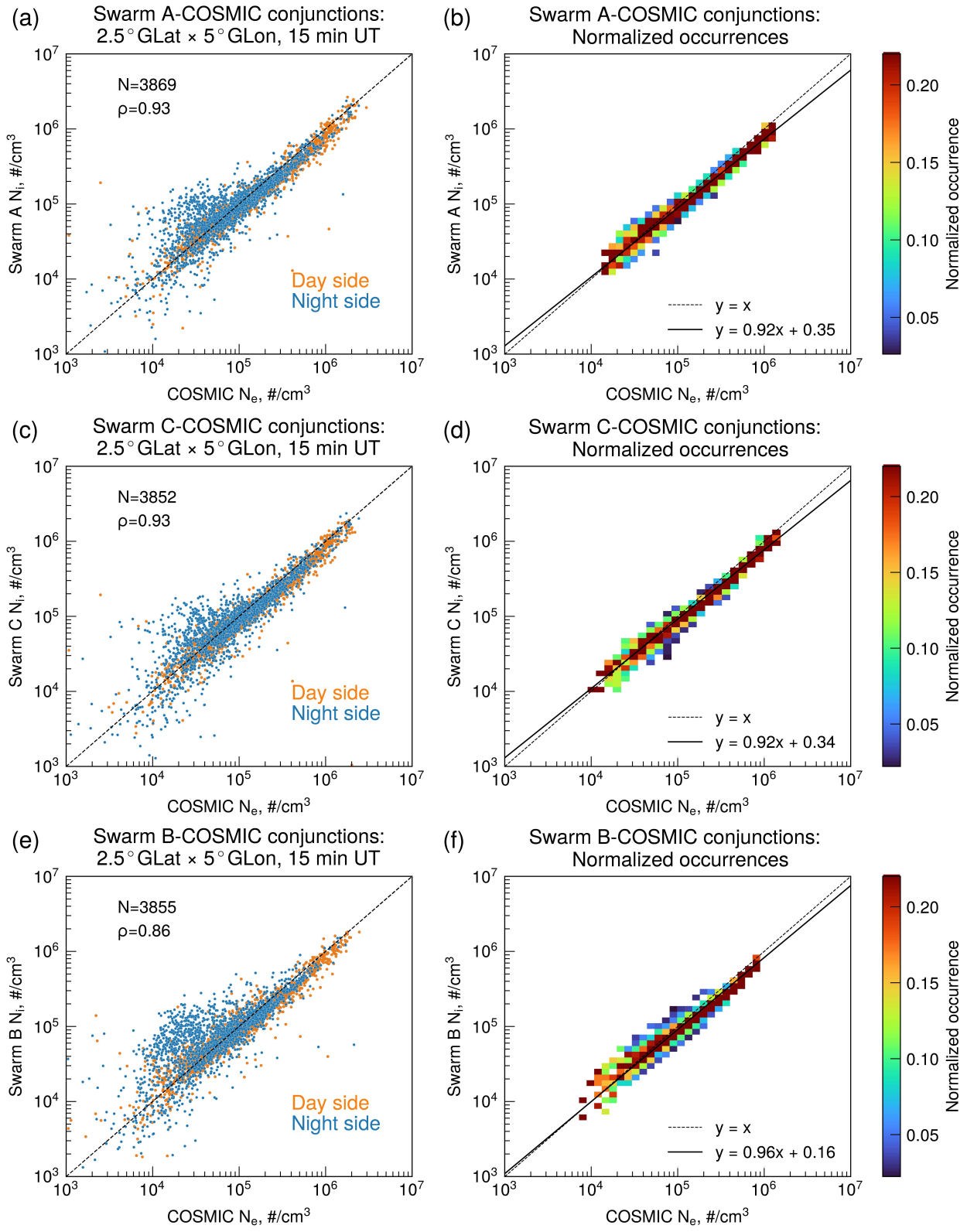

Figure 6.7. Comparison of Swarm ion densities and COSMIC-RO electron densities. (a, c, e) Scatter plot of Swarm A, C, and B versus COSMIC-RO densities. The conjunctions at night side (18-06 LT) are shown in blue, and those at day side (06-18 LT) are plotted in orange. The dashed line represents one-to-one ratio between the data sets. (b, d, f) Normalized occurrence plot of Swarm A,C, and B vs COSMIC-RO densities based on conjunctions occurring at all MLats at day side and at |MLat| > 45° at night side. The trend fitting procedure is the same as in Figure 6.3.

in orange, and conjunctions occurring at nighttime (18-06 hours LT) plotted in blue. Figure 6.7b shows the normalized occurrences for conjunctions occurring at all MLats during daytime and at |MLat| > 45° during nighttime (explained below and in the Supporting information in Smirnov et al. [2021]), with the solid black line representing the linear fit to the probability distribution maxima. In Figure 6.7a one can observe a somewhat larger scatter at night side, with a number of points located above the one-to-one line and an overall larger bias than in previous comparisons with GRACE, CHAMP and C/NOFS. The correlation between Swarm-A and COSMIC densities is high, equal to 0.93, while the percentage bias is on the order of -14%. Figure 6.7c,d demonstrates a similar comparison for Swarm-C. In case of Swarm-C, a roughly similar number of conjunctions was identified, and the bias with respect to COSMIC is -15%. Figure 6.7e,f shows the comparison between Swarm-B and COSMIC densities. It is of note that Swarm-B follows an orbit higher by approximately 50 km than those of Swarm-A and -C satellites, and is approximately at 510 km altitude. The correlation between the Swarm-B and COSMIC plasma densities is lower than that for A and C satellites and equals 0.86. Conversely, the mean value of bias is lower for Swarm-B compared to A and C spacecraft and equals -8%. However, it should be noted that the difference between COSMIC and Swarm-B depends on local time, with an underestimation by Swarm-B on the day side and a stronger overestimation on the night side. The standard deviation for Swarm-B is ~ 55%, which is higher than for spacecraft A and C for which it is approximately 45%.

In previous studies, Swarm ion densities were compared to the incoherent scatter radar and ionosonde observations during December 2013 to June 2016 period, and the offset between Swarm and ground-based measured densities was noticeable both for ionosondes and ISRs [Lomidze et al., 2018]. The corrections were introduced based on conjunctions between each of the Swarm satellites and the ISRs. The adjusted Swarm ion densities were compared to COSMIC RO observations at ~2000 collocations. In the present study, we employ a longer data set for the comparison (from 2013 to 2019), which yields approximately two times more conjunctions with COSMIC. The correction factors, introduced by Lomidze et al. [2018], were uniform with respect to local time and latitude, as Swarm satellites were found to underestimate densities by a factor of ~1.1 at conjunctions with the ISRs. Figures 6.7a,c,e indicate that at night side, all three

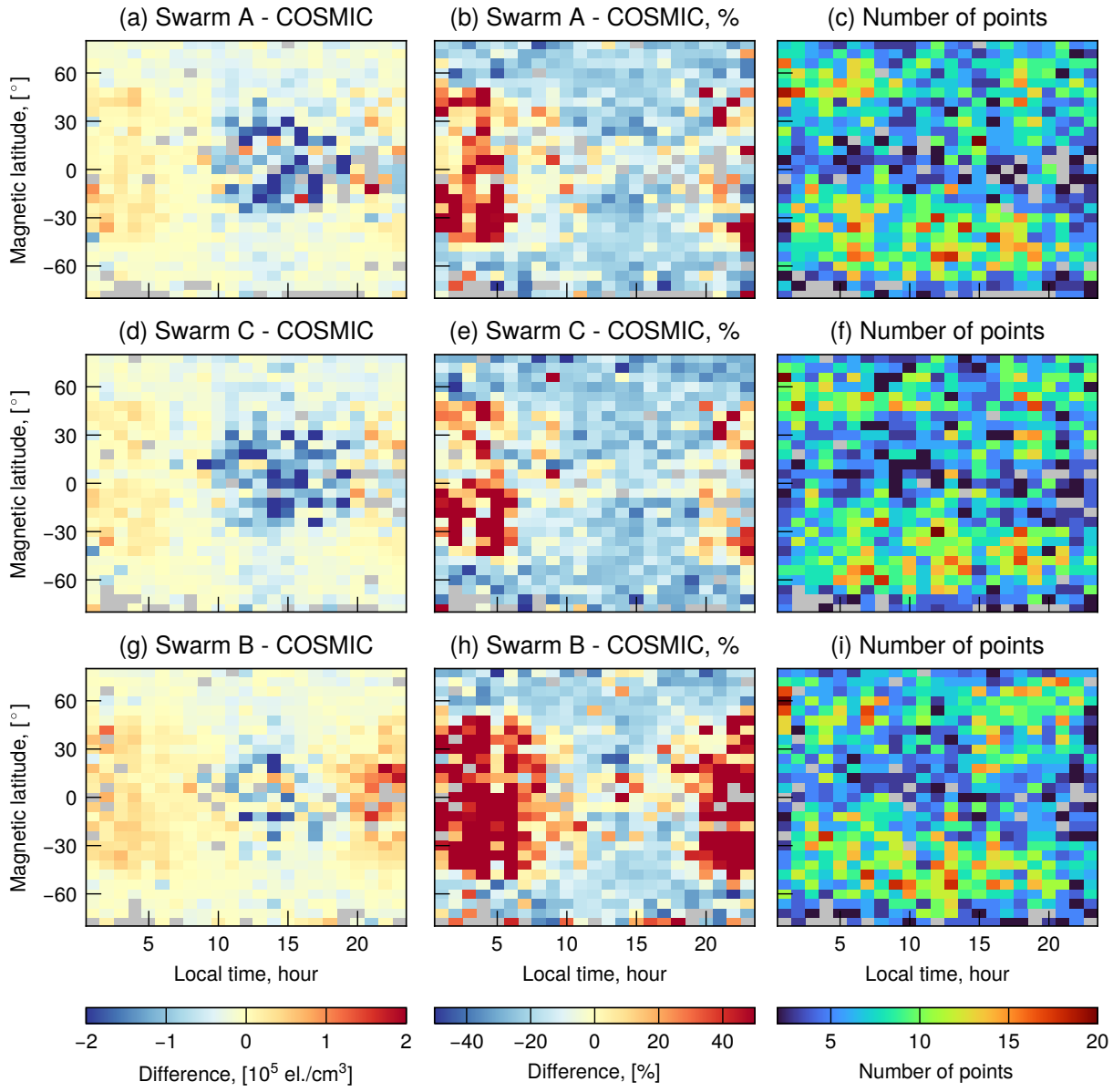

Figure 6.8. Comparison between Swarm-LP and COSMIC-RO plasma densities, binned by 5° MLat-1h LT. Mean bias (a, d, g), mean percentage bias (b, e, h), and number of conjunctions in bin (c, f, i) for Swarm A, C and B satellites, respectively.

Swarm satellites may overestimate densities, and this effect is most prominent for Swarm-B. In order to investigate this LT-dependent difference in more detail, we analyze the distribution of the mean and relative (%) bias between Swarm and COSMIC as a function of magnetic latitude and local time.

Figure 6.8 demonstrates the difference between Swarm and COSMIC plasma density data at conjunctions, binned by 5°MLat - 1h LT. In the comparison with GRACE-KBR (Fig. 6.4), COSMIC was found to slightly underestimate the crests of the equatorial ionization anomaly as a result of the Abel inversion. In Figure 6.8, the opposite effect is observed. Swarm densities are lower than COSMIC around the geomagnetic equator at day time (6-18 LT). At middle latitudes, Swarm measurements are also lower than COSMIC observations. Furthermore, on the night side (18-06 LT) all three Swarm satellites exhibit higher densities than COSMIC. Figure 6.8 (panels a and d) indicates that these patterns are highly consistent for Swarm-A and Swarm-C satellites, due to their similar orbital height (460 km) and small spatial separation of 1.4° GLon. The EIA underestimation on the day side is of the order of $1 - 2 \cdot 10^5$ el./cm^3, which corresponds to approximately $10 - 30\%$ relative bias. On the night side, the overestimation is in the order of $0.3 \cdot 10^5$ el./cm^3 (10-20%).

The overestimation of Swarm-B at night side (Figure 6.8g) is higher than in case of A and C satellites ($\sim 0.5 \cdot 10^5$ el./cm^3 compared to $\sim 0.3 \cdot 10^5$ el./cm^3). While at middle and equatorial latitudes Swarm-A densities were larger than COSMIC observations by 10-20%, the overestimation by Swarm-B is often > 40%. As noted above, Swarm A and C follow a roughly similar orbit, and their elevation after the commissioning phase is ~460 km. Swarm-B is at an altitude of ~ 510km. One of the potential reasons for the larger electron density overestimation by Swarm-B compared to A and C satellites is related to lower oxygen-hydrogen transition heights at low and mid-latitudes during nighttime compared to daytime (~ 500km versus >850 km)[e.g., Heelis et al., 2009; Aponte et al., 2013]. The Swarm LPs are assumed to measure O^+ ions exclusively and therefore at heights below ~500 km where this assumptions holds the difference between Swarm A and C with respect to COSMIC is lower than at 510 km where the H^+ and He^+ play an important role in the ion composition and would lead to larger overestimation at night side by Swarm-B. Another effect that can contribute to this overestimation is the influence of the space-

Table 6.4. Metrics evaluated before and after the adjustment.

Comparison	Median bias before correction, el./cm^3	Median bias before correction, %	Median bias after correction, el./cm^3	Median bias after correction, %
CHAMP/PLP - COSMIC/RO	-14055	-12	-3538	-3
CNOFS/CINDI - COSMIC/RO	-5458	-7	45	0
Swarm A/LP - COSMIC/RO	-11915	-11	-1192	-1
Swarm B/LP - COSMIC/RO	-4046	-5	1371	2
Swarm C/LP - COSMIC/RO	-14142	-13	-3236	-3

craft potential, which depends on the illumination of the solar cells and could lead to density overestimation at night side.

In Table 6.4 we demonstrate the median bias before and after applying the proposed calibration coefficients, evaluated on all conjunctions. It can be seen that for all three Swarm satellites the proposed corrections reduce bias to the 1-3% range. Therefore, it can be concluded that while the linear correction factors for Swarm match the probability distribution maxima and eliminate most of the bias, a more detailed analysis on the separation between oxygen and hydrogen/helium ions within Swarm's Electric Field Instrument (EFI) and effects of the spacecraft potential needs to be performed in future studies.

6.5 Summary and conclusions

Over the last 20 years, the Earth's ionosphere has become a data-rich environment, as the total number of plasma density measurements is approaching several billion points. This wealth of observations, provided by direct in situ (LP, RPA) and remote sensing (RO) techniques, presents new opportunities for the large-scale empirical modeling. Yet, in the topside ionosphere, the altitude distribution of data remains far from uniform. To compensate for the uneven coverage it is essential to utilize all of the available observations in modeling, and therefore these data sources need to be adjusted to the same reference frame. Since the measurement and calibration techniques are specific to each instrument, a reliable long-term plasma density data set is needed to perform the intercalibration.

In this study we use the GRACE-KBR electron density measurements as a reference data set.

The KBR data have been thoroughly validated by the incoherent scatter radar measurements and represent a practically calibration-free data set, providing electron densities at 400-500 km altitude over ~1.5 solar cycles in 2002-2017. As a reference data set for the intercalibration, satellite electron density observations by GRACE-KBR have notable advantages over the ground-based ISR data, such as, for instance, the global coverage in the topside ionosphere.

The comparison of the radio occultation data from COSMIC mission to the reference KBR data set shows an excellent agreement between them. Although COSMIC slightly underestimates the EIA crests, the two missions are highly consistent with a mean percentage difference of $\sim 2\%$. This indicates that first, the two missions can be used in combination for empirical modeling, and second, other data sources can be compared to COSMIC. We find that CHAMP-PLP densities are lower than those provided by COSMIC by $\sim 11\%$, and introduce simple intercalibration factors between them. C/NOFS-CINDI ion densities are generally in good agreement with COSMIC, although the trend between them is different from the one-to-one line. Swarm ion densities have been compared to COSMIC at a large number of conjunctions (> 3800 collocations for each of the satellites). It was found that while the corrections from previous studies, as well as the calibration factors introduced here based on the probability distribution functions, remove most of the bias, the difference between Swarm and COSMIC has local-time signatures which are strongest for Swarm-B. Amongst the potential reasons for such an effect, there are LT-variations in upper transition height, or influence of the spacecraft potential, and the more detailed investigations can be performed in future studies.

The calibration factors introduced in this study can have wide applications in ionospheric research. They eliminate the systematic leveling differences between the most prominent and widely used ionospheric data sets. Hence, these data can be used jointly to set up new highly accurate models of electron density, for example, those based on machine learning. Furthermore, they allow a more precise in-orbit model validation for the existing models. Moreover, the ongoing follow-up mission GRACE-FO allows density retrieval by the same technique as the original GRACE mission, and can be used as a reference for calibrating the recently launched or planned ionospheric constellations. Several active missions carry instruments that operate on the same principles as the ones analyzed in the present study. Therefore, an approach developed in

this study could be extended to other new in situ data sets, for instance the NASA's ICON-RPA data, allowing for a complex ionospheric monitoring by the GRACE-FO, ICON, and the active RO missions, such as CSES, Spire, and COSMIC-2.

Acknowledgments

This research is supported by the Helmholtz Pilot Projects Information & Data Science II, MAchine learning based Plasma density model project (MAP) - ZT-I-0022. COSMIC data can be downloaded via UCAR (https://cdaac-www.cosmic.ucar.edu/); this material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement 1852977. N. Pedatella acknowledges support from AFOSR grant FA9550-16-1-0050 and US National Science Foundation grant AGS-1522830. The GRACE electron density data have been provided in the framework of the Topside Ionosphere Radio Observations from multiple LEO-missions (TIRO) project funded by ESA via the Swarm DISC, Sub-Contract No. SW-CO-DTU-GS-126, and are accessible at [Xiong et al., 2021]. The Level 2 CHAMP PLP data are publicly available at Rother and Michaelis [2019] and through the Information System and Data Centre (ISDC) of GFZ Potsdam (https://isdc.gfz-potsdam.de/champ-isdc/). Swarm LP data were obtained from ESA (https://earth.esa.int/web/guest/swarm/data-access). C/NOFS data were obtained through NASA's Space Physics Data Facility (SPDF, spdf.gsfc.nasa.gov). The F10.7 index was downloaded from the OMNIWeb database (omniweb.gsfc.nasa.gov).

— 7 —

A novel neural network model of Earth's topside ionosphere *

Abstract

The Earth's ionosphere affects the propagation of signals from the Global Navigation Satellite Systems (GNSS). Due to the non-uniform coverage of available observations and complicated dynamics of the region, developing accurate models of the ionosphere has been a long-standing challenge. Here, we present a Neural network-based model of Electron density in the Topside ionosphere (NET), which is constructed using 19 years of GNSS radio occultation data. The NET model is tested against in-situ measurements from several missions and shows excellent agreement with the observations, outperforming the state-of-the-art International Reference Ionosphere (IRI) model by up to an order of magnitude, especially at 100-200 km above the F2-layer peak. This study provides a paradigm shift in ionospheric research, by demonstrating that ionospheric densities can be reconstructed with very high fidelity. The NET model depicts the effects of numerous physical processes governing the topside dynamics and can have wide applications in ionospheric research.

*This chapter has been published as Smirnov, A., Shprits, Y., Prol, F., Lühr, H., Berrendorf, M., Zhelavskaya, I., and Xiong, C. (2023): A novel neural network model of Earth's topside ionosphere. *Scientific Reports*, 13(1), 1303.

7.1 Introduction

Earth's ionosphere is a partially ionized region of the upper atmosphere, spanning from 60 to ~1000 km in altitude [Rishbeth and Garriott, 1969]. The ionosphere is driven by a large number of competing processes and represents a highly dynamic medium that can change substantially in a matter of several minutes. A high number of free electrons in the ionosphere affects the propagation of radio signals, including those of the GNSS [Xu and Xu, 2016]. Around 80% of the ionospheric TEC comes from the part located above the F-layer peak, known as the topside ionosphere [Bilitza, 2009]. Therefore, it is critically important to have accurate models of electron density in the topside ionosphere.

There are several approaches to model electron density in the ionosphere. The physics-based simulations, which obtain numerical solutions of the fundamental equations describing the ionospheric plasma, require sophisticated modeling codes that include coupling with the neutral atmosphere and magnetosphere [Hoque et al., 2022; Huba et al., 2000]. Running such simulations is computationally expensive and thus problematic for operational purposes. An alternative approach to model electron density is through empirical modeling, where the relation between the input and output variables is described based on the statistical representation of observations [Radicella and Nava, 2020]. Two of the most prominent empirical models of the ionosphere are the NeQuick model [Nava et al., 2008; Coïsson et al., 2006] and the IRI model [Bilitza et al., 2006; Bilitza, 2009; Bilitza et al., 2011]. These models mainly describe the climatology of the ionosphere and reproduce regular variations of the ionospheric parameters. It is worth noting that recent versions of the ionospheric models, including the IRI-2016, are becoming increasingly oriented toward weather-like predictions, compared to earlier climatological descriptions [Bilitza et al., 2017]. The empirical models of electron density typically use a layered structure of the ionosphere, with the layer-peak parameters serving as anchor points of the density profiles [Radicella and Nava, 2020]. In particular, two of the most important parameters are the peak density of the F2 layer (NmF2) and the corresponding altitude of the peak (hmF2), which have received a lot of attention in literature and can be well reproduced by the existing empirical models. In the topside ionosphere however, the models exhibit notable discrepancies from ob-

servations due to the highly non-uniform data coverage both in terms of solar activity and, most importantly, in altitude [Lühr and Xiong, 2010; Cherniak and Zakharenkova, 2016; Kashcheyev and Nava, 2019]. Therefore, accurate topside modeling has remained a significant challenge.

Over the last 2 decades, the ionosphere has become a data-rich environment. One of the most efficient ways to make use of the vast amounts of data for empirical modeling is by applying ML techniques. In recent years, a number of ML-based electron density models in the Earth's ionosphere have been developed [Sai Gowtam and Tulasi Ram, 2017; Gowtam et al., 2019; Li et al., 2021; Liu et al., 2019; Habarulema et al., 2021; Tulasi Ram et al., 2018]. Several models provide the F2-peak parameters [Sai Gowtam and Tulasi Ram, 2017; Liu et al., 2019], while others reproduce three dimensional electron density distributions [Gowtam et al., 2019; Li et al., 2021]. However, most existing ML-based models binned the data into spatial cells in terms of geographic latitude and longitude and thus do not provide continuous output [Habarulema et al., 2021; Tulasi Ram et al., 2018; Gowtam et al., 2019]. Habarulema et al. [2021] noted that these models exhibit discontinuities at the borders of spatial cells and require either 3D interpolation or filtering to produce physically meaningful output. Furthermore, it has been noted that one of the main limitations of the existing machine learning models is their lack of validation through independent data sources [Habarulema et al., 2021].

In this study, we present an empirical model of electron density in the topside ionosphere based on Chapman functions with a linear dependence of scale height on altitude. Our model consists of 4 parameters, namely the electron density of the F2-peak (NmF2), the peak height (hmF2), and 2 parameters of the linear scale height decay (slope dH_s/dh and intercept H_0), derived from 19 years of GNSS RO data (see Figure 7.1a). We use neural networks to model these parameters and develop a continuous model that yields highly accurate reconstructions of the topside ionosphere. We perform an extensive validation of the model on in-situ data from three independent missions not used for the model training. The developed Neural network-based model of Electron density in the Topside ionosphere (NET) model is in a remarkable agreement with in-situ measurements and outperforms the IRI-2016 model by up to an order of magnitude, especially at 100-200 km above the F2-layer peak and in local-winter hemispheres.

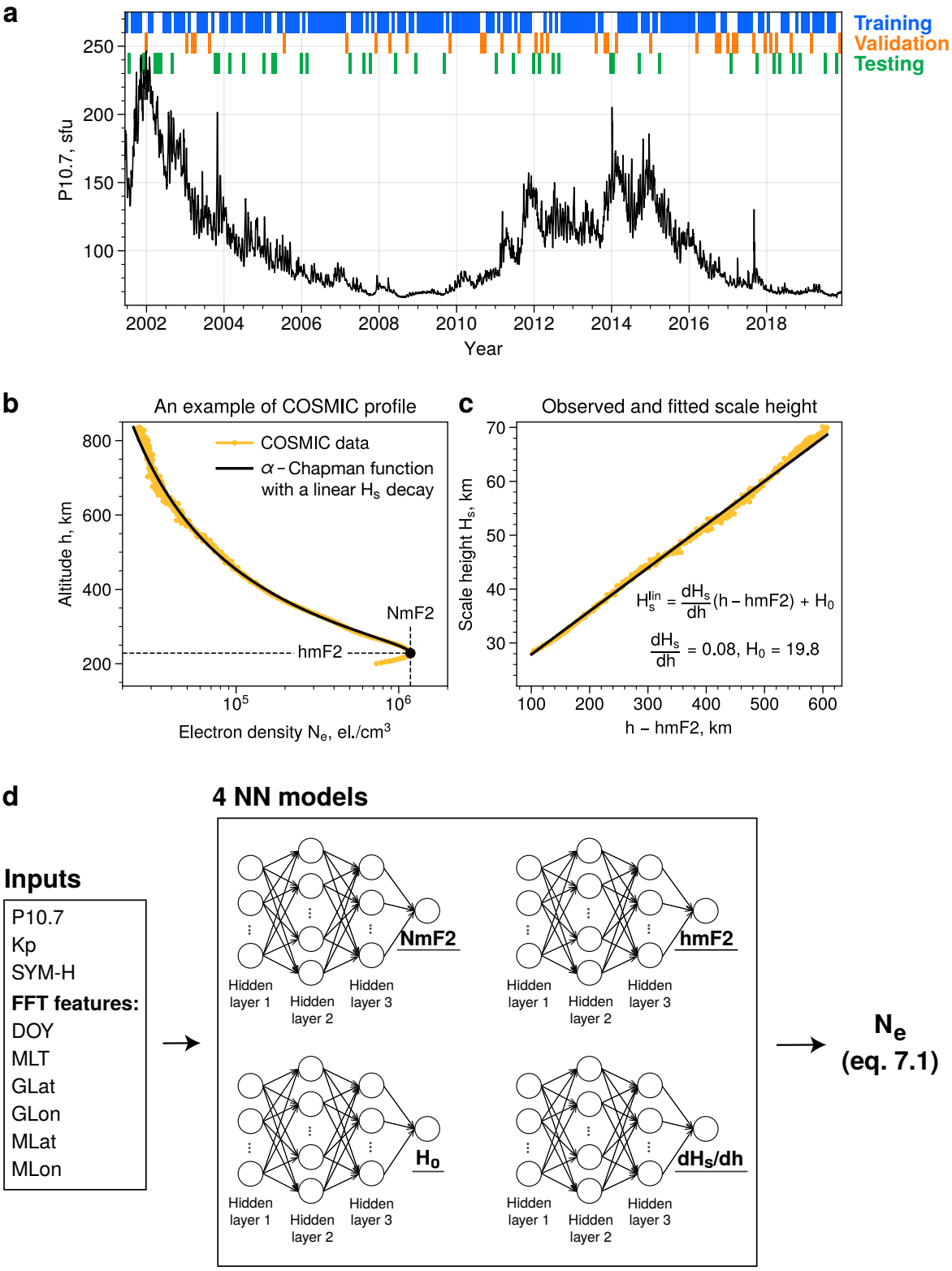

Figure 7.1. (a) Distribution of the P10.7 index and the data splitting; (b) An example of the COSMIC profile (orange) and the fitted data using the alpha-Chapman function with a linear decay of scale height with altitude ; (c) Observed scale height (orange) and the linear fit (black); (d) Schematics of the model workflow.

7.2 Methods

7.2.1 Data set

Over the last two decades, the GNSS radio occultations proved to be an invaluable tool in the ionospheric research. The RO measurements are a remote sensing technique that allows retrieval of the high-resolution electron density profiles (EDPs). Schreiner et al. [2007] estimated the precision of the RO observations to be $\sim 10^3$el./cm^3. Currently, the electron density observations provided by the RO technique constitute a major three dimensional data source in the topside ionosphere. The EDPs are retrieved using an Abel inversion, which can lead to certain artifacts, for instance the underestimation of the Equatorial Ionization Anomaly (EIA) crests, arising from the underlying assumption of spherical symmetry [Yue et al., 2010b]. However, the RO data have been extensively validated both in conjunctions with the ground-based ISRs and with satellite in-situ observations. In particular, the high quality of the RO EDPs in the topside has recently been demonstrated by comparing the COSMIC measurements to electron density observations from the GRACE-KBR system [Smirnov et al., 2021]. The KBR data were calibrated by the ISRs [Xiong et al., 2015] and therefore were used as a reference data set for these comparisons. Smirnov et al. [2021] demonstrated that the radio occultation electron densities from the COSMIC mission were in very good agreement with GRACE data with bias of <2%. The study has shown that the RO observations can thus be an important data source for empirical modeling, especially in the topside ionosphere due to their 3D coverage and large volumes of provided data.

Several constellations have provided EDP data using the radio occultation technique, starting from the early days of the GPS/MET satellite. In this study, we use data from the COSMIC, CHAMP and GRACE missions. The details of spatial and temporal coverage of these missions can be found in Smirnov et al. [2021]. In general, the COSMIC mission has provided an enormous data set of topside electron densities, exceeding 4.5 million profiles. However, the COSMIC mission operated during the declining phase of the solar cycle 23 and during the entire cycle 24 which corresponded to historically low levels of activity. The GRACE and CHAMP

missions, on the other hand, also covered more active conditions during solar cycle 23. Therefore, in order to ensure a better solar cycle coverage and provide more data corresponding to active conditions, we include data from the GRACE and CHAMP missions. It should also be noted that these missions operated at much lower altitudes than COSMIC and do not provide enough coverage of high altitudes to fit the topside EDPs. Therefore, the additional GRACE and CHAMP data were used for retrieving NmF2 and hmF2, while the topside shape parameters were fitted on COSMIC EDP observations. In this study, we use level 2 electron density profiles, provided through the "IonPrf" product. The COSMIC EDP data were subject to quality control. We removed the topside profiles that extended over $> 5^\circ$ GLat and $> 10^\circ$ GLon, as well as EDPs that exhibited positive electron density gradients at higher altitudes. Furthermore, the profiles where the derivatives of electron density exceeded the magnitudes of electron density, and the profiles with deviations from the linear alpha-Chapman fit of more than 100% were removed, in order to exclude irregular EDPs.

The developed model needs to be validated not only on the data from the missions that were used for training, but also on independent data sources. Using independent missions for validation gives a good indication of the generalization ability of the model at different altitudes, locations and timescales. In this study, we employ the GRACE-KBR observations, which provide in-situ electron density values with a spatial resolution around 200 km along the GRACE orbit. This dataset has been intercalibrated with incoherent scatter radars and can be considered as a "gold standard" [Xiong et al., 2015]. Furthermore, we use in-situ measurements of electron density by the CHAMP-PLP and the full ion densities measured by the C/NOFS-CINDI instrument, intercalibrated by Smirnov et al. [2021]. The GRACE mission operated in 2002-2017 and covered altitudes from 400 to 500 km with global coverage. The CHAMP mission covered lower altitudes, from around 400 km at the beginning of the mission lifespan to 300 km at the end of the mission. The C/NOFS mission represents an important source of data for model validation, because it covered equatorial latitudes and also provided a vast altitude coverage from 300 km up to 800 km.

7.2.2 Fitting the topside profiles with a linear alpha-Chapman function

The vertical structure of the ionosphere has been a topic of continuous interest since the early works of Appleton and Beynon [1947]. The study by Bent et al. [1972] was one of the first papers that described the functional dependence of electron density in the topside as a combination of a parabolic and exponential terms. Since then, a variety of mathematical descriptions of the topside ionosphere has been developed. A systematic review of the possible functions to fit the topside profiles was presented by Fonda et al. [2005]. In particular, they showed that the Chapman function gave the best agreement with observations based on the topside sounder data. The Chapman function is based on first principles and has been employed in numerous studies to approximate the topside ionosphere. One of the important parameters of the ionospheric plasma, included in the Chapman equation, is the effective scale height, which represents a vertical distance over which the electron density decreases by a factor of e [Rishbeth and Garriott, 1969] and thus serves as a shape factor of the topside profiles.

The part of the topside ionosphere close to the F2 layer peak is dominated by atomic oxygen, while at higher altitudes the light ions become the dominant plasma constituents [Kelley, 2009]. Different ion species have different scale heights, and this information needs to be incorporated into the empirical models. In the early works, these differences were neglected and the scale height was often assumed constant for simplicity [Reinisch et al., 2004]. This approximation works well in the lower topside (around the F2-peak) but faces obvious problems at higher altitudes. There can be several approaches to account for the varying scale height. One reasonable approach is to use a multi-layer model, for example assuming a Chapman function with a constant scale height near hmF2 but adding an exponential term for higher altitudes [Limberger et al., 2013]. Another method, first proposed by Rishbeth and Garriott [1969], is to use a single layer formulation which includes an empirical relation of scale height to altitude. One of the approximations which gained a significant popularity assumes a linear decay of scale height with altitude [Rishbeth and Garriott, 1969; Fox, 1994; Olivares-Pulido et al., 2016; Hernández-Pajares et al., 2017; Prol et al., 2018, 2019], and in this study is referred to as the linear alpha-Chapman

function. It has been demonstrated that this method produced results largely identical to the multi-layer model and could well approximate both the radio occultation data [Olivares-Pulido et al., 2016] and the topside sounder observations up to the altitude of around 1500 km [Prol et al., 2019].

In this study, we approximate the COSMIC radio occultation profiles using the linear alpha-Chapman function of the form:

$$\begin{cases} N_e(h) = \mathrm{NmF2} \cdot \exp(0.5(1 - z - \exp(-z))), \\ z = \dfrac{h - \mathrm{hmF2}}{H_s(h)}, \\ H_s(h) = \dfrac{\mathrm{dH_s}}{\mathrm{dh}}(h - \mathrm{hmF2}) + \mathrm{H_0}, \end{cases} \tag{7.1}$$

where N_e is electron density as a function of altitude h; NmF2 and hmF2 represent the peak electron density of F2-layer and the altitude of the peak, respectively; H_s is an effective scale height, in our case depending linearly on altitude, $\mathrm{dH_s/dh}$ and $\mathrm{H_0}$ show the slope and intercept of this linear trend, respectively. The F2-peak density and height were obtained from the data, while the $\mathrm{dH_s/dh}$ and $\mathrm{H_0}$ values were retrieved by fitting Equation 7.1 to COSMIC EDPs using curve-fitting routines implemented in the `scipy` Python library.

Figure 7.1b,c shows an example of the COSMIC profile, fitted using the linear alpha-Chapman function. In panel (b), the COSMIC data are shown in orange, a large black dot denotes the peak of the F2-layer and is determined from the data, and the solid black line gives the topside profile fitted to Equation 7.1. In panel (c), we demonstrate the scale height retrieved from the data using the method of Olivares-Pulido et al. [2016], and the corresponding linear fit from Equation 7.1. In Figure 7.1b it can be seen that the linear approximation reproduces the COSMIC profile well. In fact, Olivares-Pulido et al. [2016] showed that this functional dependence can approximate the vast majority of RO profiles with a correlation of over 98%. We note that the topside profiles are usually fitted to the linear alpha-Chapman method at altitudes higher than 100 km above hmF2, since the scale height may exhibit a stronger nonlinear decrease when approaching the F2-peak [Olivares-Pulido et al., 2016; Prol et al., 2018].

The linear alpha-Chapman approximation yields 4 parameters, namely the F2-peak density

and height, and 2 parameters of the linear scale height decay. In Figure 7.1d, we show the schematics of the model workflow. Each of the four parameters is modeled separately with a feedforward neural network, and the final values of electron density can be retrieved by substituting outputs of the 4 sub-models into Equation 7.1. It should be noted that NmF2 is predicted in logarithmic scale, due to the fact that the peak electron densities can span over several orders of magnitude, while the other parameters are retrieved in linear scale.

7.2.3 Neural networks

One of the most efficient ways to utilize large volumes of data for empirical modeling is by using machine learning. In particular, artificial neural networks are one of the most popular techniques to find complex non-linear relationships between the input and output variables, and have been used in many applications such as classification, regression, and image recognition. In this study, we employ multi-layer perceptrons, which are a type of the fully connected feedforward neural networks. MLPs try to find nonlinear mappings between the input and output parameters by optimizing weights and biases of the neurons contained in the hidden layers. The neural network typically comprises an input layer, one or several hidden layers, and an output layer. Every link between a neuron in one layer to each of the neurons in the next layer has a corresponding trainable weight. It shows how strongly that particular neuron influences nodes it is connected with. When the network is initialized, the weights are selected randomly. Each connection also has an associated bias which allows to better adjust the model. At first, the values contained in the input layer nodes are summarized, passed through an activation function which adds a non-linearity into the model, and used as inputs to the neurons of the first hidden layer. Activation functions are an essential part of MLPs, as only with them one can model non-linear functions. This procedure is the same for all hidden layers. Finally, the outputs of the last hidden layer are summed up and given to the output node, usually without a final activation function. The weights and biases of each layer and neuron are typically optimized by gradient-based methods, where stochastic gradient estimates can be efficiently computed using back-propagation schemes.

Due to the expressiveness introduced by the large number of trainable parameters, MLPs can

easily overfit to training data without capturing the underlying phenomena. One way to mitigate this is via regularization techniques, for instance the Dropout method [Srivastava et al., 2014], which randomly removes a subset of neurons in training of each minibatch and adds regularizing effect by increasing the stochasticity. Another frequently employed technique is an addition of gaussian noise between the layers during the network training. It has been shown that adding small amounts of noise during the training process helps a NN converge to a smooth function of the inputs and also imposes a regularization effect. Specifically, the networks trained with additional noise are less able to memorize the training set, as the model perceives the training samples infused with noise as constantly changing. This results in more robust networks with lower generalization errors. In this study, we employ both the gaussian noise and dropout techniques and treat the dropout rate and noise magnitude as hyperparameters.

In the present study, we develop a set of models to predict 4 parameters that can be represented as numerical variables and therefore address a regression problem. As inputs, we use the solar flux index P10.7, geomagnetic indices Kp and SYM-H, as well as satellite position in geographic and geomagnetic coordinates, MLT and day of year (DOY). It is of note that the COSMIC EDPs can span over several degrees in latitude and longitude, and therefore the geographic positions of the top points of the profiles were used to train the dH_s/dh and H_0 models, while during the model testing the profiles were assumed vertical. The P10.7 represents a smoother version of the 10.7 cm radio flux index (F10.7), and is derived as an average of the current F10.7 value and that over the previous 81 days [Xiong et al., 2022; Bilitza and Xiong, 2021]. This index has been used in a variety of ionospheric models, including the IRI model, and was found to give better performance than the raw F10.7 values [Richards et al., 1994], which can exhibit spikes, especially during the strong solar storms. We also use the planetary Kp index that shows the averaged state of the geomagnetic field disturbances and serves as a good proxy of convection which is an important mechanism in producing several ionospheric phenomena, for instance, the polar patches. Furthermore, we use the SYM-H index, which shows the strength of the geomagnetic storms, to account for the storm-time events.

Several of the cyclic input features have artificial boundaries between the highest and lowest values. For instance, the local time has a discontinuity at 24-00 hours, which can create artifacts

in the model output. In order to avoid this discontinuity, it is common practice in empirical modeling to replace values of these input features with their sine and cosine values [Bortnik et al., 2016a]. Furthermore, it has been shown that using higher orders of the sine and cosine functions for the positional inputs can significantly increase the model accuracy. Tancik et al. [2020] demonstrated that these simple feature transformations allowed the MLP models to learn the high-frequency dependencies in low-dimensional regression tasks and greatly enhanced the model performance for image regression problems. This technique became known as the Fourier features method [Tancik et al., 2020]. In this study, we also apply this method to several features, namely the LT, DOY, geomagnetic and geographic latitude and longitude (see Figure 7.1d). Selecting the correct FFT order can have a non-negligible effect on model training. For each of the features, we select the FFT order that results in the best model performance (the corresponding selection of the FFT orders is described in the Supplementary information of Smirnov et al. [2023]).

7.2.4 Data splitting

This study uses a supervised learning algorithm, namely a multilayer perceptron, to model electron density in the topside ionosphere. In order to train the supervised model, it is necessary to split the data into the training, validation and test subsets. The training set is used to fit the model, the validation set helps select the neural network hyperparameters and gives a more unbiased estimate of the network performance during each training iteration. In particular, the error on the validation set can be used for early stopping regularization, where the model training terminates as soon as the validation error does not decrease anymore. The test set is withheld during the model training and validation, and is only used once the models have been created to evaluate their generalization ability on the unseen data.

There are several ways to split the data into the training, validation and test data sets. If no time-dependence is assumed in the data, it is possible to split data entries between the three sets randomly, e.g., by using 70-80% of the data for training, and 10% for both validation and testing. In case of time-series and physics problems, this splitting technique should be avoided

as it can lead to what is referred to as the data leakage. For instance, when most of the points constitute the training set, one can linearly interpolate between those points to derive values that belong to the typically smaller validation and test sets. This would result in a biased estimate of the model performance [Camporeale, 2019; Smirnov et al., 2020a]. Another potential technique is to use the K-fold CV, where the data are split in the time-domain into K continuous parts, and the model is re-fitted K times each time withholding one of the parts and trained on K-1 intervals (for details, see e.g., Smirnov et al. [2020a]). At the same time, using the K-fold CV for hyperparameter tuning leads to an exponential increase in training time, as the models need to be re-trained numerous times for every hyperparameter trial. Therefore, another option is to split the data in the time domain into a number of intervals that are long enough to contain independent events of shorter timescales, therefore preventing the data leakage. This is described, for instance, in Chu et al. [2021]. In this study, we split the data into continuous 27-day segments, and randomly choose 70% of those intervals for training, 15% for validation and 15% for testing. Such a splitting allows us to validate the model on a variety of solar activity and geomagnetic conditions, while at the same time avoiding the data leakage. The data splitting is illustrated in Figure 7.1a, where one colored stripe corresponds to a single 27-day segment.

7.2.5 Hyperparameter tuning/ model selection

In this study, we use feedforward neural networks implemented in the Keras Python library [Chollet et al., 2015]. The parameters which define the network structure and training procedures are typically referred to as the hyperparameters and include the number of the hidden layers, number of neurons in each of the layers, the activation function, optimizer, dropout rate and so on. They are not directly optimized by gradient descent. Instead, we use the tree-structured Parzen estimator algorithm [Bergstra et al., 2011] implemented in the Optuna Python library to tune the number of hidden layers and neurons, dropout rate and the magnitude of gaussian noise. The preliminary hyperparameter trials indicated that a 3 layer neural network achieved a very good performance and the reduction of the MSE by adding the fourth and fifth layers was <1 %. Therefore, we later fixed a number of layers to 3 and optimized the other parameters. The

summary of the parameters, their search domains and the optimized values based on the Optuna trials are given in the Supplementary Table S1 in Smirnov et al. [2023].

7.2.6 Comparison with the International Reference Ionosphere (IRI) model

First established as a joint project of the Committee on Space Research (COSPAR) and International Union of Radio Science (URSI) in 1968 [Rawer et al., 1978], the International Reference Ionosphere is, perhaps, the most famous model of the ionosphere. The IRI has been continuously improved [Bilitza et al., 2011] and in 2014 was accepted by the International Standards Organisation (ISO) as the international standard of ionospheric specification [Radicella and Nava, 2020]. The IRI describes electron density and temperature, ion composition, ion temperature and drifts at altitudes from ~50 up to around 2000 km. For electron density modeling, NmF2 and hmF2 can be considered the most important parameters. There are several options for the topside shape functions, including the IRI-2001 version and its correction [Bilitza, 2004], as well as the topside parametrization based on the NeQuick model [Coïsson et al., 2006; Nava et al., 2008]. Since the 2007 version of the IRI, the NeQuick topside has been adopted as the default option for the topside ionosphere [Bilitza et al., 2006]. In this study, we use the IRI-2016 model with the foF2 specified by the URSI model, hmF2 given by AMTB-2013 model [Altadill et al., 2013], and the default NeQuick topside, to compare to the predictions by the developed NET model based on several in-situ and radio occultation data sets. In particular, it is crucial to quantify the degree of improvement achieved by the NET model compared to the IRI, which can be done using the skill score (SS) metric. This metric can be written as follows:

$$SS = 1 - \frac{\sum_{i=1}^{N}(m_i - o_i)^2}{\sum_{i=1}^{N}(b_i - o_i)^2}, \tag{7.2}$$

where m denotes the NET model values, o stands for observations, and b represents the baseline (IRI) model output. This metric quantifies the improvement over a baseline model, and is also sometimes referred to as the prediction efficiency [Morley et al., 2018; Murphy, 1988]. The skill score values are analyzed in detail in the Supplementary information from Smirnov et al.

[2023] for different seasons, magnetic latitudes and local times, and it is demonstrated that the NET model outperforms the IRI-2016 by up to 70-80%, with the most significant improvement achieved in the local-winter hemispheres (Supplementary Figure S4 in Smirnov et al. [2023]).

7.3 Results

7.3.1 Modeling 4 parameters of the linear alpha-Chapman function using neural networks

The developed NET model consists of 4 sub-models reproducing parameters of the linear alpha-Chapman equation based on location, season, local time, and solar and geomagnetic activity (Figure 7.1). It is, firstly, necessary to evaluate the ability of the model to recreate these four parameters correctly. A comparison between the values predicted by the NET model and COSMIC observations on the training, validation and test sets (Figure S1 in the Supplementary Information of Smirnov et al. [2023]) yields that the model reproduces all of the parameters well, with correlation coefficient ranging from 0.8 for the dH_s/dh model up to 0.96 for the NmF2 model. Furthermore, the correlation coefficient values are nearly identical on the training, validation and test sets which indicates a very good generalization ability of the model and low degree of overfitting. We now move to analyzing whether the sub-models of each parameter preserve realistic structures of the ionosphere.

Figure 7.2 shows a comparison between the observed and predicted values of the four parameters based on 3 months of COSMIC data sampled from test and validation sets. The values are binned by 1 hour MLT and 5 degrees MLat. The interval in question covers the times from late November 2013 until end of February 2014, which correspond to the December solstice (D-season) conditions. NmF2 and hmF2 parameters describe the dynamic variability of the F2-peak, while the H_0 and dH_s/dh interpreted together can be indicative of the processes in the topside. In particular, Prol et al. [2018] showed that H_0 generally shows the electron content directly above the F2 peak, while changes in dH_s/dh account for electron density decay at higher altitudes.

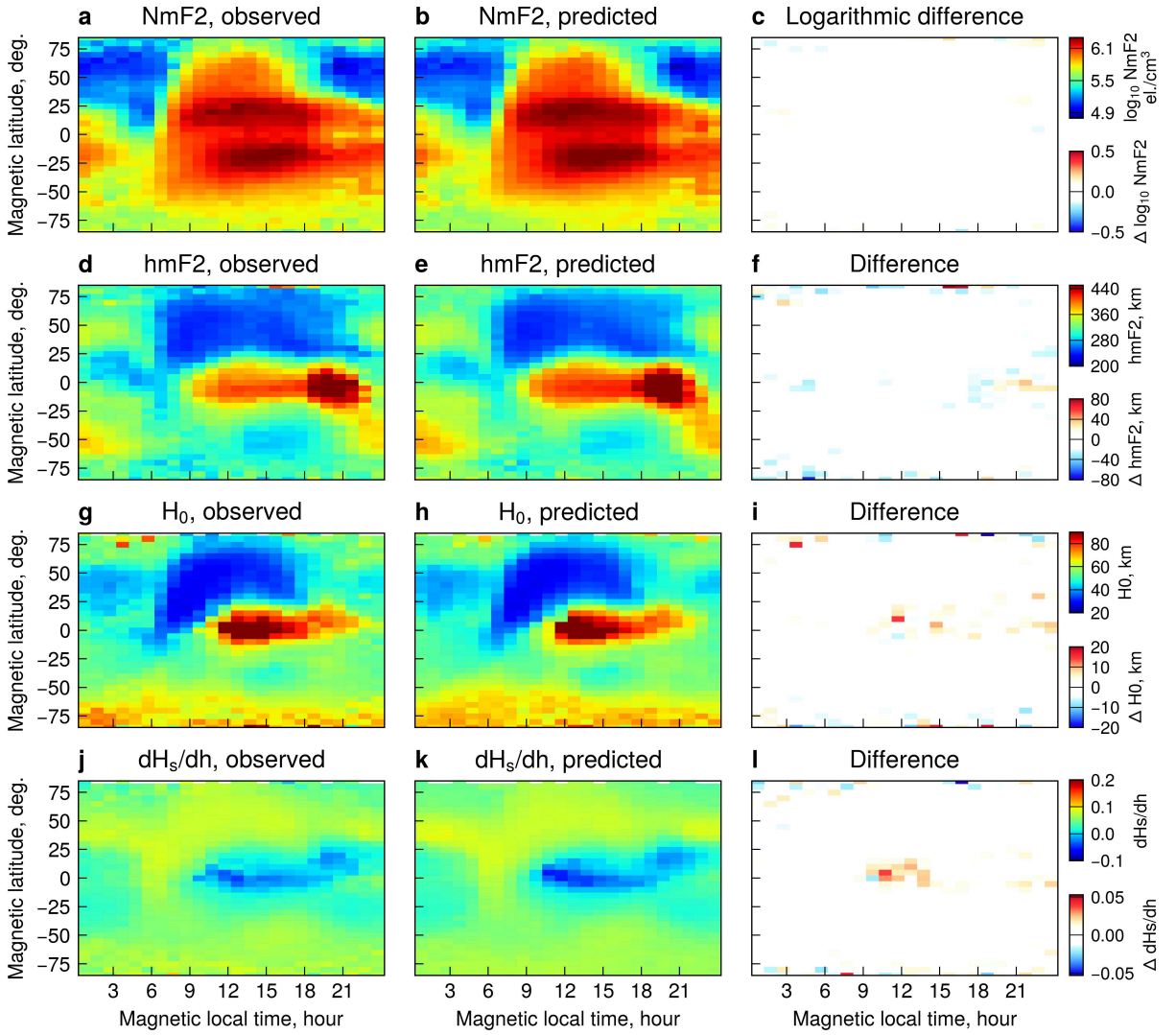

Figure 7.2. Maps of the four parameters observed by COSMIC and predicted using the NET model, binned by magnetic latitude and local time. The bins with <2 data points were removed. The data cover the time interval from 2013-11-11 until 2014-02-27, corresponding to D-season conditions, and are sampled from the validation and test sets.

Specifically, higher H_0 values correspond to the "thicker" profiles around the peak, while the increased dH_s/dh values indicate steepening of the profiles and slower decay at high altitudes.

Physical phenomena as seen in the NET model results

Several known regions, corresponding to different processes that govern the ionospheric dynamics around December solstices, can be identified in the results of NET (Figure 7.2). NmF2

exhibits two crests of ionization around 15-25°MLat in both hemispheres, separated by an electron density trough. This feature is known as the Equatorial Ionization Anomaly (EIA) (Figure 7.2a,b); it usually develops after sunrise and shows a gradual decay after midnight [Rishbeth and Garriott, 1969; Kelley, 2009]. The EIA formation can be explained as follows. The existence of zonal electric fields around the equator, where the magnetic field lines are nearly horizontal, gives rise to the vertical **E**×**B** drift. During the daytime, the zonal electric fields are directed eastwards which leads to the upward transport of plasma by the **E**×**B** drift. During the sunrise, an increase in solar illumination ionizes the neutral particles in the thermosphere. These newly ionized particles are transported to higher altitudes by the **E**×**B** drift, and an enhancement of hmF2 develops at ~ 06 MLT around the equator. Higher altitudes have lower recombination rates, and therefore the ionized particles pertain there for longer times and start diffusing downward under the gravity and pressure gradient forces. This diffusion is constrained by the magnetic field lines and leads to the formation of the two crests of ionization, and this is known as the equatorial fountain effect [Kelley, 2009]. A global maximum of hmF2 is manifested around the equator at ~19 MLT (Figure 7.2d,e). This corresponds to the sunset hours, when the eastward electric fields exhibit pre-reversal enhancements (PRE) [Farley et al., 1986] leading to large upward **E**×**B** velocities which lift the F2 peak to even higher altitudes (> 460 km). During the nighttime (21-06 hours MLT), the zonal electric fields reverse their direction and move westwards creating the downward **E**×**B** transport. The F2-peak thus subsides to lower altitudes (~ 280 km). Due to higher neutral densities and stronger recombination rates there, NmF2 starts slowly decaying exhibiting minima around 05 MLT (Figure 7.2a,b).

Figure 7.2 corresponds to the D-season conditions with the local summer in the southern hemisphere. NmF2 exhibits a strong hemispheric asymmetry, with larger electron densities in the southern hemisphere, due to the Earth's tilt and higher solar irradiation (Figure 7.2a,b). Furthermore, hmF2 also shows a summer-winter asymmetry with values in the northern hemisphere lower by about 100 km compared to the southern hemisphere. Around the solstices, there are strong winds blowing from the summer into winter hemispheres at altitudes around the F2-peak [Kelley, 2009]. In the D-season, the winds are directed from the southern into the northern hemisphere. These winds have a component parallel to the magnetic field and transport ion-

ization upward in the local-summer hemisphere, while pushing the F2-peak downward in the local-winter hemisphere, resulting in the asymmetric structure of hmF2. Furthermore, one interesting feature in Figure 7.2g,h is the increase in H_0 at polar latitudes in the southern hemisphere, which indicates that electron density profiles are convex around the peak. This pattern shows little diurnal variation and likely comes from the fact that during solstices, polar regions in the local-summer hemispheres are sunlit at all MLTs due to absence of dark nights [Prol et al., 2018]. One also observes several regions of enhanced scale height gradient (Figure 7.2j,k), namely two mid-latitude stripes ($\sim 40^\circ$ MLat in both hemispheres) which remain at all local times, and a sunrise peak (~05-07 MLT) around the magnetic equator. The sunrise peak is likely connected with the so-called morning overshoot of electron temperature. It happens due to energy exchange between the newly ionized photoelectrons and ambient electrons, which is more efficient in regions of low electron density. Once the ionization builds up, the resulting cooling decreases the temperature and the peak disappears at around 08 MLT [Stolle et al., 2011]. As the temperature is related to scale height, this feature pertains in the scale height gradient. The dH_s/dh peaks at middle and high latitudes are likely due to the downward fluxes of protons injected into the topside from the plasmasphere [Prol et al., 2018], and become more evident at low solar activity (see Supplementary Figure S3 in Smirnov et al. [2023]).

All of the above-mentioned processes are very well depicted by the NET model, as the differences between the NET predictions and COSMIC observations of the 4 parameters are close to zero at all magnetic latitudes and local times. This means that the NET model is capable of reproducing ionospheric dynamics caused by a wide variety of drivers, including the electrodynamic processes, neutral winds, field-aligned transport from the magnetosphere, etc. Furthermore, the combined interpretation of the four model parameters can yield insights into the physical processes that control the dynamics of the topside ionosphere.

7.3.2 Model testing on COSMIC electron densities

We now move to evaluating the model performance on COSMIC electron density values. Figure 7.3 shows a comparison between electron density predictions by the IRI and NET models to the

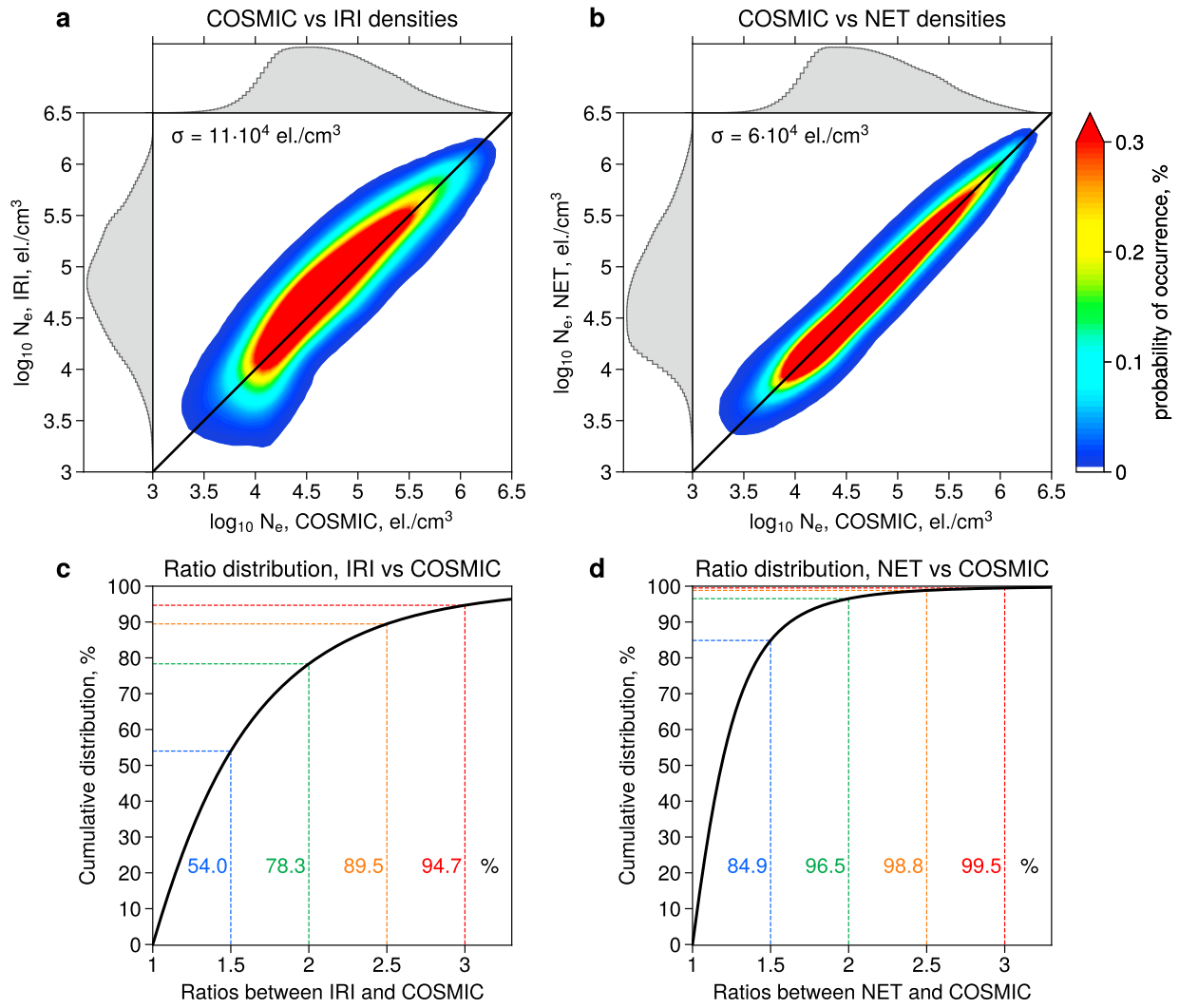

Figure 7.3. 2D histograms of electron density observed by COSMIC on the test set versus those predicted by the IRI model (a), and the developed NET model (b). (c) Cumulative distribution of ratios between the IRI model and the COSMIC data on the test set; (d) Cumulative distribution of ratios between the developed NET model and the COSMIC data on the test set.

COSMIC data for the entire test set. In panel (a), we demonstrate the 2D histogram of electron densities observed by COSMIC and predicted by the IRI-2016 model with the topside specified by the NeQuick option. The IRI gives unbiased predictions for very low and very high electron densities, as these points generally lie close to the one-to-one correspondence line. However, for intermediate densities ($\sim 10^5$ el./cm^3) there is an overestimation of electron densities by the IRI, which results in a curved shape of a 2D probability distribution. Furthermore, the IRI model

produces a close-to-gaussian distribution of electron density values, while the distribution of the COSMIC data appears more flat-top and is skewed to the right. The comparison of the NET predictions to COSMIC data is shown in Figure 7.3b. The corresponding 2D probability distribution is centered around the one-to-one line, which means that the model gives unbiased predictions in the topside ionosphere when compared to COSMIC data. The overall bias of the NET model on the test set is approximately 3 times smaller than for the IRI (-0.6·10^4el./cm^3 compared to $\sim 1.8 \cdot 10^4$el./cm^3). Furthermore, in Figure 7.3 one can see that the 2D probability distribution appears narrower for the NET model. To quantify the degree of spread of the 2D distributions, we evaluate the standard deviation of the difference between the observed and predicted electron densities. The standard deviation for the NET model is $6.5 \cdot 10^4$el./cm^3, while for the IRI this value equals $11 \cdot 10^4$el./cm^3, which is approximately 1.6 times larger. This indicates that the dynamics of electron density is on average captured better by the NET model. Another useful metric, often employed when evaluating the model predictions, is the ratio between the model predictions and observations. Figure 7.3c,d shows cumulative distributions of the ratios between the IRI and NET models to COSMIC data. We use the linear version of the ratio for values > 1, while taking the inverted ratios if the values are < 1. The cumulative distributions of the ratios, shown in Figure 7.3c,d, yield how often the model predictions lie within a given factor from the data. One can see that 96.5% of the time the NET electron densities lie within a factor of 2 from the observations, while the IRI predictions are within a factor of two 78.3% of the time. Moreover, most of the NET predictions (over 84.9%) are within a factor of 1.5 from the data, which is higher than for the IRI (54%).

The statistics shown in Figure 7.3 combine all altitudes and locations corresponding to the test intervals of the COSMIC data. In order to investigate the performance of the developed NET model in more detail, we bin the data in several dimensions and evaluate the metrics locally, compared to the previous global comparisons. We, firstly, bin the model residuals by altitude relative to hmF2 (Figure 7.4). It is of note that for this comparison we do not distinguish between different locations, and such binning is provided separately in the Supplementary Figure S4 of [Smirnov et al., 2023]. Figure 7.4a indicates that the IRI-2016 model with the NeQuick topside underestimates the F2-peak density, and overestimates electron densities in the topside.

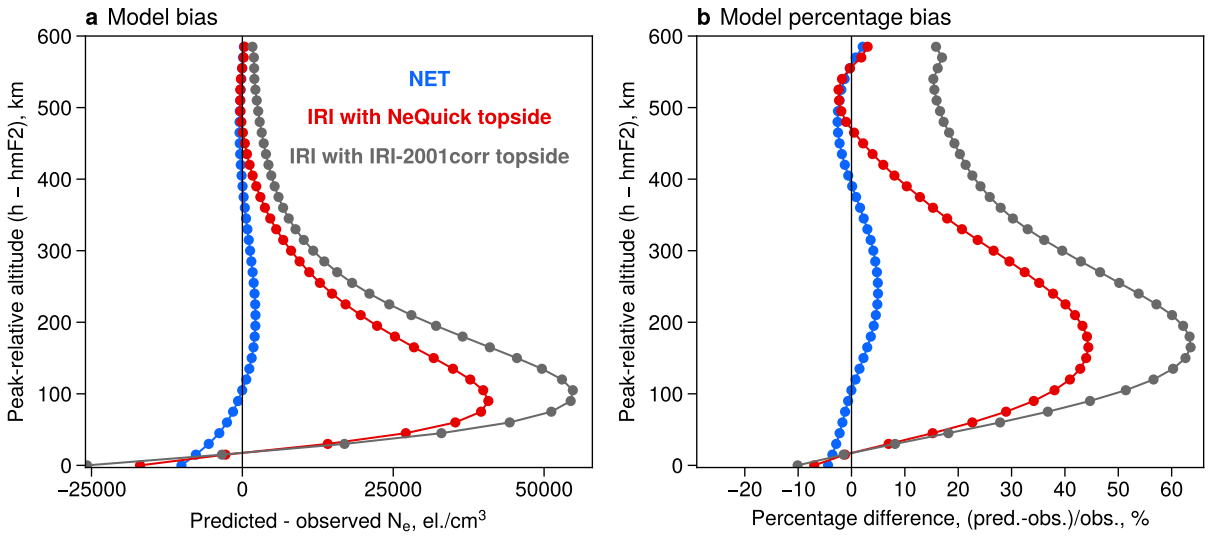

Figure 7.4. Median bias (a) and median percentage bias (b) versus altitude relative to the F2-peak, calculated on the test set of the COSMIC data. Biases of the developed NET model are plotted in blue.Vertical residuals of the IRI-2016 model are shown in red for the NeQuick topside option, and in grey for the IRI-2001corr topside shape.

In particular, the strongest overestimation by the IRI, of up to 40.000 el./cm^3, comes from the region around 100 km above the F2 peak height. Similarly, the developed NET model also underestimates the peak densities, although the bias is roughly 1.5 times smaller than for the IRI. In the topside ionosphere, however, the vertical residuals of the NET model are significantly lower than for the IRI. This is also illustrated in Figure 7.4b that shows the vertical percentage biases. The largest percentage error for the IRI is located at around 150-200 km altitude from the peak. This overestimation reaches >40%, while at higher altitudes, the residuals are smaller and converge to become almost unbiased at 500 km above hmF2. In case of the NET model, the residuals are very small at all altitudes, and do not exceed 5-7%. It is worth noting that in the region where the IRI-2016 exhibits the largest error, the NET model becomes almost unbiased and only shows deviations from the data in the order of several percent. In Figure 7.4, we also demonstrate the median bias by the IRI model when the topside is given by the IRI-2001corr shape function, which is based on the early formulation of Bent et al. [1972]. In Figure 7.4b, one can see that at altitudes $\sim$ 150 km above the peak, the bias is significantly higher for the IRI-2001corr topside option (>60%) than for the NeQuick (~40%).

The seasonal behavior of the models is investigated in Figure S4 in the Supplementary Information from Smirnov et al. [2023]. The Figure shows that the NET model gives largely unbiased predictions for all seasons, as the average ratios between the model predictions and COSMIC observations are very close to 1. In case of the IRI, the model overestimates densities in the local-winter hemispheres for December and June solstices, and overestimates the equatorial ionization anomaly crests during equinoxes. This overestimation corresponds to large values of skill score (up to 80%) and highlights the regions where the developed NET model most significantly outperforms the IRI (see Smirnov et al. [2023], Supplementary Figure S4). Overall, these results demonstrate that the NET model gives unbiased and highly accurate predictions of electron density in the topside ionosphere, remains unbiased for all seasons, and consistently outperforms the IRI model by up to an order of magnitude, especially around 200 km above hmF2 and during the local winters.

7.3.3 Model testing on independent observations

In the previous subsection, the model performance was analyzed on electron densities from the COSMIC mission. Due to the fact that the NET model was trained on COSMIC data, it is crucial to perform an additional validation on the fully independent data sources. The purpose of such validation is to ensure that the model not only reproduces the COSMIC line-of-sight profiles which may smear out some of the ionospheric structures but is also capable of resolving the finer morphology of the topside ionosphere. Therefore, in this section we test the model on 3 completely independent missions, namely, the GRACE, CHAMP and C/NOFS. These missions provide data of the highest quality and have been used in a variety of ionospheric studies, both for empirical modeling and for case studies analyzing specific space weather events.

An example of the model comparison to GRACE-KBR data for one of the days corresponding to equinoctial conditions is shown in Figure 7.5. Note that this day was not used for model training and belongs to the test set. In panel (a), we show electron densities observed by GRACE and predicted by the NET model. The model values are in excellent agreement with the observations, as the NET model correctly captures both the high electron density values corresponding

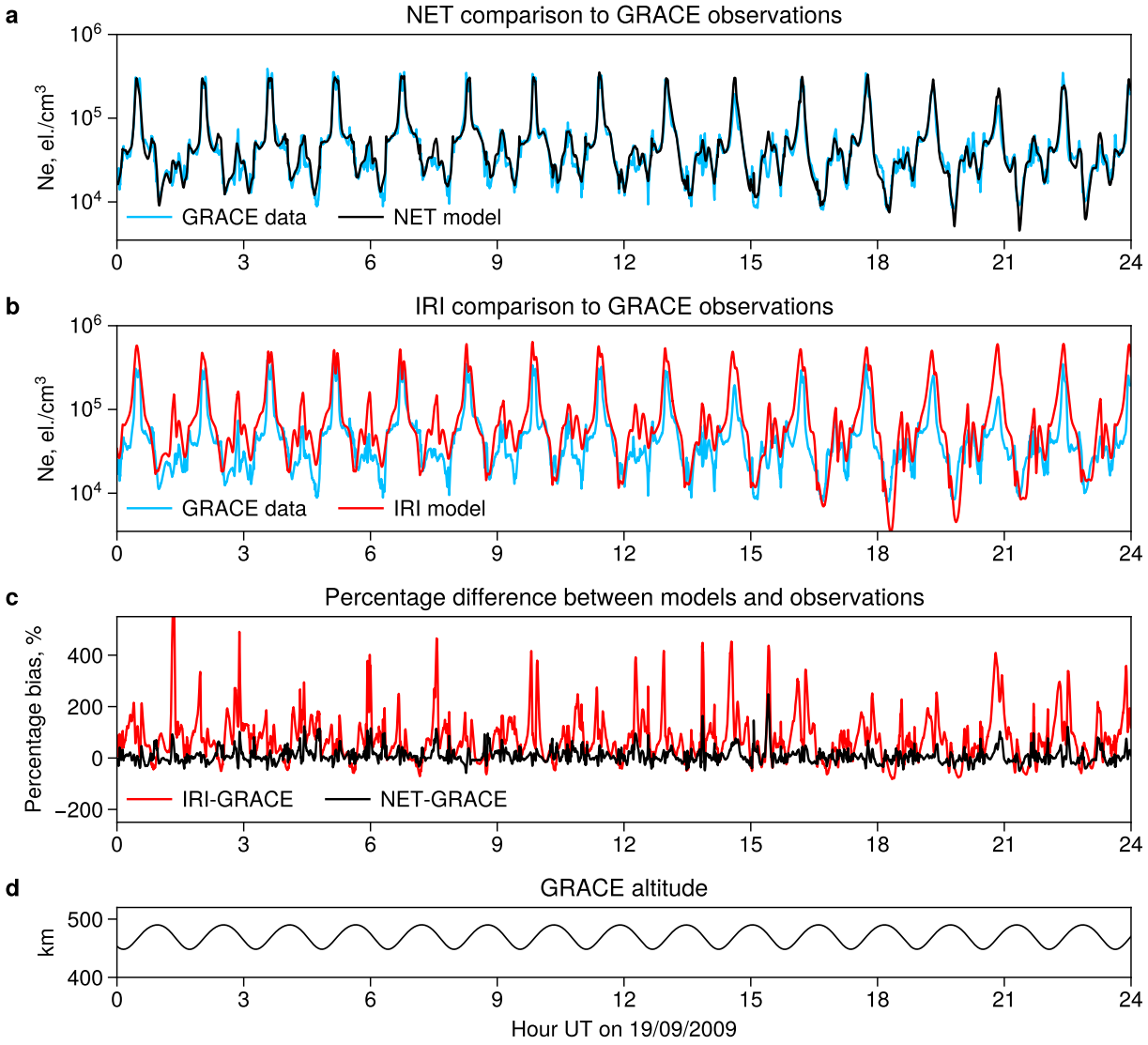

Figure 7.5. GRACE/KBR densities (shown in blue), compared to the NET predictions (panel a), and IRI predictions (subplot b) on 19 September 2009 (an example from the test interval). The percentage differences between the models and observations are given in panel (c). The GRACE altitude is shown in panel (d).

to equatorial latitudes and the low densities typically occurring at higher latitudes. Figure 7.5b gives a comparison between the IRI predictions and GRACE-KBR observations. It is evident that the IRI overestimates electron densities in the equatorial region. This is in good agreement with our previous comparison between the IRI model and COSMIC data (Supplementary Figure S4 in Smirnov et al. [2023]), which showed that the IRI overestimates the EIA crests by approximately a factor of 2. Furthermore, Figure 7.5b also shows that the IRI strongly overestimates

the regions of low densities, which correspond to higher latitudes. In fact, the density depletions shown in Figure 7.5 are due to the midlatitude ionospheric trough, which is known to be challenging region for empirical modeling of the ionosphere. Figure 7.5c gives the percentage bias for both models. One can see that the IRI model overestimates electron densities by up to 400% in the polar regions, and that the model is in general biased towards overestimation. The NET model, on the other hand, shows much lower percentage bias which is centered at zero. It should be noted that throughout the period demonstrated in Figure 7.5, GRACE altitude was about 450-480 km, which is approximately 100-150 km above the F2 peak. In Figure 7.4, this region was highlighted as the most problematic for the IRI with an overestimation of the COSMIC densities by almost 50%. The percentage bias demonstrated in Figure 7.5 agrees well with our findings in Figure 7.4, as the IRI tends to overestimate the GRACE-KBR electron densities in the same way as COSMIC. The NET model shows mostly unbiased predictions along the GRACE altitude.

Table 7.1. Metrics for comparisons between NET and IRI models and observations on 3 independent missions

Metric	**CHAMP-PLP**	**GRACE-KBR**	**C/NOFS-CINDI**
Median bias, NET, [el./cm^3]	11.209	460	-988
Median bias, IRI, [el./cm^3]	34.021	8.708	14.422
Standard deviation, NET, [el./cm^3]	182.585	127.390	101.820
Standard deviation, IRI, [el./cm^3]	218.160	165.048	161.739
Median log bias, NET	0.04	0.003	-0.01
Median log bias, IRI	0.1	0.05	0.12
% of values within a factor of 2, NET	88	89	91
% of values within a factor of 2, IRI	78	79	75
Correlation, NET, [%]	92	92	92
Correlation, IRI, [%]	87	85	86
Skill score of NET over IRI, [%]	30	40	60

Additional examples of the model testing are given in the Supplementary information of Smirnov et al. [2023]; Figure S5 gives an example of a 27 day period from the test set, showing the satellite passes of GRACE and predictions by the NET and IRI models. It can be seen that the NET model reproduces the ionospheric structures very well, and is even able to capture

electron densities in the midlatitude ionospheric trough, which has been shown to be a traditional challenge for ionospheric modeling [Xiong et al., 2013]. The metrics evaluated on the test set show that the NET model predictions are in very good agreement with observations (Table 7.1), as ~90% of the NET electron densities lie within a factor of 2 from the measurements based on CHAMP, GRACE and C/NOFS missions. Furthermore, the NET model exhibits very low values of bias, e.g., $\sim -10^3$el./cm for C/NOFS, which is approximately an order of magnitude smaller than for the IRI. This indicates that the developed NET model is not only capable of reproducing the COSMIC data which were used for model training but also yields highly accurate predictions on completely independent measurements by 3 instruments operating on different observational principles.

7.4 Discussion

The Earth's ionosphere is a region of paramount importance for a variety of scientific and industrial applications. In particular, the ionospheric delays are one of the largest error sources for GNSS navigation. Since these delays are proportional to the total electron density along the ray path, it is crucial to have accurate and reliable models of ionospheric electron densities. The bottomside ionosphere accounts for around 20% of the total electron content, while the main contribution to the TEC magnitudes comes from the topside [Bilitza, 2009]. Modeling of the topside ionosphere has remained a significant challenge, due to data sparsity in both spatial and temporal domains. As a result, the existing ionospheric models show substantial differences from the observations, evident from comparisons to both in-situ measurements and integrated TEC magnitudes [Lühr et al., 2004; Kashcheyev and Nava, 2019; Cherniak and Zakharenkova, 2016]. In this study, we use radio occultation data, which constitute a large data set of electron densities with uniform coverage of the topside, and develop a continuous NN-based model of electron density in the topside ionosphere. The developed NET model yields highly accurate predictions at all altitudes, locations, seasons and solar activity levels and outperforms the international standard of the ionosphere – the IRI – by up to 80%.

The comparison of the IRI electron density predictions to COSMIC EDPs revealed the typical

structure of the vertical biases of the IRI in the topside ionosphere. In particular, our results show that the most significant deviations ($\sim$ 40%) of IRI-2016 with the NeQuick topside from the data are located around 100-200 km above the F2-peak. At higher altitudes, the residuals decrease to the order of several percent. The vertical residuals of the IRI model, shown in Figure 7.4, are likely due to the shape function used to parametrize the topside in the NeQuick model. The NeQuick treats the semi-Epstein scale height as a function of three empirical parameters, namely H_0, which shows the scale height around hmF2, g, representing the gradient of scale height near the peak, and r that controls the asymptotic behavior of the profiles at infinity (for details, see e.g. Nava et al. [2008]). Recent studies have highlighted the rigidity of this parametrization, which is due to the fact that two of the parameters are kept constant ($r = 0.125$ and $g = 100$) while H_0 is modeled as a function of the bottomside profile thickness [Pignalberi et al., 2021; Themens et al., 2018; Pignalberi et al., 2020]. The processes that control the bottomside dynamics are different from those in the topside, and this can yield inaccuracies in the H_0 parametrization. Themens et al. [2018] demonstrated that the NeQuick tends to overestimate H_0 values during the solar minimum conditions and underestimate H_0 during the solar maximum. These results go well with our comparison to COSMIC data shown in Figure 7.4. When H_0 is overestimated, the overall profile shape becomes too convex compared to the data, and the most significant deviations would be found at round 150-200 km above hmF2 (see Figure 6 in Themens et al. [2018]). Furthermore, Pignalberi et al. [2021] recently demonstrated that the topside gradient g, which has a fixed value in the NeQuick, also exhibits strong variations based on geophysical conditions. The study indicated that the NeQuick parametrization could by improved by creating additional sub-models of H_0, r and g. In the current study, we use a flexible parametrization of the topside scale height, allowing the scale height gradient to vary with location, local time, season, solar and geomagnetic activity, and therefore the NET model remains unbiased around 100-200 km above hmF2. Another possible reason for the vertical residuals seen in Figure 7.4 is a potential overestimation of hmF2 by the IRI. Bilitza et al. [2012] demonstrated that overestimated values of hmF2 would shift the profiles upward, which could also create the vertical residual shape seen in the current study in Figure 7.4. We have also investigated the performance of the IRI-2001corr topside parametrization. It was found that the vertical residuals also exhibited

overestimation around 100-200 km above the F2-peak, which was larger than for the NeQuick topside (~60% compared to ~40%). It is evident that the topside specification of the IRI has greatly improved over the years. At altitudes ~ 150km above the peak, the mistakes decreased from ~60% to ~40%, while the main improvement by using the NeQuick option is achieved in the upper topside, where the errors reduced from > 20% to a few percent.

The NET model was found to give unbiased predictions of the topside ionosphere based on the COSMIC data. In particular, at altitudes around 150 km above the F2-peak, which correspond to the largest bias of the IRI-2016, the NET model only exhibits errors in the order of a few percent. Furthermore, we analyzed the seasonal behavior of the models for different local times and magnetic latitudes (Supplementary Figure S4 in Smirnov et al. [2023]). We calculated the skill score values, which highlighted the regions where the NET model outperformed the IRI. The strongest improvement over the IRI (up to 80%) was achieved in the local-winter hemispheres during solstices, and in both EIA crests during equinoxes. Furthermore, we performed model testing on 3 independent data sources, namely the GRACE-KBR, C/NOFS-CINDI and CHAMP-PLP observations. The model was found to give highly accurate predictions on all of these independent data sources, reproducing even the fine structures of the ionosphere including the midlatitudinal ionospheric trough. In fact, over 90% of the NET predictions lie within a factor of 2 from the observations, both in comparisons with radio occultation and in-situ measurements. It should be noted that several existing ML-based models had previously been tested on similar data sets. For instance, Gowtam et al. [2019] tested the ANNIM-3D model on the CHAMP-PLP data and observed correlation ranging from 60% to 79%, based on different locations, while for the NET model the correlations evaluated on the CHAMP data exceeded 90% (Table 7.1).

The NET model is based on observations from the COSMIC mission, which span at altitudes from hmF2 up to around 800-850 km altitude. The linear alpha-Chapman approximation can resolve electron density profiles up to around 1500 km in altitude [Prol et al., 2019], and therefore our results can be used to reconstruct electron densities up to that altitude. Furthermore, it is possible to connect the model presented here to the plasmaspheric altitudes. Recently, Prol et al. [2022] demonstrated that the scale height in the plasmasphere exhibits a quadratic dependence on altitude, and found a functional dependence which allowed to connect the radio occultation

profiles with in-situ observations of electron density in the plasmasphere made by the Van Allen Probes mission. Using the methodology of Prol et al. [2022], it is possible to connect the model presented in this study to the plasmasphere, making it a full topside option valid up to the GNSS altitudes and beyond. Therefore, the model developed in this study combined with its plasmaspheric extension can be incorporated into the IRI as a novel option for specifying the topside ionosphere and plasmasphere.

7.5 Conclusions

In this study, we developed a new empirical model of electron density in the topside ionosphere (NET) based on neural networks. The model uses the geographic and geomagnetic coordinates, local time, day of year, and solar and geomagnetic indices to predict the 4 parameters of the linear alpha-Chapman equation, namely the F2-peak density and height, and 2 parameters of the linear decay of scale height with altitude. The model has been trained and tested on $\sim$ 19 years of radio occultation data and undergone additional validation on in-situ observations by 3 independent missions. The NET model gives highly accurate and unbiased predictions for a variety of geophysical conditions, and outperforms the current topside options included into the International Reference Ionosphere model by up to 80%, especially at altitudes $\sim 100 - 200$ km above the F2-peak and in the local-winter hemispheres. In fact, the NET predictions are within a factor of 2 from the observations ~90% of the time. The NET model can have wide applications in ionospheric research, for instance, in wave propagation studies, for calibrating the new electron density data sets with unknown baseline offsets, for tomographic reconstructions in the form of a background model, as well as to analyze specific space weather events and perform long-term ionospheric reconstructions. Furthermore, the developed model can be connected to plasmaspheric altitudes and thus can become a novel topside option for the IRI. The developed framework allows the seamless incorporation of new data and new data sources. The retraining of the model can be done on a standard PC and can be performed on a regular basis. Another contribution of this work can be incorporation of measurement uncertainty into the training which will allow inclusion of a variety of data sources with various observational errors.

Acknowledgements

The data and model files and example codes are available at https://doi.org/10.5281/zenodo.7504879, and the latest model updates are provided on https://github.com/arsmirnov95/NET_topside_model. The CHAMP, GRACE and COSMIC radio occultation data were obtained from University Corporation for Atmospheric Research (UCAR) through the COSMIC Data Analysis and Archival Center (CDAAC) via the portal https://cdaac-www.cosmic.ucar.edu/. The authors are grateful to Dieter Bilitza for helpful suggestions and discussion. AS acknowledges support from the International Space Science Institute (ISSI - Bern, Switzerland) through the ISSI team on "Imaging the Invisible: Unveiling the Global Structure of the Earth's Dynamic Magnetosphere". This study is supported by the Helmholtz Pilot Projects Information & Data Science II, MAchine learning based Plasma density model project (MAP) ZT-I-0022. FP was supported by the INdoor navigation from CUBesAT Technology (INCUBATE) project, under the grant from the Technology Industries of Finland Centennial Foundation and Jane and Aatos Erkko Foundation.

— 8 —

Conclusions and future work

This dissertation analyzes the dynamics of particle populations in the near-Earth space environment. Specifically, the research presented in this thesis is concerned with the particle populations that are most relevant to satellite operations, namely, the energetic electrons in the radiation belts and ring current, which are known to damage the onboard electronics and can lead to satellite malfunctions or even loss of spacecraft, and the low-energy plasma in the ionosphere, which delays the GNSS signals and constitutes one of the main sources of error in navigation and positioning applications. The behavior of these populations is analyzed using statistical approaches, and the resulting findings are used to create new state-of-the-art data-driven models of the respective regions of geospace.

We used a large dataset of omnidirectional electron flux observations from the CXD detector aboard twenty GPS satellites to create a model of electron flux in MEO orbit for hundreds-of-keV electrons (Chapter 2). While many previous models had been created for high-energy electrons, the behavior of 120-600 keV electrons was considered very difficult to quantify, both in physics-based and empirical models. We have demonstrated with this study that these populations, too, can be predicted very well using ML techniques. Using the LightGBM machine learning algorithm, we developed a model that accurately predicts electron fluxes at MEO. Importantly, the model does not use any previous flux observations as input and only depends on the magnetospheric conditions as reflected in the solar wind driving and geomagnetic indices.

The MERLIN electron flux model reproduces the omnidirectional flux. However, in order to use these predictions for physics-based simulations (e.g., as boundary conditions), it is necessary to have information about the pitch angle distributions of electrons at different energies and locations. Therefore, we then turned to analyzing the equatorial pitch angle distributions using Van Allen Probes data. Due to the ample energy coverage of the Van Allen Probes, it is possible to track multiple processes affecting the PAD dynamics across different energies and locations,

and to understand which parameters can be used as meaningful predictors of the PAD dynamics. We have proposed a new technique of analyzing the PADs is based on Fourier series expansion. While several previous studies based their analyses on Legendre polynomials, we noted several drawbacks of this approach; for instance, the Legendre polynomials cannot be easily integrated over the solid angle to derive the omnidirectional flux, and furthermore, the absolute flux levels are distributed across the PAD coefficients, which means that the PAD coefficients reflect not only the shape of the distribution but also the flux levels. On the other hand, the Fourier sine series expansion allows one to isolate the absolute flux level into a separate coefficient (A0), so that the three coefficients before the sine terms are directly attributed to the main PAD shapes (pancake, butterfly and cap) and can be used to judge their contributions. In addition, this approximation can be analytically integrated to derive the omnidirectional flux. The use of the Fourier series approximation has led to two studies: the first one conducting a statistical analysis of the three Fourier coefficients during geomagnetic storms (Chapter 3), and the second one devoted to modeling of PADs in terms of these coefficients (Chapter 4).

In order to understand the dynamics of electron PADs, we performed a superposed epoch analysis of 129 strong geomagnetic storms throughout the entire Van Allen Probes era. We analyzed the PAD coefficients obtained from the Fourier sine series fits, and investigated their dynamics for different energies (from ∼30 keV to ∼1.6 MeV) and radial distances for day- and nighttime MLTs. We found that the Fourier coefficients reflect the PAD dynamics very well, for example, the cap PADs are well represented by this approximation, and it was found that the cap PADs transform into the pancake PADs during the main phase of the storm, and then reform during the recovery phase. The other two main PAD types, namely the butterfly and pancake PADs, can also be analyzed using this approximation, and their behavior is consistent with previous work: the pancake PADs are dominant on the dayside and become narrower with increasing activity, while the butterfly distributions are prevalent on the nightside and their characteristic 90°-minimum intensifies during the main phase.

Using the results of the superposed epoch analysis, we then constructed a model of the equatorial electron PADs. While previous models mainly constituted statistical averages for spatial and activity bins, our goal was to create, for the first time, a model that would include a continu-

ous dependence on L, MLT and activity. The developed model reproduces the normalized PAD shapes and is driven by the solar wind dynamic pressure. We also tried several other predictor variables, such as the geomagnetic indices, solar wind velocity, but found the dynamic pressure to be a better proxy, as the resulting model was in better agreement with the previous findings from the superposed epoch analysis. We analyzed the model behavior for different energies and dynamic pressure levels and found that it provided a good representation of the physical phenomena, capturing the changes in dynamics for all three main types of pitch angle distributions in the outer radiation belt. We also developed two techniques for using this model in combination with electron flux observations. For many missions, only omni- or unidirectional electron flux values are provided, and it is therefore interesting to use the developed model to fill in the information about PADs for these observations. We developed a suitable technique for cases when 2 unidirectional flux observations are available (which is the case for POES and MetOP LEO satellites), as well as for datasets where only one data point is available at each time. Both techniques provided high quality reconstructions of high-PA observations from points with low equatorial PAs. This model is applicable to a variety of tasks in inner-magnetospheric research. For example, it can be used to convert data sets with no pitch angle information to phase space densities and use these observations as boundary conditions for physics-based radiation belt modeling. Furthermore, the combination of available LEO observations with the PAD model can be used for radiation belt forecast, which has already been implemented in the radiation belt forecasting system at GFZ.

One of the missions that provides, for the most part, omnidirectional electron flux is the Cluster mission. In particular, there are two channels on the RAPID/IES instrument aboard Cluster that can provide pitch angle distributions. At the same time, the Cluster observations could be valuable for radiation belt studies, especially for setting up the outer boundary conditions. Therefore, we converted the electron fluxes measured by Cluster/RAPID/IES into the phase space densities in terms of the adiabatic invariants (Chapter 5). The developed data product has been provided to ESA and has become one of the standard products of the Cluster Science Archive. Furthermore, by analyzing the solar cycle variations of the phase space densities for low-μ electrons, we found that their behavior is similar to that of the AE index.

The Earth's ionosphere is an important medium, not least because it delays GNSS signals and

can introduce positioning errors of up to tens of meters for single-frequency positioning. The ionosphere has traditionally been challenging to model. The physics-based models are based on simplified assumptions and are generally computationally expensive to run, while the empirical models suffer from uneven data coverage and the use of functional approximations that could only reproduce the climatology of the ionosphere. In order to move towards new-generation weather-like ionospheric models trained on a variety of measurements, it is necessary to ensure that they are based on data sources that are well-calibrated and have sufficient coverage of all geographic locations and solar activity levels, as well as to use state-of-the-art modeling techniques.

We first addressed the task of providing an intercalibration for the existing most commonly used ionospheric data sets (Chapter 6). Although many previous studies evaluated the agreement between different data sources, the measurements were not systematically calibrated, and no general reference frame to calibrate the instruments had been provided. The gold standard data set would be that of the incoherent scatter radars, but they do not provide enough spatial coverage to evaluate systematic differences with other data sources as a function of altitude, MLT, and magnetic latitude. In this dissertation, we have used the GRACE-KBR electron density data as such a reference, due to the fact that they are in excellent agreement with the ISR observations and also provide global coverage. However, the GRACE-KBR data are orbit-bound, and in order to calibrate in situ data sets at different altitudes, a secondary 3D reference had to be selected. We showed that the COSMIC-RO observations were in excellent agreement with the GRACE-KBR data and could therefore be used as such a secondary reference. We then calibrated the CHAMP-PLP and C/NOFS-CINDI data to this joint reference frame. Interestingly, the Swarm Langmuir Probe observations were in good agreement with the COSMIC data during daytime (with a correlation of 95%), but showed substantial scatter at nighttime (with a much lower correlation of 74%). To investigate the possible reason behind this discrepancy, we binned the conjunctions in magnetic latitude and local time. This analysis revealed that the daytime densities were lower in Swarm than in COSMIC, while on the night side Swarm LPs overestimated electron densities by up to 40-50%. Ours was the first study to explicitly demonstrate the nighttime overestimation by Swarm LPs and suggest the influence of light ions as a potential cause. This effect has been

confirmed in multiple articles, both in comparison with empirical models [Bilitza and Xiong, 2021], when comparing the LP data to in situ observations by Swarm Faceplate (FP) [Catapano et al., 2022; Xiong et al., 2022; Pignalberi et al., 2022], and in conjunction with the ISRs [Fast et al., 2023].

As a next step, the intercalibrated observations could be used to develop and validate a new model of the topside ionosphere (Chapter 7). Artificial neural networks were selected as a modeling tool, since they allow to produce a continuous model output and can learn highly complex dependencies from the data. To reduce the data dimensions, we employed radio occultation data and fitted the topside profiles using alpha-Chapman function with a linear decay of scale height with altitude. We developed a model based on RO observations and validated it against the intercalibrated in situ measurements, to compare the performance with the International Reference Ionosphere. We analyzed the vertical model residuals of the IRI model and showed that the most significant deviations from the data are found about 100-200 km above the F2-layer peak, which has not been demonstrated before. The developed NET model was found to be in excellent agreement with observations for all independent missions, as for each of the data sets the model predictions were within a factor of 2 from the data > 90% of the time. We have shown that the ionospheric electron densities can be reconstructed with very high fidelity, even in regions traditionally considered challenging for ionospheric modeling, such as the midlatitude trough. The NET model can also be extended to plasmaspheric altitudes, and therefore can be incorporated as a new topside option for the IRI.

Over the last few decades, enormous volumes of data have been provided across different regions of geospace. These data include satellite and ground-based measurements, as well as remote sensing observations. This wealth of data provides unprecedented opportunities for improving the existing models of the near-Earth space and for the development of new modeling tools. In particular, with ever increasing amount of data, one should select approaches that would make use of all useful information and therefore select powerful modeling tools. In this dissertation, several machine learning-based models have been created. The models developed in this thesis have contributed to our understanding of what constitutes the most meaningful predictor variables for the respective problems. Our results also revealed where the most prominent mod-

els have limitations, which will likely lead to their re-evaluation and improvement. The rapid advancements in machine learning, as well as the availability of large computational resources, enabled the research presented here to emerge at this point in time. There is little doubt that the ongoing machine learning revolution will further improve our capabilities of modeling and understanding of the near-Earth space environment.

Future work

The work presented in this dissertation lays the foundation for several follow-up studies, and the results obtained in this thesis will be extended in the future. The potential directions for future studies are outlined below.

There are several ways to expand the research on radiation belts and ring current electrons presented in Chapters 2-5. In particular, one can extend the MERLIN electron flux model to provide a wider energy and spatial coverage. Since the model is based on GPS data, it is effectively bound to the GPS altitude, even though the satellites pass through a range of L-shells (L= 4 – 9) due to their high inclination. This limits the sampled equatorial pitch angles at high L-shells to a range of a few degrees, and this limitation can be overcome by using the data from equatorial spacecraft, such as the Van Allen Probes. Furthermore, an even stronger improvement would be to include the equatorial pitch angles into the model inputs, which would allow one to model not only the omnidirectional flux but also the pitch angle distributions. The output of such a model would be much more convenient for converting electron fluxes to phase space densities. Another approach, which could find many applications in inner-magnetospheric research, would be to first convert electron fluxes measured by the Van Allen Probes to phase space densities, and then model the PSD as a function of adiabatic invariants and geophysical conditions. Preliminary work in this area has recently been performed by Wing et al. [2022]; their model showed good performance at high L-shells but achieved near-zero prediction efficiency around the most critical area in the heart of the outer belt. Therefore, there are many possibilities for developing new empirical PSD models, and especially for selecting the input features that could improve the model performance.

Results from the machine learning models can be combined with physics-based radiation belts simulations. For example, the MERLIN model can be used to compute electron flux values for the GEM challenge events, and the results can be compared with other state-of-the-art models, including the physics-based ones (e.g., VERB, BAS-RBM, DREAM-3D). These comparisons could highlight areas and conditions that are better described by the respective models, and would also indicate where the other models have limitations. This would likely lead to mutual improvement of all models and methodologies involved in such a validation effort. Another way to use the ML-based results in combination with physics-based models is to use the ML predictions as boundary conditions for numerical simulations. For instance, the MERLIN electron flux model can be combined with the developed PAD model and can be used for such a task. Moreover, the development of a ML-based phase space density model would be ideally suited for this task. The outer boundary of the radiation belt simulations can also be extended to higher L-shells by using data from the Cluster mission. In particular, electron observations from the Cluster/RAPID/IES detector have already been converted to PSD in terms of adiabatic invariants, and therefore these data can be used to develop a PSD model at large radial distances.

The ionospheric modeling results presented in this dissertation can be extended in the future. First, the data intercalibration efforts undertaken in this dissertation (Chapter 6), should be continued as more data become available. Several new data sets have recently been released to the public, including the Langmuir probe observations from the International Space Station, and they can be adjusted to the reference frame that was developed in this thesis. Moreover, plasma density observations by the Defense Meteorological Satellite Program (DMSP), frequently used in ionospheric studies, may require calibration, and the framework developed in this thesis can provide a good reference for this task. In addition, the planned missions, such as the Geospace Dynamics Constellation (GDC), can be adjusted to this reference and can contribute to the development and validation of data-driven ionospheric models. Furthermore, we are currently working on the correction algorithm for the Swarm LP ion density data. The idea behind this study is based on the fact that Swarm satellites are equipped with a Faceplate, which is similar in design to a cylindrical LP and allows electron density derivation. Prior to 2023, the Faceplate operated for several orbits per day, and therefore it cannot be used as a replacement for the LP

data set (note that there are now plans to increase the FP operating times). However, machine learning techniques are very well suited for finding dependencies from sparse observations, and therefore one can use neural networks to learn the ratio between FP and LP observations. The LP data can then be corrected using these modeled factors, and this can be done for the entire duration of the mission, even when FP data are not available, and will likely help to resolve the nighttime overestimation by Swarm LPs.

The data-driven modeling of the ionosphere started in this dissertation (Chapter 7) offers many opportunities for further investigation. In this thesis, we focused on modeling the topside ionosphere and used a Chapman function with an assumption of a linear scale height decay with altitude. This approximation works well for the topside ionosphere, but may not be able to resolve a change in ionospheric profiles that occurs at the upper transition height, after which the dependence becomes approximately quadratic [Prol et al., 2019, 2022]. Therefore, in future studies we plan to include the satellite altitude as one of the inputs and use all of the available intercalibrated observations from the CHAMP, GRACE, COSMIC, C/NOFS and Swarm missions, allowing the neural network to learn the profile shape without making any underlying assumptions. Furthermore, the current data set used for modeling covers altitudes up to ~800 km, and in order to extend the model to higher altitudes we plan to use the topside sounder observations from the 1960s-1970s. These measurements cover altitudes up to around 3000 km, and also sample much more active solar cycle conditions, compared to the recent solar cycles 23 and 24, which would further enhance the model capabilities.

Model development should be closely followed by validation on independent data, and several observational techniques can be used for this task. After updating the model and incorporating the topside sounder observations, it will be possible to validate the model using VTEC observations obtained from satellite altimetry. Recently, Azpilicueta and Nava [2021] showed that the VTEC data provided by the Jason-1 satellite are almost bias-free (with a total bias of ~1 TECU). Therefore, one can integrate the vertical profiles provided by the updated NET model to derive VTEC between the ground and the satellite altitude of ~1336 km, and use the Jason-1 altimeter data for independent validation. It should also be noted that the NET model developed in this dissertation can participate in the independent validation efforts of the Community Co-

ordinated Modeling Center (CCMC). The CCMC frequently analyzes the available ionospheric models [e.g., Shim et al., 2008, 2023], and provides the accuracy rankings. Moreover, the NET model, which is already openly accessible, can be added to the pool of CCMC models to provide on-demand runs.

GNSS VTEC maps also constitute a useful data set for model validation. However, since the values are integrated from the ground up to GPS altitude, it is necessary to include a plasmaspheric extension into the NET model. One such technique was recently developed by Prol et al. [2022]. The authors used Van Allen Probes data in combination with COSMIC-RO observations to investigate the scale height behavior in the plasmasphere, and showed that the RO profiles can be well linked to the plasmaspheric heights using a statistical model. It is possible to combine the NET model with this plasmaspheric extension, integrate the density profiles in altitude, and validate the results by the VTEC maps. The next step would be to include the plasmaspheric observations into the NET, and create a combined ionosphere-plasmasphere model. It should be noted that this would require a calibration of the available plasmaspheric observations (e.g., from IMAGE, Cluster and Polar missions), and there is a gap in altitude between the available data sets in the ionosphere and plasmasphere, from about 1000 km up to around 5000-6000 km altitude [Prol et al., 2022]. The recently launched Arase constellation [Miyoshi et al., 2018] may help to partially fill in this gap, as it provides electron densities inferred from the upper-hybrid resonance frequency from around 560 km up to ~32000 km.

Bibliography

Albert, J. M., Starks, M. J., Horne, R. B., Meredith, N. P., and Glauert, S. A. (2016). Quasi-linear simulations of inner radiation belt electron pitch angle and energy distributions. *Geophysical Research Letters*, 43(6):2381–2388.

Alken, P., Thébault, E., Beggan, C. D., Amit, H., Aubert, J., Baerenzung, J., Bondar, T., Brown, W., Califf, S., Chambodut, A., et al. (2021). International geomagnetic reference field: the thirteenth generation. *Earth, Planets and Space*, 73(1):1–25.

Allison, H. J., Horne, R. B., Glauert, S. A., and Del Zanna, G. (2018). Determination of the equatorial electron differential flux from observations at low Earth orbit. *Journal of Geophysical Research: Space Physics*, 123(11):9574–9596.

Allison, H. J. and Shprits, Y. Y. (2020). Local heating of radiation belt electrons to ultra-relativistic energies. *Nature Communications*, 11(1):4533.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Altadill, D., Magdaleno, S., Torta, J., and Blanch, E. (2013). Global empirical models of the density peak height and of the equivalent scale height for quiet conditions. *Advances in Space Research*, 52(10):1756–1769.

Aponte, N., Brum, C. G., Sulzer, M. P., and González, S. A. (2013). Measurements of the O+ to H+ transition height and ion temperatures in the lower topside ionosphere over Arecibo for equinox conditions during the 2008–2009 extreme solar minimum. *Journal of Geophysical Research: Space Physics*, 118(7):4465–4470.

Appleton, E. and Beynon, W. (1947). The application of ionospheric data to radio communication problems: Part II. *Proceedings of the Physical Society (1926-1948)*, 59(1):58.

Arras, C., Jacobi, C., Wickert, J., Heise, S., and Schmidt, T. (2010). Sporadic E signatures revealed from multi-satellite radio occultation measurements. *Advances in Radio Science*, 8(GHJ. 1-1/1-):225–230.

Artemyev, A., Agapitov, O., Mozer, F., and Spence, H. (2015). Butterfly pitch angle distribution of relativistic electrons in the outer radiation belt: Evidence of nonadiabatic scattering. *Journal of Geophysical Research: Space Physics*, 120(6):4279–4297.

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.

Astafyeva, E. (2019). Ionospheric detection of natural hazards. *Reviews of Geophysics*, 57(4):1265–1288.

Auret, L. and Aldrich, C. (2011). Empirical comparison of tree ensemble variable importance measures. *Chemometrics and Intelligent Laboratory Systems*, 105(2):157–170.

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.

Azpilicueta, F. and Nava, B. (2021). On the TEC bias of altimeter satellites. *Journal of Geodesy*, 95:1–15.

Baker, D., Belian, R., Higbie, P., Klebesadel, R., and Blake, J. (1987). Deep dielectric charging effects due to high-energy electrons in earth's outer magnetosphere. *Journal of Electrostatics*, 20(1):3–19.

Baker, D., Erickson, P., Fennell, J., Foster, J., Jaynes, A., and Verronen, P. (2018). Space weather effects in the Earth's radiation belts. *Space Science Reviews*, 214:1–60.

Baker, D., Kanekal, S., Hoxie, V., Henderson, M., Li, X., Spence, H. E., Elkington, S., Friedel, R., Goldstein, J., Hudson, M., et al. (2013). A long-lived relativistic electron storage ring embedded in Earth's outer Van Allen belt. *Science*, 340(6129):186–190.

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 Halloween solar storm in 2003. *Nature*, 432(7019):878–881.

Baker, D., McPherron, R., Cayton, T., and Klebesadel, R. (1990). Linear prediction filter analysis of relativistic electron properties at 6.6 RE. *Journal of Geophysical Research: Space Physics*, 95(A9):15133–15140.

Baker, D., Pulkkinen, T., Li, X., Kanekal, S., Ogilvie, K., Lepping, R., Blake, J., Callis, L., Rostoker, G., Singer, H., et al. (1998). A strong CME-related magnetic cloud interaction with the Earth's magnetosphere: ISTP observations of rapid relativistic electron acceleration on May 15, 1997. *Geophysical Research Letters*, 25(15):2975–2978.

Bala, R., Reiff, P., and Landivar, J. (2009). Real-time prediction of magnetospheric activity using the Boyle Index. *Space Weather*, 7(4):1–17.

Balikhin, M. A., Boynton, R. J., Walker, S. N., Borovsky, J. E., Billings, S. A., and Wei, H.-L. (2011). Using the narmax approach to model the evolution of energetic electrons fluxes at geostationary orbit. *Geophysical Research Letters*, 38(18).

Baumjohann, W. and Treumann, R. A. (2012). *Basic space plasma physics*. World Scientific.

Benson, R. F. and Bilitza, D. (2009). New satellite mission with old data: Rescuing a unique data set. *Radio Science*, 44(01):1–14.

Bent, R., Llewellyn, S., and Schmid, P. (1972). A highly successful empirical model for the worldwide ionospheric electron density profile. *DBA Systems, Melbourne, Florida, USA*.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24.

Bergstra, J., Yamins, D., and Cox, D. D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. pages 13–20.

Biermann, L. (1951). Kometenschweife und solare korpuskularstrahlung. *Zeitschrift fur Astrophysik*, 29:274.

Bilitza, D. (2004). A correction for the IRI topside electron density model based on Alouette/ISIS topside sounder data. *Advances in Space Research*, 33(6):838–843.

Bilitza, D. (2009). Evaluation of the IRI-2007 model options for the topside electron density. *Advances in Space Research*, 44(6):701–706.

Bilitza, D. (2018). IRI the International Standard for the Ionosphere. *Advances in Radio Science*, 16:1–11.

Bilitza, D., Altadill, D., Truhlik, V., Shubin, V., Galkin, I., Reinisch, B., and Huang, X. (2017). International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions. *Space weather*, 15(2):418–429.

Bilitza, D., Brown, S. A., Wang, M. Y., Souza, J. R., and Roddy, P. A. (2012). Measurements and IRI model predictions during the recent solar minimum. *Journal of Atmospheric and Solar-Terrestrial Physics*, 86:99–106.

Bilitza, D., McKinnell, L.-A., Reinisch, B., and Fuller-Rowell, T. (2011). The international reference ionosphere today and in the future. *Journal of Geodesy*, 85(12):909–920.

Bilitza, D., Reinisch, B. W., Radicella, S. M., Pulinets, S., Gulyaeva, T., and Triskova, L. (2006). Improvements of the International Reference Ionosphere model for the topside electron density profile. *Radio Science*, 41(05):1–8.

Bilitza, D. and Xiong, C. (2020). A solar activity correction term for the IRI topside electron density model. *Advances in Space Research*.

Bilitza, D. and Xiong, C. (2021). A solar activity correction term for the IRI topside electron density model. *Advances in Space Research*, 68(5):2124–2137.

Blake, J., Carranza, P., Claudepierre, S., Clemmons, J., Crain, W., Dotan, Y., Fennell, J., Fuentes, F., Galvan, R., George, J., et al. (2013). The magnetic electron ion spectrometer (MagEIS)

instruments aboard the radiation belt storm probes (RBSP) spacecraft. *The Van Allen Probes Mission*, pages 383–421.

Borovsky, J. E. (2014). Feedback of the magnetosphere. *Science*, 343(6175):1086–1087.

Borovsky, J. E. and Denton, M. H. (2008). A statistical look at plasmaspheric drainage plumes. *Journal of Geophysical Research: Space Physics*, 113(A9).

Borovsky, J. E. and Shprits, Y. Y. (2017). Is the Dst index sufficient to define all geospace storms? *Journal of Geophysical Research: Space Physics*, 122(11):11–543.

Bortnik, J., Chu, X., Ma, Q., Li, W., Zhang, X., Thorne, R. M., Angelopoulos, V., Denton, R. E., Kletzing, C. A., Hospodarsky, G. B., et al. (2018). Artificial neural networks for determining magnetospheric conditions. In *Machine learning techniques for space weather*, pages 279–300. Elsevier.

Bortnik, J., Li, W., Thorne, R., and Angelopoulos, V. (2016a). A unified approach to inner magnetospheric state prediction. *Journal of Geophysical Research: Space Physics*, 121(3):2423–2430.

Bortnik, J., Li, W., Thorne, R. M., and Angelopoulos, V. (2016b). A unified approach to inner magnetospheric state prediction. *Journal of Geophysical Research: Space Physics*, 121(3):2423–2430.

Boscher, D., Bourdarie, S., O'Brien, P., Guild, T., and Shumko, M. (2012). Irbem-lib library.

Boyd, A., Spence, H., Huang, C.-L., Reeves, G., Baker, D., Turner, D., Claudepierre, S., Fennell, J., Blake, J., and Shprits, Y. (2016). Statistical properties of the radiation belt seed population. *Journal of Geophysical Research: Space Physics*, 121(8):7636–7646.

Boyd, A. J., Spence, H. E., Claudepierre, S., Fennell, J. F., Blake, J., Baker, D., Reeves, G., and Turner, D. (2014). Quantifying the radiation belt seed population in the 17 march 2013 electron acceleration event. *Geophysical Research Letters*, 41(7):2275–2281.

Boynton, R., Balikhin, M., Sibeck, D., Walker, S., Billings, S., and Ganushkina, N. (2016). Electron flux models for different energies at geostationary orbit. *Space Weather*, 14(10):846–860.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, 17(8):1166–1207.

Carbary, J., Mitchell, D., Paranicas, C., Roelof, E., Krimigis, S., Krupp, N., Khurana, K., and Dougherty, M. (2011). Pitch angle distributions of energetic electrons at Saturn. *Journal of Geophysical Research: Space Physics*, 116(A1).

Carver, M. R., Sullivan, J. P., Morley, S. K., and Rodriguez, J. V. (2018). Cross calibration of the GPS constellation CXD proton data with GOES EPS. *Space Weather*, 16(3):273–288.

Castillo Tibocha, A. M., de Wiljes, J., Shprits, Y. Y., and Aseev, N. A. (2021). Reconstructing the dynamics of the outer electron radiation belt by means of the standard and ensemble Kalman filter with the verb-3D code. *Space Weather*, 19(10):e2020SW002672.

Catapano, F., Buchert, S., Qamili, E., Nilsson, T., Bouffard, J., Siemes, C., Coco, I., D'Amicis, R., Tøffner-Clausen, L., Trenchi, L., et al. (2022). Swarm Langmuir probes' data quality validation and future improvements. *Geoscientific Instrumentation, Methods and Data Systems*, 11(1):149–162.

Chamberlain, J. (1961). *Physics of the aurora and airglow*. ACADEMIC PRESS, New York & London.

Chapman, S. (1931). The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth. *Proceedings of the Physical Society*, 43(1):26.

Chapman, S. and Bartels, J. (1940). Geomagnetism, vol. ii: Analysis of the data, and physical theories. *Geomagnetism*.

Chen, F. F. et al. (1984). *Introduction to plasma physics and controlled fusion*, volume 1. Springer.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, Y., Friedel, R., and Reeves, G. (2006). Phase space density distributions of energetic electrons in the outer radiation belt during two geospace environment modeling inner magnetosphere/storms selected storms. *Journal of Geophysical Research: Space Physics*, 111(A11).

Chen, Y., Friedel, R., Reeves, G., Onsager, T., and Thomsen, M. (2005). Multisatellite determination of the relativistic electron phase space density at geosynchronous orbit: Methodology and results during geomagnetically quiet times. *Journal of Geophysical Research: Space Physics*, 110(A10).

Chen, Y., Friedel, R. H., Henderson, M. G., Claudepierre, S. G., Morley, S. K., and Spence, H. E. (2014). REPAD: An empirical model of pitch angle distributions for energetic electrons in the Earth's outer radiation belt. *Journal of Geophysical Research: Space Physics*, 119(3):1693–1708.

Chen, Y., Reeves, G. D., Fu, X., and Henderson, M. (2019). PreMevE: New predictive model for megaelectron-volt electrons inside earth's outer radiation belt. *Space Weather*, 17(3):438–454.

Cherniak, I. and Zakharenkova, I. (2016). NeQuick and IRI-Plas model performance on topside electron content representation: spaceborne GPS measurements. *Radio Science*, 51(6):752–766.

Cherniak, I. and Zakharenkova, I. (2019). Evaluation of the IRI-2016 and NeQuick electron content specification by COSMIC GPS radio occultation, ground-based GPS and Jason-2 joint altimeter/GPS observations. *Advances in Space Research*, 63(6):1845–1859.

Cherniak, I., Zakharenkova, I., Braun, J., Wu, Q., Pedatella, N., Schreiner, W., Weiss, J.-P., and Hunt, D. (2021). Accuracy assessment of the quiet-time ionospheric F2 peak parameters as derived from COSMIC-2 multi-GNSS radio occultation measurements. *Journal of Space Weather and Space Climate*, 11:18.

Cherniak, I. V. and Zakharenkova, I. (2014). Validation of FORMOSAT-3/COSMIC radio occultation electron density profiles by incoherent scatter radar data. *Advances in Space Research*, 53(9):1304–1312.

Cherniak, I. V., Zakharenkova, I., Krankowski, A., and Shagimuratov, I. (2012). Plasmaspheric electron content derived from GPS TEC and FORMOSAT-3/COSMIC measurements: Solar minimum condition. *Advances in space research*, 50(4):427–440.

Chollet, F. et al. (2015). Keras. `https://keras.io`.

Chu, X., Ma, D., Bortnik, J., Tobiska, W. K., Cruz, A., Bouwer, S. D., Zhao, H., Ma, Q., Zhang, K., Baker, D. N., et al. (2021). Relativistic electron model in the outer radiation belt using a neural network approach. *Space Weather*, 19(12):e2021SW002808.

Chuo, Y.-J., Lee, C.-C., Chen, W.-S., and Reinisch, B. W. (2011). Comparison between bottomside ionospheric profile parameters retrieved from FORMOSAT3 measurements and ground-based observations collected at Jicamarca. *Journal of atmospheric and solar-terrestrial physics*, 73(13):1665–1673.

Clark, G., Paranicas, C., Santos-Costa, D., Livi, S., Krupp, N., Mitchell, D., Roussos, E., and Tseng, W.-L. (2014). Evolution of electron pitch angle distributions across Saturn's middle magnetospheric region from MIMI/LEMMS. *Planetary and Space Science*, 104:18–28.

Claudepierre, S., Blake, J. B., Boyd, A., Clemmons, J., Fennell, J., Gabrielse, C., Looper, M., Mazur, J., O'Brien, T., Reeves, G., et al. (2021). The Magnetic Electron Ion Spectrometer: A Review of On-Orbit Sensor Performance, Data, Operations, and Science. *Space science reviews*, 217(8):1–67.

Claudepierre, S. G. and O'Brien, T. P. (2020). Specifying high-altitude electrons using low-altitude leo systems: The SHELLS model. *Space Weather*, 18(3):e2019SW002402.

Claudepierre, S. G., O'Brien, T. P., Looper, M. D., Blake, J. B., Fennell, J. F., Roeder, J. L., Clemmons, J. H., Mazur, J. E., Turner, D. L., Reeves, G. D., and Spence, H. E. (2019). A Revised Look at Relativistic Electrons in the Earth's Inner Radiation Zone and Slot Region. *Journal of Geophysical Research: Space Physics*, 124(2):934–951.

Coïsson, P., Radicella, S., Leitinger, R., and Nava, B. (2006). Topside electron density in IRI and NeQuick: Features and limitations. *Advances in Space Research*, 37(5):937–942.

Colak, T. and Qahwaji, R. (2009). Automated solar activity prediction: a hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. *Space Weather*, 7(6).

Coley, W., Heelis, R., Hairston, M., Earle, G., Perdue, M., Power, R., Harmon, L., Holt, B., and Lippincott, C. (2010). Ion temperature and density relationships measured by CINDI from the C/NOFS spacecraft during solar minimum. *Journal of Geophysical Research: Space Physics*, 115(A2).

Constable, C. (2016). Earth's electromagnetic environment. *Surveys in Geophysics*, 37(1):27–45.

Costa, E., Roddy, P., and Ballenthin, J. (2014). Statistical analysis of C/NOFS planar Langmuir probe data. In *Annales Geophysicae*, volume 32, pages 773–791. Copernicus GmbH.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

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.

Davies, K. (1965). *Ionospheric radio propagation*, volume 80. US Department of Commerce, National Bureau of Standards.

de La Beaujardière, O. et al. (2004). C/NOFS: A mission to forecast scintillations. *Journal of Atmospheric and Solar-Terrestrial Physics*, 66(17):1573–1591.

de Lima, R. P., Chen, Y., and Lin, Y. (2020). Forecasting Megaelectron-Volt Electrons Inside Earth's Outer Radiation Belt: PreMevE 2.0 Based on Supervised Machine Learning Algorithms. *Space Weather*, 18(2).

DeForest, S. E. and McIlwain, C. E. (1971). Plasma clouds in the magnetosphere. *Journal of Geophysical Research*, 76(16):3587–3611.

Deisenroth, M. P., Faisal, A. A., and Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.

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, M., Taylor, M., Rodriguez, J., and Henderson, M. (2019). Extension of an empirical electron flux model from 6 to 20 Earth radii using Cluster/RAPID observations. *Space Weather*, 17(5):778–792.

Denton, M., Thomsen, M., Jordanova, V., Henderson, M., Borovsky, J., Denton, J., Pitchford, D., and Hartley, D. (2015). An empirical model of electron and ion fluxes derived from observations at geosynchronous orbit. *Space Weather*, 13(4):233–249.

Ding, C., Wang, D., Ma, X., and Li, H. (2016). Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustainability*, 8(11):1100.

Drozdov, A. Y., Aseev, N., Effenberger, F., Turner, D. L., Saikin, A., and Shprits, Y. Y. (2019). Storm time depletions of multi-MeV radiation belt electrons observed at different pitch angles. *Journal of Geophysical Research: Space Physics*, 124(11):8943–8953.

Druzhkov, P., Erukhimov, V., Zolotykh, N. Y., Kozinov, E., Kustikova, V., Meerov, I., and Polovinkin, A. (2011). New object detection features in the OpenCV library. *Pattern Recognition and Image Analysis*, 21(3):384.

Dungey, J. W. (1961). Interplanetary magnetic field and the auroral zones. *Physical Review Letters*, 6(2):47.

Escoubet, C., Schmidt, R., and Goldstein, M. (1997). Cluster–science and mission overview. *Space Science Reviews*, 79(1-2):11–32.

Farley, D., Bonelli, E., Fejer, B. G., and Larsen, M. (1986). The prereversal enhancement of the zonal electric field in the equatorial ionosphere. *Journal of Geophysical Research: Space Physics*, 91(A12):13723–13728.

Fast, H., Koustov, A., and Gillies, R. (2023). Validation of Swarm Langmuir Probes by Incoherent Scatter Radars at High Latitudes. *Remote Sensing*, 15(7):1846.

Fennell, J., Claudepierre, S., Blake, J., O'Brien, T., Clemmons, J., Baker, D., Spence, H. E., and Reeves, G. (2015). Van Allen Probes show that the inner radiation zone contains no MeV electrons: ECT/MagEIS data. *Geophysical Research Letters*, 42(5):1283–1289.

Fennell, J. F., Koons, H. C., Chen, M. W., and Blake, J. (2000). Internal charging: A preliminary environmental specification for satellites. *IEEE transactions on plasma science*, 28(6):2029–2036.

Fok, M.-C. and Moore, T. (1997). Ring current modeling in a realistic magnetic field configuration. *Geophysical research letters*, 24(14):1775–1778.

Fok, M.-C., Wolf, R., Spiro, R., and Moore, T. (2001). Comprehensive computational model of Earth's ring current. *Journal of Geophysical Research: Space Physics*, 106(A5):8417–8424.

Fonda, C., Coïsson, P., Nava, B., Radicella, S. M., et al. (2005). Comparison of analytical functions used to describe topside electron density profiles with satellite data. *Annals of Geophysics*.

Forsythe, V. V., Duly, T., Hampton, D., and Nguyen, V. (2020). Validation of ionospheric electron density measurements derived from Spire CubeSat constellation. *Radio Science*, 55(1):e2019RS006953.

Foukal, P. V. (2004). Solar astrophysics, 2nd. *Solar Astrophysics*, page 480.

Fox, M. W. (1994). A simple, convenient formalism for electron density profiles. *Radio science*, 29(6):1473–1491.

Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.

Friedel, R., Reeves, G., and Obara, T. (2002). Relativistic electron dynamics in the inner magnetospherea review. *Journal of Atmospheric and Solar-Terrestrial Physics*, 64(2):265–282.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Friis-Christensen, E., Lühr, H., and Hulot, G. (2006). Swarm: A constellation to study the Earth's magnetic field. *Earth, planets and space*, 58(4):351–358.

Gannon, J., Li, X., and Heynderickx, D. (2007). Pitch angle distribution analysis of radiation belt electrons based on Combined Release and Radiation Effects Satellite Medium Electrons A data. *Journal of Geophysical Research: Space Physics*, 112(A5).

Ganushkina, N., Jaynes, A., and Liemohn, M. (2017). Space weather effects produced by the ring current particles. *Space Science Reviews*, 212:1315–1344.

Ganushkina, N. Y., Dandouras, I., Shprits, Y., and Cao, J. (2011). Locations of boundaries of outer and inner radiation belts as observed by Cluster and Double Star. *Journal of Geophysical Research: Space Physics*, 116(A9).

Ganushkina, N. Y., Liemohn, M., and Dubyagin, S. (2018). Current systems in the Earth's magnetosphere. *Reviews of Geophysics*, 56(2):309–332.

Ganushkina, N. Y., Sillanpää, I., Welling, D., Haiducek, J., Liemohn, M., Dubyagin, S., and Rodriguez, J. (2019). Validation of inner magnetosphere particle transport and acceleration model (IMPTAM) with long-term GOES MAGED measurements of keV electron fluxes at geostationary orbit. *Space Weather*, 17(5):687–708.

Garrett, H. B. (1981). The charging of spacecraft surfaces. *Reviews of Geophysics*, 19(4):577–616.

Ginet, G., O'Brien, T., Huston, S., Johnston, W., Guild, T., Friedel, R., Lindstrom, C., Roth, C., Whelan, P., Quinn, R., et al. (2013). AE9, AP9 and SPM: New models for specifying the trapped energetic particle and space plasma environment. In *The Van Allen Probes Mission*, pages 579–615. Springer.

Glauert, S. A., Horne, R. B., and Meredith, N. P. (2014). Three-dimensional electron radiation belt simulations using the bas radiation belt model with new diffusion models for chorus, plasmaspheric hiss, and lightning-generated whistlers. *Journal of Geophysical Research: Space Physics*, 119(1):268–289.

Glauert, S. A., Horne, R. B., and Meredith, N. P. (2018). A 30-year simulation of the outer electron radiation belt. *Space Weather*, 16(10):1498–1522.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings.

Goldstein, J., Pascuale, S., and Kurth, W. (2019). Epoch-based model for stormtime plasmapause location. *Journal of Geophysical Research: Space Physics*, 124(6):4462–4491.

Gonzalez, W., Joselyn, J.-A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B., and Vasyliunas, V. (1994). What is a geomagnetic storm? *Journal of Geophysical Research: Space Physics*, 99(A4):5771–5792.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Goss, A., Schmidt, M., Erdogan, E., Görres, B., and Seitz, F. (2019). High-resolution vertical total electron content maps based on multi-scale B-spline representations. *Annales geophysicae*, 37(4).

Gowtam, V. S., Tulasi Ram, S., Reinisch, B., and Prajapati, A. (2019). A new artificial neural network-based global three-dimensional ionospheric model (ANNIM-3D) using long-term ionospheric observations: Preliminary results. *Journal of Geophysical Research: Space Physics*, 124(6):4639–4657.

Greeley, A. D., Kanekal, S. G., Sibeck, D. G., Schiller, Q., and Baker, D. N. (2021). Evolution of Pitch Angle Distributions of Relativistic Electrons During Geomagnetic Storms: Van Allen Probes Observations. *Journal of Geophysical Research: Space Physics*, 126(2):e2020JA028335.

Green, J., O'Brien, T., Claudepierre, S., and Boyd, A. (2021). Removing orbital variations from low altitude particle data: Method and application. *Space Weather*, 19(2):e2020SW002638.

Green, J. C. and Kivelson, M. (2004). Relativistic electrons in the outer radiation belt: Differentiating between acceleration mechanisms. *Journal of Geophysical Research: Space Physics*, 109(A3).

Gu, X., Zhao, Z., Ni, B., Shprits, Y., and Zhou, C. (2011). Statistical analysis of pitch angle distribution of radiation belt energetic electrons near the geostationary orbit: CRRES observations. *Journal of Geophysical Research: Space Physics*, 116(A1).

Gubbins, D. and Herrero-Bervera, E. (2007). *Encyclopedia of geomagnetism and paleomagnetism*. Springer Science & Business Media.

Habarulema, J. B. and Carelse, S. A. (2016). Long-term analysis between radio occultation and ionosonde peak electron density and height during geomagnetic storms. *Geophysical Research Letters*, 43(9):4106–4111.

Habarulema, J. B., Okoh, D., Burešová, D., Rabiu, B., Tshisaphungo, M., Kosch, M., Häggström, I., Erickson, P. J., and Milla, M. A. (2021). A global 3-D electron density reconstruction model based on radio occultation data and neural networks. *Journal of Atmospheric and Solar-Terrestrial Physics*, 221:105702.

Hajj, G., Kursinski, E., Bertiger, W., Leroy, S., Meehan, T., Romans, L., and Schofield, J. (1996). Initial results of GPS-LEO occultation measurements of Earth's atmosphere obtained with the GPS-MET experiment. In *GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications*, pages 144–153. Springer.

Hao, Y., Zong, Q.-G., Zhou, X.-Z., Rankin, R., Chen, X., Liu, Y., Fu, S., Spence, H. E., Blake, J. B., and Reeves, G. D. (2017). Relativistic electron dynamics produced by azimuthally localized poloidal mode ULF waves: Boomerang-shaped pitch angle evolutions. *Geophysical Research Letters*, 44(15):7618–7627.

Hargreaves, J. K. (1992). *The solar-terrestrial environment: an introduction to geospace-the science of the terrestrial upper atmosphere, ionosphere, and magnetosphere*. Cambridge university press.

Hartley, D. and Denton, M. (2014). Solving the radiation belt riddle. *Astronomy & Geophysics*, 55(6):6–17.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hastings, D. and Garrett, H. (2004). Spacecraft-environment interactions. *Spacecraft-Environment Interactions*.

Heelis, R., Coley, W., Burrell, A., Hairston, M., Earle, G., Perdue, M., Power, R., Harmon, L., Holt, B., and Lippincott, C. (2009). Behavior of the O+/H+ transition height during the extreme solar minimum of 2008. *Geophysical Research Letters*, 36(18).

Hernández-Pajares, M., Garcia-Fernàndez, M., Rius, A., Notarpietro, R., von Engeln, A., Olivares-Pulido, G., Aragón-Àngel, À., and García-Rigo, A. (2017). Electron density extrapolation above F2 peak by the linear Vary-Chap model supporting new Global Navigation Satellite Systems-LEO occultation missions. *Journal of Geophysical Research: Space Physics*, 122(8):9003–9014.

Hernández-Pajares, M., Juan, J. M., Sanz, J., Aragón-Àngel, À., García-Rigo, A., Salazar, D., and Escudero, M. (2011). The ionosphere: effects, GPS modeling and the benefits for space geodetic techniques. *Journal of Geodesy*, 85(12):887–907.

Hobiger, T. and Jakowski, N. (2017). Atmospheric signal propagation. In *Springer handbook of global navigation satellite systems*, pages 165–193. Springer.

Hoque, M. M. and Jakowski, N. (2007). Higher order ionospheric effects in precise gnss positioning. *Journal of geodesy*, 81:259–268.

Hoque, M. M., Jakowski, N., and Prol, F. S. (2022). A new climatological electron density model for supporting space weather services. *Journal of Space Weather and Space Climate*, 12:1.

Horne, R. B., Glauert, S. A., Meredith, N. P., Koskinen, H., Vainio, R., Afanasiev, A., Ganushkina, N. Y., Amariutei, O. A., Boscher, D., Sicard, A., et al. (2013). Forecasting the Earth's radiation belts and modelling solar energetic particle events: Recent results from SPACECAST. *Journal of Space Weather and Space Climate*, 3:A20.

Horne, R. B., Meredith, N. P., Thorne, R. M., Heynderickx, D., Iles, R. H., and Anderson, R. R. (2003). Evolution of energetic electron pitch angle distributions during storm time electron acceleration to megaelectronvolt energies. *Journal of Geophysical Research: Space Physics*, 108(A1):SMP–11.

Horne, R. B. and Pitchford, D. (2015). Space weather concerns for all-electric propulsion satellites. *Space Weather*, 13(8):430–433.

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.

Hunsucker, R. D. and Hargreaves, J. K. (2007). *The high-latitude ionosphere and its effects on radio propagation*. Cambridge University Press.

Ilie, R. (2020). Theoretical description. In *Ring Current Investigations*, pages 53–98. Elsevier.

Jakowski, N., Wehrenpfennig, A., Heise, S., Reigber, C., Lühr, H., Grunwaldt, L., and Meehan, T. (2002). GPS radio occultation measurements of the ionosphere from CHAMP: Early results. *Geophysical Research Letters*, 29(10):95–1.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Jaynes, A. N., Baker, D. N., Singer, H. J., Rodriguez, J. V., Loto'aniu, T., Ali, A., Elkington, S. R., Li, X., Kanekal, S. G., Claudepierre, S. G., et al. (2015). Source and seed populations for relativistic electrons: Their roles in radiation belt changes. *Journal of Geophysical Research: Space Physics*, 120(9):7240–7254.

Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., and Rehman, M. U. (2019). A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *IEEE Access*, 7:28309–28318.

Kamide, Y. and Akasofu, S.-I. (1983). Notes on the auroral electrojet indices. *Reviews of Geophysics*, 21(7):1647–1656.

Kamide, Y. and Chian, A. (2007). *Handbook of the solar-terrestrial environment*. Springer.

Kamiya, K., Seki, K., Saito, S., Amano, T., and Miyoshi, Y. (2018). Formation of butterfly pitch angle distributions of relativistic electrons in the outer radiation belt with a monochromatic Pc5 wave. *Journal of Geophysical Research: Space Physics*, 123(6):4679–4691.

Kashcheyev, A. and Nava, B. (2019). Validation of NeQuick 2 model topside ionosphere and plasmasphere electron content using COSMIC POD TEC. *Journal of Geophysical Research: Space Physics*, 124(11):9525–9536.

Katsavrias, C., Aminalragia-Giamini, S., Papadimitriou, C., Daglis, I. A., Sandberg, I., and Jiggens, P. (2022). Radiation Belt Model Including Semi-Annual Variation and Solar Driving (Sentinel). *Space Weather*, 20(1):e2021SW002936.

Katsavrias, C., AminalragiaGiamini, S., Papadimitriou, C., Sandberg, I., Jiggens, P., Daglis, I., and Evans, H. (2021). On the Interplanetary Parameter Schemes Which Drive the Variability of the Source/Seed Electron Population at GEO. *Journal of Geophysical Research: Space Physics*, 126(6):e2020JA028939. e2020JA028939 2020JA028939.

Kavanagh, A. J., Cobbett, N., and Kirsch, P. (2018). Radiation belt slot region filling events: Sustained energetic precipitation into the mesosphere. *Journal of Geophysical Research: Space Physics*, 123(9):7999–8020.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, pages 3146–3154.

Kellerman, A. and Shprits, Y. (2012). On the influence of solar wind conditions on the outer-electron radiation belt. *Journal of Geophysical Research: Space Physics*, 117(A5).

Kelley, M., Wong, V., Aponte, N., Coker, C., Mannucci, A., and Komjathy, A. (2009). Comparison of COSMIC occultation-based electron density profiles and TIP observations with Arecibo incoherent scatter radar data. *Radio Science*, 44(4).

Kelley, M. C. (2009). *The Earth's ionosphere: Plasma physics and electrodynamics*. Academic press.

Kitamura, K., Nakamura, Y., Tokumitsu, M., Ishida, Y., and Watari, S. (2011). Prediction of the electron flux environment in geosynchronous orbit using a neural network technique. *Artificial Life and Robotics*, 16(3):389–392.

Kivelson, M. and Russell, C. (1995). *Introduction to space physics*. Cambridge university press.

Klimenko, M., Klimenko, V., Zakharenkova, I., and Cherniak, I. V. (2015). The global morphology of the plasmaspheric electron content during Northern winter 2009 based on GPS/COSMIC observation and GSM TIP model results. *Advances in Space Research*, 55(8):2077–2085.

Knudsen, D., Burchill, J., Buchert, S., Eriksson, A., Gill, R., Wahlund, J.-E., Åhlén, L., Smith, M., and Moffat, B. (2017). Thermal ion imagers and Langmuir probes in the Swarm electric field instruments. *Journal of Geophysical Research: Space Physics*, 122(2):2655–2673.

Koller, J., Reeves, G. D., and Friedel, R. H. (2008). Lanl* v1. 0: A radiation belt drift shell model suitable for real-time and reanalysis applications. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Konstantinidis, K. and Sarris, T. (2015). Calculations of the integral invariant coordinates I and L* in the magnetosphere and mapping of the regions where I is conserved, using a particle tracer (ptr3D v2. 0), LANL*, SPENVIS, and IRBEM. *Geoscientific Model Development*, 8(9).

Koskinen, H. E. and Kilpua, E. K. (2022). *Physics of Earth's radiation belts: Theory and observations*. Springer Nature.

Kronberg, E., Rashev, M., Daly, P., Shprits, Y. Y., Turner, D., Drozdov, A., Dobynde, M., Kellerman, A. C., Fritz, T., Pierrard, V., et al. (2016). Contamination in electron observations of the silicon detector on board Cluster/RAPID/IES instrument in Earth's radiation belts and ring current. *Space Weather*, 14(6):449–462.

Kronberg, E. A., Ashour-Abdalla, M., Dandouras, I., Delcourt, D. C., Grigorenko, E. E., Kistler, L. M., Kuzichev, I. V., Liao, J., Maggiolo, R., Malova, H. V., et al. (2014). Circulation of heavy ions and their dynamical effects in the magnetosphere: Recent observations and models. *Space Science Reviews*, 184:173–235.

Kronberg, E. A., Daly, P. W., Grigorenko, E. E., Smirnov, A. G., Klecker, B., and Malykhin, A. Y. (2021). Energetic Charged Particles in the Terrestrial Magnetosphere: Cluster/RAPID Results. *Journal of Geophysical Research: Space Physics*, 126(9):e2021JA029273. e2021JA029273 2021JA029273.

Kronberg, E. A., Daly, P. W., and Vilenius, E. (2015). Calibration report of the RAPID measurements in the Cluster Science Archive (CSA). Technical report, CAA-EST-CR-RAP.

Lai, P.-C., Burke, W. J., and Gentile, L. (2013). Topside electron density profiles observed at low latitudes by COSMIC and compared with in situ ion densities measured by C/NOFS. *Journal of Geophysical Research: Space Physics*, 118(5):2670–2680.

Landau, L. and Lifshits, E. (1960). *Mechanics*, volume 1. CUP Archive.

Lanzerotti, L., LaFleur, K., Maclennan, C., and Maurer, D. (1998). Geosynchronous spacecraft charging in January 1997. *Geophysical research letters*, 25(15):2967–2970.

Lee, H.-B., Jee, G., Kim, Y. H., and Shim, J. (2013). Characteristics of global plasmaspheric TEC in comparison with the ionosphere simultaneously observed by Jason-1 satellite. *Journal of Geophysical Research: Space Physics*, 118(2):935–946.

Lei, J., Syndergaard, S., Burns, A. G., Solomon, S. C., Wang, W., Zeng, Z., Roble, R. G., Wu, Q., Kuo, Y.-H., Holt, J. M., et al. (2007). Comparison of COSMIC ionospheric measurements with ground-based observations and model predictions: Preliminary results. *Journal of Geophysical Research: Space Physics*, 112(A7).

Li, W. and Hudson, M. (2019). Earth's Van Allen radiation belts: From discovery to the Van Allen Probes era. *Journal of Geophysical Research: Space Physics*, 124(11):8319–8351.

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).

Li, W., Zhao, D., He, C., Shen, Y., Hu, A., and Zhang, K. (2021). Application of a Multi-Layer Artificial Neural Network in a 3-D Global Electron Density Model Using the Long-Term Observations of COSMIC, Fengyun-3C, and Digisonde. *Space Weather*, 19(3):e2020SW002605.

Li, X. (2004). Variations of 0.7–6.0 MeV electrons at geosynchronous orbit as a function of solar wind. *Space Weather*, 2(3).

Li, X., Baker, D., O'Brien, T., Xie, L., and Zong, Q. (2006). Correlation between the inner edge of outer radiation belt electrons and the innermost plasmapause location. *Geophysical research letters*, 33(14).

Li, X., Baker, D., Temerin, M., Reeves, G., Friedel, R., and Shen, C. (2005). Energetic electrons, 50 keV to 6 MeV, at geosynchronous orbit: Their responses to solar wind variations. *Space Weather*, 3(4).

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., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, 16(12):2079–2102.

Limberger, M. (2015). *Ionosphere modeling from GPS radio occultations and complementary data based on B-splines*. PhD thesis, Technische Universität München.

Limberger, M., Liang, W., Schmidt, M., Dettmering, D., and Hugentobler, U. (2013). Regional representation of F2 Chapman parameters based on electron density profiles. In *Annales Geophysicae*, volume 31, pages 2215–2227. Copernicus GmbH.

Ling, A., Ginet, G., Hilmer, R., and Perry, K. (2010). A neural network-based geosynchronous relativistic electron flux forecasting model. *Space Weather*, 8(9):1–14.

Linnainmaa, S. (1970). The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master's thesis, Masters Thesis (in Finnish), Univ. Helsinki.

Liu, J., Lin, C., Lin, C., Tsai, H., Solomon, S., Sun, Y., Lee, I., Schreiner, W., and Kuo, Y. (2010). Artificial plasma cave in the low-latitude ionosphere results from the radio occultation inversion of the FORMOSAT-3/COSMIC. *Journal of Geophysical Research: Space Physics*, 115(A7).

Liu, Z., Fang, H., Hoque, M. M., Weng, L., Yang, S., and Gao, Z. (2019). A new empirical model of NmF2 based on CHAMP, GRACE, and COSMIC radio occultation. *Remote Sensing*, 11(11):1386.

Lomidze, L., Knudsen, D. J., Burchill, J., Kouznetsov, A., and Buchert, S. C. (2018). Calibration and validation of Swarm plasma densities and electron temperatures using ground-based radars and satellite radio occultation measurements. *Radio Science*, 53(1):15–36.

Lühr, H., Rother, M., Köhler, W., Ritter, P., and Grunwaldt, L. (2004). Thermospheric up-welling in the cusp region: Evidence from champ observations. *Geophysical Research Letters*, 31(6).

Lühr, H. and Xiong, C. (2010). IRI-2007 model overestimates electron density during the 23/24 solar minimum. *Geophysical Research Letters*, 37(23).

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pages 4765–4774.

Lyons, L. and Williams, D. (1975a). The storm and poststorm evolution of energetic (35–560 kev) radiation belt electron distributions. *Journal of Geophysical Research*, 80(28):3985–3994.

Lyons, L. R. and Thorne, R. M. (1973). Equilibrium structure of radiation belt electrons. *Journal of Geophysical Research (1896-1977)*, 78(13):2142–2149.

Lyons, L. R., Thorne, R. M., and Kennel, C. F. (1972). Pitch-angle diffusion of radiation belt electrons within the plasmasphere. *Journal of Geophysical Research*, 77(19):3455–3474.

Lyons, L. R. and Williams, D. J. (1975b). The quiet time structure of energetic (35–560 keV) radiation belt electrons. *Journal of Geophysical Research (1896-1977)*, 80(7):943–950.

Lyons, L. R. and Williams, D. J. (2013). *Quantitative aspects of magnetospheric physics*, volume 23. Springer Science & Business Media.

Maloney, K. O., Schmid, M., and Weller, D. E. (2012). Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. *Methods in Ecology and Evolution*, 3(1):116–128.

Mauk, B., Fox, N. J., Kanekal, S., Kessel, R., Sibeck, D., and Ukhorskiy, a. A. (2012). Science objectives and rationale for the Radiation Belt Storm Probes mission. *The Van Allen Probes Mission*, pages 3–27.

Mayaud, P. (1993). *Derivation, Meaning, and Use of Geomagnetic Indices*, volume 77. AGU AMERICAN GEOPHYSICAL UNION.

McNamara, L., Cooke, D., Valladares, C., and Reinisch, B. (2007). Comparison of CHAMP and Digisonde plasma frequencies at Jicamarca, Peru. *Radio Science*, 42(02):1–14.

Melbourne, W., Davis, E., Duncan, C., Hajj, G., Hardy, K., Kursinski, E., Meehan, T., Young, L., and Yunck, T. (1994). The application of spaceborne GPS to atmospheric limb sounding and global change monitoring.

Meredith, N. P., Horne, R. B., Glauert, S. A., Thorne, R. M., Summers, D., Albert, J. M., and Anderson, R. R. (2006). Energetic outer zone electron loss timescales during low geomagnetic activity. *Journal of Geophysical Research: Space Physics*, 111(A5).

Meredith, N. P., Horne, R. B., Johnstone, A. D., and Anderson, R. R. (2000). The temporal evolution of electron distributions and associated wave activity following substorm injections in the inner magnetosphere. *Journal of Geophysical Research: Space Physics*, 105(A6):12907–12917.

Merrill, R. T. and McElhinny, M. W. (1983). *The Earth's magnetic field: Its history, origin and planetary perspective*, volume 401. Academic Press London.

Millan, R. and Thorne, R. (2007). Review of radiation belt relativistic electron losses. *Journal of Atmospheric and Solar-Terrestrial Physics*, 69(3):362–377.

Min, K., Bortnik, J., and Lee, J. (2013). A novel technique for rapid l calculation: Algorithm and implementation. *Journal of Geophysical Research: Space Physics*, 118(5):1912–1921.

Minzner, R. (1977). The 1976 standard atmosphere and its relationship to earlier standards. *Reviews of geophysics*, 15(3):375–384.

Miyoshi, Y., Shinohara, I., Takashima, T., Asamura, K., Higashio, N., Mitani, T., Kasahara, S., Yokota, S., Kazama, Y., Wang, S.-Y., et al. (2018). Geospace exploration project ERG. *Earth, Planets and Space*, 70:1–13.

Moldwin, M. B., Zou, S., and Heine, T. (2016). The story of plumes: the development of a new conceptual framework for understanding magnetosphere and ionosphere coupling. In *Annales Geophysicae*, volume 34, pages 1243–1253. Copernicus Publications Göttingen, Germany.

Moosavi, S. M. and Ghassabian, S. (2018). Linearity of calibration curves for analytical methods: A review of criteria for assessment of method reliability. *Calibration and Validation of Analytical Methods - A Sampling of Current Approaches; IntechOpen Ltd.: London, UK*, pages 109–127.

Morimoto, J. and Ponton, F. (2021). Virtual reality in biology: could we become virtual naturalists? *Evolution: Education and Outreach*, 14(1):7.

Morley, S. K., Brito, T. V., and Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, 16(1):69–88.

Morley, S. K., Sullivan, J. P., Henderson, M. G., Blake, J. B., and Baker, D. N. (2016). The global positioning system constellation as a space weather monitor: Comparison of electron measurements with van allen probes data. *Space Weather*, 14(2):76–92.

Motoba, T., Ohtani, S., Gkioulidou, M., Mitchell, D. G., Ukhorskiy, A., Takahashi, K., Lanzerotti, L. J., Claudepierre, S., Spence, H., and Reeves, G. D. (2020). Pitch angle dependence

of electron and ion flux changes during local magnetic dipolarization inside geosynchronous orbit. *Journal of Geophysical Research: Space Physics*, 125(2):e2019JA027543.

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly weather review*, 116(12):2417–2424.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Nava, B., Coisson, P., and Radicella, S. (2008). A new version of the NeQuick ionosphere electron density model. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(15):1856–1862.

Ng, A. (2000). *CS229 Lecture notes*. Stanford University.

Ni, B., Yan, L., Fu, S., Gu, X., Cao, X., Xiang, Z., and Zhang, Y. (2020). Distinct formation and evolution characteristics of outer radiation belt electron butterfly pitch angle distributions observed by Van Allen Probes. *Geophysical Research Letters*, 47(4):e2019GL086487.

Ni, B., Zou, Z., Gu, X., Zhou, C., Thorne, R. M., Bortnik, J., Shi, R., Zhao, Z., Baker, D. N., Kanekal, S. G., et al. (2015). Variability of the pitch angle distribution of radiation belt ultrarelativistic electrons during and following intense geomagnetic storms: Van Allen Probes observations. *Journal of Geophysical Research: Space Physics*, 120(6):4863–4876.

Ni, B., Zou, Z., Li, X., Bortnik, J., Xie, L., and Gu, X. (2016). Occurrence characteristics of outer zone relativistic electron butterfly distribution: A survey of Van Allen Probes REPT measurements. *Geophysical Research Letters*, 43(11):5644–5652.

O'Brien, T. and Moldwin, M. (2003). Empirical plasmapause models from magnetic indices. *Geophysical Research Letters*, 30(4).

Olifer, L., Mann, I. R., Morley, S. K., Ozeke, L. G., and Choi, D. (2018). On the role of last closed drift shell dynamics in driving fast losses and van allen radiation belt extinction. *Journal of Geophysical Research: Space Physics*, 123(5):3692–3703.

Oliphant, T. E. et al. (2006). *A guide to NumPy*, volume 1. Trelgol Publishing USA.

Olivares-Pulido, G., Hernández-Pajares, M., Aragón-Àngel, A., and Garcia-Rigo, A. (2016). A linear scale height Chapman model supported by GNSS occultation measurements. *Journal of Geophysical Research: Space Physics*, 121(8):7932–7940.

Olsen, N. and Stolle, C. (2017). Magnetic signatures of ionospheric and magnetospheric current systems during geomagnetic quiet conditions – An overview. *Space Science Reviews*, 206:5–25.

Orlova, K., Shprits, Y., and Spasojevic, M. (2016). New global loss model of energetic and relativistic electrons based on Van Allen Probes measurements. *Journal of Geophysical Research: Space Physics*, 121(2):1308–1314.

Osthus, D., Caragea, P., Higdon, D., Morley, S., Reeves, G., and Weaver, B. (2014). Dynamic linear models for forecasting of radiation belt electrons and limitations on physical interpretation of predictive models. *Space Weather*, 12(6):426–446.

Pandya, M., Bhaskara, V., Ebihara, Y., Kanekal, S. G., and Baker, D. N. (2020). Evolution of Pitch Angle-Distributed Megaelectron Volt Electrons During Each Phase of the Geomagnetic Storm. *Journal of Geophysical Research: Space Physics*, 125(1):e2019JA027086.

Parker, E. N. (1958). Dynamics of the interplanetary gas and magnetic fields. *The Astrophysical Journal*, 128:664.

Paulikas, G. and Blake, J. (1979). Effects of the solar wind on magnetospheric dynamics: Energetic electrons at the synchronous orbit. *Quantitative Modeling of Magnetospheric Processes, Geophys. Monogr. Ser*, 21:180–202.

Pavon-Carrasco, F. J. and De Santis, A. (2016). The South Atlantic anomaly: The key for a possible geomagnetic reversal. *Frontiers in Earth Science*, 4:40.

Pedatella, N., Yue, X., and Schreiner, W. (2015). Comparison between GPS radio occultation electron densities and in situ satellite observations. *Radio Science*, 50(6):518–525.

Petit, G. and Luzum, B. (2010). IERS Conventions (2010). Technical report, www.iers.org.

Pignalberi, A., Pezzopane, M., Coco, I., Piersanti, M., Giannattasio, F., De Michelis, P., Tozzi, R., and Consolini, G. (2022). Inter-Calibration and Statistical Validation of Topside Ionosphere Electron Density Observations Made by CSES-01 Mission. *Remote Sensing*, 14(18):4679.

Pignalberi, A., Pezzopane, M., and Nava, B. (2021). Optimizing the NeQuick Topside Scale Height Parameters Through COSMIC/FORMOSAT-3 Radio Occultation Data. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

Pignalberi, A., Pezzopane, M., Nava, B., and Coïsson, P. (2020). On the link between the topside ionospheric effective scale height and the plasma ambipolar diffusion, theory and preliminary results. *Scientific Reports*, 10(1):1–17.

Pinto, V., Zhang, X.-J., Mourenas, D., Bortnik, J., Artemyev, A., Lyons, L., and Moya, P. (2020). On the confinement of ultrarelativistic electron remnant belts to low L-shells. *Journal of Geophysical Research: Space Physics*.

Prol, F. S., Hernández-Pajares, M., Camargo, P. d. O., and Muella, M. T. d. A. H. (2018). Spatial and temporal features of the topside ionospheric electron density by a new model based on GPS radio occultation data. *Journal of Geophysical Research: Space Physics*, 123(3):2104–2115.

Prol, F. S., Smirnov, A. G., Hoque, M. M., and Shprits, Y. Y. (2022). Combined model of topside ionosphere and plasmasphere derived from radio-occultation and Van Allen Probes data. *Scientific Reports*, 12(1):1–11.

Prol, F. S., Themens, D. R., Hernández-Pajares, M., de Oliveira Camargo, P., and Muella, M. T. d. A. H. (2019). Linear Vary-Chap topside electron density model with topside sounder and radio-occultation data. *Surveys in Geophysics*, 40(2):277–293.

Radicella, S. M. and Nava, B. (2020). Empirical ionospheric models. In *The Dynamical Ionosphere*, pages 39–53. Elsevier.

Rawer, K., Bilitza, D., and Ramakrishnan, S. (1978). Goals and status of the International Reference Ionosphere. *Reviews of geophysics*, 16(2):177–181.

Reeves, G., Henderson, M., McLachlan, P., Belian, R., Friedel, R., and Korth, A. (1996). Radial propagation of substorm injections. In *International conference on substorms*, volume 389, page 579.

Reeves, G., McAdams, K., Friedel, R., and O'Brien, T. (2003). Acceleration and loss of relativistic electrons during geomagnetic storms. *Geophysical Research Letters*, 30(10).

Reeves, G., Morley, S., and Cunningham, G. (2013). Long-term variations in solar wind velocity and radiation belt electrons. *Journal of Geophysical Research: Space Physics*, 118(3):1040–1048.

Reeves, G. D., Chen, Y., Cunningham, G., Friedel, R., Henderson, M. G., Jordanova, V., Koller, J., Morley, S., Thomsen, M., and Zaharia, S. (2012). Dynamic radiation environment assimilation model: DREAM. *Space Weather*, 10(3):1–25.

Reeves, G. D., Friedel, R. H., Larsen, B. A., Skoug, R. M., Funsten, H. O., Claudepierre, S. G., Fennell, J. F., Turner, D. L., Denton, M. H., Spence, H. E., et al. (2016). Energy-dependent dynamics of keV to MeV electrons in the inner zone, outer zone, and slot regions. *Journal of Geophysical Research: Space Physics*, 121(1):397–412.

Reeves, G. D., Morley, S. K., Friedel, R. H., Henderson, M. G., Cayton, T. E., Cunningham, G., Blake, J. B., Christensen, R. A., and Thomsen, D. (2011). On the relationship between relativistic electron flux and solar wind velocity: Paulikas and Blake revisited. *Journal of Geophysical Research: Space Physics*, 116(A2).

Reidy, J., Horne, R., Glauert, S., Clilverd, M., Meredith, N., Woodfield, E., Ross, J., Allison, H. J., and Rodger, C. (2021). Comparing electron precipitation fluxes calculated from pitch angle diffusion coefficients to LEO satellite observations. *Journal of Geophysical Research: Space Physics*, 126(3):e2020JA028410.

Reiff, P., Hill, T., and Burch, J. (1977). Solar wind plasma injection at the dayside magnetospheric cusp. *Journal of Geophysical Research*, 82(4):479–491.

Reigber, C., Lühr, H., Schwintzer, P., and Wickert, J. (2004). Earth Observation with CHAMP. *Results from Three Years*.

Reinisch, B. W., Huang, X.-Q., Belehaki, A., Shi, J.-K., Zhang, M.-L., and Ilma, R. (2004). Modeling the IRI topside profile using scale heights from ground-based ionosonde measurements. *Advances in Space Research*, 34(9):2026–2031.

Ren, C. X., Dorostkar, O., Rouet-Leduc, B., Hulbert, C., Strebel, D., Guyer, R. A., Johnson, P. A., and Carmeliet, J. (2019). Machine learning reveals the state of intermittent frictional dynamics in a sheared granular fault. *Geophysical Research Letters*, 46(13):7395–7403.

Richards, P., Fennelly, J., and Torr, D. (1994). EUVAC: A solar EUV flux model for aeronomic calculations. *Journal of Geophysical Research: Space Physics*, 99(A5):8981–8992.

Ripoll, J.-F., Reeves, G. D., Cunningham, G. S., Loridan, V., Denton, M., Santolík, O., Kurth, W., Kletzing, C., Turner, D., Henderson, M., et al. (2016). Reproducing the observed energy-dependent structure of Earth's electron radiation belts during storm recovery with an event-specific diffusion model. *Geophysical Research Letters*, 43(11):5616–5625.

Rishbeth, H. and Garriott, O. K. (1969). Introduction to ionospheric physics. *Introduction to ionospheric physics*.

Roederer, J. G. (1967). On the adiabatic motion of energetic particles in a model magnetosphere. *Journal of Geophysical Research*, 72(3):981–992.

Roederer, J. G. and Zhang, H. (2016). *Dynamics of magnetically trapped particles*. Springer.

Rother, M. and Michaelis, I. (2019). CH-ME-2-PLPT-CHAMP electron density and temperature time series in low time resolution (Level 2).

Rother, M., Schlegel, K., Lühr, H., and Cooke, D. (2010). Validation of CHAMP electron temperature measurements by incoherent scatter radar data. *Radio Science*, 45(06):1–10.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Sai Gowtam, V. and Tulasi Ram, S. (2017). An Artificial Neural Network-Based Ionospheric Model to Predict NmF2 and hmF2 Using Long-Term Data Set of FORMOSAT-3/COSMIC Radio Occultation Observations: Preliminary Results. *Journal of Geophysical Research: Space Physics*, 122(11):11–743.

Sakaguchi, K., Nagatsuma, T., Reeves, G. D., and Spence, H. E. (2015). Prediction of MeV electron fluxes throughout the outer radiation belt using multivariate autoregressive models. *Space Weather*, 13(12):853–867.

Schreiner, W., Rocken, C., Sokolovskiy, S., Syndergaard, S., and Hunt, D. (2007). Estimates of the precision of GPS radio occultations from the COSMIC/FORMOSAT-3 mission. *Geophysical Research Letters*, 34(4).

Schreiner, W. S., Weiss, J., Anthes, R. A., Braun, J., Chu, V., Fong, J., Hunt, D., Kuo, Y.-H., Meehan, T., Serafino, W., et al. (2020). COSMIC-2 Radio Occultation Constellation: First Results. *Geophysical Research Letters*, 47(4):e2019GL086841.

Schulz, M. and Lanzerotti, L. (1974). *Particle diffusion in the radiation belts*. Springer-Verlag, New York.

Scotto, C., de Gonzalez, M. M., Radicella, S., and Zolesi, B. (1997). On the prediction of f1 ledge occurrence and critical frequency. *Advances in Space Research*, 20(9):1773–1775.

Selesnick, R. and Blake, J. (2002). Relativistic electron drift shell splitting. *Journal of Geophysical Research: Space Physics*, 107(A9):SMP–27.

Selesnick, R. S. and Albert, J. (2019). Variability of the proton radiation belt. *Journal of Geophysical Research: Space Physics*, 124(7):5516–5527.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American statistical association*, 63(324):1379–1389.

Shi, R., Summers, D., Ni, B., Fennell, J. F., Blake, J. B., Spence, H. E., and Reeves, G. D. (2016). Survey of radiation belt energetic electron pitch angle distributions based on the Van Allen Probes MagEIS measurements. *Journal of Geophysical Research: Space Physics*, 121(2):1078–1090.

Shim, J., Scherliess, L., Schunk, R., and Thompson, D. (2008). Spatial correlations of day-to-day ionospheric total electron content variability obtained from ground-based GPS. *Journal of Geophysical Research: Space Physics*, 113(A9).

Shim, J., Song, I.-S., Jee, G., Kwak, Y.-S., Tsagouri, I., Goncharenko, L., McInerney, J., Vitt, A., Rastaetter, L., Yue, J., et al. (2023). Validation of Ionospheric Specifications During Geomagnetic Storms: TEC and foF2 During the 2013 March Storm Event-II. *Space Weather*, 21(5):e2022SW003388.

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., Subbotin, D., Drozdov, A., Usanova, M. E., Kellerman, A., Orlova, K., Baker, D. N., Turner, D. L., and Kim, K.-C. (2013). Unusual stable trapping of the ultrarelativistic electrons in the Van Allen radiation belts. *Nature Physics*, 9(11):699–703.

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., Thorne, R., Friedel, R., Reeves, G., Fennell, J., Baker, D., and Kanekal, S. (2006). Outward radial diffusion driven by losses at magnetopause. *Journal of Geophysical Research: Space Physics*, 111(A11).

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.

Shumko, M., Sample, J., Johnson, A., Blake, B., Crew, A., Spence, H., Klumpar, D., Agapitov, O., and Handley, M. (2018). Microburst scale size derived from multiple bounces of a microburst simultaneously observed with the FIREBIRD-II CubeSats. *Geophysical Research Letters*, 45(17):8811–8818.

Sibeck, D., McEntire, R., Lui, A., Lopez, R., and Krimigis, S. (1987). Magnetic field drift shell splitting: Cause of unusual dayside particle pitch angle distributions during storms and substorms. *Journal of Geophysical Research: Space Physics*, 92(A12):13485–13497.

Sillanpää, I., Ganushkina, N. Y., Dubyagin, S., and Rodriguez, J. (2017). Electron fluxes at geostationary orbit from GOES MAGED data. *Space Weather*, 15(12):1602–1614.

Simms, L. E., Engebretson, M. J., Pilipenko, V., Reeves, G. D., and Clilverd, M. (2016). Empirical predictive models of daily relativistic electron flux at geostationary orbit: Multiple regression analysis. *Journal of Geophysical Research: Space Physics*, 121(4):3181–3197.

Simms, L. E., Pilipenko, V., Engebretson, M. J., Reeves, G. D., Smith, A., and Clilverd, M. (2014). Prediction of relativistic electron flux at geostationary orbit following storms: Multiple regression analysis. *Journal of Geophysical Research: Space Physics*, 119(9):7297–7318.

Smirnov, A., Berrendorf, M., Shprits, Y., Kronberg, E. A., Allison, H. J., Aseev, N. A., Zhelavskaya, I. S., Morley, S. K., Reeves, G. D., Carver, M. R., et al. (2020a). Medium

energy electron flux in Earth's outer radiation belt (MERLIN): A machine learning model. *Space Weather*, 18(11):e2020SW002532.

Smirnov, A., Kronberg, E., Aseev, N., Shprits, Y., and Kellerman, A. (2019a). Calculations of Adiabatic Invariants and Phase Space Density for Cluster mission (Software, Version v1. 0): Zenodo.

Smirnov, A., Kronberg, E. A., Latallerie, F., Daly, P. W., Aseev, N., Shprits, Y., Kellerman, A., Kasahara, S., Turner, D., and Taylor, M. (2019b). Electron intensity measurements by the Cluster/RAPID/IES instrument in Earth's radiation belts and ring current. *Space Weather*, 17(4):553–566.

Smirnov, A., Shprits, Y., Allison, H., Aseev, N., Drozdov, A., Kollmann, P., Wang, D., and Saikin, A. (2022a). Storm-time evolution of the equatorial electron pitch angle distributions in Earth's outer radiation belt. *Frontiers in Astronomy and Space Science*, Special issue: "Plasma Waves in Space Physics: Carrying On the Research Legacies of Peter Gary and Richard Thorne".

Smirnov, A., SHPRITS, Y., Allison, H. J., Aseev, N., Drozdov, A., Kollmann, P., Wang, D., and Saikin, A. (2022b). An empirical model of electron pitch angle distributions in the outer radiation belt based on Van Allen Probes data.

Smirnov, A., Shprits, Y., Prol, F., Lühr, H., Berrendorf, M., Zhelavskaya, I., and Xiong, C. (2023). A novel neural network model of Earth's topside ionosphere. *Scientific Reports*, 13(1):1303.

Smirnov, A., Shprits, Y., Zhelavskaya, I., Lühr, H., Xiong, C., Goss, A., Prol, F. S., Schmidt, M., Hoque, M., Pedatella, N., et al. (2021). Intercalibration of the plasma density measurements in Earth's topside ionosphere. *Journal of Geophysical Research: Space Physics*, 126(10):e2021JA029334.

Smirnov, A., Shprits, Y. Y., Allison, H., Aseev, N., Drozdov, A., Kollmann, P., Wang, D., and Saikin, A. (2022c). An empirical model of the equatorial electron pitch angle distributions in Earth's outer radiation belt. *Space Weather*, page e2022SW003053.

Smirnov, A. G., Kronberg, E. A., Daly, P. W., Aseev, N. A., Shprits, Y. Y., and Kellerman, A. C. (2020b). Adiabatic invariants calculations for Cluster mission: A long-term product for radiation belts studies. *Journal of Geophysical Research: Space Physics*, 125(2):e2019JA027576.

Somov, B. V. (2012). Adiabatic invariants in astrophysical plasma. *Plasma Astrophysics, Part I: Fundamentals and Practice*, pages 115–127.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Stolle, C., Liu, H., Truhlík, V., Lühr, H., and Richards, P. (2011). Solar flux variation of the electron temperature morning overshoot in the equatorial F region. *Journal of Geophysical Research: Space Physics*, 116(A4).

Su, Z., Zheng, H., and Wang, S. (2009). Evolution of electron pitch angle distribution due to interactions with whistler mode chorus following substorm injections. *Journal of Geophysical Research: Space Physics*, 114(A8).

Subbotin, D. and Shprits, Y. (2009). Three-dimensional modeling of the radiation belts using the Versatile Electron Radiation Belt (VERB) code. *Space Weather*, 7(10).

Summers, D., Ma, C., and Mukai, T. (2004). Competition between acceleration and loss mechanisms of relativistic electrons during geomagnetic storms. *Journal of Geophysical Research: Space Physics*, 109(A4).

Szabó-Roberts, M., Shprits, Y. Y., Allison, H. J., Vasile, R., Smirnov, A. G., Aseev, N. A., Drozdov, A. Y., Miyoshi, Y., Claudepierre, S. G., Kasahara, S., et al. (2021). Preliminary Statistical Comparisons of Spin-Averaged Electron Data from Arase and Van Allen Probes Instruments. *Journal of Geophysical Research: Space Physics*, page e2020JA028929.

Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., and Ng, R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547.

Tapping, K. (2013). The 10.7 cm solar radio flux (f10. 7). *Space weather*, 11(7):394–406.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes mathematicae*, 12(85):173.

Themens, D. R., Jayachandran, P., Bilitza, D., Erickson, P. J., Häggström, I., Lyashenko, M. V., Reid, B., Varney, R. H., and Pustovalova, L. (2018). Topside electron density representations for middle and high latitudes: A topside parameterization for E-CHAIM based on the NeQuick. *Journal of Geophysical Research: Space Physics*, 123(2):1603–1617.

Tsutai, A., Mitsui, C., and Nagai, T. (1999). Prediction of a geosynchronous electron environment with in situ magnetic field measurements. *Earth, planets and space*, 51(3):219–223.

Tsyganenko, N. and Sitnov, M. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *Journal of Geophysical Research: Space Physics*, 110(A3).

Tsyganenko, N. A. (1989). A magnetospheric magnetic field model with a warped tail current sheet. *Planetary and Space Science*, 37(1):5–20.

Tu, W., Cunningham, G., Chen, Y., Henderson, M., Camporeale, E., and Reeves, G. (2013). Modeling radiation belt electron dynamics during gem challenge intervals with the dream3d diffusion model. *Journal of Geophysical Research: Space Physics*, 118(10):6197–6211.

Tu, W., Cunningham, G., Chen, Y., Morley, S., Reeves, G., Blake, J., Baker, D., and Spence, H. (2014). Event-specific chorus wave and electron seed population models in dream3d using the van allen probes. *Geophysical Research Letters*, 41(5):1359–1366.

Tulasi Ram, S., Sai Gowtam, V., Mitra, A., and Reinisch, B. (2018). The improved two-dimensional artificial neural network-based ionospheric model (ANNIM). *Journal of Geophysical Research: Space Physics*, 123(7):5807–5820.

Turner, D., Claudepierre, S., Fennell, J., O'Brien, T., Blake, J., Lemon, C., Gkioulidou, M., Takahashi, K., Reeves, G., Thaller, S., et al. (2015a). Energetic electron injections deep into the inner magnetosphere associated with substorm activity. *Geophysical Research Letters*, 42(7):2079–2087.

Turner, D., Kilpua, E., Hietala, H., Claudepierre, S., O'Brien, T., Fennell, J., Blake, J., Jaynes, A., Kanekal, S., Baker, D., et al. (2019). The response of Earth's electron radiation belts to geomagnetic storms: Statistics from the Van Allen Probes era including effects from different storm drivers. *Journal of Geophysical Research: Space Physics*, 124(2):1013–1034.

Turner, D., O'Brien, T., Fennell, J., Claudepierre, S., Blake, J., Kilpua, E., and Hietala, H. (2015b). The effects of geomagnetic storms on electrons in Earth's radiation belts. *Geophysical Research Letters*, 42(21):9176–9184.

Vampola, A. (1997). Outer zone energetic electron environment update. In *Conference on the High Energy Radiation Background in Space. Workshop Record*, pages 128–136. IEEE.

Vampola, A. (1998). Outer zone energetic electron environment update. In *Conference on the high energy radiation background in space. Workshop record*, pages 128–136. IEEE.

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.

Walt, M. (2005). *Introduction to geomagnetically trapped radiation*. Cambridge University Press.

Wang, X., Cheng, W., Zhou, Z., Xu, S., Yang, D., and Cui, J. (2019). Comparison of CSES ionospheric RO data with COSMIC measurements. In *Annales Geophysicae*, volume 37, pages 1025–1038. Copernicus GmbH.

West, H., Buck, R., and Walton, J. (1973). Electron pitch angle distributions throughout the magnetosphere as observed on Ogo 5. *Journal of Geophysical Research*, 78(7):1064–1081.

Wilken, B., Axford, W., Daglis, I., Daly, P., Güttler, W., Ip, W., Korth, A., Kremser, G., Livi, S., Vasyliunas, V., et al. (1997). RAPID: The imaging energetic particle spectrometer on Cluster. *The Cluster and Phoenix Missions*, pages 399–473.

Williams, W. (1960). Space telemetry systems. *Proceedings of the IRE*, 48(4):685–690.

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., Turner, D. L., Ukhorskiy, A. Y., Johnson, J. R., Sotirelis, T., Nikoukar, R., and Romeo, G. (2022). Modeling radiation belt electrons with information theory informed neural networks. *Space Weather*, 20(8):e2022SW003090.

Xiong, C., Jiang, H., Yan, R., Lühr, H., Stolle, C., Yin, F., Smirnov, A., Piersanti, M., Liu, Y., Wan, X., et al. (2022). Solar flux influence on the in-situ plasma density at topside ionosphere measured by Swarm satellites. *Journal of Geophysical Research: Space Physics*, 127(5):e2022JA030275.

Xiong, C., Lühr, H., and Ma, S. (2013). The subauroral electron density trough: Comparison between satellite observations and IRI-2007 model estimates. *Advances in Space Research*, 51(4):536–544.

Xiong, C., Lühr, H., Ma, S., and Schlegel, K. (2015). Validation of GRACE electron densities by incoherent scatter radar data and estimation of plasma scale height in the topside ionosphere. *Advances in Space Research*, 55(8):2048–2057.

Xiong, C., Lühr, H., and Stolle, C. (2021). GRACE Electron Density derived from the K-Band Ranging System (KBR).

Xiong, C., Park, J., Lühr, H., Stolle, C., and Ma, S. (2010). Comparing plasma bubble occurrence rates at CHAMP and GRACE altitudes during high and low solar activity. In *Annales Geophysicae*, volume 28, pages 1647–1658.

Xu, G. and Xu, Y. (2016). *GPS: theory, algorithms and applications*. Springer.

Yizengaw, E., Moldwin, M., Galvan, D., Iijima, B., Komjathy, A., and Mannucci, A. (2008). Global plasmaspheric TEC and its relative contribution to GPS TEC. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(11-12):1541–1548.

Yu, Y., Koller, J., Zaharia, S., and Jordanova, V. (2012). L* neural networks from different magnetic field models and their applicability. *Space Weather*, 10(2).

Yue, X., Schreiner, W. S., Lei, J., Rocken, C., Kuo, Y.-H., and Wan, W. (2010a). Climatology of ionospheric upper transition height derived from cosmic satellites during the solar minimum of 2008. *Journal of atmospheric and solar-terrestrial physics*, 72(17):1270–1274.

Yue, X., Schreiner, W. S., Lei, J., Sokolovskiy, S. V., Rocken, C., Hunt, D. C., and Kuo, Y.-H. (2010b). Error analysis of Abel retrieved electron density profiles from radio occultation measurements. In *Annales Geophysicae*, volume 28, pages 217–222. Copernicus GmbH.

Zhang, J., Mucs, D., Norinder, U., and Svensson, F. (2019). LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity–Application to the Tox21 and Mutagenicity Data Sets. *Journal of Chemical Information and Modeling*, 59(10):4150–4158.

Zhao, H., Baker, D., Jaynes, A., Li, X., Elkington, S., Kanekal, S., Spence, H., Boyd, A., Huang, C.-L., and Forsyth, C. (2017). On the relation between radiation belt electrons and solar wind parameters/geomagnetic indices: Dependence on the first adiabatic invariant and L. *Journal of Geophysical Research: Space Physics*, 122(2):1624–1642.

Zhao, H., Friedel, R., Chen, Y., Baker, D., Li, X., Malaspina, D., Larsen, B., Skoug, R., Funsten, H., Reeves, G., et al. (2021). Equatorial pitch angle distributions of 1–50 keV electrons in Earth's inner magnetosphere: an empirical model based on the Van Allen Probes observations. *Journal of Geophysical Research: Space Physics*, 126(1):e2020JA028322.

Zhao, H., Friedel, R., Chen, Y., Reeves, G., Baker, D., Li, X., Jaynes, A., Kanekal, S., Claudepierre, S., Fennell, J., et al. (2018). An empirical model of radiation belt electron pitch angle distributions based on Van Allen Probes measurements. *Journal of Geophysical Research: Space Physics*, 123(5):3493–3511.

Zhao, H., Ni, B., Li, X., Baker, D., Johnston, W., Zhang, W., Xiang, Z., Gu, X., Jaynes, A., Kanekal, S., et al. (2019). Plasmaspheric hiss waves generate a reversed energy spectrum of radiation belt electrons. *Nature Physics*, 15(4):367–372.

Zhao, X., Hao, Y., Zong, Q.-G., Zhou, X.-Z., Yue, C., Chen, X., Liu, Y., Blake, J. B., Claudepierre, S. G., and Reeves, G. D. (2020). Origin of electron boomerang stripes: Localized ULF wave-particle interactions. *Geophysical Research Letters*, 47(17):e2020GL087960.

Zhelavskaya, I. S., 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.

Zong, Q., Rankin, R., and Zhou, X. (2017). The interaction of ultra-low-frequency pc3-5 waves with charged particles in Earth's magnetosphere. *Reviews of Modern Plasma Physics*, 1(1):1–90.

Appendices

— A —

Chapman production function *

Using theoretical considerations, Sydney Chapman in 1931 deduced a function that can be used to describe a simplified ionized layer. The Sun plays the key role in the description of the Chapman layer. The corresponding double-exponent function, which has been applied heavily in ionospheric research, is based on the following assumptions [e.g., Hunsucker and Hargreaves, 2007]:

1. the atmosphere contains only 1 type of gas;
2. the atmosphere is plainly stratified;
3. the radiation is absorbed proportionally to concentration of gas particles;
4. the ionizing radiation is monochromatic.

Let S_∞ be the intensity of incident radiation, which is at a solar zenith angle χ, acting on the top of the atmosphere (Figure 1.12). When the ionization penetrates into the atmosphere and is absorbed, its intensity diminishes. We denote the energy flux at altitude h as S and the corresponding flux at height $h + dh$ as $S + dS$, σ as the absorption cross section of the gas and N as the number density. In a cylinder of the unit cross section (Figure 1.12) with axis parallel to the ionizing beam, the absorbed energy (dS) can be expressed as:

$$dS = S\sigma N dh \sec\chi, \tag{A.1}$$

which means that the absorbed energy is proportional to the cross section and the intensity of the ionizing radiation. Dividing equation 1.32 by S and integrating it, gives:

$$\int_\infty^h \frac{dS}{S} = \sec\chi \int_\infty^h \sigma N dh = -\tau \sec\chi, \tag{A.2}$$

where $\tau = -\int_\infty^h N\sigma dh$ represents the *optical depth* of the atmosphere down to altitude h.

*This derivation follows Davies [1965] and Rishbeth and Garriott [1969].

At very high altitudes, one can assume that $S \to S_\infty$ ($\tau \to 0$), which gives:

$$S = S_\infty \exp(-\tau \sec\chi). \tag{A.3}$$

In a unit volume $dh \sec\chi$, the absorbed energy can be written as

$$\frac{dS}{dh \sec\chi} = N\sigma S = N\sigma S_\infty \exp(-\tau \sec\chi), \tag{A.4}$$

and the number of electrons produced in the unit volume can be expressed as

$$q(\chi, h) = N\sigma S_\infty \eta \exp(-\tau \sec\chi), \tag{A.5}$$

where η is the *ionization efficiency*, i.e., number of electrons created per unit of the absorbed energy.

Considering that $p = NkT$ and $dh = -dp/Nmg$, the optical depth becomes

$$\tau = \sigma \int_h^\infty N dh = \frac{\sigma}{mg} \int_h^\infty dp = \frac{\sigma p}{mg} = \frac{\sigma NkT}{mg} = \sigma NH. \tag{A.6}$$

Similarly to Equation 1.37, it is possible to define the reduced height with an increment $dz = dh/H$, so that $z = \int_{h_0}^{h} dh/H$. It is convenient to choose the height of unit optical depth as a reference height h_0. Then,

$$z = -\ln\tau \leftrightarrow \tau = \exp(-z). \tag{A.7}$$

Substitution of Equations A.7 and A.6 into Equation A.5 yields:

$$q(\chi, z) = q_0 \exp(1 - z - \sec\chi \exp(-z)), \tag{A.8}$$

with $q_0 = S_\infty \eta / \epsilon H$ representing the production rate for the overhead Sun ($z = 0$ when $\tau = 1$). Using Equation 1.37, one can rewrite Equation A.8 as:

$$q(\chi, h) = q_0 \exp\left(1 - \frac{h - h_0}{H} - \sec\chi \exp\left(-\frac{h - h_0}{H}\right)\right), \tag{A.9}$$

known as the *Chapman production function*.

Taking a derivative of the production function (Equation A.8) and equating it to zero yields the reduced height of the maximum production as:

$$z_m = \ln\sec\chi. \tag{A.10}$$

Let us now use Equation A.10 to derive a scaling property for the shape of the Chapman profile depending on the solar zenith angle. In particular, it is of interest to shift the profile using:

$$\hat{z} = z - \ln \sec \chi. \tag{A.11}$$

One can, therefore, express the reduced height as $z = \hat{z} + \ln \sec \chi$, and substitute it into the Chapman production function, which yields:

$$\begin{aligned} q(\chi, \hat{z}) &= \hat{q_0} \exp(1 - (\hat{z} + \ln \sec \chi) - \exp(-(\hat{z} + \ln \sec \chi)) \sec \chi) \\ &= q_0 \exp(1 - \hat{z} - \ln \sec \chi - \exp(-\hat{z}) \cdot \underbrace{\exp(-\ln \sec \chi)}_{=1/\sec \chi} \sec \chi) \\ &= q_0 \exp(-\ln \sec \chi) \exp(1 - \hat{z} - \exp(-\hat{z})) \\ &= q_0 \cos \chi \exp(1 - \hat{z} + \exp(-\hat{z})). \end{aligned} \tag{A.12}$$

Furthermore, one can define $q_m(\chi) = q_0 \cos \chi$ and write an alternative form of the Chapman equation:

$$q(\chi, z) = q_m(\chi) \exp(1 - z + \exp(-z)). \tag{A.13}$$

The equation above (A.13) demonstrates that the shape of the Chapman production function does not depend on χ. Instead, it is the anchor point of the Chapman layer with coordinates (q, z) that is shifted depending on the solar zenith angle (see also Figure 1.13), namely, the amplitude of the maximum ionization is scaled by a factor of $\cos \chi$, and the peak height is shifted by $\ln \sec \chi$.

— B —

Backpropagation in minibatches

In Chapter 1, the neural network gradient descent was described, for simplicity, on a single training example, both for the forward and backward passes. In this Appendix section, the update rule for the neural networks will be generalized for the minibatch training.

Let the number of input features be $n = N_0$, same as before. However, instead of passing a single example through the network, we will now consider a minibatch of B samples. Therefore, instead of the input vector $\mathbf{x} \in \mathbb{R}^n$ used previously, the input matrix will be denoted as $\mathbf{X} \in \mathbb{R}^{n \times B}$. The forward pass of the network for a layer l can be defined as:

$$\begin{cases} \mathbf{Z}^{[l]} &= \mathbf{W}^{[l]}\mathbf{A}^{[l-1]} + \mathbf{b}^{[l]}, \\ \mathbf{A}^{[l]} &= g^{[l]}(\mathbf{Z}^{[l]}), \end{cases} \tag{B.1}$$

for $l = 1, ..., r$. In Equation B.1, the number of neurons in layer l is N_l, and since we pass all the samples within a minibatch through the network, the dimensions of the matrices $\mathbf{A}$ and $\mathbf{Z}$ will be $N_l \times B$. For the last layer (r), the weight matrix $\mathbf{W}^{[r]} \in \mathbb{R}^{1 \times N_{r-1}}$, and the output $\hat{\mathbf{Y}} \in \mathbb{R}^{1 \times B}$. It should be noted that in this Section we keep the same lettering as in Section 1, but the letters are capitalized since here they represent matrices. They are comprised of column vectors that correspond to measurement instances.

The loss function for a batch of samples can be defined as:

$$J = \frac{1}{B}\sum_{i=1}^{B}\frac{1}{2}(\hat{Y}_i - Y_i)^2, \tag{B.2}$$

and represents an averaged value of loss over the minibatch.

The backpropagation is, in general, performed same as before. The delta-matrix for the last layer r is defined as:

$$\boldsymbol{\Delta}^{[r]} = \frac{\partial J}{\partial \mathbf{Z}^{[r]}}, \tag{B.3}$$

where $\boldsymbol{\Delta}^{[r]} \in \mathbb{R}^{1 \times B}$. For an intermediate layer l, the delta-matrix is expressed as:

$$\boldsymbol{\Delta}^{[l]} = \frac{\partial J}{\partial \mathbf{Z}^{[l]}} = \left(\left(\mathbf{W}^{[l+1]}\right)^{\top} \boldsymbol{\Delta}^{[l+1]}\right) \odot g'\left(\mathbf{Z}^{[l]}\right), \tag{B.4}$$

and $\boldsymbol{\Delta}^{[l]} \in \mathbb{R}^{N_l \times B}$.

The partial derivatives in the case of minibatch backpropagation are defined as:

$$\frac{\partial J}{\partial \mathbf{W}^{[l]}} = \frac{1}{B} \boldsymbol{\Delta}^{[l]} \left(\mathbf{A}^{[l-1]} \right)^\top, \text{ and } \quad \frac{\partial J}{\partial \mathbf{b}^{[l]}} = \frac{1}{B} \sum_{i=1}^{B} \boldsymbol{\Delta}_i^{[l]}, \tag{B.5}$$

where $\dfrac{\partial J}{\partial \mathbf{W}^{[l]}} \in \mathbb{R}^{N_l \times N_{l-1}}$, $\dfrac{\partial J}{\partial \mathbf{b}^{[l]}} \in \mathbb{R}^{N_l \times 1}$, and the i-th column of the delta-matrix $\boldsymbol{\Delta}_i^{[l]} \in \mathbb{R}^{N_l \times 1}$.

Finally, the weights and biases are updated as follows:

$$\begin{cases} \mathbf{W}^{[l]} := \mathbf{W}^{[l]} - \alpha \dfrac{\partial J}{\partial \mathbf{W}^{[l]}}, \\ \mathbf{b}^{[l]} := \mathbf{b}^{[l]} - \alpha \dfrac{\partial J}{\partial \mathbf{b}^{[l]}}. \end{cases} \tag{B.6}$$