frontiers | Research Topics

Generation-to-generation communications in space physics

Edited by
Joseph E. Borovsky, Elena E. Grigorenko, Jorge Luis Chau, Yoshizumi Miyoshi, Maria Usanova, Georgia Adair De Nolfo, Antonella Greco and Noora Partamies

Published in
Frontiers in Astronomy and Space Sciences

ISSN 1664-8714
ISBN 978-2-8325-2553-1
DOI 10.3389/978-2-8325-2553-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Generation-to-generation communications in space physics

Topic editors
Joseph E. Borovsky — Space Science Institute, United States
Elena E. Grigorenko — Space Research Institute (RAS), Russia
Jorge Luis Chau — Leibniz Institute of Atmospheric Physics (LG), Germany
Yoshizumi Miyoshi — Nagoya University, Japan
Maria Usanova — University of Colorado Boulder, United States
Georgia Adair De Nolfo — Goddard Space Flight Center, National Aeronautics and Space Administration, United States
Antonella Greco — University of Calabria, Italy
Noora Partamies — The University Centre in Svalbard, Norway

Citation
Borovsky, J. E., Grigorenko, E. E., Chau, J. L., Miyoshi, Y., Usanova, M., De Nolfo, G. A., Greco, A., Partamies, N., eds. (2023). *Generation-to-generation communications in space physics*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-2553-1

Table of contents

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Editorial
PUBLISHED 11 May 2023
DOI 10.3389/fspas.2023.1195579

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Julio Navarro,
University of Victoria, Canada

*CORRESPONDENCE
Joseph E. Borovsky,
jborovsky@spacescience.org

RECEIVED 28 March 2023
ACCEPTED 08 May 2023
PUBLISHED 11 May 2023

CITATION
Borovsky JE, Chau JL, De Nolfo GA, Greco A, Grigorenko EE, Miyoshi Y, Partamies N and Usanova ME (2023), Editorial: Generation-to-Generation Communications in space physics. *Front. Astron. Space Sci.* 10:1195579. doi: 10.3389/fspas.2023.1195579

Editorial: Generation-to-Generation Communications in space physics

Joseph E. Borovsky[1]*, Jorge Luis Chau[2], Georgia Adair De Nolfo[3], Antonella Greco[4], Elena E. Grigorenko[5], Yoshizumi Miyoshi[6], Noora Partamies[7] and Maria E. Usanova[8]

[1]Space Science Institute, Boulder, CO, United States, [2]Leibniz Institute of Atmospheric Physics, Rostock, Germany, [3]NASA/Goddard Space Flight Center, Greenbelt, MD, United States, [4]Universita della Calabria, Arcavacata di Rende, Rende, Italy, [5]Space Research Institute (RAS), Moscow, Russia, [6]Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Japan, [7]The University Centre in Svalbard, Longyearbyen, Norway, [8]University of Colorado, Boulder, CO, United States

KEYWORDS

space physics, ionosphere, magetosphere, solar wind, solar physics, heliosphere, aurora, system science

Editorial on the Research Topic
Generation-to-Generation Communications in space physics

Space physics has been an active area of international research for more than 60 years. Spacecraft measurements require enormous resources: space research is a community effort. Generation after generation of researchers enter the field and their careers evolve shaped by personal experience, mentors, and collaborators. The senior generations of scientists retire and their decades of irreplaceable human wisdom become difficult to access.

One goal of this Frontiers Research Topic (Research Topic) was to document some of the lifetime wisdom of the senior and mid-career leaders in the field of space physics and to reveal it to the generations that follow: advice, mistakes, proud moments, lessons learned, mentors, influential colleagues, concerns for the future. A second goal of this Research Topic was to hear the voices of early career scientists and glimpse their visions for the future of space physics.

This unique Frontiers Research Topic, collected into a Frontiers Research Topic, contains a lot of wisdom, advice, history, and stories of experience (cf. Figure 1). The editors hope these published open-access papers will provide entertaining, enlightening, and valuable advice to both young and experienced researchers in space physics. This editorial contains brief summaries of the 26 papers of this Frontiers Research Topic, plus 17 other papers.

The authors of this editorial (who were also the editors of the Frontiers Research Topic) would like to inform the readers that a number of manuscripts submitted to this Research Topic, although compliant with the goals of the Research Topic, did not pass the initial-validation screening of Frontiers, typically failing because they were deemed too personal or not focused on a specific scientific Research Topic. Seventeen of these manuscripts are published in the open-access AGU journal Perspectives of Earth and Space Scientists https://agupubs.onlinelibrary.wiley.com/journal/26376989. These 17 additional articles also contain a great deal of wisdom, history, personal lessons, humor, and advice. We consider those articles in the Perspectives journal to be part of this "Generation-to-Generation" Research

FIGURE 1
Accoutrements of space physics.

Topic and we highly recommend that the readers of this Frontiers Research Topic examine those Perspectives articles which are in the spirit of this Frontiers Research Topic. After this editorial overviews (alphabetically by author) the 26 articles in this Frontiers Research Topic, this editorial overviews (alphabetically by author) the 17 articles in Perspectives.

Akasofu describes his methodology for solving scientific problems. In particular, focuses on his approach and experience to correct well-accepted theories that contradict observations.

Andre (2022) discusses the need for a "wider perspective" for space sciences wherein results in space science should be presented in a way that makes those results useful to other fields of research such as astrophysics, plasma physics, and astrobiology. Mats Andre also elaborates on the importance of diversity on a research team.

Antonova overviews for the younger generations of space-physics researchers the development of her career and her views about auroral processes. argues that understanding these aurora processes could be a key to understanding other magnetospheric processes.

Clauer advises young scientists to pick something for a vocation that is fun and exciting, advice he himself received from his mentors. highlights the importance of intuition over equations and he tells of the great scientific satisfaction of friendships in research.

Gonzalez-Esparza discusses the opportunities and challenges of developing space-research programs and facilities in developing countries. argues that observational infrastructure is crucial to creating a local scientific community and advises us that this infrastructure should be pursued as a national priority.

Haerendel reviews the development of the field of research focusing on the plasma physics of the aurora, particularly focusing on discrete auroral arcs. looks at the present state of the understanding of the aurora and points out that there are many open questions, particularly involving auroral generation mechanisms.

Hysell discusses several examples of unexpected discoveries in equatorial aeronomy made when experiments did not go according to plan. tells us that it is important to be able to "pivot" when plans fail and to make discoveries from the plan failures.

Kahler compares the advantages of solar imaging with the disadvantages of SEP *in situ* single-point measurements and the lack of synoptic measurements of SEP events, leading to SEP science being "second class" in heliophysics. puts forth the hope for future SEP imaging techniques.

Kronberg discusses lessons learned from improving particle data analysis focusing on the importance of calibration, statistics, and machine learning. also addresses future directions for space research (such as combining data and models, and looking in three dimensions) and Elena summarizes for us best practices in data analysis.

Lockwood discusses, using personal examples and famous cases, how mistakes can be an important driver for scientific progress. One wise lesson that puts forth is to not be overly fearful of mistakes or failures: in a subsection "Learning How to Handle and Exploit Mistakes" Mike says to avoid them but also learn from them.

Lockwood reminds us that we should be developing a system science perspective of the Earth's magnetosphere. points out that even with a full understanding of how a system works, a complex system (such as the Earth's magnetosphere or a flock of starlings) may not be predictable.

Lübken reflects on his own work and the work of others on the Earth's mesosphere/lower thermosphere. describes physical processes acting in the mesosphere/lower thermosphere (MLT) and discusses open questions dealing with that region. He makes the point that young scientists in MLT can achieve visibility in the science community more quickly than in many other fields of physics.

McGranaghan recommends that the space-physics community embrace complexity and systems science. discusses the importance in overcoming the disconnects between different scientific communities and describes cultural and scientific grand challenges for space physics such as the need to construct "participatory ecosystems of knowledge sharing, governance, and trust".

Mobius describes how multiple diverse observations have provided information about the interstellar wind in the vicinity of our Sun. describes the picture assembled from the synergy of (a) *in situ* measurements of interstellar pickup ions and (b) remote neutral-particle imaging to provide us knowledge about the environment of the Milky Way galaxy that our heliosphere resides in.

Palmroth recounts the story of the development of Vlasiator, the first global hybrid-Vlasov code for simulating the solar-wind-driven magnetosphere-ionosphere system. shares with us the advice that long-term code development should target the development to the computer resources of the future, not the present, an argument validated by the success of Vlasiator.

Pedatella overviews the need for and development of whole-atmosphere models. points out how these whole-atmosphere models help us to understand the role of terrestrial weather on the variability of the ionosphere-thermosphere and Nick discusses future whole-atmosphere modeling and the science that those future models will yield.

Reames overviews the evolution of the field of solar-energetic-particle research (SEP) of which he was a key participant. recounts our progress in the understanding of SEPs, leading past outdated models of SEP origins to the current knowledge of particle events driven by shock waves and solar jets.

Richards discusses two problems: 1) the model-data discrepancies in the ionospheric photoelectron-flux spectra and 2) the lack of thermospheric neutral-wind data. presents a personal account of the solutions to these two problems, which sometimes involved "being in the right place at the right time".

Roederer reviews the history of the "climate revolution" of the 1980s, discussing the split of the research community into those favoring an Earth system science approach and those focusing on whatever has the greatest impact on society. also discusses the early history of "space weather" and he puts forth a number of insightful lessons learned.

Rostoker presents his lessons from a career in space physics, particularly lessons gleaned from his work to understand magnetospheric substorms. points out that frameworks for the understanding of substorms are only as good as the existing data, and the relevant magnetospheric data is notoriously sparse.

Sanchez-Cano discusses the solar-wind, magnetosphere, ionosphere, atmosphere system for Mars. argues that understanding the ionosphere as a natural sink of energy from both the solar wind and the atmosphere may be a key to understanding this system. To that end she describes several proposed new Mars missions.

Shiokawa recounts the development of the idea of Earthward flow braking in the near-Earth plasma sheet and its relationship to substorm phenomena. addresses the colleagues who helped to develop this idea and discusses a number of lessons learned that would be valuable for early-career scientists.

Sivadas hypothesizes that working to maximize our number of publications and our number of citations can inhibit scientific discovery. Using a simple mathematical toy model explores and demonstrates this undesirable possibility.

Sonnerup recounts his experiences in graduate school, what he learned and did not learn, and discusses some of the research topics that he has worked on through the years. recalls some of the people with whom he has worked, particularly the people who were most influential.

Tsyganenko discusses the modeling of the Earth's magnetospheric magnetic field from measurements from multiple spacecraft taken at different times and under different conditions. overviews the lessons learned in his half-century-long efforts in building the well-known magnetospheric models.

Wang argues that solar physics has a great advantage over other fields of astrophysics because of the enormous amounts of high-quality solar data, claiming also that much of that solar data is underutilized. wonders why solar physicists are not more skeptical of theoretical models, particularly the "fashionable ones".

Borovsky (2022) recounts his pathway into a science career and discusses the three favorite papers that he wrote, each paper focused on one hypothesis *versus* another hypothesis. Even as a senior scientist, Joe Borovsky has a hidden argument that he is still just a kid from Detroit.

Fairfield (2022) overviews his science career spanning the entire history of magnetospheric physics. Don Fairfield emphasizes the importance of understanding yourself and praises the influence of fortunate circumstances.

Fisk (2023) was originally invited to the Frontiers Research Topic "Generation-to-Generation" but instead under the guidance of the editors was directly submitted to the "Perspectives of Earth and Space Scientists" journal. In this article Len Fisk argues that there is a need for "enhanced vitality" in heliophysics, i.e., a need for paradigm shifts. With three examples of shifts that were not accepted, Len warns that complacency will limit the respect that is given to heliophysics and coronal physics.

Fuselier (2022) uses personal examples to make a case about the importance of mentoring in space physics. Stephen Fuselier's key advice is to "identify and rely on your mentors". Stephen also points out that great advancements in space physics have been enabled by open access to spacecraft data.

Gombosi (2022) recounts his scientific and professional career and the important influences of the people he worked with. Tamas Gombosi tells of his personal experiences (sometimes behind the scenes) bringing together the space communities of the East and the West during the Cold War.

Huba (2023) was originally invited to the Frontiers Research Topic "Generation-to-Generation" but instead under the guidance of the editors was directly submitted to the "Perspectives of Earth and Space Scientists" journal. In this article Joe Huba recounts the story of how he and Glenn Joyce developed the SAMI2 numerical model, bending the rules to attain success.

Kennel (2022) reviews his career of 50 years of serving on NASA advisory panels, recounting influential personalities on those panels and the accomplishments they made. Charlie Kennel describes how those panels led to the Voyager mission to the outer planets, to the Hubble Space Telescope, to the space shuttle, to the space station, and to the James Webb Space Telescope.

Kessel (2022) recounts her research career from graduate student to NASA scientist to NASA Program Manager and Program Scientist. Mona Kessel's research spanned diverse Research Topic such as shock physics, ULF waves, magnetospheric reconnection, ion measurements, and creating standardized formats for spacecraft data.

Koskinen (2022) argues that, as in a nonlinear dynamical system, a small perturbation to a scientific career can completely change its evolution. After recounting several examples from his own career, Hannu Koskinen gives the advice to young scientists: "watch your opportunities and be well-prepared".

Liemohn (2023) provides a great deal of wisdom to the space-physics community. Mike Liemohn's perspectives involve advice on conducting research, on improving research leadership, and on becoming involved in diversity, equity, and inclusion improvements. On the personal level Mike asks us to "work hard and be kind".

Mitchell (2022) recounts his science career from lowly manual labor to designing and flying space-flight scientific instrumentation. Don Mitchell talks about the advantages of "playing to your strengths," which appears as "taking the easy route": easy for Don but probably difficult for others.

Mozer (2022) tells the story of his career from undergraduate student to senior researcher and talks about some of the influential people that he met. Forrest Mozer emphasizes the importance of viewing other researchers as collaborators and not as competitors, looking forward to their advancements.

Pellinen-Wannberg (2023) describes her multifaceted career path and how taking risks repeatedly paid off. Asta Pellinen-Wannberg has particular advice for female scientists, and she has a warning to us all to keep the space environment clean.

Sibeck (2022) recounts the development of his research career, focusing on his history of interaction with scientists from Eastern Europe after the end of the cold war. discusses the international cooperation in space physics and in particular the integration of the Russian-Czech Interball-Magion project into the International Solar-Terrestrial Physics (ISTP) project.

Smith (2022) talks about the importance of mentoring, opening the article with a brief history of his training and about mistakes he made. Throughout the article, Chuck Smith lays out the clear lessons he learned in the development of his scientific career: lessons for the apprentice and lessons for the mentor. He ends with an encouraging message: "Do not ever forget that we got into this career for the joy of the work".

Thomsen (2022) recounts her launch into a scientific career in space physics, overcoming discouragement which almost prevented the career from happening. offers lessons she has learned, which include "a commitment to scientific significance, integrity, communication, and humility".

Wolf (2023) yields advice for a young scientist and a description of the development of the Rice Convection Model for the Earth's magnetosphere. Dick Wolf gives the advice "do not be afraid to be different" and "do not hide your problems". Describing his interactions with Nobel-Prize winners, he points out that even famous professors can be wrong.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

JB was supported at the Space Science Institute by the NSF GEM Program via grant AGS-2027569 and by the NASA HERMES Interdisciplinary Science Program via grant 80NSSC21K1406.

Acknowledgments

The authors of this editorial would like to thank Thom Moore and Tuija Pulkkinen for their help with editorial duties and the authors would like to thank the many reviewers of these Generation-to-Generation Frontiers articles Luke Barnard, Gian Luca Delzanno, Mick Denton, Scott England, Colin Forsyth, Huishan Fu, Hitoshi Fujiwara, Christine Gabrielse, Tamas Gombosi, Maxime Grandin, Mike Hartinger, Katie Herlingshaw, Nickolay Ivchenko, Emilia Kilpua, Hannu Koskinen, Lou Lanzerotti, Mike Liemohn, Bob Lysak, Ryan McGranaghan, Bob McPherron, Marco Milla, Daniel Okoh, Evgeny Panov, Don Reames, Gordon Rostoker, Mike Schulz, David Sibeck, Nithin Sivadas, Jim Slavin, Juan Valdivia, Erdal Yigit, Lev Zelenyi, and Lulu Zhao. Special thanks to Bill Peterson acting as editor of Perspectives of Earth and Space Scientists for rescuing a number of the submissions to this Research Topic. Memories to author and friend Steve Kahler 1939-2023.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Andre, M. (2022). Space physics: The need for a wider perspective. *Front. Astron. Space Sci.* 9, 937742. doi:10.3389/fspas.2022.937742

Borovsky, J. E. (2022). Science: Through the years. *Perspect. Earth Space Sci.* 3, e2022CN000171. doi:10.1029/2022CN000171

Fairfield, D. H. (2022). What influences a career: A personal reflection. *Perspect. Earth Space Sci.* 3, e2022CN000181. doi:10.1029/2022CN000181

Fisk, L. A. (2023). On the vitality, or lack thereof, in heliospheric physics. *Perspect. Earth Space Sci.* 4, e2022CN000192.

Fuselier, S. A. (2022). The importance of mentoring in space physics: Examples from a successful career. *Perspect. Earth Space Sci.* 3, e2022CN000178. doi:10.1029/2022CN000178

Gombosi, T. I. (2022). Manifesto of a space scientist. *Perspect. Earth Space Sci.* 3, e2022CN000174. doi:10.1029/2022CN000174

Gonzalez-Esparza, J. (2022). Space physics career in a developing country: Opportunities and challenges. *Front. Astron. Space Sci.* 9, 952000. doi:10.3389/fspas.2022.952000

Huba, J. D. (2023). On the development of the Sami2 ionosphere model. *Perspect. Earth Space Sci.* 4, e2022CN000195. doi:10.1029/2022cn000195

Kennel, C. F. (2022). 50 years in space science. *Perspect. Earth Space Sci.* 3, e2022CN000168. doi:10.1029/2022CN000168

Kessel, M. (2022). Tools for the long run. *Perspect. Earth Space Sci.* 3, e2022CN000173. doi:10.1029/2022CN000173

Koskinen, H. (2022). Bifurcations in scientific careers. *Perspect. Earth Space Sci.* 3, e2022CN000183. doi:10.1029/2022CN000183

Liemohn, M. W. (2023). Conducting space physics research: Wind the frog, work hard, and be nice. *Perspect. Earth Space Sci.* 4, e2022CN000197. doi:10.1029/2022CN000197

Mitchell, D. G. (2022). Following the path of least resistance. *Perspect. Earth Space Sci.* 3, e2022CN000175. doi:10.1029/2022CN000175

Mozer, F. (2022). My name is forrest mozer. *Perspect. Earth Space Sci.* 3, e2022CN000169. doi:10.1029/2022CN000169

Pellinen-Wannberg, A. (2023). Even a blind chicken sometimes finds a grain of corn - in my case in space. *Perspect. Earth Space Sci.* 4, e2022CN000176. doi:10.1029/2022cn000176

Roerderer, J. G. (2022). First shots of the climate revolution: An untold story. *Front. Astron. Space Sci.* 9, 1025466.

Rostoker, G. (2022). Lessons from a career in space physics. *Front. Astron. Space Sci.* 9, 912036. doi:10.3389/fspas.2022.912036

Sanchez-Cano, B. (2023). Mars' ionosphere: The key for systematic exploration of the red planet. *Front. Astron. Space Sci.* 9, 1101945. doi:10.3389/fspas.2022.1101945

Sibeck, D. G. (2022). Eastern approaches: Interball in the International Solar-Terrestrial Program. *Perspect. Earth Space Sci.* 3, e2022CN000194. doi:10.1029/2022CN000194

Smith, C. W. (2022). Professional development, mentoring, and outreach in scientific research. *Perspect. Earth Space Sci.* 3, e2022CN000188. doi:10.1029/2022CN000188

Thomsen, M. F. (2022). Launch notes. *Perspect. Earth Space Sci.* 3, e2022CN000172. doi:10.1029/2022CN000172

Wolf, R. A. (2023). Advice for a young scientist. *Perspect. Earth Space Sci.* 4, e2022CN000177. doi:10.1029/2022cn000177

frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 09 February 2022
doi: 10.3389/fspas.2022.852798

In Praise of Mistakes

Mike Lockwood *

Department of Meteorology, University of Reading, Reading, United KIngdom

Mistakes are a key driver of scientific progress. We should do all we can to eliminate them, partly to keep the literature record clean, but also because expunging what is wrong often leads us to understand what is right.

Keywords: origins of ideas and concepts, scientific breakthroughs, handling mistakes, science ethics, peer review, science and society

OPEN ACCESS

Edited by:
Joseph E. Borovsky,
Space Science Institute, United States

Reviewed by:
Michael Hartinger,
Space Science Institute, United States
Louis J. Lanzerotti,
New Jersey Institute of Technology, United States

***Correspondence:**
Mike Lockwood
m.lockwood@reading.ac.uk

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *11 January 2022*
Accepted: *21 January 2022*
Published: *09 February 2022*

Citation:
Lockwood M (2022) In Praise of Mistakes. Front. Astron. Space Sci. 9:852798. doi: 10.3389/fspas.2022.852798

"I HAVE NOT FAILED, NOT ONCE. I HAVE DISCOVERED TEN THOUSAND WAYS THAT DON'T WORK"

This is one of great many reported wordings of a quote attributed to Thomas Edison about how he developed the light bulb into a viable and efficient device. It chimes with an insight that came to me in the most unlikely of locations and ways. One evening in Longyearbyen on Svalbard, the world's northernmost permanent human settlement, after a day teaching space physics at UNIS, the world's northernmost university, I happened to share the hotel sauna (and a crate of beer therein) with a surgeon from Oslo who had been flown in to perform a specialist operation at Longyearbyen hospital. We got talking about humans that we admired and when asked to select three, I chose Richard Feynman, Marie Curie and Ernest Shackleton for, respectively, charisma, humanity and leadership. More of Feynman and Curie later, but the surgeon was particularly delighted by my choice of Shackleton, saying how much he admired his determination and skill in rescuing every single member of his ill-fated expedition and how little respect he had for the other famous British Antarctic explorer, Robert Falcon Scott, whom he considered to have been arrogant and foolhardy. Why this conversation was so revealing to me was that it made clear something I have come to believe to be an important distinction in science—the difference between *heroic failures* and *dismal failures*. Heroic failures are to be cherished and celebrated because they often reveal more than do successes—it is unedifying dismal failures that must be avoided.

My point is that we should not be overly fearful of mistakes or failure in any area of science or we will not even try to push boundaries. Of course, we should check everything as deeply and carefully as we possibly can to avoid dismal failure, but it is vital that we do not allow that to cause excessive delay in publishing. Charles Darwin made his seminal observations on the voyage of the Beagle between 1836 and 1838 yet he did not publish them in full until 1859. His friends, botanist Joseph Hooker and geologist Charles Lyell, both warned him in 1856 to pause refining and developing his ideas and publish because other scientists were starting to think along the same lines. In the end, Darwin had to rush to publish one of the most important books in all science (Darwin, 1859) at a time of great personal trauma with the death of his son from scarlet fever and when his daughters were seriously ill with diphtheria. The reason for the rush was a letter he received from the Malay Archipelago from Alfred Russel Wallace that laid out the same themes as were in Darwin's now famous, but then still unpublished, book. The ideas of Darwin and Wallace were jointly announced to the world at the Linnean Society of London in 1858 (Darwin and Wallace, 1858).

I remember a few rare evenings riding my bicycle home, filled with the elation of a scientific epiphany. Somewhat ludicrously, I would ride with extra care on such special evenings—thinking I was the first human to ever realize this, albeit small, contribution to science and fearing it could be lost or delayed if something happened to me. Later, of course, it usually transpired that there were space scientists in other parts of the world who were pretty much there or thereabouts in coming to the same conclusion. That means that if I had waited, even for just one more round of checking, I risked being written out of the story as told by the literature record. For that is, the way of it: a great deal of the real development of science goes unreported. The notebooks of many famous scientists tell a very different story to their publications: the publications give the impression of a logical progression towards an infallible and inevitable conclusion, whereas the notebooks reveal a series of mistakes and failures with b backward steps and f forward ones (which is fine as long as $f > b$). Fortune favors the brave, and if we are right and have really moved things forward we will win the acclaim of our peers in the world of science if we publish–and if we are wrong they will surely tell us.

PEER REVIEW

Peer review is science's way of correcting its mistakes. In my view, it is by far the greatest British contribution to science. I say "British," but such things are never so simple as crude nationalism pretends, and peer review has incrementally evolved over the past 350 years in many places and in both procedure and aims (Moxham & Fyfe, 2018). In addition, although first introduced in 1665 by the founding Editor of Philosophical Transactions of the Royal Society of London (Henry Oldenburg), he was actually an immigrant to Britain who was born and educated as Heinrich Oldenburg in Bremen, Germany (Rix, 1893).

Peer review is repeatedly accused of being "no longer fit for purpose," often by those with a vested interest in overturning a part of the scientific consensus that it has generated. And that is, the key point—peer review generates and documents scientific consensus. Surely it can be fallible and/or inefficient and/or slow—but it is the only method that we have ever had of organizing scientific findings into a consensus. You just need to look at the information wild west of the internet to see what can happen without it—misinformation abounds and swamps reliable information. So next time you are left angry by an unfair and poor review of your paper (it happens!) you have to just accept it as the price that we all pay for generating scientific consensus and find a way around your unhelpful reviewer.

Peer review is harsh and sometimes cruel—mistakes will be commented on, and quite rightly so because the literature record must be kept clean, clear and correct. One of the most brutal aspects about being a scientist, and therefore subject to constant peer review and comment, is the imperative that it drives to be ruthlessly honest with yourself. You may feel your idea may be more elegant than reality, but if it doesn't fit all the facts then it is, to some level, wrong and you must concede that. It is vital to face up to a mistake as early as possible—we are all fallible but we do great damage to both our reputation and our effectiveness as a scientist by persisting in defending something that is, wrong. Science as a whole, like the individuals who progress it, learns from its mistakes.

MUTATIONS

Let me use an analogy. In nature, mutations are the key element of Darwinian evolution and they are nothing more than DNA replication mistakes (e.g., Pray, 2008) and look what evolution has achieved in turning single-cell organisms into the rich variety of life on Earth today. This has more than a trivial parallel to the development of science. In evolution, the principle of "survival of the fittest" decides which mutations thrive and which fail. Similarly in science, if a concept, theory or equation survives peer scrutiny and is adopted then it thrives otherwise, like an unsuccessful mutation, it dies away.

If one can recognize a mistake, it is often a path to unexpected progress. This can come about in a number of ways. An erroneous result causes scientists to conduct further experiments and often identify a previously unsuspected truth. Mistakes often bring to your attention new areas, techniques and theories that you had not realized were relevant and so drive lateral thinking and serendipitous discovery. Trial-and-error is a very common path to progress. A more subtle point is that making a mistake and then coming to understand why it is wrong is invaluable in helping you identify what is right.

The truth is that mistakes and failure are embedded in the very fabric of science. If we do not test our own ideas to destruction, somebody else will. I note that Marie Curie once said "There are sadistic scientists who hurry to hunt down errors instead of establishing the truth." Sadly, that is, a valid comment as there are individuals who use the excuse of the progress of science, when their real motivation is one-upmanship. As scientists we must rise above that, by not indulging in it ourselves and not being perturbed or deflected (into either agreement or disagreement) if it is targeted at us. It is rare, but scientists are, after all, human beings. The best scientists praise progress wherever it comes from and feel empathy with those whose ideas fall by the wayside. In other words, they try to class failures as heroic rather than dismal. My experience has been that the positivity and shared goals of the global space science community massively outweigh destructive petty rivalries. Maybe I have just been lucky, but I think not. I genuinely believe that the ethos of science drives better behavior.

LEARNING HOW TO HANDLE AND EXPLOIT MISTAKES

We should try to avoid mistakes but we should not fear them. After all, the greatest minds in science have all made mistakes. Stuart Firestein is a Professor in Biological Sciences at Columbia University and his excellent 2015 book "Failure: Why Science Is So Successful" reviews a great many historical and contemporary examples. I will give just one example here—a mistake by arguably the most famous, influential and astonishing scientists of all time, Albert Einstein. In 1916 he published his general theory of relativity, a truly astounding intellectual achievement, the equations of which included the “cosmological constant,” to make the universe static by counteracting contraction under gravity—something Einstein regarded as necessary at the time. Then in 1929 Edwin Hubble showed that the universe is expanding, and Einstein removed the cosmological constant from his equations, reputedly calling it his “greatest blunder.” Ironically, his blunder has turned out to be removing it as we now know that the universe is not only expanding, but the expansion is accelerating. To describe why that is, happening, scientists have effectively re-introduced the cosmological constant into general relativity and reinterpreted it as the energy density of space, or vacuum energy, that arises in quantum mechanics, and the related concept of dark energy (e.g., O’Raifeartaigh et al., 2018), which means it is negative where Einstein saw it as positive. Einstein’s mistake showed us the way forward when it turned out not to be a mistake even though he had proposed the very opposite of what is needed!

Serendipity—and maintaining a memory of a mistake—played a large part in my career. It was a mistake that involved Richard Feynman. In 1979 he came out to New Zealand to give a series of evening lectures on quantum electrodynamics at Auckland University while I was a post-doc there. They were brilliant lectures—entertaining, informative, clear, funny and fascinating—and he was, quite simply, the most charismatic speaker that I have ever seen. The Wednesday lecture posed a problem for me. It was a must-see event, but it was also our third wedding anniversary. The agreed solution was an early-evening meal at our favorite restaurant, followed by a quick dash over Albert Park to the University for the lecture (never let it be said I didn’t know how to show a girl a good time!). I was researching the spatial pattern of field-aligned flows of ions in the polar ionosphere at the time (the “polar wind”) and that day I had come across a review paper entitled “Ion velocity distributions in the high-latitude ionosphere” (St-Maurice and Schunk, 1979). I was already running late, so without looking at it, I hurriedly ran off a photocopy (all 36 pages of it) and rushed home. While I was waiting for Celia to get ready, I had a look at it and immediately realized I’d made a mistake: it was about bizarre ion distribution functions driven by collisions between ions and atoms and not about the spatial distributions of flows as I had hoped. I remember that when she emerged ready for her half-a-night out on the town, I dropped the paper in the waste bin, rather theatrically, saying “that was a waste of time and paper, I will never work on that!” I was wrong. Eight years later we were studying ion flows in the cusp using the EISCAT radar, searching for the ionospheric signatures of flux transfer events—bursts of solar wind magnetosphere coupling (we did find them eventually: Lockwood et al., 1993) and the derived ion temperatures were puzzling me. The radar measured the line-of sight component of the ion velocity and I was trying to use the ion temperature (derived assuming a Maxwellian distribution) to infer the magnitude of the velocity to get the vector flow, but that inferred velocity was often far too high making it look like fast flows were always nearly perpendicular to the radar beam! Joe Doupnik from Utah had joined us as a consultant when we were setting up the UK EISCAT activity and I remembered his wise mantra “always check your raw data—look at the radar spectra.” I did, and I immediately recognized a characteristic form that I had very briefly glimpsed that evening 8 years before in St-Maurice and Schunk’s brilliantly predictive paper. My mistake had led to the discovery of radar echoes from non-thermal ions in the auroral ionosphere, which generated a whole series of papers for both me and my colleagues. It showed me how you search for one thing and you often find something quite different and the key that unlocked that particular door was remembering a mistake I had made 8 years earlier!

The very next day, there was another seminal moment in my life. I met Feynman himself at lunch in the senior common room. He was charming and kind to me, a young researcher who had completed his PhD under a year earlier, and having a world-famous Nobel prize winner take a real interest in my work was wonderful for my confidence. I told him it was all coming together rather nicely except one aspect and he said “follow the bit that doesn’t work, young man, that’s the good bit”—advice that later led me to quite a few realizations (including the non-thermal ion distributions) and that I still give to my PhD students.

ONE MISTAKE WE MUST NEVER MAKE

My third choice was Maria Salomea Skłodowska (Marie Curie). For me she is not only an inspiration, but also a lesson in why equality of opportunity is so vitally important. Such extraordinary talent is so rare that it must never be ignored and I would invoke her as a perfect example of what could have been lost to humankind by gender discrimination—just as Srinivasa Ramanujan shows what we would lose to ethnic discrimination, John Dalton to religious discrimination, Michael Faraday to class discrimination, Linus Pauling to economic discrimination, Albert Einstein to racial discrimination, Alan Turing to sexuality discrimination and Stephen Hawking to disability discrimination. Talent for science can and does come from anywhere and we must never make the mistake of failing to recognize and nurture it because of a prejudice. What I love about Marie Curie is her skill, her total dedication to, and passion for, her research, and also how determined she was to use it for the good of humanity. She once wrote one of the most inspiring quotes that I know: “You cannot hope to build a better world without improving the individuals. To that end, each of us must work for his own improvement and, at the same time, share a general responsibility for all humanity, our particular duty being to aid those to whom we think we can be most useful” (Curie, 1923).

In 1903 Marie Curie gained a PhD from the Sorbonne, becoming the first woman ever to receive a doctorate in France. In the same year, she was the first women to be awarded a Nobel prize, despite the French nomination only citing her husband Pierre and Henri Becquerel for their studies of radioactivity: great credit goes to the Swedish mathematician Magnus Mittag–Leffler who made sure Marie was included in the award even though she had not been proposed for it. Of that incident Marie Curie's friend, the British engineer, mathematician, physicist, inventor, and suffragette Hertha Ayrton quipped "Errors are notoriously hard to kill, but an error that ascribes to a man what was actually the work of a woman has more lives than a cat." In 1911, Curie won her second Nobel prize, becoming the first person ever to do so. Being an entirely new area of study, neither of the Curies could have had any idea just how dangerous ionizing radiations from radioactivity are to human health. Famously, Marie's notebooks and personal effects are still so radioactive they are kept in lead containers and still only handled with extensive precautions. She died aged 66 in 1934, of bone marrow failure but probably not induced by her work with radioactive materials, as is commonly thought. A study of her exhumed body in 1995 found that it was almost certainly over-exposure to X-rays that killed her (Butler, 1995). The reason for that is, that she also developed, funded and operated mobile battlefield X-ray units to save the lives of thousands of wounded French soldiers in World War 1. She operated these with her daughter Irène Joliot-Curie who went on to also win a Nobel prize (for Chemistry), the only mother-and-daughter pair to ever do so. Sadly Irène, like her mother, died of lukemia aged 59 and almost certainly for the same reason. The cause of death of Marie and her daughter makes the way that she was repeatedly demonized and hounded by the hypocritical, racist and misogynistic press in her adopted country even more disgraceful and shameful.

I partly mention Marie Curie here in the interest of balance in discussing mistakes. In 2003 some class notes, written in 1907 by a 13-year-old student called Isabelle Chavannes, were saved from destruction by her great-nephew. They tuned out to be verbatim descriptions of lessons given by Marie Curie. From them we learn that Curie taught her students to avoid mistakes, saying "The secret is not to work too quickly" (Chavannes, 1907). Whilst noting all that I have said above about speed of publication being vital, Marie Curie was, of course, absolutely correct.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The work of the author is supported by the UK Science and Technology Facilities Council consolidated grant number ST/R000921/1.

REFERENCES

Butler, D. (1995). X-rays, Not Radium, May Have Killed Curie. *Nature* 377 (6545), 96. doi:10.1038/377096b0

Chavannes, I. (1907). *Leçons de Marie Curie: Physique élémentaire pour les enfans de nos amis*. Paris, France: EDP Science. 2-86883-635-6. Reprinted, 2003.

Curie, M. (1923). *Pierre Curie, with Autobiographical Notes*. translated by Charlotte and Vernon Kellogg. New York, USA: Macmillan.

Darwin, C. R. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. First edition. London, UK: John Murray.

Darwin, C. R., and Wallace, A. R. (1858). On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. [Read 1 July]. *J. Proc. Linn. Soc. LondonZoology* 3 (20 August), 45–50. doi:10.1111/j.1096-3642.1858.tb02500.x

Firestein, S. (2015). *Failure: Why Science Is So Successful*. Oxford, UK: Oxford University Press. 9780199390106.

Lockwood, M., Denig, W. F., Farmer, A. D., Davda, V. N., Cowley, S. W. H., and Lühr, H. (1993). Ionospheric Signatures of Pulsed Reconnection at the Earth's Magnetopause. *Nature* 361 (6411), 424–428. doi:10.1038/361424a0

Moxham, N., and Fyfe, A. (2018). The Royal Society and the Prehistory of Peer Review, 1665-1965. *Hist. J.* 61 (4), 863–889. doi:10.1017/S0018246x17000334

O'Raifeartaigh, C., O'Keeffe, M., Nahm, W., and Mitton, S. (2018). One Hundred Years of the Cosmological Constant: from "Superfluous Stunt" to Dark Energy. *Epj H* 43 (1), 73–117. doi:10.1140/epjh/e2017-80061-7

Pray, L. A. (2008). Mutation and DNA Replication. *Nat. Edu.* 1 (1), 214. doi:10.1007/springerreference_101430

Rix, H. (1893). Henry Oldenburg, First Secretary of the Royal Society. *Nature* 49, 9–12. doi:10.1038/049009a0

St-Maurice, J.-P., and Schunk, R. W. (1979). Ion Velocity Distributions in the High-Latitude Ionosphere. *Rev. Geophys.* 17 (1), 99–134. doi:10.1029/RG017i001p00099

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 24 February 2022
doi: 10.3389/fspas.2022.856188

The Joined-up Magnetosphere

*Mike Lockwood**

Department of Meteorology, University of Reading, Reading, United Kingdom

Systems science is a relatively new way of studying the magnetosphere. This perspective outlines the need for it and how it can contribute to our understanding and so give more reliable forecasts, predictions, and space weather climatologies.

Keywords: magnetosphere, systems science, response times, state indicators, coupling functions

DESCRIPTION, UNDERSTANDING AND PREDICTION

In 1970 Hannes Alfvén won the Nobel Prize for his formulation of magnetohydrodynamics (MHD), the fluid theory of a magnetoplasma. With some assumptions, MHD gives us a self-consistent and predictive description of magnetospheric behavior by quantifying the 3-dimensional structure of the electric and magnetic fields, $\underline{\mathbf{E}}$ and $\underline{\mathbf{B}}$, but with particles only represented by the moments of their (assumed Maxwellian) distributions. However, we should not lose sight of the fact that the magnetosphere is a massive, highly complex multibody system of interacting charged particles.

In lectures to new students, I often compare and contrast magnetospheric behavior with the murmuration of a flock of starlings (see **Figure 1**). Numerical modelling and systems analysis has revealed that each bird in the flock is following the simple interaction law of copying the flight of its seven nearest neighbors, the number that gives the optimum trade-off between group cohesion and individual effort (Young et al., 2013). Any changes in flight by one bird, induced either through error in replicating the flight direction of adjacent birds or by variability in atmospheric conditions (e.g., a gust of wind), will propagate through the flock because of the birds' reaction time and hence the flocks behave as critical systems, poised to respond maximally to environmental perturbations (Cavagna et al., 2010). In the magnetosphere, there are far too many charged particles to numerically model all the mutual interactions but the smaller number of birds in a flock mean it can be modelled that way, as well as by using an analytic theory such as the Vicsek model (Ginellia, 2016), which is the equivalent of MHD for the magnetosphere. These models readily reproduce the types of evolving forms that are seen in bird murmuration[1], which are so varied that they can, by chance, take on ironically coincidental forms[2]. This means that we have a description of the objects (the birds) and how they interact. From this we can predict the types of collective behavior of the flock that are seen—but we cannot define the perturbations well enough to allow us to predict exactly what a given bird flock will do. Hence for complex interacting systems the collective behaviors may not be predictable, even though we have a good description of the objects and good understanding of how they interact.

A flock of birds has some parallels to the magnetosphere, but is the magnetosphere as unpredictable? The answer, fortunately, is no because it does not have the equivalent of the error that individual starlings make in assessing the flight of their nearest neighbors. It does,

[1]A real-time movie of a starling murmuration recorded by Paula McCracken in County Galway, Ireland on 21 January 2022 can be seen at https://www.personal.reading.ac.uk/~ym901336/movies/murmuration.mp4.

[2]For example, "Very Impressive Starling Murmurations" by Daniel Biber—a series of images captured in a 10-s window on the 31 December 2016 near Sant Pere Pescador in Catalonia, Spain. https://www.worldphoto.org/zh/sony-world-photography-awards/winners-galleries/2018/professional/shortlisted/natural-world/very).

OPEN ACCESS

Edited by:
Noora Partamies,
The University Centre in Svalbard, Norway

Reviewed by:
Katie Herlingshaw,
The University Centre in Svalbard, Norway

****Correspondence:***
Mike Lockwood
m.lockwood@reading.ac.uk

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *16 January 2022*
Accepted: *07 February 2022*
Published: *24 February 2022*

Citation:
Lockwood M (2022) The Joined-up Magnetosphere. Front. Astron. Space Sci. 9:856188. doi: 10.3389/fspas.2022.856188

FIGURE 1 | Starling murmuration photographed at the village of Rigg, near Gretna in the Scottish borders, on 25 November 2013. The text discusses similarities and differences between the dynamics of such bird flocks and of the magnetosphere. One coincidental similarity of these two non-equilibrium, multi-body, interacting systems is the potential to cause disruption to power grids. Magnetosphere-ionosphere current systems cause Geomagnetically Induced Currents (GICs) that can give disruption and damage to power supply networks and the weight of the very large number of starlings roosting on the power grid cables in this picture also caused local power disruptions (See The Guardian newspaper (UK), Tuesday 21 January 2014 https://www.theguardian.com/environment/gallery/2014/jan/21/starling-mumuration-season-in-pictures). Photo: Owen Humphreys/PA.

however, have the equivalent of highly variable atmospheric conditions, because it is constantly buffeted by the solar wind and subject to variations in solar wind-magnetosphere coupling by the variability of the north-south component of the interplanetary magnetic field (IMF). This means that systems analysis, of the kind used to study a flock of birds and swarming phenomena in general, does have applications in understanding the magnetosphere.

If we start from the equations of special relativity and the Coulomb force law, we can readily derive the Lorentz equation, which gives us mathematical expressions for $\underline{\mathbf{E}}$ and $\underline{\mathbf{B}}$, defining what they are. The electric field $\underline{\mathbf{E}}$ is a net effect of all the charged particles in the cosmos on the point in question, dependent only on their charge and position, whereas the magnetic field $\underline{\mathbf{B}}$ is the net effect of all the moving charged particles and dependent on their charge, position and motion. As the effects of both decrease rapidly with increased distance between the charged particle and the point (the inverse square laws), the effect of local charges is much greater than distant ones. Hence the (limited) parallel to the flock of birds. We see this distance-dependent connectivity directly in the magnetosphere where, for example, the *Dst, Sym-H* and *SMR* indices are computed from perturbations to the ground-level geomagnetic field at low-latitudes. They are influenced most by the relatively nearby ring current, but also by the more-distant magnetopause currents (*Siscoe* et al., 1968) and by the tail currents *Turner* et al., 2000; *Asikainen* et al., 2010).

Every charged particle in a magnetoplasma influences every other through its contribution to $\underline{\mathbf{E}}$ and $\underline{\mathbf{B}}$, giving a collective behavior which means changes spread through the magnetosphere at propagation speeds set by the speeds of the variety of wave phenomena that are possible. The typical speeds of the information propagation, the spatial scales of the magnetosphere and the temporal variability of the driving interplanetary conditions mean that the magnetosphere, like the bird flock, is often not an equilibrium system and some part, if not all of it, will always still be responding to a prior influence.

Jim Dungey often used to complain about "overly mechanistic and insufficiently holistic thinking" about the magnetosphere. His point was that scientists would often argue that "the change in A generates B which changes C, etc." when A, B, and C were so coupled that the chain of causality running through them was not necessarily the one that people thought. To make matters worse, these chains often contained element pairs that were actually two different descriptions of the same thing, arising from the way people think about Maxwell's equations. JD's "socio-educational" theory was that we are first taught electromagnetism through the dynamo and the motor, where the mechanical construction causes there to be a direction of causality in the equals symbol of Faraday's law. But that is not the meaning of the equals sign in Faraday's law in general which, in Oliver Heaviside's brilliant formulation of Maxwell's equations, tells us that the curl of the electric field and the rate of change of the magnetic field are two different descriptions of exactly the same thing—not that one generates the other. Another example is Ampère's law (the relevant form of another of Maxwell's equation in a plasma because the displacement current can be neglected compared to the free currents for all but exceptionally high frequency phenomena).

$$(\nabla\times\ \underline{\mathbf{B}}) = \mu o\,\underline{\mathbf{J}} \tag{1}$$

JD's point was that **Equation 1** doesn't tell us that a magnetic shear generates a current, nor that a current generates a magnetic shear—it tells us that they are two different descriptions of exactly the same thing: you cannot have one without the other. It means that, for example, the cross-tail current disruption during a substorm onset and the dipolarisation in the near-Earth geomagnetic field (and the associated Earthward convection surge of the frozen-in plasma) are just different ways of

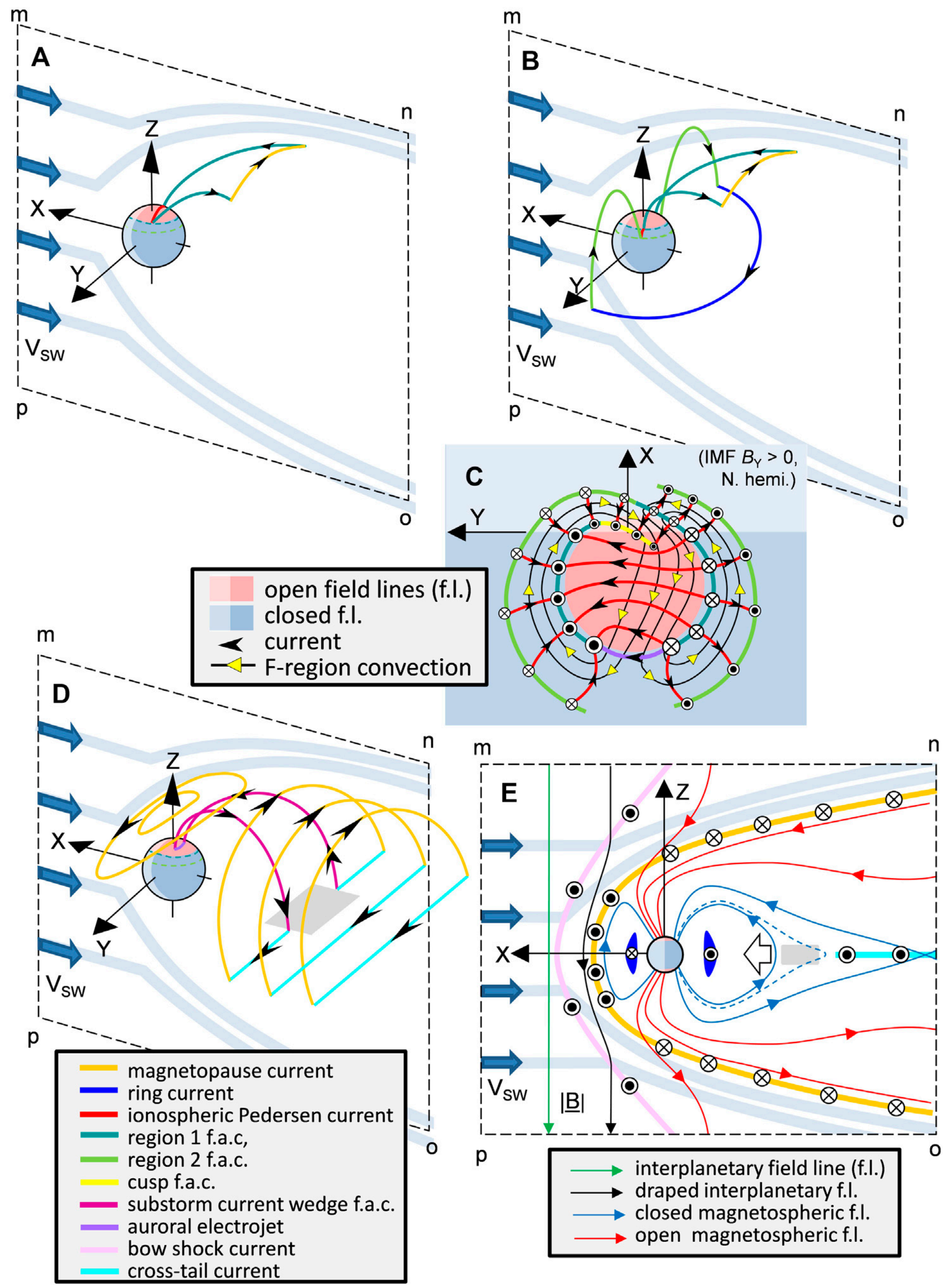

FIGURE 2 | Five different views of the magnetosphere-ionosphere system. Parts **(A)**,**(B)** and **(D)** show the main current loops (in the northern hemisphere only for clarity), viewed from mid-latitudes in the northern hemisphere in the pre-midnight sector. The X, Y, and *Z* axes of the Geocentric Solar Ecliptic (GSE) reference frame are shown and the blue block arrows show the undisturbed solar wind flow, upstream of the bow shock. Pink areas in the ionosphere are where geomagnetic field lines are open (i.e., they thread the magnetopause) and light blue areas are where field lines are closed (i.e., they do not thread the magnetopause). Solar wind coupling occurs because closed field lines are opened by magnetic reconnection in the dayside magnetopause and closed again some time later by reconnection in the cross-tail current sheet. The key gives the color codes of the types of current in each segment of each loop. Note that the currents are shown as single filaments for clarity but in reality form extended sheets. Part **(A)** shows the "polar cap circuit" in which the field aligned currents (f.a.c.) that bring momentum and energy from the shocked solar wind (magnetosheath) to the ionosphere (the "Region 1" currents) enable a current loop completed by dissipative ionospheric Pedersen currents and magnetopause currents. The return sunward convection in the auroral ionosphere requires dusk-to-dawn Pedersen currents across the auroral oval and **(B)** shows the "polar cap/ring current circuit" that makes this possible: this circuit enhances the nightside ring current via the "Region 2" f. a.c.s at lower latitudes and is completed by a Region 1 f. a.c. and magnetopause current loop. The pattern of ionospheric currents and flows is shown in **(C)** which is a view from above the north pole for the case of IMF $B_Y > 0$. Part **(D)** shows the magnetopause and cross-tail current loops, the inner part of the latter being deflected into the ionosphere in the "current wedge" during substorm expansion phases when the central, near-Earth edge of the cross-tail current (in the grey area) is disrupted. This means there is a dipolarisation within the wedge and a sunward convection surge of frozen-in plasma. Part **(E)** is a cross-section view of the magnetosphere from its dusk flank and shows the currents and the corresponding magnetic field lines in the meridian (XZ) plane. The white arrow shows a convection surge and the dashed and solid blue lines the position of nightside closed field lines before and after dipolarisation.

describing precisely the same phenomenon: one does not cause the other (see **Figure 2E**). One reason why this matters is that a causal link would involve a lag and a growth time, neither of which exist when they are really two different descriptions of the same thing.

Another reason why everything in the magnetosphere is connected to the other parts of the magnetosphere is quasi-neutrality of the plasma which means currents are very close to divergence-free, i.e., sources and sinks of current are very small and so they must flow in closed loops, as shown in **Figure 2**. It is remarkable to think that the flux circulation of magnetospheric convection could not happen without not only ion-neutral and electron-neutral collisions in the upper atmosphere that generate Pedersen conductivity, but also a completely different phenomenon in the gradient and curvature drift of trapped particles, that give the ring current. This is because both the ionospheric Pedersen current and the ring current are involved in closing the global field-aligned currents that transfer energy and momentum from the sources in the magnetopause and cross-tail current sheet to the upper atmosphere (see **Figure 2**).

Its behavior may not be as unpredictable as a flock of birds, but the magnetosphere is a joined-up, complex, interacting, many-bodied system and the concepts of Self-Organized Criticality (SOC) have been applied to it in some areas (see review by Watkins et al., 2016).

THE PHASES OF EXPLORATION OF NEAR-EARTH SPACE

I am lucky enough to sit between two generations of magnetospheric scientists. Nowadays I meet a generation of young scientists armed with the latest computational, modelling and visualization techniques that are fed by multi-point and multi-instrument data of unprecedented resolution, accuracy and scope. But I also met a generation before mine that thought its way through to some basic understanding of the terrestrial space environment with unbelievably limited resources and the crudest of observations.

Understanding this joined-up, time-dependent magnetosphere has come in (overlapping) stages. The first stage was to follow the seminal paper of Chapman and Ferraro (1931) and consider temporal changes as a series of equilibrium steady-states. This established some mechanisms, connections and behaviors but kept others well concealed. This approach can, of course, be valid if the changes are slower than the response time constants (Milan et al., 2021) but, as one of my big heroes outside science, Peter Gabriel, sang: "We've tried making movies from a handful of stills"[3].

As more data from near-Earth interplanetary space and the magnetosphere was obtained, statistical studies grew in importance. To reduce noise and to help spot trends, data from long intervals were averaged together. This inherently puts time derivates to zero and so gives an average, but steady-state view of the magnetosphere.

It was realized that steady state hides a great many mechanisms at work in the magnetosphere because one has no chance of identifying the chain of causality in a coupled system. Case studies were used to study the temporal evolutions of the magnetosphere. Examples included geomagnetic storms, sudden compressions by the solar wind, the substorm cycle and, on temporal scales down to a few minutes, studies of ULF wave phenomena and of flux transfer events (bursts in magnetic reconnection in the dayside magnetopause). Much was done from studying transient responses in the magnetosphere - but there is a serendipitous element to such studies. Not all behaviors were revealed because some events are very rare and even if they did happen we did not always have the instrumentation that we needed in the right places at the right time. These studies revealed mechanisms but could not give understanding of their overall importance and occurrence. There are classes of statistical studies that can help in this, such as superposed epoch analysis (a.k.a. "Chree analysis", and in meteorology termed "compositing") and statistical studies that divide the data by the phase of a sequence (e.g., by the phase of the substorm cycle) but these can only inform about a pre-determined sequence (which defines over which intervals to average data).

The next phase of studying the joined-up, time-dependent magnetosphere was the development of numerical modelling. Global MHD models and targeted kinetic models allow us to generate unusual, even unrealistic, scenarios that can reveal facets of the response of the magnetosphere that we had not seen, or maybe seen but not recognized, in the observations. The difficulty here is understanding what is an artefact of the modelling (due to numerical errors) and what is a realistic magnetospheric behavior. To make it even more confusing, a key mechanism, magnetic reconnection, is both! This makes it vital that modelling is constantly referenced back to observational case and statistical studies.

Many studies implicitly used the idea that the magnetosphere would return towards a steady state after perturbation. It was realized that, particularly at sunspot maximum, the temporal variability of the solar wind and IMF was great enough to mean that the magnetosphere would always be returning towards a steady state, in which case the magnetosphere became a non-equilibrium system. It is often assumed that northward-IMF conditions, that persist for exactly half the time, give the steady-state towards which the magnetosphere relaxes. However, we know steady state is never achieved because, although the magnetic flux in the geomagnetic tail weakens during northward IMF, we have never seen the tail completely disappear. Quiet times are when the IMF points northward, so the solar wind flow gives a dusk-to-dawn electric field that is applied across the magnetosphere. However, the persistence of the tail means that the magnetic shear between the tail lobes across the cross-tail current sheet remains and magnetic reconnection can only give a dawn-to-dusk electric field (it may be weak or even be zero but it cannot match the dusk-to-dawn electric field applied by the solar wind). This means that there is a curl of $\underline{\mathbf{E}}$ within the magnetosphere which, by Faraday's law, means that there is a rate

[3]Peter Gabriel, "Slowburn" first solo album 1977 (a.k.a. "Car").

of change in the magnetic field, i.e., non-steady-state conditions (Lockwood, 2004; Lockwood, 2019). Hence, rather than northward IMF being a steady-state, it is an interval a slow decline of the open flux caused by continued low-level reconnection in the cross tail current sheet and by open flux closure by lobe reconnections taking place simultaneously or sequentially in both hemispheres (Lockwood and Moen, 1999). Northward IMF conditions can persist for long enough for the magnetosphere to become very close to becoming fully closed (Wang et al., 2022) but has never yet been seen to last so long that the tail completely disappears. This argues that, rather than the magnetosphere sometimes not being an equilibrium system, that is never an equilibrium system.

This brings us to what I see as a new phase of study that we are just embarking upon, systems analysis of the magnetosphere, the analysis of the joined-up magnetosphere (Borovsky & Valdivia, 2018). This will often aided by new machine learning techniques (Camporeale, 2019).

THE JOINED-UP MAGNETOSPHERE

This idea is inspired by the development of Earth Systems Science, ESS, in environmental sciences (*Steffen* et al., 2020). Driven by anthropogenic changes to the natural world, such as ozone depletion, greenhouse trapping of heat by increased carbon dioxide, deforestation and pollution, there was a greater realization in the 1980s that the various aspects of terrestrial environmental science, such as the oceans, the atmospheric layers, the biomass and its many ecosystems, the soil, the ice sheets, were all interconnected—and that the mechanisms and feedbacks of those interconnections were important but not understood. It was also realized that the interactions were perturbed by both human activities and natural events such as volcanoes, large meteorite impacts and solar change and there were a variety of timescales on which the subsequent interactions took place. The early development of numerical climate models highlighted limitations in our understanding of behavior of the holistic system.

That is not dissimilar to the state of magnetospheric science until quite recently—we knew of the various regions, we knew their main behaviors, we knew a lot of mechanisms—but we had only a limited holistic view of the whole system. Hence it was almost the complete opposite of the study of bird murmuration, where we can visually see the holistic behavior but needed research to understand the mechanisms that were at work. We do have numerical models of the various regions of geospace and we had learned how to couple them together to some extent, although very often the links have been rather crude parameterizations rather using fully self-consistent numerical modelling and changes in scale lengths and times are a problem.

ESS studies the relationships between various global parameters that quantify the state of all or key parts of the system such as the Global Mean Air Surface Temperature (GMAST), the global warming index or the global radiative forcing, the volume of ice in the polar cap ice sheets, the global mean sea level, the carbon dioxide mixing ratio, ocean acidity, total ocean heat content, the drought severity index and potentially other global or region aggregates of factors such as mean soil water content, tundra extent, vegetation growth rates, stratospheric aerosol content and biomass dryness.

We have a number of analogous indices and indicators that tell us about the state of aspects of the magnetosphere. These include: the transpolar voltage Φ_{PC} (physically more meaningful when split into the magnetopause and tail reconnection voltages, Φ_D and Φ_N but these are not so routinely measured) which tell us about magnetic flux transfer through the system; the total open solar flux F_{PC}, given by the area of the ionospheric polar caps $A_{PC} = F_{PC}/B_i$, which tells us about the amount of energy stored in the geomagnetic tail; the dayside auroral *AU* (or *SMU*) geomagnetic indices that tell us about reconnection in the dayside magnetopause, and the nightside auroral *AL* (or *SML*) geomagnetic indices, that tell us about reconnection in the tail (Lockwood and McWilliams, 2021a); the *Dst*, *Sym-H* or *SMR* indices (or their time derivatives) that are mainly dominated by the ring current and so tell us the fluxes of energetic particles in that region; the Polar Cap Indices *PCN* and *PCS* that quantify of the current and electric field in the polar caps; the mid-latitude range indices *kp* (*ap*) and *am* and mid-latitude hourly indices such as *IDV* and *IHV* that tell us about the disturbance level of the geomagnetic field on a global basis; and the TIROS or *mPe* and *mPi* precipitation power indices (for electrons and ions respectively). There are also a great many near-continuous measurement series of specific variables that monitor a specific aspect of the magnetosphere.

In the past, these have been used in correlative studies with combinations of interplanetary parameters ("coupling functions", designed to quantify solar wind control of the magnetosphere; see reviews by Lockwood and McWilliams (2021b) and Lockwood (2022)) and one index or indicator which was tacitly taken to quantify the state of the whole magnetosphere. The corollary that one index can do this is that the coupling function should predict all magnetospheric indices and indicators equally well. Newell et al. (2007) proposed is that there is a "universal coupling function" that best predicts all terrestrial space weather indices and indicator. However, this idea runs counter to the method now routinely used to reconstruct interplanetary parameters from historic observations of geomagnetic activity which exploit the finding that different geomagnetic indices have different responses to interplanetary parameters and so combinations of historic index observations can be used to infer the separate interplanetary parameters (e.g., Lockwood et al., 2014) and this led to the reconstruction of a space weather climatology over the last 400 years from geomagnetic, sunspot and cosmogenic isotopes data (Lockwood et al., 2018). That different, seemingly-global terrestrial indices correlate best with different coupling functions has been demonstrated directly using simultaneous datasets by Lockwood and McWilliams (2021b). The limitation of coupling functions is that they cannot account for factors such as variable feedback loops between different aspects and parts of the magnetosphere or the effects of preconditioning space weather conditions on the prior state of the magnetosphere (Borovsky, 2021).

Borovsky and Osmane (2019) present a composite correlation system analysis scheme that points to the way forward, being aimed at exploiting all the best correlations with interplanetary parameters and various state indices of the magnetosphere. The challenge now is to make use of all best knowledge of the interconnections between different parts and aspects of the magnetosphere to generate optimum prediction schemes. These would include allowance for different lags - both the lags between the interplanetary parameters and the magnetospheric state indicators and the lags associated with the interaction chains within the magnetosphere. This development should go hand-in-hand with the findings from machine learning techniques about the correlations and interaction chains (Camporeale, 2019). There are a number of technical developments to explore. For example, correlations might need weighting according to significance, and the difference between correlations evaluated allowing for the fact that the different parameters are inter-correlated. Alternatively, or additionally, time-integral correlations may appropriate in some or all cases (Lockwood et al., 2016; Borovsky, 2017). These steps should mean that the prediction schemes can allow for one of the biggest limitations of coupling functions, namely pre-conditioning by the existing state of the magnetosphere (Lockwood and McWilliams, 2021b; Lockwood, 2022). This should allow us to make more reliable and accurate forecasts of space weather systems, with better defined uncertainties. In turn, this will aid the operators of the various operational systems perturbed or damaged by space weather effects.

Like the previous overlapping "phases" of magnetospheric discovery, the development of what one could term GSS ("Geospace System Science") will both contribute to and learn from new theory, case studies, statistical studies, and numerical modelling. One important benefit of better understanding of magnetospheric dynamics is in the development of a space climatology and, in particular, predictions of extreme event occurrence probabilities, which is information urgently needed to allow more cost-effective design of robust systems such as power grids, spacecraft, and communications and navigation systems.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The work of the author is supported by the UK Science and Technology Facilities Council consolidated grant number ST/R000921/1.

REFERENCES

Asikainen, T., Maliniemi, V., and Mursula, K. (2010). Modeling the Contributions of Ring, Tail, and Magnetopause Currents to the correctedDstindex. *J. Geophys. Res.* 115, a–n. doi:10.1029/2010JA015774

Borovsky, J. E. (2021). Is Our Understanding of Solar-Wind/Magnetosphere Coupling Satisfactory? *Front. Astron. Space Sci.* 8, 634073. doi:10.3389/fspas.2021.634073

Borovsky, J. E., and Osmane, A. (2019). Compacting the Description of a Time-dependent Multivariable System and its Multivariable Driver by Reducing the State Vectors to Aggregate Scalars: the Earth's Solar-Wind-Driven Magnetosphere. *Nonlin. Process. Geophys.* 26, 429–443. doi:10.5194/npg-26-429-2019

Borovsky, J. E. (2017). Time-integral Correlations of Multiple Variables with the Relativistic-Electron Flux at Geosynchronous Orbit: The strong Roles of the Substorm-Injected Electrons and the Ion Plasma Sheet. *J. Geophys. Res.* 122, 11961–11990. doi:10.1002/2017ja024476

Borovsky, J. E., and Valdivia, J. A. (2018). The Earth's Magnetosphere: A Systems Science Overview and Assessment. *Surv. Geophys.* 39, 817–859. doi:10.1007/s10712-018-9487-x

Camporeale, E. (2019). The challenge of Machine Learning in Space Weather: Nowcasting and Forecasting. *Space Weather* 17, 1166–1207. doi:10.1029/2018SW002061

Cavagna, A., Cimarelli, A., Giardina, I., Parisi, G., Santagati, R., Stefanini, F., et al. (2010). Scale-free Correlations in Starling Flocks. *Proc. Natl. Acad. Sci.* 107 (26), 11865–11870. doi:10.1073/pnas.1005766107

Chapman, S., and Ferraro, V. C. A. (1931). A New Theory of Magnetic Storms. *J. Geophys. Res.* 36 (3), 171–186. doi:10.1029/TE036i003p00171

Ginelli, F. (2016). The Physics of the Vicsek Model. *Eur. Phys. J. Spec. Top.* 225, 2099–2117. doi:10.1140/epjst/e2016-60066-8

Lockwood, M. (2004). Solar Outputs, their variations and their effects of Earth. *The Sun, Solar Analogs and the Climate. Proc. Saas-Fee Advanced Course* 34. Handel, Switzerland: Nature, 107–304.

Lockwood, M. (2019). Does adding solar wind Poynting flux improve the optimum solar wind-magnetosphere coupling function? *J. Geophys. Res. Space Phys.* 124, 5498–5515. doi:10.1029/2019JA026639

Lockwood, M., and McWilliams, K. A. (2021a). A survey of 25 years' transpolar voltage data from the SuperDARN radar network and the Expanding–Contracting Polar Cap model. *J. Geophys. Res. Space Phys.* 126, e2021JA029554. doi:10.1029/2021JA029554

Lockwood, M., and McWilliams, K. A. (2021b). On Optimum Solar Wind-Magnetosphere Coupling Functions for Transpolar Voltage and Planetary Geomagnetic Activity. *JGR Space Phys.* 126, e2021JA029946. doi:10.1029/2021JA029946

Lockwood, M., and Moen, J. (1999). Reconfiguration and Closure of Lobe Flux by Reconnection during Northward IMF: Possible Evidence for Signatures in Cusp/cleft Auroral Emissions. *Ann. Geophys.* 17, 996–1011. doi:10.1007/s00585-999-0996-2

Lockwood, M., Nevanlinna, H., Barnard, L., Owens, M. J., Harrison, R. G., and Scott, A. P. C. J. (2014). Reconstruction of Geomagnetic Activity and Near-Earth Interplanetary Conditions over the Past 167 Yr - Part 4: Near-Earth Solar Wind Speed, IMF, and Open Solar Flux. *Ann. Geophys.* 32, 383–399. doi:10.5194/angeo-32-383-2014

Lockwood, M., Owens, M. J., Barnard, L. A., Bentley, S., Scott, C. J., and Watt, C. E. (2016). On the Origins and Timescales of Geoeffective IMF. *Space Weather* 14, 406–432. doi:10.1002/2016SW001375

Lockwood, M., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., and Bentley, S. (2018). Space Climate and Space Weather over the Past 400 years: 2. Proxy Indicators of Geomagnetic Storm and Substorm Occurrence. *J. Space Weather Space Clim.* 8, A12. doi:10.1051/swsc/2017048

Lockwood, M. (2022). *Solar Wind – Magnetosphere Coupling Functions: Pitfalls, Limitations and Applications*. Space Wea. 20, e2021SW002989. doi:10.1029/2021SW002989

Milan, S. E., Carter, J. A., Sangha, H., Bower, G. E., and Anderson, B. J. (2021). Magnetospheric Flux Throughput in the Dungey Cycle: Identification of Convection State during 2010. *J. Geophys. Res. Space Phys.* 126, e2020JA028437. doi:10.1029/2020JA028437

Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., and Rich, F. J. (2007). A Nearly Universal Solar Wind-Magnetosphere Coupling Function Inferred from 10 Magnetospheric State Variables. *J. Geophys. Res.* 112, a–n. doi:10.1029/2006JA012015

Siscoe, G. L., Formisano, V., and Lazarus, A. J. (1968). Relation between Geomagnetic Sudden Impulses and Solar Wind Pressure Changes-An Experimental Investigation. *J. Geophys. Res.* 73 (15), 4869–4874. doi:10.1029/JA073i015p04869

Steffen, W., Richardson, K., Rockström, J., Schellnhuber, H. J., Dube, O. P., Dutreuil, S., et al. (2020). The Emergence and Evolution of Earth System Science. *Nat. Rev. Earth Environ.* 1, 54–63. doi:10.1038/s43017-019-0005-6

Turner, N. E., Baker, D. N., Pulkkinen, T. I., and McPherron, R. L. (2000). Evaluation of the Tail Current Contribution toDst. *J. Geophys. Res.* 105 (A3), 5431–5439. doi:10.1029/1999JA000248

Wang, X.–Y., Zhang, Q.–H., Wang, C., Zhang, Y.–L., Tang, B.–B., Xing, Z.–Y., et al. (2022). *A Small and Nearly Fully Closed Earth's Magnetosphere for Strongly Northward Interplanetary Magnetic Field*. PNAS, US National Academy of Sciences submitted.

Watkins, N. W., Pruessner, G., Chapman, S. C., Crosby, N. B., and Jensen, H. J. (2016). 25 Years of Self-Organized Criticality: Concepts and Controversies. *Space Sci. Rev.* 198, 3–44. doi:10.1007/s11214-015-0155-x

Young, G. F., Scardovi, L., Cavagna, A., Giardina, I., and Leonard, N. E. (2013). Starling Flock Networks Manage Uncertainty in Consensus at Low Cost. *Plos Comput. Biol.* 9 (1), e1002894. doi:10.1371/journal.pcbi.1002894

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 13 April 2022
doi: 10.3389/fspas.2022.892965

Second-Class Citizen in the Heliophysics Community

*Stephen Kahler**

Space Vehicles Directorate, AFRL/RVBX, Kirtland AFB, Albuquerque, NM, United States

The study of solar energetic particles (SEPs) is an important area of solar research and space weather. An SEP event extends over large regions of the heliosphere, involves energy ranges varying by decades, and evolves over various time and spatial scales and with ion composition, but with SEP observations limited to *in situ* detections on a few spacecraft for any given event, we are unable to observe these properties synoptically. Solar studies in general are the beneficiaries of imaging and remote sensing observations over practically all wavelengths and timescales from ground and space based detectors that drive increasingly highly sophisticated models. I see this divide as creating a two-class system for researchers, with us SEP researchers as second class members. Following a brief review of my experience with solar imagery and failed ideas on remote imaging of SEP events, I review two remarkable developments that give hope for some new SEP imaging technique. Finally, I discuss two poorly understood questions of impulsive and gradual SEP events that I think can be feasibly approached with current modeling techniques.

Keywords: solar energetic particles, solar flares (1496), coronal mass ejection, solar corona and wind, heliospheric magnetic field

OPEN ACCESS

Edited by:
Joseph E. Borovsky,
Space Science Institute, United States

Reviewed by:
Lulu Zhao,
University of Michigan, United States
Donald Reames,
University of Maryland, United States

Correspondence:
Stephen Kahler
stephen.kahler@us.af.mil

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *09 March 2022*
Accepted: *24 March 2022*
Published: *13 April 2022*

Citation:
Kahler S (2022) Second-Class Citizen in the Heliophysics Community. Front. Astron. Space Sci. 9:892965. doi: 10.3389/fspas.2022.892965

INTRODUCTION

Solar-heliophysics encompasses a broad range of topics and research techniques. Over the past several decades I believe there has been a growing broad division in the community between those who work with remote observations and imaging techniques, and others, like me, who are confined to getting their primary solar data only from *in situ* observations. I have in mind observations of solar energetic particles (SEPs) that can be observed only at several heliospheric locations, currently confined to the ecliptic plane. It is great that we have Parker Solar Probe and Solar Orbiter, but the point is that no matter how many SEP detectors we have in space, we are still just drawing samples of a broadly distributed and evolving phenomenon, the SEP event (I'm thinking E > 10 MeV ions and maybe E > 0.5 MeV electrons). There may be multiple sources of unknown size scales from various seed populations and spatial distributions of unknown numbers of SEPs over wide (or maybe narrow) ranges of longitude and latitude for a given SEP event, and the story gets more complex as we ask about different SEP energy ranges and ions of different rigidities. We continue to depend on statistics of many SEP events just to get a rough handle on their basic energy and spatial distributions. The blind man is much better informed about his proverbial elephant than we are about SEP events in the heliosphere. In the meantime from the remote observers, I am dazzled by solar images and model representations of increasing spatial, temporal, and thermal or energy resolution with ever more detailed physics.

Maybe years of frustration are catching up with me. It didn't always seem this way. I began with analysis of SEP proton events in grad school, then at NRL I worked with OGO-V X-ray flares from the NRL full-sun proportional counter. A definite change of focus from SEPs to flare X-rays, but the basic tool was still analysis of time series of detector counting rates at different energy bands. The object was to get temporal variations of X-ray spectra (flare temperatures and emission measures) or SEP energies and their characteristic values. While one case was *in situ* observations (SEPs) and the other remote observations of full disk emission from the Sun, that seemed like a minor distinction. It was all solar flare physics.

SOLAR IMAGES CONFRONT AUTHOR

I remember my first encounter with a full disk solar X-ray rocket image, proudly presented for my consideration in an early meeting with my new boss, Pippo Vaiana, the American Science and Engineering (AS&E) physicist in charge of solar observations within the Ricardo Giacconi X-ray astrophysics group. I had a sense of panic that I would somehow have to make a big transition from working with simple time series data to getting physics out of those dark photographic blobs representing solar active regions. Then came the AS&E X-ray telescope on Skylab with lots of solar X-ray images on film. Locations and evolution of coronal holes or numbers and distributions of bright points seemed like straightforward approaches to take from solar X-ray images, and after a poor start (Kahler, Krieger, and Vaiana, 1975), I got used to analyzing X-ray flare images and later analyzed X-ray coronal hole boundaries (e.g., Kahler and Hudson, 2002). Continued interest in SEP events led to correlations of radio bursts and CMEs with SEP events, which did not require image analysis. I began collaborating with Don Reames and Ed Cliver about 1982, again looking only at CME or radio burst listings, not needing image analysis. Work using SEP events as probes of magnetic clouds followed (Kahler et al., 2011), again no remote images needed.

THE DIVIDE OF IMAGING VERSUS *IN SITU*

During the past 2 decades it has been impressive to see the great successes of solar imaging missions. SOHO images greatly eclipsed the pioneering Skylab images, thanks to the revolution from film detectors to CCDs, with ever greater fields of view and spatial resolutions as images of Hinode, SDO, IRIS, PSP, and other missions are presented and analyzed in detail. Advances in physical models combining detailed calculations with quantitative images of magnetic field lines and ionic radiation have led to deeper appreciation of the physical processes in space and time in the solar atmosphere and interplanetary space. **Figure 1** shows several recent examples in which authors have combined models and data to extend imagery to heliospheric SEP ion distributions.

In stark contrast, the observation of SEPs can be made only with *in situ* detectors, from which we can produce intensity-time profiles (**Figure 2A**), evolving or fluence energy spectra, and spatial distributions of SEP events by compiling event averages (Lario et al., 2006; Cohen et al., 2017; see **Figure 2B**), but a host of questions about spatial/temporal/compositional/energetic evolution of SEP events remains untouched and unapproachable. As a researcher of SEP events, I am frustrated and envious of my first-class colleagues who traffic in spiffy, eye-catching solar/heliospheric images even more spectacular than those of **Figure 1**. If you are a magazine or journal science editor hoping to engage your reader with a single image from heliospheric SEP physics, do you go with an example from **Figure 1** or from **Figure 2**? Are the SEP distributions of **Figure 1** right on or badly off the mark? We'll never know because we can't image the SEP events we now study.

This is not my first whine on the physical barrier to becoming a first-class heliospheric research citizen with images of real SEP particle distributions, evolving in time, and color-coded for energy, maybe even (I'm dreaming here) composition. With co-author B. Ragot we set the goal (**Figure 2C**) of exploring possible ways that SEPs might interact with the SW to produce neutral radiation that can be imaged by a detector and maybe deconvolved to produce 3-D spatial reconstructions. We (Kahler and Ragot, 2008) found that 4–7 MeV ion de-excitation from SEP collisions with heliospheric ^{16}O and ^{12}C would be far too weak for observation, but that π^0-decay γ-rays as detectable signatures of $E \geq 300$ MeV nuc^{-1} SEP ions was possible in large events.

Further candidate remote SEP signatures of positron-decay 0.511 MeV line emission from $E > 300$ MeV protons; neutrons and the 2.23 MeV neutron-capture line from $E > 30$ MeV nuc^{-1} ions; synchrotron emission from $E > 0.3$ MeV electrons; and transition radiation (TR) from $E < 100$ keV electrons and from ions were discussed in Kahler and Ragot (2009). TR arises any time a particle crosses an inhomogeneous medium with variation in refractive index and has likely been observed in decimetric bursts of turbulent flares (Fleishman et al., 2005) and in type II bursts from narrow density structures in wakes of CMEs (Chernov et al., 2007). It is best generated by electrons in dense regions where $\omega_p >> \omega_B$, but by protons only in tenuous regions of $\omega_p << \omega_B$ (Fleishman and Kahler, 1992). A common theme is that it is not enough to detect such radiation, but it must be imaged to distinguish populations trapped in the corona from those of interplanetary space.

TWO HOPEFUL SURPRISES

SEP Event ENAs

At the time of our second paper (Kahler and Ragot, 2009) energetic neutral atoms (ENAs) were known as messengers of distant energetic particle populations and the basic tool of the then recently launched (2008) Interstellar Boundary Explorer (IBEX) mission (McComas et al., 2009) to explore the heliospheric termination shock and heliosheath. We acknowledged, but did not explore, ENAs as possible probes of SEPs, so it was an exciting surprise to learn that a SEP event on 5 December 2006 had been detected on STEREO with ENAs by Mewaldt et al. (2009). This was an appropriately big deal at the time and a quick glimpse of ENAs as a promising basis of SEP imaging, as they propagated directly to Earth through a thick sludge of heliospheric magnetic turbulence that retarded the arrival of the charged SEPs composing the main event. The

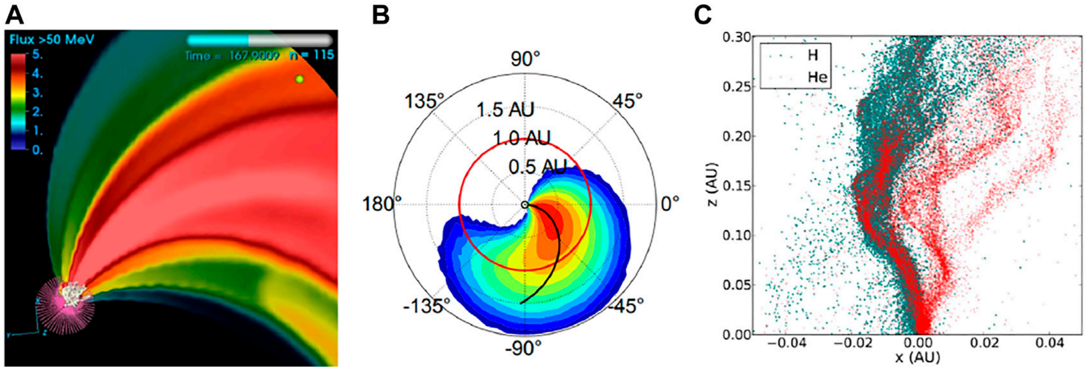

FIGURE 1 | (A): E > 50 MeV proton distribution of the 14 July 2000 event 3 h after injection with the STAT model of Linker et al. (2019). **(B)**: Equatorial log color-coded distribution of 10 MeV protons 3 h after impulsive injection based on the FP-FLRP model of Laitinen et al. (2016). **(C)**: model images of H (turquoise) and 4He (red) SEP spatial distributions following spatially separated but simultaneous impulsive injections (Guo et al., 2022).

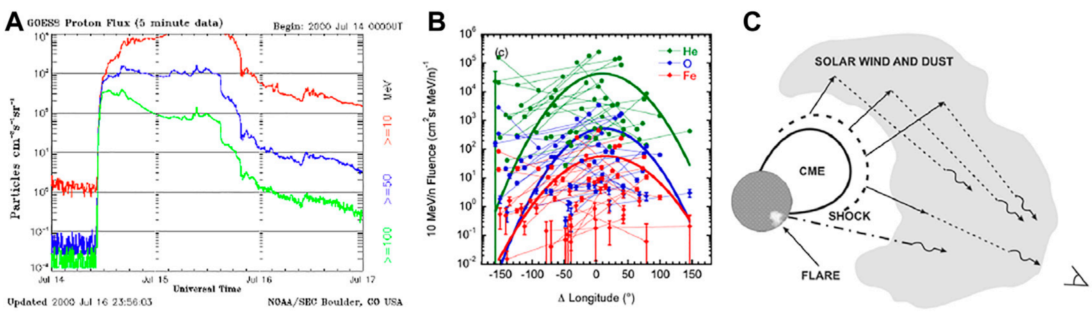

FIGURE 2 | (A): GOES profiles of the 14 July 2000 SEP event modeled in **Figure 1**. **(B)**: Longitude of event fluences for 10 MeV nuc^{-1} He, O, and Fe with best fit curves (Cohen et al., 2017). **(C)**: schematic of remote observation of SEPs against a grey background of SW and dust (Kahler and Ragot, 2008).

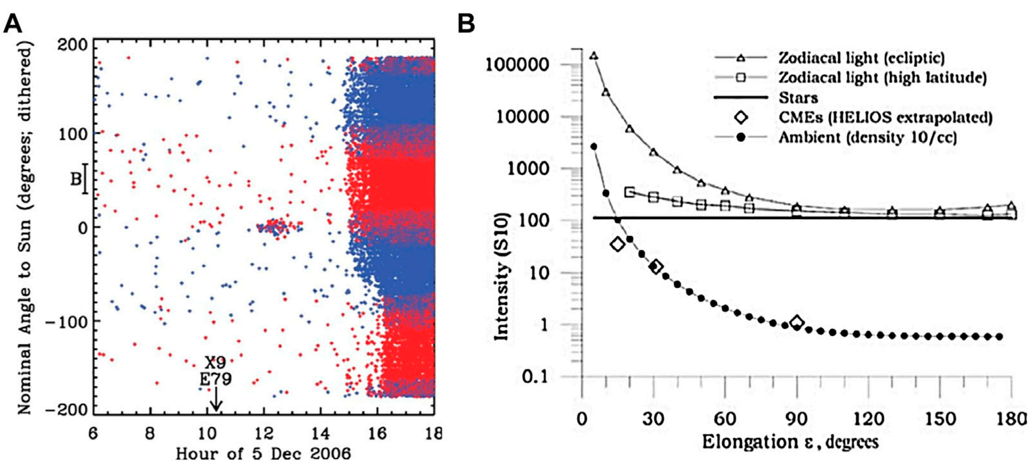

FIGURE 3 | (A): Plot of the measured angle to the Sun for individual 1.6–12 MeV proton events on 5 December 2006 (red = LET-A; blue = LET-B). Note the group of counts within ±10º of the Sun from ~1130 UT to ~1350 UT, well before the SEP onset at ~1445 UT (Mewaldt et al., 2009). **(B)**: Surface brightness in S10 units versus solar elongation angle for zodiacal and star light, and of expected CME brightness extrapolated from Helios measurements. A calculation of an ambient medium having a density of 10 e– cm^{-3} at one AU and an inverse-square density drop off with solar distance is also shown (Jackson et al., 2004).

STEREO A/B Low Energy Telescopes (LET-A and LET-B) were not designed to select neutrals, so it was the timing profile and directional information that confirmed the presence of SEPs (**Figure 3A**). While the LETs are not ENA detectors, separating charged and neutral particles, there were favorable conditions to observe that event. It was a very big (2000 pfu for E > 10 MeV proton) event, able to produce an observable ENA flux at one AU, and it originated on the east limb, allowing a sufficient temporal delay of the onset of the far larger SEP event. The ENAs were further confined to E < 5 MeV, as the cross section for the charge exchange cross section with ambient O^{+6} coronal/SW ions drops rapidly with energy (Mewaldt et al., 2009). Finally, the SW density drops at least as r^{-2}, presumably negating their use as heliospheric probes of SEPs, although charge exchange calculations of protons with atomic H, O^{6+}, and C^{4+} by Wang et al. (2014) suggest that ENA detectors of sufficiently low background could detect particle acceleration in the low corona. The Earth's dipole magnetic field may be such a detector, channeling high energy charged particles to the poles and converting low energy (>0.8 MeV) ENAs into quasi-trapped magnetic equatorial protons (Mason et al., 2021).

White Light Interplanetary CMEs

Imaging the SW was thought impossible until Helios B white-light photometer observations revealed the passages of CMEs through its heliospheric field of view (Jackson, 1985). Observing Thomson-scattering of solar white-light photons was also at one time considered a difficult challenge, but because of a very steady zodiacal and stellar white-light background the Solar Mass Ejection Imager imaged CMEs two orders of magnitude fainter than that background (Jackson et al., 2004; **Figure 3B**). The combination of interplanetary scintillation (IPS) and white-light observations now enable the SW velocities and densities to be reconstructed throughout the inner heliosphere (Jackson et al., 1988, 2020); see https://ips.ucsd.edu), a feat considered impossible at one time and suggesting that there may yet be hope for some new way to image SEPs in space.

Coronal/Interplanetary SEP Imaging

The hope and plea here is that somebody somewhere will get a brilliant idea to detect some kind of neutral radiation from energetic ions and electrons distributed throughout the heliosphere as a SEP event. The odds are really long, but the rewards are enormous. We (currently second-class research citizens stuck with our *in situ* observations) will be able to join our fellow first-class citizens in proudly displaying images of SEP events and making direct comparisons with increasingly sophisticated model outputs. The advances in understanding where SEPs originate relative to shocks and coronal and solar wind features, followed by their transport histories will greatly accelerate our understanding of SEP physics.

A PLEA FOR TWO NEEDED SEP MODELS

I will end this story with a modest request to the SEP modeling community for two efforts addressing currently neglected targets that I think well within the capabilities of several SEP models. Rather than the usual procedure of starting with SEP events observed at one AU to estimate injection spectra and numbers, the models would start with injected SEP profiles and calculate resulting SEP numbers and energies observed in one AU detectors.

Total Numbers in ^{3}He Events

For nearly 50 years (Reames, 2021) SEP events have been observed with substantial enhancements (>100 ×) of ^{3}He/^{4}He over the coronal/SW abundance of 5×10^{-4} in the few MeV/nuc energy range. Those events are generally small and impulsive, with source regions in coronal flares and jets. The ^{3}He acceleration process was first explained by absorption of electromagnetic ion cyclotron waves, but currently favored (Reames, 2021) is magnetic reconnection in confined coronal volumes, which may account for upper limits to the observed ^{3}He event fluence distribution observed at one AU (Ho et al., 2005; **Figure 4**). The ^{3}He acceleration process appears to occur differently from that of ^{4}He (Ho et al., 2019) and may even completely strip a coronal source region of all ^{3}He ions (Reames, 2021).

Well-developed models of jets (Panesar et al., 2016, 2017; Wyper et al., 2018) and extensive observations of ^{3}He coronal sources (Bučík, 2020; Bučík et al., 2021) make it imperative that we compare a calculated ^{3}He ion injection population with corresponding one AU observations to determine the accelerated fraction of the source ^{3}He as a measure of the strength of the acceleration process. This has not been attempted since the Reames (1999) estimate assuming a source region area of 3,000 km^2, density of $\rho = 10^{10}$ cm^{-3}, scale height of 10^4 km, ^{3}He/^{4}He = 5×10^{-5} for a total 5×10^{31} ^{3}He in the volume. He assumed a large one AU event of 10^5 cm^{-2} ^{3}He (see **Figure 4A**) resulting from uniform injection in a 20° cone followed by scatter-free propagation and concluded that >10% of the source ^{3}He was accelerated. Surely we (the modelling community intended) can do much better than that. The basic goal is to connect the number of ^{3}He in the source region to the number accelerated and injected from the corona. A size estimate of a reconnection region of an observed jet source could serve as the basis of an input ^{3}He number with a nearly delta function injection in space and time assumed for the accelerated population.

Nearly all 0.02–2 MeV/nuc ^{3}He-rich events are associated with type III radio bursts (Wang et al., 2012; Bučík et al., 2016; Bučík et al., 2018; Bučík et al., 2021), which are used for timing ^{3}He injections but could also aid substantially in the ^{3}He source volume estimate. If we are lucky, the coronal injection regions of the type III-burst electrons are shared by the ^{3}He ions, so the coronal size and location of a ^{3}He event and its extent into the heliosphere will be defined by that of the type III radio burst. It is not yet feasible to image type III radio bursts, but that is exactly one of the goals of the NASA SunRISE mission, due for launch in 2023 (Kasper et al., 2019). SunRISE consists of six small spacecraft at supra-geosynchronous orbit with radio telescopes operating the in the 0.1–22 MHz region, which extends from 10 Rs to one AU.

Shock SEPs for CME Energetics

Fast CMEs are the drivers of shocks that accelerate coronal and SW seed particles to energies sometimes reaching GeV energies (Reames, 2020). That SEP energy is ultimately derived from the

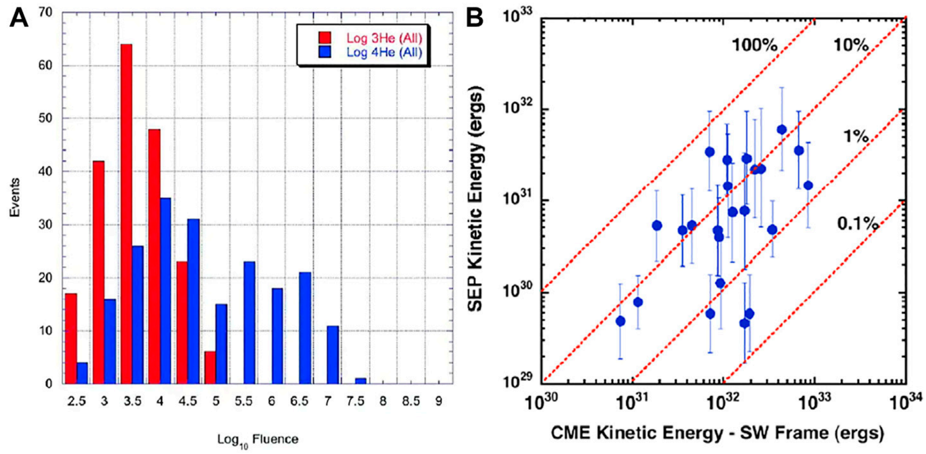

FIGURE 4 | (A): Plot of ^{3}He events versus fluence (cm^2 sr MeV/nucleon)$^{-1}$ observed 1997 to 2003 on ACE (Ho et al., 2005). **(B)**: Plot of SEP event energies versus the associated CME kinetic energies for 23 events (Mewaldt et al., 2008).

kinetic energy of the CME, so an important question is the conversion efficiency of the CME energy into that of SEPs. Mewaldt et al. (2008) examined this question for the 50 biggest E > 30 MeV SEP events of 1996–2003 using associated CME speeds and masses given for 23 of those events in Gopalswamy et al.(2004, 2005). The CME energies are estimated to be accurate within a factor of 2, but the hard part of the comparison was to estimate the total SEP energies, which Mewaldt et al. (2008) calculated with fluence spectra observed from 10 keV/nuc to 1,000 MeV/nuc. Included in their energy calculation were numbers of H and He particle crossings at one AU and adiabatic energy losses for each crossing. The source longitude and latitude distributions were assumed to fall off exponentially from central meridian with e-folding drop-offs of 25° east of CM and 45° to the west and 35° with latitude. For six events the abundances of He and heavier ions were measured, and for the remaining 17 events protons were assumed to be 75 ± 7% of the total energy. Assuming that the shock properties depend on the CME speed relative to the SW, Mewaldt et al. (2008) subtracted an assumed SW speed of 400 km/s from the CME speeds to calculate the CME kinetic energies. The resulting comparison is reproduced in **Figure 4B**, where the median efficiency is 6.5%.

A similar improved comparison based on simulations of one AU SEP scatterings and energy losses by Chollet et al. (2010) and Gaussian spatial distributions of SEP events adopted by Lario et al. (2006) was carried out by Emslie et al. (2012) for 20 SEP events with results (their **Figure 2B**) comparable to those of **Figure 4B**. Another comparison of 94 SEP and CME energies by Kahler and Vourlidas (2013) used CME speeds at measured centers of mass rather than the leading edge speeds and a rotationally symmetric exponential distribution with an e-folding angle of 45° for SEP events with spectra determined by only the 2 and 20 MeV H fluences. Their Figure 7 also showed high SEP efficiencies, including some exceeding unity.

The preceding works all used spatial, spectral, and transport assumptions to convert one AU observed SEP fluences to total numbers and energies of the produced SEPs. The results can be very model-dependent, however. With a simple CME latitude correction Gopalswamy et al. (2021) increased the number of interplanetary E > 500 MeV protons by about an order of magnitude in five of 14 SEP events calculated by de Nolfo et al. (2019) in their study of solar sustained gamma-ray events. In general, however, the assumption parameters are not tested in these SEP calculations. I propose that modelers go the other way, starting with a CME shock model producing SEP events of known spatial, temporal, and energy distributions. The model, with full transport properties, would then track and predict both the total accelerated SEP energies and the intensities and fluences observed at a designated one AU detector. In this scheme the shock longitudinal and latitudinal widths and acceleration timescale variations with energy could all be tracked. The advantage of this approach is that the SEP distributions and energies are known and can be compared with resulting one AU SEP observations and CME energies. Multiple model runs can then establish the uncertainties of the reverse process of estimating SEP energies solely from the one AU observations. The SEP efficiencies of CMEs are too important to be left in their current state of understanding.

CONCLUSION

I am resigned to continue my SEP investigations as a second-class citizen of the heliospheric research community, operating in the slow lane of *in situ* observations, while hoping for a better future through some great discovery. In the meantime I would be delighted to see acceptance of my challenges of the preceding Section.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was funded by AFOSR task 22RVCOR006.

ACKNOWLEDGMENTS

The views expressed are those of the author and do not reflect the official guidance or position of the United States Government, the Department of Defense or of the United States Air Force. My work has been supported by AFRL and AFOSR and has benefited from interactions with many colleagues, especially DR. I thank JB for inviting this Perspective, which has benefited from reviewer comments and discussions within the International Space Science Institute (ISSI) Team ID 425 "Origins of 3He-rich SEPs".

REFERENCES

Bučík, R. (2020). ^{3}He-Rich Solar Energetic Particles: Solar Sources. *Space Sci. Rev.* 216 (2), 24. doi:10.1007/s11214-020-00650-5

Bučík, R., Innes, D. E., Mason, G. M., and Wiedenbeck, M. E. (2016). Association of 3He-Rich Solar Energetic Particles with Large-Scale Coronal Waves. *ApJ* 833 (1), 63. doi:10.3847/1538-4357/833/1/63

Bučík, R., Innes, D. E., Mason, G. M., Wiedenbeck, M. E., Gómez-Herrero, R., and Nitta, N. V. (2018). 3He-rich Solar Energetic Particles in Helical Jets on the Sun. *ApJ* 852 (2), 76. doi:10.3847/1538-4357/aa9d8f

Bučík, R., Mulay, S. M., Mason, G. M., Nitta, N. V., Desai, M. I., and Dayeh, M. A. (2021). Temperature in Solar Sources of 3He-Rich Solar Energetic Particles and Relation to Ion Abundances. *ApJ* 908 (2), 243. doi:10.3847/1538-4357/abd62d

Chernov, G. P., Kaiser, M. L., Bougeret, J.-L., Fomichev, V. V., and Gorgutsa, R. V. (2007). Fine Structure of Solar Radio Bursts Observed at Decametric and Hectometric Waves. *Sol. Phys.* 241 (1), 145–169. doi:10.1007/s11207-007-0258-y

Chollet, E. E., Giacalone, J., and Mewaldt, R. A. (2010). Effects of Interplanetary Transport on Derived Energetic Particle Source Strengths. *J. Geophys. Res.* 115 (A6), A06101. doi:10.1029/2009ja014877

Cohen, C. M. S., Mason, G. M., and Mewaldt, R. A. (2017). Characteristics of Solar Energetic Ions as a Function of Longitude. *ApJ* 843 (2), 132. doi:10.3847/1538-4357/aa7513

de Nolfo, G. A., Bruno, A., Ryan, J. M., Dalla, S., Giacalone, J., Richardson, I. G., et al. (2019). Comparing Long-Duration Gamma-Ray Flares and High-Energy Solar Energetic Particles. *ApJ* 879 (2), 90. doi:10.3847/1538-4357/ab258f

Emslie, A. G., Dennis, B. R., Shih, A. Y., Chamberlin, P. C., Mewaldt, R. A., Moore, C. S., et al. (2012). Global Energetics of Thirty-Eight Large Solar Eruptive Events. *ApJ* 759 (1), 71. doi:10.1088/0004-637x/759/1/71

Fleishman, G. D., and Kahler, S. W. (1992). Microwave Transition Radiation in Solar Flares and in Astrophysics. *ApJ* 394, 688. doi:10.1086/171622

Fleishman, G. D., Nita, G. M., and Gary, D. E. (2005). Evidence for Resonant Transition Radiation in Decimetric Continuum Solar Bursts. *ApJ* 620 (1), 506–516. doi:10.1086/427022

Gopalswamy, N., Yashiro, S., Krucker, S., Stenborg, G., and Howard, R. A. (2004). Intensity Variation of Large Solar Energetic Particle Events Associated with Coronal Mass Ejections. *J. Geophys. Res. Space Phys.* 109 (12), A12105. doi:10.1029/2004ja010602

Gopalswamy, N., Yashiro, S., Liu, Y., Michalek, G., Vourlidas, A., Kaiser, M. L., et al. (2005). Coronal Mass Ejections and Other Extreme Characteristics of the 2003 October-November Solar Eruptions. *J. Geophys. Res. Space Phys.* 110, A09S15. doi:10.1029/2004ja010958

Gopalswamy, N., Yashiro, S., Mäkelä, P., Xie, H., and Akiyama, S. (2021). The Common Origin of High-Energy Protons in Solar Energetic Particle Events and Sustained Gamma-Ray Emission from the Sun. *ApJ* 915 (2), 82. doi:10.3847/1538-4357/ac004f

Guo, F., Zhao, L., Cohen, C. M. S., Giacalone, J., Leske, R. A., Wiedenbeck, M. E., et al. (2022). Variable Ion Compositions of Solar Energetic Particle Events in the Inner Heliosphere: A Field Line Braiding Model with Compound Injections. *ApJ* 924 (1), 22. doi:10.3847/1538-4357/ac3233

Ho, G. C., Mason, G. M., and Allen, R. C. (2019). 3He-Rich Solar Energetic Particle Events with No Measurable 4He Intensity Increases. *Sol. Phys.* 294 (3), 33. doi:10.1007/s11207-019-1420-z

Ho, G. C., Roelof, E. C., and Mason, G. M. (2005). The Upper Limit on 3He Fluence in Solar Energetic Particle Events. *ApJ* 621 (2), L141–L144. doi:10.1086/429251

Jackson, B. V., Buffington, A., Cota, L., Odstrcil, D., Bisi, M. M., Fallows, R., et al. (2020). Iterative Tomography: A Key to Providing Time- Dependent 3-D Reconstructions of the Inner Heliosphere and the Unification of Space Weather Forecasting Techniques. *Front. Astron. Space Sci.* 7, 76. doi:10.3389/fspas.2020.568429

Jackson, B. V., Buffington, A., Hick, P. P., Altrock, R. C., Figueroa, S., Holladay, P. E., et al. (2004). The Solar Mass-Ejection Imager (SMEI) Mission. *Sol. Phys.* 225 (1), 177–207. doi:10.1007/s11207-004-2766-3

Jackson, B. V. (1985). Imaging of Coronal Mass Ejections by the HELIOS Spacecraft. *Solar Phys.* 100 (1/2), 563–574. doi:10.1007/978-94-009-4588-3_27

Jackson, B. V., Rompolt, B., and Švestka, Z. (1988). Solar and Interplanetary Observations of the Mass Ejection on 7 May, 1979. *Sol. Phys.* 115 (2), 327–343. doi:10.1007/bf00148732

Kahler, S. W., Haggerty, D. K., and Richardson, I. G. (2011). Magnetic Field-Line Lengths in Interplanetary Coronal Mass Ejections Inferred from Energetic Electron Events. *ApJ* 736 (2), 106. doi:10.1088/0004-637x/736/2/106

Kahler, S. W., and Hudson, H. S. (2002). Boundary Structures and Changes in Long-lived Coronal Holes. *ApJ* 574 (1), 467–476. doi:10.1086/340937

Kahler, S. W., Krieger, A. S., and Vaiana, G. S. (1975). Morphological Evolution of X-ray Flare Structures from the Rise through the Decay Phase. *ApJ* 199 (2), L57. doi:10.1086/181848

Kahler, S. W., and Ragot, B. R. (2008). Remote Sensing of Gamma-Ray Emission from Solar Energetic Proton Interactions with the Solar Wind. *ApJ* 675 (1), 846–852. doi:10.1086/526416

Kahler, S. W., and Ragot, B. R. (2009). Viewing Radiation Signatures of Solar Energetic Particles in Interplanetary Space. *Adv. Space Res.* 43 (10), 1484–1490. doi:10.1016/j.asr.2009.01.013

Kahler, S. W., and Vourlidas, A. (2013). A Comparison of the Intensities and Energies of Gradual Solar Energetic Particle Events with the Dynamical Properties of Associated Coronal Mass Ejections. *ApJ* 769 (2), 143. doi:10.1088/0004-637x/769/2/143

Kasper, J. C., Lazio, J., Romero-Wolf, A., Bain, H. M., Bastian, T., Cohen, C., et al. (2019). *The Sun Radio Interferometer Space Experiment (SunRISE) Mission Concept.* San Francisco: American Geophysical Union. Fall Meeting 2019, abstract #SH33A-02.

Laitinen, T., Kopp, A., Effenberger, F., Dalla, S., and Marsh, M. S. (2016). Solar Energetic Particle Access to Distant Longitudes through Turbulent Field-Line Meandering. *A&A* 591, A18. doi:10.1051/0004-6361/201527801

Lario, D., Kallenrode, M. B., Decker, R. B., Roelof, E. C., Krimigis, S. M., Aran, A., et al. (2006). Radial and Longitudinal Dependence of Solar 4-13 MeV and 27-37 MeV Proton Peak Intensities and Fluences: Helios and IMP 8 Observations. *ApJ* 653, 1531–1544. doi:10.1086/508982

Linker, J. A., Caplan, R. M., Schwadron, N., Gorby, M., Downs, C., Torok, T., et al. (2019). Coupled MHD-Focused Transport Simulations for Modeling Solar Particle Events. *J. Phys. Conf. Ser.* 1225 (1), 012007. doi:10.1088/1742-6596/1225/1/012007

Mason, G. M., Greenspan, M. E., Kanekal, S. G., Leske, R. A., Looper, M. D., Mazur, J. E., et al. (2021). Evidence for Energetic Neutral Hydrogen Emission from Solar Particle Events. *ApJ* 923 (2), 195. doi:10.3847/1538-4357/ac2fa2

McComas, D. J., Allegrini, F., Bochsler, P., Bzowski, M., Collier, M., Fahr, H., et al. (2009). IBEX-interstellar Boundary Explorer. *Space Sci. Rev.* 146 (1-4), 11–33. doi:10.1007/s11214-009-9499-4

Mewaldt, R. A., Cohen, C. M. S., Giacalone, J., Mason, G. M., Chollet, E. E., Desai, M. I., et al. (2008). How Efficient Are Coronal Mass Ejections at Accelerating Solar Energetic Particles? PARTICLE ACCELERATION and TRANSPORT IN the HELIOSPHERE and BEYOND: 7th Ann. *Internat. Astrophys. Conf. AIP Conf. Proc.* 1039, 111. doi:10.1063/1.2982431
Mewaldt, R. A., Leske, R. A., Stone, E. C., Barghouty, A. F., Labrador, A. W., Cohen, C. M. S., et al. (2009). STEREO Observations of Energetic Neutral Hydrogen Atoms during the 2006 December 5 Solar Flare. *ApJ* 693 (1), L11–L15. doi:10.1088/0004-637x/693/1/l11
Panesar, N. K., Sterling, A. C., Moore, R. L., and Chakrapani, P. (2016). Magnetic Flux Cancelation as the Trigger of Solar Quiet-Region Coronal Jets. *ApJ* 832 (1), L7. doi:10.3847/2041-8205/832/1/l7
Panesar, N. K., Sterling, A. C., and Moore, R. L. (2017). Magnetic Flux Cancellation as the Origin of Solar Quiet-Region Pre-jet Minifilaments. *ApJ* 844 (2), 131. doi:10.3847/1538-4357/aa7b77
Reames, D. V. (2021). Fifty Years of 3He-Rich Events. *Front. Astron. Space Sci.* 8, 164. doi:10.3389/fspas.2021.760261
Reames, D. V. (2020). Four Distinct Pathways to the Element Abundances in Solar Energetic Particles. *Space Sci. Rev.* 216 (2), 20. doi:10.1007/s11214-020-0643-5
Reames, D. V. (1999). Particle Acceleration at the Sun and in the Heliosphere. *Space Sci. Rev.* 90 (3/4), 413–491. doi:10.1023/a:1005105831781
Wang, L., Li, G., Shih, A. Y., Lin, R. P., and Wimmer-Schweingruber, R. F. (2014). Simulation of Energetic Neutral Atoms from Solar Energetic Particles. *ApJ* 793 (2), L37. doi:10.1088/2041-8205/793/2/l37
Wang, L., Lin, R. P., Krucker, S., and Mason, G. M. (2012). A Statistical Study of Solar Electron Events over One Solar Cycle. *ApJ* 759 (1), 69. doi:10.1088/0004-637x/759/1/69
Wyper, P. F., DeVore, C. R., and Antiochos, S. K. (2018). A Breakout Model for Solar Coronal Jets with Filaments. *ApJ* 852 (2), 98. doi:10.3847/1538-4357/aa9ffc

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 14 April 2022
doi: 10.3389/fspas.2022.883050

Planned Science and Scientific Discovery in Equatorial Aeronomy

*D. L. Hysell**

Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, United States

This paper discusses the relationship between planning and discovery in science using examples drawn from equatorial aeronomy in general and research at the Jicamarca Radio Observatory in particular. The examples reveal a pattern of discoveries taking place despite rather than because of careful planning.

Keywords: equatorial aeronomy, incoherent scatter, radar, instability, irregularity, discovery science, serendipity

Edited by:
Jorge Luis Chau,
Leibniz Institute of Atmospheric Physics (LG), Germany

Reviewed by:
Marco Milla,
Pontifical Catholic University of Peru, Peru

Correspondence:
D. L. Hysell
david.hysell@cornell.edu

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *24 February 2022*
Accepted: *11 March 2022*
Published: *14 April 2022*

Citation:
Hysell DL (2022) Planned Science and Scientific Discovery in Equatorial Aeronomy. Front. Astron. Space Sci. 9:883050. doi: 10.3389/fspas.2022.883050

1 INTRODUCTION

Edward Lorenz is famous because of a shortcut he took in 1961. Integrating a small system of differential equations numerically, he initialized the calculations using values from the middle of a prior run to save time. Surprisingly, the solution he found departed rapidly from the prior tabulated result. He traced the discrepancy to miniscule roundoff errors in the values he used for initial conditions. (His computer stored floating point numbers to six decimal places, but he had only printed and then rekeyed them to three). That small changes in the initial conditions could lead so quickly to drastic changes in the behavior of the system defied common sense. The finding led to chaos theory, raised deep questions about determinism, and doomed prospects for long-range weather forecasting, Lorenz's original problem (Lorenz, 1963).

Undoubtedly, other investigators had encountered similar phenomena in the early days of numerical computing but dismissed them, concentrating instead on the immediate problem at hand. Lorenz is remarkable for having set aside the comparatively routine task before him in favor of getting to the bottom of the "error." His keen judgment, and his freedom to move "off task," led to one of the most important scientific results of the 20th century with impacts beyond Lorenz's discipline.

History is full of scientific endeavors which were important because they did not go according to plan. A short list of examples includes Rutherford's discovery of the nucleus, Penzias' and Wilson's discovery of cosmic background radiation, and the Michelson Morley experiment. Space physics and aeronomy is no exception, and it is worth recounting a few more modest but more contemporary examples as reminders of how plans gone wrong remain hallmarks of scientific discovery.

2 EXAMPLES FROM EQUATORIAL AERONOMY AND SPACE PHYSICS

Built at about the same time as the Arecibo Observatory, the Jicamarca Radio Observatory was constructed outside Lima, Peru, to demonstrate the possibility of studying the upper atmosphere through the scattering of electromagnetic waves from free electrons (Gordon, 1958). This was to be a cost-effective way of learning about the environment that new spacecraft were being designed to inhabit. For a detailed history of early developments at Jicamarca, see (Woodman et al., 2019).

The motivation for Jicamarca's location was its proximity to the magnetic equator where it would be relatively inexpensive to build a flat antenna to illuminate the ionosphere at an angle normal to the earth's magnetic field. Early intuition about the scattering mechanism held that different ion species, gyrating about the magnetic field lines, would give distinct peaks in the autocorrelation function of the backscattered signal. The ionosphere would in effect become a giant ion mass spectrometer. The required sensitivity of the radar was expected to be substantial, and so a design employing 5 MW transmitter power and an antenna field with a 300 m-squared area was selected. The tests were not expected to take very long.

During the build-out of Jicamarca, extensive experiments were performed, and both the theory of scattering from thermal electron fluctuations in a plasma (now known as incoherent scatter) and the experimental techniques required to observe it underwent substantial development. The ion gyroresonances were not observed however. As (Farley, 1964; Dougherty, 1964) would show independently, the gyroresonances are largely destroyed by ion Coulomb collisions. Even when the Coulomb collision frequency is much less than the ion gyrofrequency, the gyroresonance cannot be observed if the deviation of the electron position over a gyroperiod is comparable to the radar wavelength divided by 4π. Later, Farley (Farley, 1967) would observe the gyroresonance for H+ ions [see also (Rodrigues et al., 2007)], but the O+ gyroresonance cannot be detected given nominal ionospheric conditions.

So early experiments at Jicamarca did not go according to plan, but were they a failure? Hardly. To begin with, the experiments provided a basis for much deeper quantitative understanding of incoherent scatter and the practical methods required to make it useful. These include an understanding of the interpretation of the scattering cross section, practical means of measuring absolute electron densities, and theory and methods for inferring ion composition and electron and ion temperatures from the autocorrelation function (Bowles et al., 1962; Bowles, 1963; Farley, 1966). Within a few years, the incoherent scatter technique had fulfilled its promises and could be used to measure the most important state parameters of ionospheric plasmas. The techniques were equally applicable at non-equatorial sites. ISR remains the most incisive tool for ground-based remote sensing of the ionosphere.

Had construction waited on a more complete knowledge of ISR theory, it is unlikely that an incoherent scatter radar would have been built at the magnetic equator. It is fortunate that it was because this led directly to an enormous range of discoveries in aeronomy that would otherwise have been greatly delayed. Some of these discoveries are at the center of contemporary research in equatorial aeronomy, space physics, and space weather.

2.1 The Temperature Ratio Problem

By the 1980s, plasma state parameter estimates including electron and ion temperature, ion composition, plasma drift, and electron number density were being measured at ISR facilities around the world and deposited in databases. The procedure involved measuring and fitting the scattered signal autocorrelation function to ISR theory. An inconsistency appeared in the results for Jicamarca, however, were the electron-to-ion temperature ratio at night was found to be consistently less than unity. This is not physically reasonable. There seemed to be a problem, but was it with the experiment or with ISR theory?

Other facilities were not reporting difficulties with the temperature ratio, and archival data from early measurements at Jicamarca did not exhibit the problem either, so it was assumed that a bias had crept into the modern experiment. A painstaking analysis of errors and biases in the methodology did not reveal a mistake, however (Pingree, 1990). Furthermore, new experiments showed that the problem nearly vanished when the angle between the radar beam and the perpendicular-to-B direction was increased. Eventually, records were unearthed indicating that the temperature ratio problem had always existed at Jicamarca but had been artificially "corrected" in the database (Clark et al., 1976)! Indeed, the temptation to simply "correct" the problem in the modern era was strong too.

The temperature radio problem ultimately pointed to a subtle problem with ISR theory at small magnetic aspect angles arising from the neglect of electron Coulomb collisions (Sulzer and Gonzalez, 1999; Woodman, 2004; Kudeki and Milla, 2011; Milla and Kudeki, 2011). The effect of Coulomb collisions on the ISR spectrum at small magnetic aspect angles is difficult to capture and has not been formulated in closed form, and numerical estimates of the experimental effects are expensive to calculate and store. There remains no completely satisfactory resolution to the problem, although interim methods allow us to infer electron and ion temperatures from ISR autocorrelation function measurements sufficiently well for most intents and purposes so long as the magnetic aspect angle is not too small.

More importantly, the problem is now being investigated using a variety of avante garde methods including large particle-in-cell (PIC) simulations, producing fundamental insights into transport properties in kinetic plasmas and effective means of exploring them [see (Longley et al., 2018; Longley et al., 2019)]. The research will very likely be more impactful in the end than the ionospheric energy balance problem that motivated electron and ion temperature ratio measurements in the first place.

2.2 Ionospheric Irregularities

Due to its frequency and its equatorial geometry, Jicamarca immediately encountered intense, ubiquitous, and unexpected radar "clutter." **Figure 1** shows a typical range-time-intensity plot illustrating coherent backscatter observed at Jicamarca over a 24-h period. The enhanced backscatter comes from plasma density irregularities excited by different mechanisms including neutral turbulence and plasma instabilities. Coherent scatter signifies free energy in the upper atmosphere and is a distinctive telltale of space weather.

Incoherent scatter cannot be measured in regions of strong clutter which comes through the main lobe and/or the sidelobes of all of Jicamarca's antenna pointing positions. This prohibits a number of desirable experiments from being performed. The scientific tradeoff of the clutter is favorable, however, since coherent scatter offers keen insights into important processes in equatorial aeronomy and space weather that might have gone unnoticed otherwise.

Near 100 km altitude, the equatorial electrojet, a strong current system confined to low magnetic latitudes, flows. The current gives rise to both configuration-space and phase-space instabilities, gradient drift and Farley Buneman instabilities

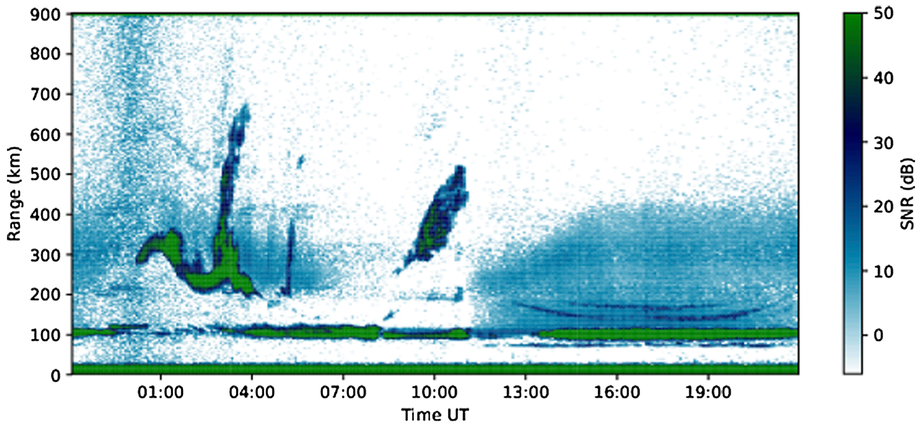

FIGURE 1 | Range time intensity plot for September 14, 2010, showing backscatter signal-to-noise ratio versus altitude and time. Note that UT = LT + 5 h. Cyan hues mainly represent incoherent scatter whereas blue and green hues represent coherent scatter (see text).

respectively (Farley, 1963; Simon, 1963). The resulting plasma density irregularities are strongly magnetic field aligned, as are irregularities in the *F* region and above. The two instabilities are also closely coupled and described by a unified dispersion relation (Fejer and Kelley, 1980).

The irregularities can be studied in detail near the magnetic equator where it is straightforward to distinguish echoes from different altitudes and where interferometry and imaging methods can be applied. Primary gradient drift waves can be observed at Jicamarca using interferometry and aperture-synthesis imaging, for example, (Kudeki et al., 1982; Hysell and Chau, 2002). Following their discovery at Jicamarca, similar instabilities were found to operate the auroral zone and, later, at middle latitudes. Farley Buneman instability has important implications outside equatorial aeronomy, altering the conductance of the auroral *E* region important for MI coupling (Liu et al., 2016) and causing heating in the solar chromosphere [e.g., (Madsen et al., 2013)].

Below the electrojet irregularities are intermittent irregularities in the mesosphere. These represent fluctuations in the index of refraction driven by mesospheric turbulence and exaggerated by the presence of free electrons in the *D* region. While they are not unique to the equatorial zone, they can be readily observed at Jicamarca due to the 50 MHz frequency and the high sensitivity of the facility (Woodman and Guillén, 1974). Neutral turbulence can be observed at Jicamarca in the mesosphere, stratosphere, and troposphere, and the discovery gave rise to the field of MST-radar techniques. Using these techniques, it is possible to measure neutral wind speeds, turbulence parameters, and turbulent fine structure (Fukao et al., 1974; Sheth et al., 2006; Guo et al., 2007; Lee et al., 2019).

Above the electrojet and throughout the *F* region, ionospheric interchange instability driven by free energy in the postsunset bottomside *F* region density gradient produces deep deformations in the bottomside that can penetrate into the topside, producing towering plumes of coherent backscatter (see **Figure 1**). The phenomenon underlies equatorial spread *F* (ESF) which is among the earliest space weather effects detected (Booker and Wells, 1938). The association with interchange instability enjoyed considerable speculation but was not established until Woodman and La Hoz (Woodman and La Hoz, 1976) produced definitive range-time-intensity (RTI) radar imagery of the process. (The authors reportedly needed to justify the recent purchase of an expensive new Versatec printer and invented RTI-style figures for this paper! The figures were very persuasive and remain the standard means of presentation).

ESF remains the cause célèbre of equatorial aeronomy as it disrupts modern radio communication, navigation, and imaging systems. Despite the fact that the underlying physics appears to be well understood, accurate forecasts remain elusive (Woodman, 2009). Jicamarca contributes to the research by measuring simultaneously both the causes (background ionospheric structure, vertical, and east-west plasma drifts via ISR) and the effects (irregularity morphology via coherent scatter and aperture synthesis imaging).

The interchange instabilities responsible for ESF are very similar to the $\mathbf{E} \times \mathbf{B}$ and current convective instabilities that create irregularities in the auroral *F* region, i.e., the irregularities monitored by the SuperDARN radar network. (The predecessor to SuperDARN, STARE, was an effort to infer high-latitude convection patterns from auroral electrojet echoes on the basis of earlier experiences from Jicamarca). Some of the other irregularities in **Figure 1**, meanwhile, are unique to the equatorial zone. Their discoveries were contrary to orthodoxy and deserve special attention.

2.2.1 150-km Echoes

Balsley (1964) identified another persistent source of radar clutter in the daytime valley region between about 140 and 170 km altitude in the early days of Jicamarca. (Royrvik and Miller, 1981; Royrvik, 1982) would associate the clutter with field-aligned plasma density irregularities, but it was not until a decade later that mysterious "150-km echoes" would receive serious scientific attention. (Mysterious because they exist in a homogeneous stratum of the ionosphere where gradient drift-type instability should not occur). (Kudeki and Fawcett, 1993) investigated the layers with a new, high-resolution mode and found them to be highly structured, exhibiting a stunning necklace shape that plunged with decreasing solar zenith angle. Most remarkably, the Doppler shifts of the echoes seemed to match the vertical plasma drifts. This suggested a means of measuring ionospheric dynamics using relatively low power radar systems going forward. The correspondence between the 150-km echo Doppler shifts at zenith with the vertical plasma drifts was established by (Woodman and Villanueva, 1995). Later, (Chau and Woodman, 2004) would show that both the vertical and zonal plasma drifts could be estimated accurately on the basis of line-of-sight 150-km echo Doppler shift measurements.

The source of the echoes remained a mystery for many more years. Two important clues helped expedite the research. One was the discovery of two distinct types of echoes—one spectrally narrow and field-aligned, and the other broad like a naturally enhanced ion line (Chau, 2004; Chau and Kudeki, 2013). The other was the observation that the layer height and intensity reacted to solar flares (Reyes, 2012) [see also (Pedatella et al., 2019)]. The echoes were clearly not due to some variant of gradient drift instability.

A pivotal finding came from (Oppenheim and Dimant, 2016) who identified energetic photoelectrons as the likely source of free energy behind the echoes. Their simulations reproduced narrow and broad spectral features and predicted both enhanced ion and electron lines. The authors tentatively associated their findings with upper hybrid instability (Basu et al., 1982; Jasperse et al., 1995). That association was made more explicitly by (Longley et al., 2020) who pointed out that the gaps in the echo necklace structure could be explained by cyclotron damping where the upper hybrid frequency matches an electron gyroharmonic frequency. As shown by (Lehmacher et al., 2020), this implies that not only plasma drifts but also plasma density profiles can be inferred from the 150-km echoes using low-power radar.

2.2.2 Bottom-Type Layers

Woodman and La Hoz (1976) identified several types of coherent scatter echoes associated with postsunset F-region instability. Among these were "bottom-type" or thin scattering layers that serve as precursors for intense plume events. The layers do not have signatures in ionograms, total electron content (TEC) measurements, radio scintillations, or airglow imagery and were neglected by the aeronomy and space-weather communities. Since "ESF" is a term taken from ionospheric sounding, if a phenomenon does not affect ionograms, it is not regarded as ESF. The bottom-type layers were just another unfortunate source of clutter.

For decades after their discovery, it was taken for granted that bottom-type layers signified marginal collisional interchange instability. There are several problems with this explanation, however. First, the layers exhibit little or no vertical development over time whereas interchange instabilities are convective instabilities and require vertical development. Second, the intensity of the layers does not vary directly with the strength of the background zonal electric field as is expected for the ionospheric interchange instability (Zargham and Seyler, 1987). Thirdly, the layers form at the base of the F region, near the valley, rather than in the steepest part of the bottomside where the growth rate for interchange instability is greatest.

Three clues shed light on the significance of the layers. The first was that fine structure in backscatter from the layers often exhibits horizontal striations (Hysell, 1992). This might be expected for gradient drift-type instability driven by zonal winds near a horizontal density gradient. The second was that the layers are not continuous but are patchy with horizontal scales of tens to hundreds of km (Hysell et al., 2005) (see **Figure 2**). The third was that they exist at altitudes and times where the plasma flow is westward–opposite the direction of the zonal neutral winds. Vertical shear flow predominates in the postsunset bottomside ionosphere, and associated with the shear flow is strong vertical current that is, usually neglected in stability analysis (Haerendel et al., 1992).

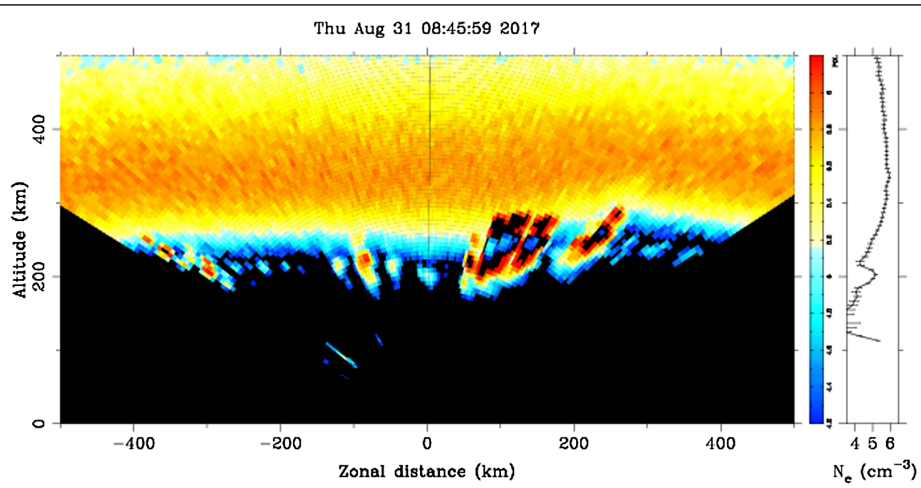

FIGURE 2 | ALTAIR radar images showing bottom-type layers at the base of the F region just prior to the onset of ESF. The data were acquired during NASA project WINDY (Hysell et al., 2020).

Hysell and Kudeki (2004) showed that the aforementioned vertical current destabilizes the ionosphere to irregularities propagating at angles intermediate between the vertical and the horizon. The instability has a larger growth rate than the ionospheric interchange instability but is confined to a narrow stratum. The instability causes the bottomside to be corrugated and unstable to wind-driven gradient drift instability in the ascending phases, explaining the bottom-type layers. Most importantly, the irregularities precondition the ionosphere to interchange instability which can grow more rapidly than it would otherwise. Numerical simulations show that this auxiliary instability is required for ESF depletions and radar plumes to develop and penetrate the topside as quickly after sunset as they do (Hysell et al., 2022).

2.2.3 High-Altitude Echoes

In 2008, a problem was discovered with some of Jicamarca's routine ISR experiments. Noise estimates were being calculated using samples from distant range gates above about 1,500 km. At night, the noise estimates would sometimes become anomalously large. It appeared that the distant range gates were being contaminated by signals of some kind. Further investigation showed that intense scattering layers were present at very high altitudes, most often between midnight and sunrise. At the time, a different method for estimating noise was introduced. The layers did not receive special attention and eventually disappeared.

However, the high-altitude echoes returned during the next solar minimum, and this time they were investigated further in a series of experiments beginning in 2018 and continuing to the present (Derghazarian et al., 2021). It was found that, during low solar flux conditions, the echoes are common in the pre-dawn sector between about 1,500 and 2,200 km altitude. The echoes exhibit Doppler shifts between about ±150 m/s and zonal drift rates of a few tens of m/s determined by multi-beam experiments. They are not obviously related to ESF.

Most importantly, the echoes exhibit sidebands upshifted and downshifted from the carrier by the lower hybrid frequency for protons. This is a remarkable result that is, not well understood. One candidate mechanism is lower hybrid drift instability (Krall and Liewer, 1971; Gladd, 1975), a streaming instability similar to modified two-stream instability, excited in this case by ion diamagnetic drift in the vicinity of existing plasma density irregularities. Another candidate is linear mode conversion and/or resonance instability in the vicinity of existing irregularities driven by lightning-induced whistlers (Lee and Kuo, 1984). The former mechanism is related to one invoked to explain small-scale irregularities in ESF (Huba and Ossakow, 1981). The latter mechanism is similar to one invoked to explain explosive spread *F* (Liao et al., 1989). In either case, pre-existing irregularities are required to bootstrap the instability. The source of these could be ESF although this is merely speculation at this point.

The high-altitude echoes represent frontier science in equatorial aeronomy with overtones for adjacent research domains like lightning research and even magnetic reconnection. For many years, however, they were just an overlooked source of radar clutter. In fact, the high-altitude echoes were recognized in the early days of Jicamarca experiments but were neglected because of more pressing problems making ISR experiments work.

3 CONCLUSION

This paper reviewed a number of discoveries in equatorial aeronomy and space physics going back to the 1960s. In each case, the discoveries came from projects and experiments that did not go according to plan. These were not merely serendipitous discoveries. Nor should they be considered negative outcomes of hypothesis tests. Rather, in most cases, they were the byproducts of the failures of assumptions that were not in doubt. Gyroresonances were supposed to be part of the incoherent backscatter spectrum. Coulomb collisions were supposed to be negligible. ESF was supposed to be driven entirely by ion interchange instability. The equatorial valley region and the inner plasmasphere were supposed to be stable. Pursuing the problems furthermore required deliberate departures from planned research. In some cases, this occurred only after long delays.

How can research be structured to make allowances for plans gone wrong and subsequent off-plan excursions? The premise seems to be at odds with contemporary trends which favor decadal planning for funding agencies, meticulous planning in research proposals, and long-range research plans for applicants for even junior positions. Perhaps, as Eisenhower famously said, plans are useless, but planning is indispensable. The important point is that it is essential to have the freedom to pivot when plans fail and to pursue the discoveries which may lie beneath.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://millstonehill.haystack.mit.edu/.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by the award NNX15AL02G from the National Aeronautics and Space Administration. The Jicamarca Radio Observatory is a facility of the Instituto Geofisíco del Perú operated with support from the NSF award AGS-1732209 through Cornell.

ACKNOWLEDGMENTS

The author is indebted to Prof. Juha Vierinen who acquired and provided the data shown in **Figure 1** and to Prof. Jorge Chau for his helpful suggestions.

REFERENCES

Balsley, B. B. (1964). Evidence of a Stratified Echoing Region at 150 Kilometers in the Vicinity of the Magnetic Equator during Daylight Hours. *J. Geophys. Res.* 69, 1925–1930. doi:10.1029/jz069i009p01925

Basu, B., Chang, T., and Jasperse, J. R. (1982). Electrostatic Plasma Instabilities in the Daytime Lower Ionosphere. *Geophys. Res. Lett.* 9, 68–71. doi:10.1029/gl009i001p00068

Booker, H. G., and Wells, H. W. (1938). Scattering of Radio Waves by theF-Region of the Ionosphere. *J. Geophys. Res.* 43, 249. doi:10.1029/te043i003p00249

Bowles, K. L. (1963). Measuring Plasma Density of the Magnetosphere. *Science* 139, 389–391. doi:10.1126/science.139.3553.389

Bowles, K. L., Ochs, G. R., and Green, J. L. (1962). On the Absolute Intensity of Incoherent Scatter Echoes from the Ionosphere. *J. Res. Natl. Bur. Stan. Sect. D. Rad. Prop.* 66D, 395. doi:10.6028/jres.066d.041

Chau, J. L., and Kudeki, E. (2013). Discovery of Two Distinct Types of Equatorial 150 Km Radar Echoes. *Geophys. Res. Lett.* 40, 4509–4514. doi:10.1002/grl.50893

Chau, J. L. (2004). Unexpected Spectral Characteristics of VHF Radar Signals from 150-km Region over Jicamarca. *Geophys. Res. Lett.* 31, L23803. doi:10.1029/2004GL021620

Chau, J. L., and Woodman, R. F. (2004). Daytime Vertical and Zonal Velocities from 150-km Echoes: Their Relevance toF-Region Dynamics. *Geophys. Res. Lett.* 31, a–n. doi:10.1029/2004GL020800

Clark, W. L., McClure, J. P., and Van Zandt, T. E. (1976). "Description and Catalog of Ionospheric F Region Data, Jicamarca Radio Observatory (November 1966 – April, 1969)," in *Tech. Rep. UAG-53, World Data Center A for Solar-Terrestrial Physics*.

Derghazarian, S., Hysell, D. L., Kuyeng, K., and Milla, M. A. (2021). High Altitude Echoes from the Equatorial Topside Ionosphere during Solar Minimum. *J. Geophys. Res. Space Phys.* 126. doi:10.1029/2020JA028424

Dougherty, J. P. (1964). Model Fokker-Planck Equation for a Plasma and its Solution. *Phys. Fluids* 7, 1788. doi:10.1063/1.2746779

Farley, D. T. (1963). A Plasma Instability Resulting in Field-Aligned Irregularities in the Ionosphere. *J. Geophys. Res.* 68, 6083–6097. doi:10.1029/jz068i022p06083

Farley, D. T. (1966). A Theory of Incoherent Scattering of Radio Waves by a Plasma: 4. The Effect of Unequal Ion and Electron Temperatures. *J. Geophys. Res.* 71, 4091–4098. doi:10.1029/jz071i017p04091

Farley, D. T. (1967). Proton Gyroresonance Observed in Incoherent Scattering from the Ionosphere. *Phys. Fluids* 10, 1584–1586. doi:10.1063/1.1762326

Farley, D. T. (1964). The Effect of Coulomb Collisions on Incoherent Scattering of Radio Waves by a Plasma. *J. Geophys. Res.* 69, 1972402–1973200. doi:10.1029/jz069i001p00197

Fejer, B. G., and Kelley, M. C. (1980). Ionospheric Irregularities. *Rev. Geophys.* 18, 401. doi:10.1029/rg018i002p00401

Fukao, S., Sato, T., Kato, S., Harper, R. M., Woodman, R. F., and Gordon, W. E. (1974). Mesospheric Winds and Waves over Jicamarca on May 23–24, 1974. *J. Geophys. Res.* 84, 4379–4386.

Gladd, N. T. (1975). The Lower Hybrid Drift Instability and the Modified Two Stream Instability in High Density Theta Pinch Environments. *Plasma Phys.* 18, 27–40.

Gordon, W. (1958). Incoherent Scattering of Radio Waves by Free Electrons with Applications to Space Exploration by Radar. *Proc. IRE* 46, 1824–1829. doi:10.1109/jrproc.1958.286852

Guo, L., Lehmacher, G., Kudeki, E., Akgiray, A., Sheth, R., and Chau, J. L. (2007). Turbulent Kinetic Energy Dissipation Rate and Eddy Diffusivity Study in the Tropical Mesosphere Using Jicamarca Radar Data. *Adv. Space Res.* 40. doi:10.1016/j.asr.2007.05.068

Haerendel, G., Eccles, J. V., and Çakir, S. (1992). Theory for Modeling the Equatorial Evening Ionosphere and the Origin of the Shear in the Horizontal Plasma Flow. *J. Geophys. Res.* 97, 1209. doi:10.1029/91ja02226

Huba, J. D., and Ossakow, S. L. (1981). On 11-cm Irregularities during Equatorial spreadF. *J. Geophys. Res.* 86, 829. doi:10.1029/ja086ia02p00829

Hysell, D. L., and Chau, J. L. (2002). Imaging Radar Observations and Nonlocal Theory of Large-Scale Plasma Waves in the Equatorial Electrojet. *Ann. Geophys.* 20, 1167–1179. doi:10.5194/angeo-20-1167-2002

Hysell, D. L., and Kudeki, E. (2004). Collisional Shear Instability in the equatorialFregion Ionosphere. *J. Geophys. Res.* 109. doi:10.1029/2004JA010636

Hysell, D. L., Larsen, M. F., Swenson, C. M., Barjatya, A., Wheeler, T. F., Sarango, M. F., et al. (2005). Onset Conditions for Equatorial spreadFdetermined during EQUIS II. *Geophys. Res. Lett.* 32, L24104. doi:10.1029/2005GL024743

Hysell, D. L., Fang, T. W., and Fuller-Rowell, T. J. (2022). Modeling Equatorial *F*-Region Ionospheric Instability Using a Regional Irregularity Model and WAM-IPE. *J. Geophys. Res.*

Hysell, D. L. (1992). *On the Hierarchy of Processes Contributing to Equatorial Spread F. Ph.D. Thesis*. Ithaca, N. Y: Cornell Univ.

Hysell, D. L., Rao, S., Groves, K. M., and Larsen, M. F. (2020). Radar Investigation of Postsunset Equatorial Ionospheric Instability over Kwajalein during Project WINDY. *J. Geophys. Res. Space Phys.* 125. doi:10.1029/2020ja027997

Jasperse, J. R., Basu, B., Retterer, J. M., Decker, D. T., and Chang, T. (1995). "High Frequency Electrostatic Plasma Instabilities and Turbulence Layers in the Lower Ionosphere," in *Space Plasmas: Coupling between Small and Medium Scale Processes (Geophysical Monograph 86* (American Geophysical Union), 77–94. doi:10.1029/gm086p0077

Krall, N. A., and Liewer, P. C. (1971). Low-frequency Instabilities in Magnetic Pulses. *Phys. Rev. A.* 4, 2094–2103. doi:10.1103/PhysRevA.4.2094

Kudeki, E., Farley, D. T., and Fejer, B. G. (1982). Long Wavelength Irregularities in the Equatorial Electrojet. *Geophys. Res. Lett.* 9, 684–687. doi:10.1029/gl009i006p00684

Kudeki, E., and Fawcett, C. D. (1993). High Resolution Observations of 150 Km Echoes at Jicamarca. *Geophys. Res. Lett.* 20, 1987–1990. doi:10.1029/93gl01256

Kudeki, E., and Milla, M. A. (2011). Incoherent Scatter Spectral Theories-Part I: A General Framework and Results for Small Magnetic Aspect Angles. *IEEE Trans. Geosci. Remote Sensing* 49 (1), 315–328. doi:10.1109/tgrs.2010.2057252

Lee, K., Kudeki, E., Reyes, P. M., Lehmacher, G. A., and Milla, M. (2019). Mesospheric Wind Estimation with the Jicamarca MST Radar Using Spectral Mainlobe Identification. *Radio Sci.* 54, 1222–1239. doi:10.1029/2019RS006892

Lee, M. C., and Kuo, S. P. (1984). Production of Lower Hybrid Waves and Field-Aligned Plasma Density Striations by Whistlers. *J. Geophys. Res.* 89 (A12), 10873–10880. 10. doi:10.1029/JA089iA12p10873

Lehmacher, G. A., Wu, H., Kudeki, E., Reyes, P. M., Hysell, D. L., and Milla, M. (2020). Height Variation of Gaps in 150-km Echoes and Whole Atmosphere Community Climate Model Electron Densities Suggest Link to Upper Hybrid Resonance. *J. Geophys. Res. Space Phys.* 125. doi:10.1029/2019JA027204

Liao, C. P., Freidberg, J. P., and Lee, M. C. (1989). Explosive Spread F Caused by Lightning-Induced Electromagnetic Effects. *J. Atmos. Terrestrial Phys.* 51, 751–758. doi:10.1016/0021-9169(89)90032-9

Liu, J., Wang, W., Oppenheim, M., Dimant, Y., Wiltberger, M., and Merkin, S. (2016). Anomalous Electron Heating Effects on the E Region Ionosphere in TIEGCM. *Geophys. Res. Lett.* 43, 2351–2358. doi:10.1002/2016GL068010

Longley, W. J., Oppenheim, M. M., and Dimant, Y. S. (2019). Nonlinear Effects of Electron-Electron Collisions on ISR Temperature Measurements. *J. Geophys. Res. Space Phys.* 124, 6313–6329. doi:10.1029/2019JA026753

Longley, W. J., Oppenheim, M. M., Fletcher, A. C., and Dimant, Y. S. (2018). ISR Spectra Simulations with Electron-Ion Coulomb Collisions. *J. Geophys. Res. Space Phys.* 123, 2990–3004. doi:10.1002/2017JA025015

Longley, W. J., Oppenheim, M. M., Pedatella, N. M., and Dimant, Y. S. (2020). The Photoelectron-Driven Upper Hybrid Instability as the Cause of 150-km Echoes. *Geophys. Res. Lett.* 47. doi:10.1029/2020GL087391

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *J. Atmos. Sci.* 20, 130–141. doi:10.1175/1520-0469(1963)020<0130:dnf>2.0.co;2

Madsen, C. A., Dimant, Y. S., Oppenheim, M. M., and Fontenla, J. M. (2013). The Multi-Species Farley-Buneman Instability in the Solar Chromosphere. *ApJ* 783. doi:10.1088/0004–637X/783/2/128

Milla, M. A., and Kudeki, E. (2011). Incoherent Scatter Spectral Theories-Part II: Modeling the Spectrum for Modes Propagating Perpendicular to **B**. *IEEE Trans. Geosci. Remote Sensing* 49 (1), 329–345. doi:10.1109/tgrs.2010.2057253

Oppenheim, M. M., and Dimant, Y. S. (2016). Photoelectron-induced Waves: A Likely Source of 150 Km Radar Echoes and Enhanced Electron Modes. *Geophys. Res. Lett.* 43, 3637–3644. doi:10.1002/2016GL068179

Pedatella, N. M., Chau, J. L., Vierinen, J., Qian, L., Reyes, P., Kudeki, E., et al. (2019). Solar Flare Effects on 150-km Echoes Observed over Jicamarca: WACCM-X Simulations. *Geophys. Res. Lett.* 46, 10951–10958. 10. doi:10.1029/2019GL084790

Pingree, J. E. (1990). *Incoherent Scatter Measurements and Inferred Energy Fluxes in the Equarotial F-Region Ionosphere*. Ithaca, N. Y: Ph.D. thesis, Cornell Univ.

Reyes, P. M. (2012). *Solar Flare Effects Observed over Jicamarca during MST-ISR Experiments*. Master's thesis, University of Illinois at Urbana-Champaign.

Rodrigues, F. S., Nicolls, M. J., Woodman, R., Hysell, D. L., Chau, J. L., and González, S. A. (2007). Ion Gyroresonance Observations at Jicamarca Revisited. *Geophys. Res. Lett.* 34, a–n. doi:10.1029/2007GL029680

Royrvik, O. (1982). Drift and Aspect Sensitivity of Scattering Irregularities in the Upper Equatorial E-Region. *J. Geophys. Res.* 87, 8338.

Royrvik, O., and Miller, K. L. (1981). Nonthermal Scattering of Radio Waves Near 150 Km above Jicamarca, Peru. *J. Geophys. Res.* 86, 180.

Sheth, R., Kudeki, E., Lehmacher, G., Sarango, M., Woodman, R., Chau, J., et al. (2006). A High-Resolution Study of Mesospheric fine Structure with the Jicamarca MST Radar. *Ann. Geophys.* 24, 1281–1293. doi:10.5194/angeo-24-1281-2006

Simon, A. (1963). Instability of a Partially Ionized Plasma in Crossed Electric and Magnetic fields. *Phys. Fluids* 6, 382. doi:10.1063/1.1706743

Sulzer, M. P., and Gonzalez, S. (1999). The Effect of Electron Coulomb Collisions on the Incoherent Scatter Spectrum in the *F* Region at Jicamarca. *J. Geophys. Res.* 104, 552. 22,535–22. doi:10.1029/1999ja900288

Woodman, R. F., Farley, D. T., Balsley, B. B., and Milla, M. A. (2019). The Early History of the Jicamarca Radio Observatory and the Incoherent Scatter Technique. *Hist. Geo Space Sci.* doi:10.5194/hgss-10-245-2019

Woodman, R. F., and Guillén, A. (1974). Radar Observations of Winds and Turbulence in the Stratosphere and Mesosphere. *J. Atmos. Sci.* 31, 493–505. doi:10.1175/1520-0469(1974)031<0493:roowat>2.0.co;2

Woodman, R. F., and La Hoz, C. (1976). Radar Observations ofFregion Equatorial Irregularities. *J. Geophys. Res.* 81, 5447–5466. doi:10.1029/ja081i031p05447

Woodman, R. F. (2004). On a Proper Electron Collision Frequency for a Fokker-Planck Collision Model with Jicamarca Applications. *J. Atmos. Solar-Terrestrial Phys.* 6617, 1521–1541. doi:10.1016/j.jastp.2004.07.001

Woodman, R. F. (2009). Spread F - an Old Equatorial Aeronomy Problem Finally Resolved? *Ann. Geophys.* 27, 1915–1934. doi:10.5194/angeo-27-1915-2009

Woodman, R. F., and Villanueva, F. (1995). "Comparison of Electric fields Measured at *F* Region Heights with 150 Km Irregularity Drift Measurements," in Proceedings of the Ninth International Symposium on Equatorial Aeronomy, Indonesia, March 20-24 (Bali). [Dataset].

Zargham, S., and Seyler, C. E. (1987). Collisional Interchange Instability, 1, Numerical Simulations of Intermediate-Scale Irregularities. *J. Geophys. Res.* 92. 10,073. doi:10.1029/ja092ia09p10073

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MM declared a past co-authorship with the author DH to the handling editor.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 19 April 2022
doi: 10.3389/fspas.2022.879290

Check for updates

When I Encountered Difficult Problems

Syun-Ichi Akasofu *

International Arctic Research Center, University of Alaska, Fairbanks, AK, United States

All of us encounter our unique difficulties as scientists. Although the kind of difficulty varies by scientist, I describe my own in this program, and explain how I responded, together with a sort of moral. They are: (a) When theorists seemed to encounter a deadlock, (b) When a well-accepted theory is confronted with a seriously contradictory observational fact, (c) When there is a generally believed morphology of a phenomenon that does not agree with my observations, (d) When a single approach or theory prevails, but I have a different idea. Obviously, my responses are just an attempt, which will be judged by future developments in space physics.

Keywords: geomagnetic storm, solar wind, sunspot, aurora, solar flare

OPEN ACCESS

Edited by:
Noora Partamies, The University Centre in Svalbard, Norway

Reviewed by:
Maxime Grandin, University of Helsinki, Finland

Correspondence:
Syun-Ichi Akasofu
sakasofu@alaska.edu

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *19 February 2022*
Accepted: *15 March 2022*
Published: *19 April 2022*

Citation:
Akasofu S-I (2022) When I Encountered Difficult Problems. Front. Astron. Space Sci. 9:879290. doi: 10.3389/fspas.2022.879290

INTRODUCTION

Space physics began to develop in the 1960s by combining parts of geomagnetism (a study of geomagnetic storms), ionospheric physics, auroral physics, planetary physics and solar physics under the advent of satellite observations. My first paper in the Journal of Geophysical Research, co-authored with Sydney Chapman, was published in 1961. I have witnessed the development of space physics from its early days.

I believe that every young scientist knows how research should be conducted in general terms. Thus, for the purpose of this particular program, I try to describe plainly my humble experiences, when I encountered difficult problems, which had been and have been lasted for a long time in space physics; how long each difficulty had or have lasted will be mentioned in each topic.

Some of the difficulties include: (a) when theorists seemed to encounter a deadlock for a long time; (b) when there is a well-accepted theory, but there seems to exist a seriously contradicting observed fact; (c) when there was a generally accepted morphology of a phenomenon, which does not seem to agree with my observations and (d) when a single theoretical approach or theory has prevailed for a long time, but I have a different idea.

Most readers must have encountered some of these or similar problems. In fact, some of the above problems have been what all space physicists share even at the present time. It is my hope that the readers will get some hints from how I responded, whether it proved to be right or wrong or will be judged by future progress in space physics.

When Theorists Seemed to Encounter a Deadlock

The "Unknown Parameter" in the Solar Wind

Sydney Chapman theorized the interaction between solar plasma flow and the Earth's dipole magnetic field in 1931; it became the basis of magnetospheric physics and space physics, including planetary physics. Chapman and others established also the concept of the standard geomagnetic storm; it begins with the storm sudden commencement (SSC) as storm onset, which is followed by the initial and main phases; a step function-like increase of SSC indicates the arrival of an intensified solar wind; here, the intensity of the plasma flow is proportional to the kinetic pressure of the solar

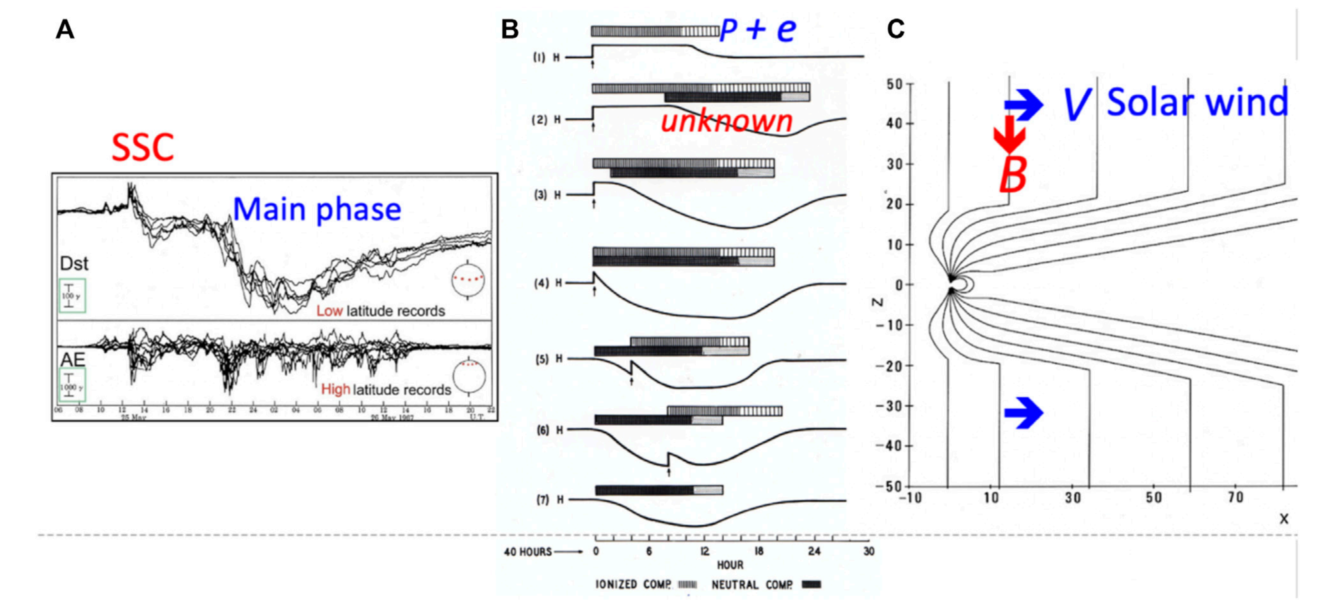

FIGURE 1 | (A) Magnetic records of a number of stations from both low latitudes (upper) and low latitudes (lower) of a geomagnetic storm (only the horizontal component); the circular insert show approximate location of the observatories. They are typical geomagnetic records, and it can be considered as a standard type of geomagnetic storm. **(B)** It illustrates a great variety of geomagnetic storms; (p + e) indicates the case, in which the solar wind consists of only protons and electrons; "unknown" indicates that it was thought that the solar wind had some unknown parameter causing the main phase. **(C)** The linked magnetic field between IMF and magnetospheric field; this interaction constitutes the solar wind-magnetosphere dynamo, proving the power fo the main phase and the aurora.

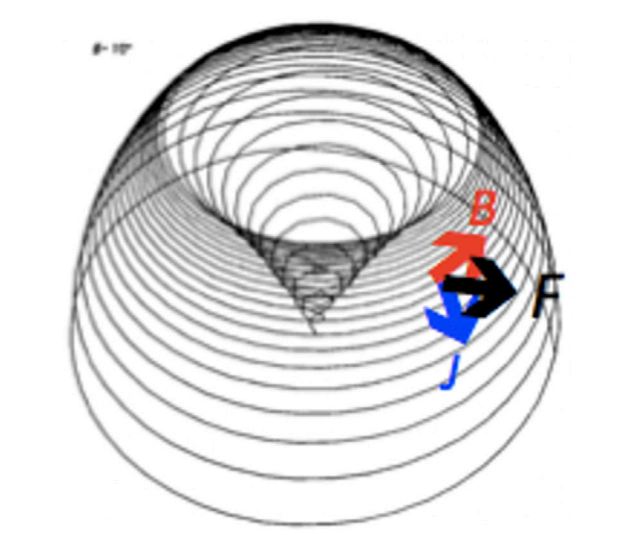

FIGURE 2 | One of the magnetic field lines generated by the solar unipolar induction current system. The current ***J,*** magnetic field direction ***B*** and the direction of the outward force ***F*** by the **(*J*** x ***B*)** force are indicated.

wind, which is also proportional to the magnitude of the SSC, being typically 25 nT (Chapman's 1931 theory); **Figure 1A**.

Since then, Chapman and many researchers had tried to find the way for the solar plasma particles to enter into the magnetosphere in order to explain the main phase of geomagnetic storms (the ring current around the Earth and the aurora). When I joined Chapman in 1959 as his graduate student, he asked me to pursue this problem further.

Since I did not know much about geomagnetic storms at that time, I decided to examine, first of all, a large number of geomagnetic storm records (magnetograms) from many stations. It was a great surprise to me that many geomagnetic storms began with a large SSC, indicating the arrival of an intensified (high kinetic pressure) of the solar wind, but were not followed by the main phase; further, many storms developed an intense main phase without SSC; there are many between the two, so that the concept of the standard geomagnetic storm is not appropriate.

Thus, I concluded that there must be an "unknown parameter" in the solar wind, which determines the development of the main phase, other than an intensified solar wind; **Figure 1B**. I recall that in one conference, I was told by a person that I was not qualified as a graduate student in space physics by not knowing that the solar wind consists of only protons and electrons. However, Chapman agreed with me about my finding, and published a joint paper (Akasofu and Chapman, 1963).

Chapman presented our result during the First International Conference on the Solar Wind, which was held at the Jet Propulsion Laboratory (JPL) in 1964. After Chapman's talk, some people expressed their doubt about our finding, but Jim Dungey suggested that our "unknown parameter" may be the southward-oriented interplanetary magnetic field (IMF) component. His suggestion was confirmed by a satellite observation a few years later.

The importance of the interplanetary magnetic field has now been well recognized in terms of the interaction between the solar wind and the magnetosphere. The solar wind blows across the linked field lines between the IMF and the magnetospheric field, constituting the solar

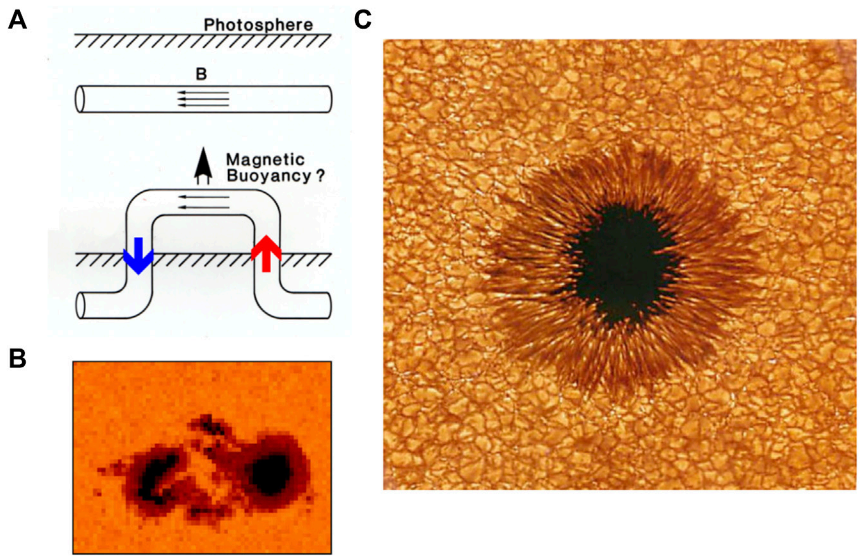

FIGURE 3 | (A) A schematic illustration of Babcock's theory of a pair of sunspots. **(B)** A typical example of a pair of sunspots (Courtesy of the Kitt Peak Solar Observatory). **(C)** A typical example of single spots (Courtesy of The Kitt Peak Solar Observatory and the NASA Sunspot Collection).

wind-magnetosphere dynamo, generating the power for geomagnetic storms and the aurora; **Figure 1C**.

A kind of moral here is that even a new graduate student can find a clue for solving a long-standing problem.

Causes of the Solar Wind

Gene Parker published the first paper on the solar wind in 1958 (he coined the term "solar wind"). However, in spite of the fact that so much has been observed and studied on details of the solar wind, its cause is still very controversial at best even after more than 60 years (cf. Viall and Borovsky, 2020).

In such a situation, one way of approaching the problem is not to pursue the problem under the basic assumptions everyone shares. In this case, many theoretical researchers' work have been based on the high temperature of the corona in order to overcome the powerful solar gravitational force; in fact, I used to say: "the corona is so hot that it blows away by itself".

The solar wind is a heliospheric phenomenon, so that a very large-scale force throughout the heliosphere is needed to propel the whole heliospheric plasma outward. Further, the needed force must be a very powerful (***J*** x ***B***) force, which can propel the whole heliospheric plasma with a speed of up to 800 km/s. The question was then what kind of current system might be able to produce such a (***J*** x ***B***) force, rather than low level processes in the corona. In 1950, Alfven (1981) worked on the solar unipolar induction current system, which can be generated by a rotating dipolar magnetic body like the sun surrounded by plasma. It has the needed configuration and the requirements. An estimate indicated that the (***J*** x ***B)*** force can propel the heliospheric plasma with a speed of 200 km/s (Lee and Akasofu, 2021). We are currently trying to search for processes to provide additional 500 km/s (**Figure 2**).

After Parker introduced his theory of the solar wind in 1958, we have not yet solved the problem. It is taking more than 60 years. It may be that a radically different approach is needed in such a situation, for example introducing the (***J*** x ***B***) force and others.

When a Well-Accepted theory was Confronted With a Seriously Contradictory Observational Fact

In space physics, there is hardly a perfect theory. There is always an observed fact, which disagrees or contradicts with well-accepted theory. Unfortunately, many researchers disregard such a fact by considering that it is just an exception or anomaly.

In solar physics, there is a well-accepted theory on sunspots (Babcock, 1961), in which a thin magnetic tube emerges from below the photospheric surface by magnetic buoyancy, and the two emerging points can be identified as a pair (positive and negative, N or S) of spots, **Figures 3A,B**. So far, such a tube of magnetic flux has not been detected yet.

On the other hand, there exist single spots. When I took up single sunspots, I was told by two prominent solar physicists and many others that a single spot may be a "broken pipe", so that it is wasting time to even consider them. A typical example of single spots is shown in **Figure 3C**.

It is puzzling why single spots have hardly been studied in the past, in spite of the fact that most standard textbooks and monographs on solar physics carry images of single spots. This may be because it has been believed from the earliest days that spots are like a magnet (which has both the N and S poles together) and because Babcock's theory can so intuitively be understood. Another reason may be that 'magnetic monopoles' are not supposed to exist, so that single spots are avoided to be considered. Thus, single sunspots have almost been disregarded in the past as a "broken pipe" at best.

However, since I was convinced that single spots do exist, I tried to examine them. After I began to study magnetic fields on

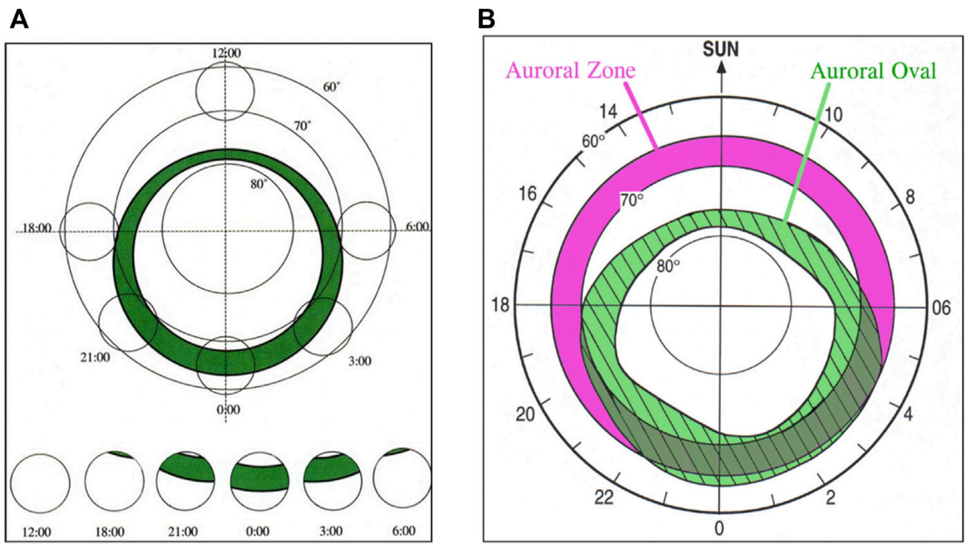

FIGURE 4 | (A) Auroral belt imagined by the author. The aurora appeared almost always near the northern and reached overhead and receded in the northern sky in morning hours. The circles indicate the field of view in Fairbanks, Alaska. **(B)** Feldstein's auroral oval (green) and the auroral zone (pink) in MLT coordinate system.

the photosphere, I found other contradictions. For example, there are weak, but large-scale fields, both positive and negative fields, which line up alternatively in longitude. They are called unipolar regions. They have generally been considered as old active regions stretched out by the non-uniform rotation of the sun. However, by examining unipolar magnetic fields, I found they grow and decay with the sunspot cycle.

Furthermore, the most important finding is that positive single spots are born in a positive unipolar region (vice versa), and a pair of spots is born *only* at the boundary (positive and negative) of neighboring unipolar regions, not in the middle of unipolar regions; these findings encouraged me to pursue further the problem. However, one of the reviewers on an early version of my paper rejected it by mentioning "a noble idea, but heretic".

After synthesizing a number of observed facts on sunspots, I suggested how positive single spots are born in positive unipolar region and how a pair of spots are formed at the boundary of neighboring (positive and negative) unipolar regions. On the basis of these findings, I suggested that a single spot is the basic unit of sunspots, rather than a pair of spots.

The point is that under my idea, I can assemble many other features of sunspots compared with the well-accepted theory, regardless of whether my attempt is right or wrong.

This is a very important point in proposing a new idea against a well-accepted theory. For example, I can now conclude why single spots can exist [positive single spots are born in a positive unipolar region (vice versa)]; unipolar regions grow and decay with the sunspot cycle (not decaying old active regions) as generally considered (Akasofu, 2020a).

Solar physicists who are specifically working on sunspots would perhaps agree with me that the formation of sunspots is not yet solved. Babcock's theory was published in 1961 and is well accepted in general, but it may be said that we do not really know what sunspots are.

When There is a Generally Believed Morphology of a Phenomenon Which Does Not agree With my Observations

Auroral Oval

The concept of the auroral zone was established by E. Loomis as early as in 1860. Since then, the auroral zone had long been believed to be the belt of the aurora. When I began to observe the aurora, I noticed that the aurora appeared always in the northern sky in the evening and shifted southward as the evening progressed; after midnight hours, the aurora shifted back to the northern sky. Since Fairbanks is located in the auroral zone, I thought that the aurora should appear always overhead any time; **Figure 4A**.

I asked my colleagues why my observation did not agree with Loomis' auroral zone. They said that it was believed that the aurora was formed at the center line of the auroral zone, 67° in geomagnetic latitude and then shifted southward after it was formed. Thus, I examined all-sky images at Fort Yukon (67°, the center line of the auroral zone), and found that the aurora also appeared first in the northern sky. Then, I found also the same even at Barrow (70°, the northern edge of the auroral zone). At that time, I was busy in preparing my Ph.D. thesis on geomagnetic storms, and thus had forgotten about my observation.

Then, I found a paper by Yasha Feldstein (1963), in which he showed the belt of the aurora is actually the auroral oval, which is deformed poleward side in daylight hours; its midday location higher than 75° in latitude. His oval could explain my early auroral observation.

Thus, I thought that one way to prove the auroral oval and Feldstein's finding was to set up a chain of all-sky cameras along the magnetic meridian line between Alaska and the northwestern tip of Greenland (located near the geomagnetic pole), which could scan the whole polar sky as the earth rotates once a day like an azimuth scan radar at an airport. This became my first grant from the National Science Foundation. This observation proved the accuracy of the auroral oval, but the result could not convince many. I found that

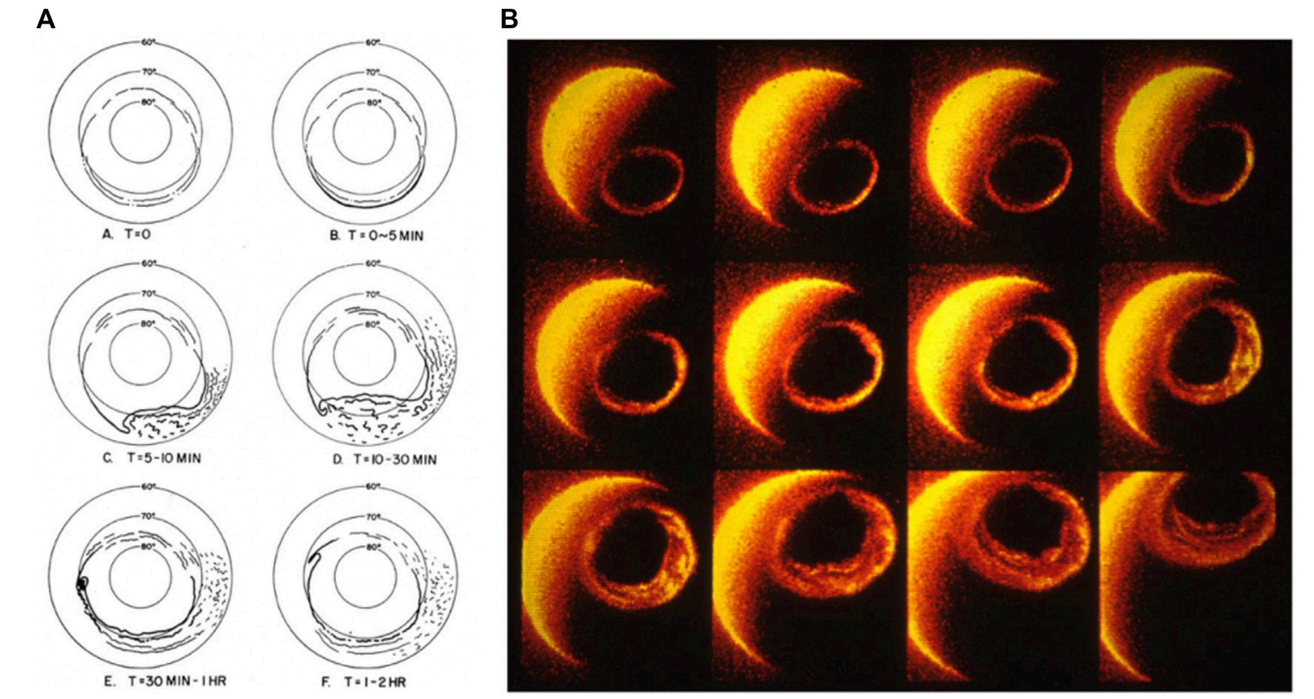

FIGURE 5 | (A) The schematic illustration of the development of auroral substorms (Akasofu, 1964). The first indication of substorm onset is a sudden brightening of an auroal arc (T = 0), followed by a rapid poleward advance; this movement causes westward traveling surge (WTS), which propagates toward the evening sky; In the morning sky, the auroral arcs disintegrate into "patches" and drift eastward. It takes about 2 h **(B)** A series of images of an auroral substorm obtained by the Explorer satellite in 1982 (Courtesy of L. Frank).

the oval nearly coincides with the outer boundary of the outer radiation belt on the ionosphere, which was just worked out by James Van Allen of the University of Iowa; this was more convincing than the meridian scanning observation. However, the final proof had to wait until the first image of the oval by the Canadian satellite ISIS II became available in 1972; the controversy faded away after the first oval image taken from the Canadian ISIS II satellite in 1972 became available. Later, I asked the NASA plane Galileo to fly under the midday oval over the Norwegian Sea (the best location to see the midday oval in the northern hemisphere); everyone on board was excited to see the red auroral arcs.

The auroral oval is very crucial in studying polar and magnetospheric phenomena, because it provides the natural frame of reference; for example, the auroral oval surround the "roots", of magnetic field lines which are linked with IMF field lines. Many auroral phenomena are different, depending on whether they are within, in or outside the oval.

After all, the aurora and auroral oval are fixed in the MLT-MLAT frame, but we observers rotate under them. Since the controversy on the oval was so intense, I thought that it was somewhat like the ancient situation, whether the Sun or the earth rotated.

A New Morphology of Auroral Activities: Auroral Substorms

Another example of well-established observation in the 1930s is that it had been very firmly believed that the aurora was quiet in evening hours, active in midnight hours and patchy in morning hours and that the Earth and its observers witnessed such a pattern once a night as the Earth rotated. When I found such a pattern occurred twice or three times in one very active night, I wondered instantaneously whether the Earth rotated three times in that night, since I was told that the believed pattern had so firmly been established for long time.

Thus, I decided to disregard what had been believed and analyzed simultaneous auroral activities in Siberia (evening sky), Alaska (midnight sky) and Canada (morning sky). It was my finding that such a pattern over the whole polar sky repeats several times in 24 h during active periods (once or twice a night) under which the earth rotates. The concept of auroral substorms was established on the basis of such a study based on a number of all-sky cameras located in the Arctic region (Akasofu, 1964); **Figure 5**. However, it took some time to convince many colleagues on my results, since they had firmly believed in the old concept. One way to observe two auroral substorms in 6 h was to fly from the East Coast to Alaska in midnight hours against the Earth's rotation. This flight was conducted several times by the NASA Galileo and the Air Force's flying ionospheric laboratory (KC135) to record two substorms. In 1982, the Explorer satellite imaged an auroral substorm high above the north polar region (**Figure 5B**); its images resembled the schematic illustrations made by the author, becoming another convincing fact.

It was fortunate that many magnetospheric phenomena observed by satellites occur also simultaneously with auroral substorms, so that the field of magnetospheric substorms had also been established soon afterward by the whole community.

On the other hand, it should be mentioned that some of my colleagues pointed out that the concept of auroral substorms has unfortunately become a “well-established concept”. It is my sincere hope that a better concept will emerge soon in order to make a new progress in auroral physics, since many fundamental issues of the aurora have not been explored yet. Even the cause of a thin curtain-like structures of the aurora, one of the basic features of the aurora, is not well understood.

When a Single Approach or theory Prevails, but I Have a Different Idea

When I met Hannes Alfven in 1966 for the first time at a conference held in Norway, he almost insisted to me that in studying space physics, I should concentrate on the electric currents, instead of magnetic field lines. This method is called the electric current approach by him.

Since he was the originator of the MHD (magnetohydrodynamic) theory (Alfven, 1950), this was a great surprise to me; I had just begun to learn MHD at that time; in the theory, a magnetic field line can be identified by plasma particles attached to it, namely the “frozen-in field” concept. This approach is called the magnetic field line approach.

By then (1966), Alfven must have already realized the limitation of the concept of “frozen-in field” in rarefied plasma in solar and magnetospheric phenomena. On the other hand, it seems that the magnetic field line approach has been the main or only one approach in space physics for a long time.

I recall that Alfven was unhappy by the fact that people did not listen to his advice, although he repeated his warning whenever possible (cf. Alfven, 1981; Alfven, 1986; even in his Nobel Prize acceptance speech). Thus, by taking this opportunity, I would like to remind the reader of his warning.

Both solar flares and auroral substorms are electromagnetic phenomena in rarefied plasma. By taking the electric current approach, it is necessary to consider, first of all, a dynamo as the power supply, and secondly transmission of the power to the location where the power is finally dissipated (observed phenomena are mostly dissipation results) in understanding the whole processes of a phenomenon. I believe that this is the standard approach to study electromagnetic phenomena in nature. Even if one wants to devote on one part of the series of this electromagnetic process, it is important to keep in mind the whole current system (Akasofu, 2020b).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

Looking back my research carrier, first of all, I owe so much to both Sydney Chapman and Hannes Alfven in building up my independent mind. I owe greatly so many of my colleagues for their discussions, regardless they agreed or disagreed. I am very happy that I am still corresponding with Tony Lui at Applied Physics Lab of the Johns Hopkins University, Bruce Tsurutani at the JPL (Jet Propulsion’ Laboratory) and Lou Lee at Academia Sinica in Taiwan after three or four decades.

REFERENCES

Akasofu, S.-I. (1964). The Development of the Auroral Substorm. *Planet. Space Sci.* 12, 273–282. doi:10.1016/0032-0633(64)90151-5

Akasofu, S. I. (2020a). A Morphological Study of Unipolar Magnetic fields and the Relationship with Sunspots. *J. Atmos. Solar-Terr. Phys.* 218 (2021), 105625. doi:10.1016/j.jastp.2021.105625

Akasofu, S. I. (2020b). Electric Current Approach Studying Both Auroral Substorms and Solar Flares Together. *Frontiers Astronom Space Sci.* 7, 2–20. doi:10.3389./fspas.202000004

Akasofu, S. I., and Chapman, S. (1963). The Development of the Main Phase of Magnetic Storms. *J. Geophys. Res.* 68, 125–129. doi:10.1029/jz068i001p00125

Alfvén, H. (1981). “Cosmic Plasma,” in *Astrophysics and Space Science Library* (Berlin/Heidelberg, Germany: Springer-Verlag). 90-277-1151-8.

Alfven, H. (1986). Double Layers and Circuits in Astrophysics. *IEEE Trans. Plasma Sci.* 14 (6), 779–793. doi:10.1109/tps.1986.4316626

Babcock, H. W. (1961). The Topology of the Sun’s Magnetic Field and the 22-YEAR Cycle. *ApJ* 133, 572–587. doi:10.1086/147060

Feldstein, Y. (1963). Some Problems Concerning the Morphology of Auroras and Magnetic Disturbances at High Latitudes. *Geomagnetism and Aeronomy* 3, 183–192.

Lee, L. C., and Akasofu, S. I. (2021). On the Causes of the Solar Wind: Part 1. Unipolar Solar Induction Currents. *J. Geophys. Space Phys.* 126, 1–8. doi:10.1029/2021ja029358

Viall, N. M., and Borovsky, J. E. (2020). Nine Outstanding Questions of Solar Wind Physics. *J. Geophys. Res. Space Phys.* 125, e2018JA026005. doi:10.1029/2018JA026005

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 20 April 2022
doi: 10.3389/fspas.2022.890864

Check for updates

A Perspective on Solar Energetic Particles

*Donald V. Reames**

Institute for Physical Science and Technology, University of Maryland, College Park, MD, United States

The author has been fortunate to observe and participate in the rise of the field of solar energetic particles (SEPs), from the early abundance studies, to the contemporary paradigm of shock acceleration in large SEP events, and element abundance enhancements that are power laws in mass-to-charge ratios from H to Pb. Through painful evolution the "birdcage" model and the "solar-flare myth" came and went, leaving us with shock waves and solar jets that can interact as sources of SEPs.

Keywords: solar energetic particles, solar jets, shock waves, solar system abundances, coronal mass ejection (CME)

OPEN ACCESS

Edited by:
Georgia Adair De Nolfo, National Aeronautics and Space Administration, United States

Reviewed by:
Emilia Kilpua, University of Helsinki, Finland

****Correspondence:***
Donald V. Reames
dvreames@gmail.com
orcid.org/0000-0001-9048-822X

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *06 March 2022*
Accepted: *31 March 2022*
Published: *20 April 2022*

Citation:
Reames DV (2022) A Perspective on Solar Energetic Particles. Front. Astron. Space Sci. 9:890864. doi: 10.3389/fspas.2022.890864

INTRODUCTION

Often the evolution of science and of scientists seems a diffusive process, a random walk through topics and talented colleagues. It is common to think of planning a course of study, a proposal, or even an entire career in advance, but perhaps it works just as well with some randomness.

NUCLEAR EMULSION

I became an undergraduate, then a graduate student, at the University of California at Berkeley. For my PhD thesis I studied nuclear interactions of heavy ions, especially O at 10 MeV amu^{-1}, using nuclear emulsion detectors under the guidance of emulsion expert Walter H. Barkas. At that time emulsions were also flown on balloons to study the composition of galactic cosmic rays (GCRs), and the first measurement of element abundances in solar energetic particles (SEPs) was made with emulsions on sounding rockets from Ft. Churchill, Manitoba by Fichtel and Guss (1961). I was impressed by these applications of emulsion to the budding Space Sciences and readily expressed my interest when Carl Fichtel contacted Barkas for possible new PhDs to hire. I began working at Goddard Space Flight Center in 1964, studying GCR abundances on the manned Gemini mission (Durgaprasad et al., 1970) and eventually extending the sounding rocket measurements of SEPs up to the element Fe (Bertsch et al., 1969).

During the 1970s, increasingly accurate SEP measurements began to be made almost continuously with *dE/dx* vs *E* particle telescopes (measuring energy loss in a thin detector vs total energy as a particle stops) on board satellites, eventually using Si solid-state detectors, and lower-resolution, labor-intensive nuclear emulsions faded from use. I marvel at how important they once were to my career: I became an astrophysicist because of my specialized knowledge of an obsolete technology which no longer exists. But by the time nuclear emulsions died I had become an astrophysicist. There was no turning back. By 1977 I was working on particle telescopes for spacecraft with Tycho von Rosenvinge and Frank McDonald.

^{3}HE-RICH EVENTS

It was well known that GCRs fragment against interstellar H during their lifetime in space, ^{4}He produces ^{3}He and ^{2}H while C, N, and O fragment to produce Li, Be, and B. Discovery of enhanced ^{3}He in SEP abundances was first considered as new evidence for fragmentation in solar flares but soon came the discovery of ~1000-fold enhancements, ^{3}He/^{4}He = 1.5 (Serlemitsos and Balasubrahmanyan, 1975) vs 4 x 10^{-4} in the solar wind, with no measurable ^{2}H, Li, Be, or B at all. These events had nothing to do with fragmentation; this was a completely new resonance phenomenon, where waves resonate with the ^{3}He gyrofrequency (e.g. Fisk, 1978; Temerin and Roth, 1992). These ^{3}He-rich events were small (Mason, 2007). In fact, working with Robert Lin, we found that they were associated with small 10–100 keV electron events (Reames et al., 1985) and with type III radio bursts (Reames and Stone, 1986) that the streaming electrons produce. Years earlier, Wild et al. (1963) had distinguished two sources of SEPs producing different radio bursts (i) rapidly streaming electrons producing type III radio bursts and (ii) type II bursts produced at shock waves known to accelerate ions as well. Lin (1970, 1974) had found that type III bursts often came from "pure" electron events. Apparently, these frequent type-III events were mostly ^{3}He-rich events.

Many early theories of ^{3}He enhancement (e.g. Fisk, 1978; see early references in Reames, 2021a, 2021c) involved selective heating of the ^{3}He by resonant wave absorption followed by acceleration later by some unspecified mechanism. Attending a dinner following a spacecraft meeting, I was sitting across from Mike Temerin who asked "What do you do?" so I began to talk about these weird ^{3}He-rich events. He thought there might be a relationship with the "ion conics" seen in the Earth's magnetosphere where mirroring ions absorb energy from impinging waves until they can finally escape, forming the conic spatial distribution. A result of this discussion was the paper by Temerin and Roth (1992) where streaming electrons produce the waves that are resonantly absorbed to preferentially accelerate ^{3}He, providing both acceleration and a connection between the ^{3}He and the abundant streaming type-III electrons. Ideas originate in many places.

These "impulsive" SEP events have associated enhancements of heavy elements which were eventually found to increase as the 3.6 power, on average, of an element's mass-to-charge ratio *A/Q*, its atomic mass *A* divided by its electronic charge *Q*, first up to Fe, and then throughout the periodic table by a factor of ~1,000 between H and Pb (Mason et al., 2004; Reames and Ng, 2004; Reames et al., 2014a). Theoretical simulations associate these strong abundance enhancements with magnetic reconnection (Drake et al., 2009) on open field lines and an important association now connects impulsive SEP events with solar jets (Kahler et al., 2001; Reames, 2002; Nitta et al., 2006, 2015; Wang et al., 2006; Bučík et al., 2018a, 2018b; Bučík, 2020). Escape on the open magnetic field lines from jets produces no nuclear fragments.

CMES

Meanwhile, the large "gradual" SEP events were found by Kahler et al. (1984) to have a 96% correlation with fast, wide coronal mass ejections (CMEs) and the shock waves that they drive (Kahler, 2001; Gopalswamy et al., 2012; Kouloumvakos et al., 2019). Shock theory had been well developed for GCRs and was already being applied to "energetic storm particles" (ESPs) that peak at the interplanetary continuation of these same shocks (Lee, 2005; Lee et al., 2012). CME-driven shock acceleration explained the broad spatial distribution of large SEP events (Cane et al., 1988; Reames, 1999, 2013) replacing the "birdcage" model that was invented to allow protons to hop from loop to loop across the face of the corona from flares (see Reames 2021a). In contrast, shocks easily cross magnetic fields, accelerating particles over a broad front. The highest energies were produced by the shocks at the corona (Zank et al., 2000; Cliver et al., 2004; Ng and Reames 2008; Desai and Giacalone, 2016) near their onsets at ~2 R_S (Reames, 2009a; 2009b). The work of Zank et al. (2000) led to the development of the iPATH models of SEP transport (e.g. Hu et al., 2018).

The growing realization of the importance of CMEs and of shock acceleration of SEPs, especially in large gradual events, was pointed out by Jack Gosling's (1993, 1994) paper "The solar-flare myth." This paper caused a great controversy with flare enthusiasts who had not followed the evolution of CME and SEP research (see Reames, 2021a or 2021b for relevant publications). We have now come to understand that flares do not contribute SEPs in space; flares are hot and bright precisely because all the energy from magnetic reconnection, including accelerated particles, is contained on closed magnetic loops or dumped into their footpoints, so only photons and neutrals escape (Mandzhavidze et al., 1999; Murphy et al., 1991, 2016). Jets are the open-field equivalents that act as a source of interplanetary SEPs (e.g. Bučík, 2020), but CME-driven shocks dominate large events.

As protons stream away from a shock they amplify Alfvén waves that scatter all ions coming behind. This strengthens the acceleration and scatters and traps lower-rigidity ions, limiting intensities at the "streaming limit" (Reames and Ng, 1998, 2010), flattening energy spectra (Reames and Ng, 2010), and altering element abundances (Reames et al., 2000). Study of this self-consistent theory of wave-particle interactions was led by Chee Keong Ng (Ng and Reames, 1994, 1995; Ng et al., 1999, 2003, 2012) and applied to the time evolution of shock acceleration (Ng and Reames, 2008). Hopefully, someone will continue and extend these careful self-consistent studies.

FIP AND *A/Q*

It had been known for many years (e.g., Webber, 1975) that the abundances of elements in SEPs had ~3x enhancements, relative to photospheric abundances, of elements with low (<10 eV) first ionization potential (FIP). This 3x enhancement is an ion-neutral fractionation during formation of the solar corona. Electromagnetic waves can affect low-FIP elements (e.g., Mg, Si, and Fe) that are initially ionized, but not high-FIP neutral atoms (e.g., O, Ne, and Ar) rising up to the corona where all become ionized. Incidentally, the FIP pattern of SEPs differs from that of the solar wind (Mewaldt et al., 2002; Reames 2018a;

Laming et al., 2019); SEPs are not just accelerated solar wind. However, FIP may help locate the different sources of SEPs and solar wind in the corona (Brooks and Yardley 2021).

Meyer (1985); (Reames, 1995, 2014) found that element abundances in SEP events, compared with photospheric abundances, consisted of a FIP effect, shared by all events, and a dependence upon *A/Q*, that varied from event to event. The FIP effect occurred during formation of the corona, while the *A/Q* dependence resulted during acceleration, much later. Breneman and Stone (1985) established a power-law dependence using average *Q* values measured by Luhn et al. (1984). However, the *Q* values of the ions depend upon source electron temperature as noted by Meyer (1985).

Source Temperatures

The relevance of temperature was also noted by Jean-Paul Meyer in impulsive SEP events (Reames et al., 1994) where ^{4}He, C, N, and O abundances appeared un-enhanced because they were all fully ionized, while Ne, Mg, and Si had comparable enhancements because they were in stable two-electron states, while Fe was further enhanced. This configuration can occur at about 3 MK. Direct charge measurements in impulsive events had shown that elements up to Si were fully ionized (Luhn et al., 1987), thus they must then be stripped *after* acceleration, as was later proven (DiFabio et al., 2008).

Much later, we have been able to determine a temperatures for each event from its abundance measurements by trying *Q* values for many temperatures to see which gives the best-fit power law of enhancements vs *A/Q* (Reames et al., 2014b); most impulsive SEP events yield ~2.5 MK and, recently, EUV temperatures in solar jet sources of impulsive events were also found to peak at ~2.5 MK (Bučík et al., 2021). Source temperatures for impulsive SEP events were mostly within the ~10% error of the determination, however, similar techniques applied to abundances of gradual SEP events (Reames, 2016a; 2018b) varied widely from 0.6–2 MK when dominated by ambient coronal ions and >2 MK when they involved reaccelerated impulsive ions. These higher-temperature gradual SEP events fit in with the growing evidence that CME-driven shock waves could sometimes preferentially reaccelerate ions from residual impulsive suprathermal ions originally from jets (Tylka et al., 2001, 2005; Desai et al., 2003; Tylka and Lee, 2006; Sandroos and Vainio, 2007; Reames, 2016b). These suprathermal ions were found to collect in pools, perhaps from multiple small jets that are difficult to resolve (Desai et al., 2003; Wiedenbeck et al., 2008, 2013; Bučík et al., 2014, 2015; Chen et al., 2015) repeatedly sampled by shocks (Reames, 2022).

Clearly, SEPs now seemed more complicated than just impulsive events from jets and gradual events from CME-driven shocks. Kahler et al. (2001) found CMEs from the jets in impulsive events that could drive fast local shocks and there were also large CMEs in gradual events could sample pools of residual impulsive ions. Reames (2020) suggested four SEP classes: 1) SEP1 impulsive events from pure magnetic reconnection in jets, 2) SEP2 events with additional acceleration when the local CME from that jet is fast enough to produce a shock, 3) SEP3 events are dominated by seed particles from preexisting impulsive suprathermal pools that are traversed by wide, fast shocks, and 4) SEP4 events are accelerated by wide, fast shocks predominantly from the ambient coronal material. The new emphasis on shocks and jets was a major change from the previous "flare myth." The abundances of SEPs from impulsive events retain their unique signature even when combined with ambient plasma and reaccelerated by shock waves.

We knew about power-law dependence upon *A/Q* in 1985 (Breneman and Stone, 1985). Powers in magnetic rigidity produce these powers in *A/Q* at a given MeV amu^{-1}. We knew about the importance of *Q* variations and temperature in determining abundances (Luhn et al., 1987; Meyer, 1985; Reames et al., 1994; Leske et al., 1995; Mason et al., 1995; Tylka et al., 1995). Yet it took ~20 years to relate this *A/Q* dependence to source temperatures in impulsive (Reames et al., 2014b) and gradual (Reames, 2016a; 2018b) events and to shocks plying various seed populations. It is true that reaccelerated impulsive ions may have changed their *A/Q* from stripping, but the patterns are dominated by their initial huge enhancement of the seed population while the *A/Q* dependence in shock acceleration is weak.

PERSPECTIVES

Where did I learn astrophysics? Not in graduate school. Early in my career I acknowledge learning astrophysics theory from colleague Reuven Ramaty. I learned specifics about electrons from co-author Robert Lin, radio emission from Robert Stone, CMEs from Stephen Kahler, and detectors from Tycho von Rosenvinge. I learned by working with these and other colleagues. Later, I learned a great deal about particle acceleration and transport from many years of discussions with Chee Ng, but I also profited greatly by working with other co-authors, by reading papers, and by endlessly looking at data. I am still learning astrophysics.

The most important contributions to SEP studies, in my opinion, were the determinations that the source of gradual events is CME-driven shock waves (Kahler et al., 1984) and that the source of impulsive SEP events is reconnection in jets (Kahler et al., 2001; Bučík, 2020). A lot of early insights were overlooked: Wild et al. (1963) already knew about the two sources of SEPs; [Meyer (1985), Figure 11] knew that source temperatures were an important determinant of abundances. Were flares taken so seriously just because they are easier to see than CMEs, shocks, or jets?

What is my most productive work? Ironically, an early review article (Reames, 1999) was not only well received as a first review of SEPs, but, it was especially helpful to me in collecting ideas that improved my own perspective. Thus, writing review articles can be as educational for the author as for the reader and I have written more as new areas evolved (Reames, 2013, 2015; 2018b, 2020; 2021b; 2021c). Textbooks are even better (Reames, 2021a). Regarding research articles, I think the recent articles on SEP temperatures cited above and the correlations of energy spectra with abundances (e.g., Reames, 2021d, 2022) will be as productive as the earlier articles on ^{3}He-rich events and type III bursts, FIP, or onset times.

In recent years I have continued to work mostly with data from the LEMT on the *Wind* spacecraft, now 27 years of data. There are detailed spectra and element abundance measurements during hundreds of SEP events, all different, from this and many other spacecraft, all freely available on the web (https://cdaweb.gsfc.nasa.gov/sp_phys/). Yet there are so few other people who look at it that I seem to have exclusive access. In contrast, there are also armies of co-authors who flock to join a few select articles. Are these topics vastly more interesting? Am I missing something wonderful, or is the issue more about funding than scientific interest? Aye, there's the rub. We few retirees, funded only by pensions, are able to graze unmolested the choicest historic pastures of data from instruments that are yet unequalled—without even writing a proposal. When possible, find time to follow the physics, rather than the crowd.

Actually, proposals are also interesting. Can you predict what you will discover in the next 3 years? I cannot. Will you doggedly follow an approved plan even if a surprising new avenue suddenly opens? Some will. Many ideas sound good on paper but later turn out to be unsupported by the data. I once calculated my "batting average" as only slightly over 0.300. Should we publish all those losers, i.e. "good ideas" that did not work? Approved proposals can also suffer from "group think." It is not a perfect system but it is hard to suggest improvement—unless you are retired.

In my opinion, abundances are a key to underlying physics of SEPs that has been poorly exploited theoretically. Why are energy spectra correlated with abundance enhancements in "pure" (SEP4) shock events (e.g. Reames 2021d, 2022)? How can they then be completely independent in "pure" (SEP1) impulsive events? Where do resonances (e.g. ^{3}He) fit into reconnection models that predict power laws in A/Q? Surely, there are opportunities for mirroring ^{3}He in reconnection regions. Is C/O somehow suppressed, on average, in SEPs, or is it overestimated in the photosphere (e.g., Reames 2021b)? Are ^{4}He/O depletions related to the high FIP of He; are there occasional He-poor regions in the solar corona (Reames 2017, 2019)? I am still trying to learn astrophysics.

We cannot produce beautiful images of the Sun with SEPs, but we have made significant progress with the data we do have. We have been doing "multi-messenger" science for 60 years with SEPs, type-II and type-III radio bursts, and CMEs, long before it became so fashionable. There is much more of it to do.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://cdaweb.gsfc.nasa.gov/sp_phys/.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I thank my many colleagues who have helped advance the study of SEPs.

REFERENCES

Bertsch, D. L., Fichtel, C. E., and Reames, D. V. (1969). Relative Abundance of Iron-Group Nuclei in Solar Cosmic Rays. *ApJ* 157, L53. doi:10.1086/180383

Breneman, H. H., and Stone, E. C. (1985). Solar Coronal and Photospheric Abundances from Solar Energetic Particle Measurements. *ApJ* 299, L57. doi:10.1086/184580

Brooks, D. H., and Yardley, S. L. (2021). The Source of the Major Solar Energetic Particle Events from Super Active Region 11944. *Sci. Adv.* 7 (10), eabf0068. doi:10.1126/sciadv.abf0068

Bučík, R. (2020). ^{3}He-rich Solar Energetic Particles: Solar Sources. *Space Sci. Rev.* 216, 24.

Bučík, R., Innes, D. E., Chen, N. H., Mason, G. M., Gómez-Herrero, R., and Wiedenbeck, M. E. (2015). Long-lived Energetic Particle Source Regions on the Sun. *J. Phys. Conf. Ser.* 642, 012002. doi:10.1088/1742-6596/642/1/012002

Bučík, R., Innes, D. E., Mall, U., Korth, A., Mason, G. M., and Gómez-Herrero, R. (2014). Multi-Spacecraft Observations of Recurrent ^{3}He-Rich Solar Energetic Particles. *ApJ* 786, 71. doi:10.1088/0004-637X/786/1/71

Bučík, R., Innes, D. E., Mason, G. M., Wiedenbeck, M. E., Gómez-Herrero, R., and Nitta, N. V. (2018a). ^{3}He-rich Solar Energetic Particles in Helical Jets on the Sun. *Astrophys. J.* 852, 76.

Bučík, R., Mulay, S. M., Mason, G. M., Nitta, N. V., Desai, M. I., and Dayeh, M. A. (2021). Temperature in Solar Sources of ^{3}He-Rich Solar Energetic Particles and Relation to Ion Abundances. *Astrophys. J.* 908, 243.

Bučík, R., Wiedenbeck, M. E., Mason, G. M., Gómez-Herrero, R., Nitta, N. V., and Wang, L. (2018b). ^{3}He-rich Solar Energetic Particles from sunspot Jets. *Astrophys. J. Lett.* 869, L21.

Cane, H. V., Reames, D. V., and von Rosenvinge, T. T. (1988). The Role of Interplanetary Shocks in the Longitude Distribution of Solar Energetic Particles. *J. Geophys. Res.* 93, 9555. doi:10.1029/ja093ia09p09555

Chen, N.-h., Bučík, R., Innes, D. E., and Mason, G. M. (2015). Case Studies of Multi-day3He-Rich Solar Energetic Particle Periods. *A&A* 580, A16. doi:10.1051/0004-6361/201525618

Cliver, E. W., Kahler, S. W., and Reames, D. V. (2004). Coronal Shocks and Solar Energetic Proton Events. *ApJ* 605, 902–910. doi:10.1086/382651

Desai, M., and Giacalone, J. (2016). Large Gradual Solar Energetic Particle Events. *Living Rev. Sol. Phys.* 13. doi:10.1007/s41116-016-0002-5

Desai, M. I., Mason, G. M., Dwyer, J. R., Mazur, J. E., Gold, R. E., Krimigis, S. M., et al. (2003). Evidence for a Suprathermal Seed Population of Heavy Ions Accelerated by Interplanetary Shocks Near 1 AU. *ApJ* 588, 1149–1162. doi:10.1086/374310

DiFabio, R., Guo, Z., Möbius, E., Klecker, B., Kucharek, H., Mason, G. M., et al. (2008). Energy-dependent Charge States and Their Connection with Ion Abundances in Impulsive Solar Energetic Particle Events. *Astrophysical J.* 687, 623–634. doi:10.1086/591833

Drake, J. F., Cassak, P. A., Shay, M. A., Swisdak, M., and Quataert, E. (2009). A Magnetic Reconnection Mechanism for Ion Acceleration and Abundance Enhancements in Impulsive Flares. *ApJ* 700, L16–L20. doi:10.1088/0004-637X/700/1/L16

Durgaprasad, N., Fichtel, C. E., Guss, D. E., Reames, D. V., O'dell, F. W., Shapiro, M. M., et al. (1970). Chemical Composition of Relativistic Cosmic Rays Detected above the Atmosphere. *Phys. Rev. D* 1, 1021–1028. doi:10.1103/physrevd.1.1021

Fichtel, C. E., and Guss, D. E. (1961). Heavy Nuclei in Solar Cosmic Rays. *Phys. Rev. Lett.* 6, 495–497. doi:10.1103/PhysRevLett.6.495

Fisk, L. A. (1978). He-3-rich Flares - A Possible Explanation. *ApJ* 224, 1048. doi:10.1086/156456

Gopalswamy, N., Xie, H., Yashiro, S., Akiyama, S., Mäkelä, P., and Usoskin, I. G. (2012). Properties of Ground Level Enhancement Events and the Associated Solar Eruptions during Solar Cycle 23. *Space Sci. Rev.* 171, 23–60. doi:10.1007/s11214-012-9890-4

Gosling, J. T. (1994). Correction to "The Solar Flare Myth". *J. Geophys. Res.* 99, 4259. doi:10.1029/94ja00015

Gosling, J. T. (1993). The Solar Flare Myth. *J. Geophys. Res.* 98, 18937–18949. doi:10.1029/93JA01896

Hu, J., Li, G., Fu, S., Zank, G., and Ao, X. (2018). Modeling a Single SEP Event from Multiple Vantage Points Using the iPATH Model. *ApJ* 854, L19. doi:10.3847/2041-8213/aaabc1

Kahler, S. W., Reames, D. V., and Sheeley, Jr., N. R., Jr. (2001). Coronal Mass Ejections Associated with Impulsive Solar Energetic Particle Events. *ApJ* 562, 558–565. doi:10.1086/323847

Kahler, S. W., Sheeley, N. R., Jr., Howard, R. A., Koomen, M. J., Michels, D. J., McGuire, R. E., et al. (1984). Associations between Coronal Mass Ejections and Solar Energetic Proton Events. *J. Geophys. Res.* 89, 9683. doi:10.1029/JA089iA11p09683

Kahler, S. W. (2001). The Correlation between Solar Energetic Particle Peak Intensities and Speeds of Coronal Mass Ejections: Effects of Ambient Particle Intensities and Energy Spectra. *J. Geophys. Res.* 106, 20947–20955. doi:10.1029/2000JA002231

Kouloumvakos, A., Rouillard, A. P., Wu, Y., Vainio, R., Vourlidas, A., Plotnikov, I., et al. (2019). Connecting the Properties of Coronal Shock Waves with Those of Solar Energetic Particles. *Astrophys. J.* 876 80. doi:10.3847/1538-4357/ab15d7

Laming, J. M., Vourlidas, A., Korendyke, C., Chua, D., Cranmer, S. R., Ko, Y.-K., et al. (2019). Element Abundances: a New Diagnostic for the Solar Wind. *ApJ* 879, 124. doi:10.3847/1538-4357/ab23f1

Lee, M. A. (2005). Coupled Hydromagnetic Wave Excitation and Ion Acceleration at an Evolving Coronal/interplanetary Shock. *Astrophys J. Suppl. S* 158, 38–67. doi:10.1086/428753

Lee, M. A., Mewaldt, R. A., and Giacalone, J. (2012). Shock Acceleration of Ions in the Heliosphere. *Space Sci. Rev.* 173, 247–281. doi:10.1007/s11214-012-9932-y

Leske, R. A., Cummings, J. R., Mewaldt, R. A., Stone, E. C., and von Rosenvinge, T. T. (1995). Measurements of the Ionic Charge States of Solar Energetic Particles Using the Geomagnetic Field. *Astrophys J. Lett.* 452, L149. doi:10.1086/309718

Lin, R. P. (1974). Non-relativistic Solar Electrons. *Space Sci. Rev.* 16, 189. doi:10.1007/bf00240886

Lin, R. P. (1970). The Emission and Propagation of 40 keV Solar Flare Electrons. I: The Relationship of 40 keV Electrons to Energetic Proton and Relativistic Electron Emission by the Sun. *Sol. Phys.* 12, 266. doi:10.1007/bf00227122

Luhn, A., Klecker, B., Hovestadt, D., Gloeckler, G., Ipavich, F. M., Scholer, M., et al. (1984). Ionic Charge States of N, Ne, Mg, Si and S in Solar Energetic Particle Events. *Adv. Space Res.* 4, 161–164. doi:10.1016/0273-1177(84)90307-7

Luhn, A., Klecker, B., Hovestadt, D., and Moebius, E. (1987). The Mean Ionic Charge of Silicon in He-3-Rich Solar Flares. *ApJ* 317, 951. doi:10.1086/165343

Mandzhavidze, N., Ramaty, R., and Kozlovsky, B. (1999). Determination of the Abundances of Subcoronal ^{4}He and of Solar Flare-Accelerated ^{3}He and ^{4}He from Gamma-ray Spectroscopy. *Astrophys. J.* 518, 918.

Mason, G. M. (2007). 3He-Rich Solar Energetic Particle Events. *Space Sci. Rev.* 130, 231–242. doi:10.1007/s11214-007-9156-8

Mason, G. M., Mazur, J. E., Dwyer, J. R., Jokipii, J. R., Gold, R. E., and Krimigis, S. M. (2004). Abundances of Heavy and Ultraheavy Ions in3He-rich Solar Flares. *ApJ* 606, 555–564. doi:10.1086/382864

Mason, G. M., Mazur, J. E., Looper, M. D., and Mewaldt, R. A. (1995). Charge State Measurements of Solar Energetic Particles Observed with SAMPEX. *ApJ* 452, 901. doi:10.1086/176358

Mewaldt, R. A., Cohen, C. M. S., Leske, R. A., Christian, E. R., Cummings, A. C., Stone, E. C., et al. (2002). Fractionation of Solar Energetic Particles and Solar Wind According to First Ionization Potential. *Adv. Space Res.* 30, 79–84. doi:10.1016/s0273-1177(02)00263-6

Meyer, J.-P. (1985). The Baseline Composition of Solar Energetic Particles. *ApJS* 57, 151. doi:10.1086/191000

Murphy, R. J., Kozlovsky, B., and Share, G. H. (2016). Evidence for Enhanced ^{3}He in Flare-Accelerated Particles Based on New Calculations of the Gamma-ray Line Spectrum. *Astrophys.J.* 833, 166. doi:10.3847/1538-4357/833/2/196

Murphy, R. J., Ramaty, R., Reames, D. V., and Kozlovsky, B. (1991). Solar Abundances from Gamma-ray Spectroscopy - Comparisons with Energetic Particle, Photospheric, and Coronal Abundances. *ApJ* 371, 793. doi:10.1086/169944

Ng, C. K., and Reames, D. V. (1994). Focused Interplanetary Transport of Approximately 1 MeV Solar Energetic Protons through Self-Generated Alfven Waves. *ApJ* 424, 1032. doi:10.1086/173954

Ng, C. K., and Reames, D. V. (1995). Pitch Angle Diffusion Coefficient in an Extended Quasi-Linear Theory. *ApJ* 453, 890. doi:10.1086/176449

Ng, C. K., and Reames, D. V. (2008). Shock Acceleration of Solar Energetic Protons: the First 10 minutes. *ApJ* 686, L123–L126. doi:10.1086/592996

Ng, C. K., Reames, D. V., and Tylka, A. J. (1999). Effect of Proton-Amplified Waves on the Evolution of Solar Energetic Particle Composition in Gradual Events. *Geophys. Res. Lett.* 26, 2145–2148. doi:10.1029/1999gl900459

Ng, C. K., Reames, D. V., and Tylka, A. J. (2003). Modeling Shock-accelerated Solar Energetic Particles Coupled to Interplanetary Alfven Waves. *ApJ* 591, 461–485. doi:10.1086/375293

Ng, C. K., Reames, D. V., and Tylka, A. J. (2012). Solar Energetic Particles: Shock Acceleration and Transport through Self-Amplified Waves. *AIP Conf. Proc.* 1436, 212. doi:10.1063/1.4723610

Nitta, N. V., Mason, G. M., Wang, L., Cohen, C. M. S., and Wiedenbeck, M. E. (2015). SOLAR SOURCES OF3He-RICH SOLAR ENERGETIC PARTICLE EVENTS IN SOLAR CYCLE 24. *ApJ* 806, 235. doi:10.1088/0004-637x/806/2/235

Nitta, N. V., Reames, D. V., DeRosa, M. L., Liu, Y., Yashiro, S., and Gopalswamy, N. (2006). Solar Sources of Impulsive Solar Energetic Particle Events and Their Magnetic Field Connection to the Earth. *ApJ* 650, 438–450. doi:10.1086/507442

Reames, D. V. (2018b). Abundances, Ionization States, Temperatures, and FIP in Solar Energetic Particles. *Space Sci. Rev.* 214, 61. doi:10.1007/s11214-018-0495-4

Reames, D. V., Cliver, E. W., and Kahler, S. W. (2014a). Abundance Enhancements in Impulsive Solar Energetic-Particle Events with Associated Coronal Mass Ejections. *Sol. Phys.* 289, 3817–3841. doi:10.1007/s11207-014-0547-1

Reames, D. V., Cliver, E. W., and Kahler, S. W. (2014b). Variations in Abundance Enhancements in Impulsive Solar Energetic-Particle Events and Related CMEs and Flares. *Sol. Phys.* 289, 4675–4689. doi:10.1007/s11207-014-0589-4

Reames, D. V. (1995). Coronal Abundances Determined from Energetic Particles. *Adv. Space Res.* 15 (7), 41–51. doi:10.1016/0273-1177(94)00018-v

Reames, D. V. (2014). Element Abundances in Solar Energetic Particles and the Solar corona. *Sol. Phys.* 289, 977–993. doi:10.1007/s11207-013-0350-4

Reames, D. V. (2022). Energy Spectra vs. Element Abundances in Solar Energetic Particles and the Roles of Magnetic Reconnection and Shock Acceleration, In press. *Solar Phys.* doi:10.1007/s11207-022-01961-2

Reames, D. V. (2021c). Fifty Years of ^{3}He-Rich Events. *Front. Astron. Space Sci.* 8, 164. doi:10.3389/fspas.2021.760261

Reames, D. V. (2020). Four Distinct Pathways to the Element Abundances in Solar Energetic Particles. *Space Sci. Rev.* 216, 20. doi:10.1007/s11214-020-0643-5

Reames, D. V. (2019). Helium Suppression in Impulsive Solar Energetic-Particle Events. *Sol. Phys.* 294, 32. doi:10.1007/s11207-019-1422-x

Reames, D. V. (2002). Magnetic Topology of Impulsive and Gradual Solar Energetic Particle Events. *Astrophys. J. Lett.* 571, L63–L66. doi:10.1086/341149

Reames, D. V., Meyer, J. P., and von Rosenvinge, T. T. (1994). Energetic-particle Abundances in Impulsive Solar Flare Events. *ApJS* 90, 649. doi:10.1086/191887

Reames, D. V., and Ng, C. K. (2004). Heavy-Element Abundances in Solar Energetic Particle Events. *ApJ* 610, 510–522. doi:10.1088/0004-637X/723/2/128610.1086/421518

Reames, D. V., and Ng, C. K. (1998). Streaming-limited Intensities of Solar Energetic Particles. *ApJ* 504, 1002–1005. doi:10.1086/306124

Reames, D. V., and Ng, C. K. (2010). Streaming-limited Intensities of Solar Energetic Particles on the Intensity Plateau. *ApJ* 723, 1286–1293. doi:10.1088/0004-637x/723/2/1286

Reames, D. V., Ng, C. K., and Tylka, A. J. (2000). Initial Time Dependence of Abundances in Solar Energetic Particle Events. *Astrophys. J. Lett.* 531, L83–L86. doi:10.1086/312517

Reames, D. V. (2021d). On the Correlation between Energy Spectra and Element Abundances in Solar Energetic Particles. *Sol. Phys.* 296, 24. doi:10.1007/s11207-021-01762-z

Reames, D. V. (1999). Particle Acceleration at the Sun and in the Heliosphere. *Space Sci. Rev.* 90, 413–491. doi:10.1023/A:1005105831781xx
Reames, D. V. (2021b). Sixty Years of Element Abundance Measurements in Solar Energetic Particles. *Space Sci. Rev.* 217, 72. doi:10.1007/s11214-021-00845-4
Reames, D. V. (2021a). "Solar Energetic Particles," in *Lec. Notes Phys*. Second Edition (Cham, Switzerland: 978 Springer Nature). open access.
Reames, D. V. (2009b). Solar Energetic-Particle Release Times in Historic Ground-Level Events. *ApJ* 706, 844–850. doi:10.1088/0004-637X/706/1/844
Reames, D. V. (2009a). Solar Release Times of Energetic Particles in Ground-Level Events. *ApJ* 693, 812–821. doi:10.1088/0004-637X/693/1/812
Reames, D. V., and Stone, R. G. (1986). The Identification of Solar He-3-Rich Events and the Study of Particle Acceleration at the Sun. *ApJ* 308, 902. doi:10.1086/164560
Reames, D. V. (2016a). Temperature of the Source Plasma in Gradual Solar Energetic Particle Events. *Sol. Phys.* 291, 911–930. doi:10.1007/s11207-016-0854-9
Reames, D. V. (2018a). The "FIP Effect" and the Origins of Solar Energetic Particles and of the Solar Wind. *Sol. Phys.* 293, 47. doi:10.1007/s11207-018-1267-8
Reames, D. V. (2017). The Abundance of Helium in the Source Plasma of Solar Energetic Particles. *Sol. Phys.* 292, 156. doi:10.1007/s11207-017-1173-5
Reames, D. V. (2016b). The Origin of Element Abundance Variations in Solar Energetic Particles. *Sol. Phys.* 291, 2099–2115. doi:10.1007/s11207-016-0942-x
Reames, D. V. (2013). The Two Sources of Solar Energetic Particles. *Space Sci. Rev.* 175, 53–92. doi:10.1007/s11214-013-9958-9
Reames, D. V., von Rosenvinge, T. T., and Lin, R. P. (1985). Solar He-3-Rich Events and Nonrelativistic Electron Events - A New Association. *ApJ* 292, 716. doi:10.1086/163203
Reames, D. V. (2015). What Are the Sources of Solar Energetic Particles? Element Abundances and Source Plasma Temperatures. *Space Sci. Rev.* 194, 303–327. doi:10.1007/s11214-015-0210-7
Sandroos, A., and Vainio, R. (2007). Simulation Results for Heavy Ion Spectral Variability in Large Gradual Solar Energetic Particle Events. *ApJ* 662, L127–L130. doi:10.1086/519378
Serlemitsos, A. T., and Balasubrahmanyan, V. K. (1975). Solar Particle Events with Anomalously Large Relative Abundance of He-3. *ApJ* 198, 195. doi:10.1086/153592
Temerin, M., and Roth, I. (1992). The Production of He-3 and Heavy Ion Enrichment in He-3-Rich Flares by Electromagnetic Hydrogen Cyclotron Waves. *ApJ* 391, L105. doi:10.1086/186408
Tylka, A. J., Boberg, P. R., Adams, J. H. J., Jr., Beahm, L. P., Dietrich, W. F., and Kleis, T. (1995). The Mean Ionic Charge State of Solar Energetic Fe Ions above 200 MeV Per Nucleon. *ApJ* 444, L109. doi:10.1086/187872
Tylka, A. J., Cohen, C. M. S., Dietrich, W. F., Lee, M. A., Maclennan, C. G., Mewaldt, R. A., et al. (2005). Shock Geometry, Seed Populations, and the Origin of Variable Elemental Composition at High Energies in Large Gradual Solar Particle Events. *ApJ* 625, 474–495. doi:10.1086/429384
Tylka, A. J., Cohen, C. M. S., Dietrich, W. F., Maclennan, C. G., McGuire, R. E., Ng, C. K., et al. (2001). Evidence for Remnant Flare Suprathermals in the Source Population of Solar Energetic Particles in the 2000 Bastille Day Event. *Astrophys. J. Lett.* 558, L59–L63. doi:10.1086/323344
Tylka, A. J., and Lee, M. A. (2006). A Model for Spectral and Compositional Variability at High Energies in Large, Gradual Solar Particle Events. *ApJ* 646, 1319–1334. doi:10.1086/505106
Wang, Y. M., Pick, M., and Mason, G. M. (2006). Coronal Holes, Jets, and the Origin of3He-rich Particle Events. *ApJ* 639, 495–509. doi:10.1086/499355
Webber, W. R. (1975). Solar and Galactic Cosmic ray Abundances - A Comparison and Some Comments. *Proc. 14th Int. Cos. Ray Conf. Munich* 5, 1597.
Wiedenbeck, M. E., Cohen, C. M. S., Cummings, A. C., de Nolfo, G. A., Leske, R. A., Mewaldt, R. A., et al. (2008). Persistent Energetic ^{3}He in the Inner Heliosphere. *Proc. 30th Int. Cosmic Ray Conf. (Mérida)* 1, 91.
Wiedenbeck, M. E., Mason, G. M., Cohen, C. M. S., Nitta, N. V., Gómez-Herrero, R., and Haggerty, D. K. (2013). Observations of Solar Energetic Particles From3he-Rich Events over A Wide Range of Heliographic Longitude. *ApJ* 762, 54. doi:10.1088/0004-637x/762/1/54
Wild, J. P., Smerd, S. F., and Weiss, A. A. (1963). Solar Bursts. *Annu. Rev. Astron. Astrophys.* 1, 291–366. doi:10.1146/annurev.aa.01.090163.001451
Zank, G. P., Rice, W. K. M., and Wu, C. C. (2000). Particle Acceleration and Coronal Mass Ejection Driven Shocks: A Theoretical Model. *J. Geophys. Res.* 105, 25079–25095. doi:10.1029/1999JA000455

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 12 May 2022
doi: 10.3389/fspas.2022.898837

From Coronal Holes to Pulsars and Back Again: Learning the Importance of Data

Y.-M. Wang *

Space Science Division, Naval Research Laboratory, Washington, DC, United States

Although wanting to become an astronomer from an early age, I ended up in solar physics purely by chance, after first working in high-energy astrophysics. I've never regretted switching from the pulsar to the solar magnetosphere, because solar physics has a great advantage over other areas of astrophysics—in the enormous amount of high-quality data available, much of it underutilized. I've often wondered why theoreticians and modelers don't spent more time looking at these data (perhaps they feel that it is cheating, like taking a peek at the answers to a difficult homework assignment?). Conversely, I wonder why observers and data analysts aren't more skeptical of the theoretical models—especially the fashionable ones.

Keywords: coronal holes, solar wind, solar dynamo, magnetograms, coronal loops, coronal heating

OPEN ACCESS

Edited by:
Georgia Adair De Nolfo, National Aeronautics and Space Administration, United States

Reviewed by:
Donald Reames, University of Maryland, College Park, United States

****Correspondence:***
Y.-M. Wang
yi.wang@nrl.navy.mil

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *17 March 2022*
Accepted: *25 April 2022*
Published: *12 May 2022*

Citation:
Wang Y-M (2022) From Coronal Holes to Pulsars and Back Again: Learning the Importance of Data. Front. Astron. Space Sci. 9:898837. doi: 10.3389/fspas.2022.898837

1 INTRODUCTION

I decided to become an astronomer after I read Fred Hoyle's "Frontiers of Astronomy" at the age of fourteen. Written in 1955, just before the start of the space age, the book covers everything from the Earth's rotation to the steady-state theory of the Universe, even touching on coronal heating along the way. I've always regarded Hoyle's writing as a model of clarity, in which he focused on the essence of each problem and discussed it in simple physical terms. This seems to be a rare ability, judging from most science writing, which often relies on jargon to disguise a lack of real understanding, or is devoid of meaningful content, like a typical NASA news release ("SDO Peers Into Huge Hole on Sun!").

Deciding which area of astronomy/astrophysics to specialize in was much more difficult than deciding to become an astronomer. I ended up in solar physics as a result of a long series of fortuitous events:

1) My first experience with solar physics was as an astronomy major at Harvard College, when Gene Avrett, the head of the tutorial section, suggested that I work with him and George Withbroe. As a result, as part of my senior thesis, I helped analyze OSO-4 spectroheliograms showing an area of greatly reduced EUV emission (which we referred to as a "hole") near the Sun's south pole. Unfortunately, in our model for the polar transition region and corona in November 1967 (Withbroe and Wang 1972), we completely missed the big thing: that the reason for the low densities and temperatures was that solar wind was escaping from our "hole." Just a year later, Skylab showed conclusively that such "holes" were sources of high-speed solar wind. I think that George afterwards greatly regretted not making the connection, which might seem obvious in hindsight.

I must admit that I did not find the solar physics done at Harvard at that time, with its emphasis on spectroscopy and plane-parallel models of the solar atmosphere, terribly exciting. However, it was probably because of my undergraduate connections that I was eventually able to return to solar physics.

2) Perhaps the most important thing I learned as an astronomy major is that doing astronomy requires a solid knowledge of physics. After resolving to switch from solar physics to "real" astrophysics, I enrolled in MIT's physics department, where I was (fortunately) required to take the same courses as the other physics graduate students, in order to pass the general exam by the end of the second year. For my research advisor, I chose the plasma physicist Bruno Coppi, who seemed to believe that all magnetized objects in the Universe were tokamaks. With almost no supervision, I wrote a very skimpy, 60-page thesis on pulsar magnetospheres, which I paid an MIT secretary to type. As she handed me the typewritten manuscript, she said: "If I had known it was so easy to write a Ph.D. thesis, I would have done one myself!" Although this lack of supervision and less-than-stellar thesis work meant that I was off to a very slow start careerwise, it at least forced me to learn to work and think independently. Many of the postdocs I encounter today seem to just parrot the ideas of their thesis supervisors rather than trying to think for themselves. This might be for funding reasons, but I sometimes wonder if it is because they're not really that interested in what they're doing.
3) After MIT, I spent almost a decade in Europe, doing mainly theoretical work on pulsar magnetospheres and accretion onto neutron stars. The only data that I was able to use were timing measurements for a handful of binary X-ray pulsars, and I eventually felt that I was running out of ideas. In 1985, while at Bonn University, I applied for another research grant from the Deutsche Forschungsgemeinschaft. The evaluation began with the words: "Herr Wang is now approaching his mid-thirties, and it is time for him to look for a long-term position. Therefore, we will be extending his funding for only two more years." I shall be forever grateful to the reviewer(s) for this warning.
4) For the third time in my career, I was able to survive only by networking. As it happened, two senior scientists from NRL, Maurice Shapiro and Ken Johnston, were visiting the adjoining Max Planck Institute for Radio Astronomy as Humboldt Fellows. Maury (a pioneer cosmic ray physicist) attended our seminars and was friendly to the younger scientists, so I asked him about the possibility of a job at NRL. Maury told me to talk to Ken, who offered me a job in radio astronomy as we had lunch at the pizza restaurant next door.
5) I then made the luckiest "mistake" of my career: I waited almost half a year before showing up at Ken Johnston's office at NRL. When I reminded him of who I was, he said: "I thought you weren't coming, and I've run out of money. But don't worry, I'll send you to Anand." Anand was the owner of a Beltway Bandit company called Applied Research Corporation. Just by chance, in the spring of 1986, Neil Sheeley obtained funding from NRL's Research Advisory Committee to hire two scientists to work on magnetic flux transport models. As Neil later told me, he chose me not because of my work in high-energy astrophysics, but because he was familar with my undergraduate paper with George Withbroe (which was actually written by George). His other hire was Ana Nash, who had just obtained her doctorate in astronomy from the University of Wisconsin, doing her thesis on molecular clouds.

Switching from pulsar magnetospheres and accretion disks to the solar magnetic field was easy; the big difference was that an enormous amount of data was now available to test models and suggest new ideas. I was also extremely fortunate to have a boss who could explain solar concepts in a clear and simple way (like Fred Hoyle) and who was also an excellent mathematician. Neil's ability to explain things well may have been due to the influence of Richard Feynman, his freshman physics honors section instructor at Caltech, who would urge his students to try to think through questions themselves rather than just reading the textbook. Another great piece of luck was that I "returned" to solar physics in time for the launches of Ulysses and SOHO.

In the remainder of this Perspective article, I describe what I would consider to be my most significant contributions to research in solar physics. These contributions all relied heavily on the type of detailed observational data that are not available in most other areas of astrophysics.

2 THE QUASI-RIGID ROTATION OF CORONAL HOLES

One of the exciting discoveries made by Skylab was that coronal holes often rotate much more rigidly than the underlying photosphere. This mysterious property was even a topic of discussion in my plasma physics courses in the 1970s. During my first few years at NRL, the rotation of solar magnetic fields was the main focus of our small group.

We first tried to understand the quasi-rigid rotation shown by large-scale photospheric field patterns. According to Stenflo (1989), these unipolar patterns consisted of magnetic flux that emerged *in situ* from a rigidly rotating source at the bottom of the convection zone. Our flux transport simulations showed instead that they consisted of flux migrating poleward and equatorward from decaying active regions (Sheeley, Nash, and Wang 1987; Wang and Sheeley 1994). To understand why the patterns maintain their shape, consider a long line of ducks crossing a stream, with the current faster on one side than the other. If the ducks just drifted with the current, the line would become increasingly sheared with time. However, if each duck (flux element) continually swims toward the far bank (Sun's pole) and the line is continually replenished by new ducks entering from the near bank (active region latitudes), a stationary pattern will be set up.

Although this mechanism might at first seem applicable to coronal holes, the physical basis for their quasi-rigid rotation turns out to be different from that of the large-scale unipolar patterns. Coronal holes consist of open magnetic flux, whose

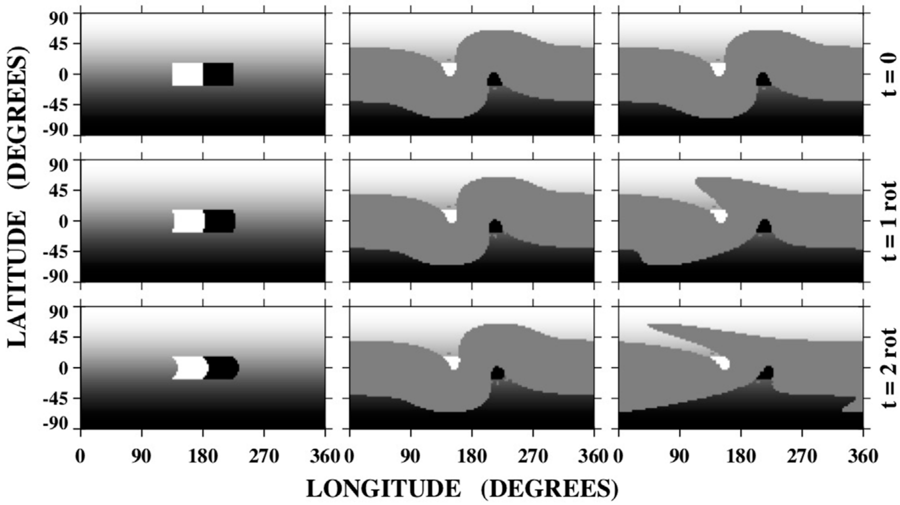

FIGURE 1 | The quasi-rigid rotation of a coronal hole, as illustrated by a configuration consisting of a bipole at the equator and an axisymmetric dipole field. The underlying photosphere rotates at the Snodgrass rate $\omega(L)=13.38-2.30\sin^2 L-1.62\sin^4 L$ deg day^{-1}, where L is latitude. Left column: Photospheric flux distribution after the lapse of 0, 1, and 2 rotations. Middle column: Corresponding open field regions, determined by applying a PFSS extrapolation to the photospheric field. Right column: Open field regions as they would appear if they rotated at the photospheric plasma rate. In a frame corotating with the equatorial bipole, the photospheric flux distribution and thus the distribution of open flux remain practically unchanged with time.

sources are the lowest-order multipoles of the photospheric field (the high-order multipoles fall off rapidly with height and are associated with smaller closed loops). By its very nature, a structure consisting of low-order multipoles, such as the outer coronal field or its footpoint areas (coronal holes), must rotate more rigidly than one consisting of high-order multipoles. To offset the effect of rotational shearing (which converts lower-order multipoles into higher-order ones), continual reconnection is required. When I presented these ideas during a discussion session at the 1992 G. S. Vaiana Memorial Symposium, I heard skeptical noises from the luminaries sitting in the front row, whose consensus opinion was that the rotation of coronal holes reflected that of its deep-seated sources.

The mixed reaction at the Vaiana Symposium motivated us to revisit the topic of coronal hole rotation and to look more closely at the Skylab observations. Comparing an image of the boot-shaped "Coronal Hole 1" with a Kitt Peak magnetogram taken one rotation earlier, I noticed that CH1 was connected to a decaying active region complex just below the equator. I then used the potential-field source-surface (PFSS) model to simulate the evolution of CH1, representing the photospheric field by an axisymmetric dipole and an east–west oriented bipole centered at the equator (In the PFSS model, the coronal field is assumed to be current-free and is constrained to be radial at a heliocentric distance of 2.5 $R_\odot$; all field lines that cross this "source surface" are considered to be "open.") To further simplify things, the only transport process that I included was differential rotation. As shown in **Figure 1**, the equatorial bipole distorts the open field areas at the poles, creating a pair of equatorward extensions that connect to the bipole. The extensions retain their vertical orientation from rotation to rotation, despite the underlying differential rotation. This result is easily understood: in a frame corotating with the equatorial bipole, the photospheric flux distribution remains almost unchanged from rotation to rotation; therefore the coronal field and distribution of open flux must also remain unchanged, at least in the current-free approximation (Wang and Sheeley 1993).

The reaction to our new explanation from Herb Gursky, then the superintendent of NRL's Space Science Division, was that we had made an interesting problem "boring." Spiro Antiochos dismissed the explanation with a wave of the hand, saying that it was "so obvious." The most insightful comment came from Eugene Parker, who said to Neil at a Chapman Conference dinner: "So you're saying it's just a property of potential fields."

A prediction of the model is that reconnection between open and closed field lines ("interchange reconnection") must be occurring continually at the boundaries of coronal holes. The reconnection site is high in the corona at the streamer cusps (and thus relatively difficult to observe), not near the photosphere as sometimes assumed. A possible observational signature of the reconnection is the heliospheric plasma sheet itself, which (as seen in white-light coronagraph images) consists of narrow raylike structures; these rays may represent newly reconnected, open field lines along which streamer material escapes into the heliosphere (Wang et al., 1998).

Judging from the fact that most studies of coronal and coronal hole rotation continue to invoke subsurface phenomena, it would appear that our explanation is still not widely accepted, perhaps because it is so "boring."

3 THE FLUX TRANSPORT DYNAMO

Shortly before we started our flux transport simulations, helioseismologists had detected the presence of a ~10–20 m s^{-1} poleward flow at the solar surface, and also shown that the radial gradient in the subsurface differential rotation was relatively small and opposite in sign to that assumed in earlier dynamo models, such as that of Leighton (1969). This led us to modify Leighton's model by adding meridional circulation and including only the latitudinal gradient in the differential rotation. We also required that the toroidal flux emerging at the surface be eventually resubmerged and annihilated by merging with its opposite-hemisphere counterpart at the equator, rather than being expelled from the Sun as assumed by Babcock (1961) and Leighton (1969). In our two-level model, the poleward-migrating flux at the photosphere was linked to the equatorward-migrating subsurface flux through flux conservation, with the implicit assumption that continual reconnection acts to reduce any magnetic stresses that build up (as in the corona). Our main result was that a ~1 m s^{-1} equatorward flow at the bottom of the convection zone would give rise to an equatorward progression of flux emergence and a reversal of the polar fields with an 11 year period (Wang, Sheeley, and Nash 1991). The ~10 m s^{-1} poleward surface flow resulted in highly concentrated "topknot" polar fields, consistent with magnetograph observations.

The only reaction to our model came from Arnab Choudhuri, who wrote: "We do not know what to make of this paper. The authors seem to be unaware of any of the dynamo research that has been done in the last 2 decades." Four years later, Choudhuri et al. (1995) published a paper entitled "The solar dynamo with meridional circulation" without mentioning our earlier work. Their model differed in that it was based on classical mean-field dynamo theory rather than the Babcock–Leighton picture, which takes proper account of magnetograph observations showing that the polar fields are formed from the surface transport of active region fields. It was only much later that Charbonneau (2010) called attention to our work as the "first post-helioseismic dynamo model based on the Babcock–Leighton mechanism."

4 "UNIPOLAR" PLAGES, LOOPLIKE FINE STRUCTURE, AND CORONAL HEATING

The coronal heating problem has been "solved" many times, with the current paradigm being the nanoflare model of Parker (1988), according to which footpoint motions generate tangential discontinuities in the corona (see, e.g., Dahlburg et al., 2016; Pontin and Hornig 2020). But if "nanoflares" and energy dissipation are occurring well above the loop footpoints, it is surprising that the upper parts of the loops appear so smooth and featureless. In contrast, the footpoint areas show a great deal of topologically complex fine structure (Aschwanden et al., 2007). Until now, however, a strong argument against heating via reconnection with small bipoles is that active region plages generally show almost no minority-polarity flux, even in high-resolution magnetograms.

However, by comparing SDO/AIA images of plages with HMI magnetograms, we have found numerous cases where the EUV images show small, looplike structures with horizontal dimensions of 3″–6″ (2-4 Mm), but the magnetograms (with 0.5″ pixels) show no minority-polarity flux at all (Wang 2016; Wang, Ugarte-Urra, and Reep 2019). The same holds for the "unipolar" network inside coronal holes, where the cores of Fe IX 17.1 nm plumes contain clusters of small loops which often do not have corresponding minority-polarity signatures in the magnetograms (Wang, Warren, and Muglach 2016). Because the loops have horizontal extents greatly exceeding the HMI pixel size, the failure to detect the minority-polarity flux is probably not simply the result of inadequate spatial resolution, but suggests a problem with instrument sensitivity in the presence of a strongly dominant polarity.

In 17.1 nm images, plage areas have a spongy or reticulated appearance. This so-called moss is generally interpreted as the transition region of the overlying hot loops (e.g., Berger et al., 1999). However, De Pontieu et al. (2003) found that the moss emission is only weakly correlated with the distribution of Ca II K emission, a proxy for the photospheric field strength. In my view, the 17.1 nm moss consists mainly of small, barely resolved loops that are continually being churned by the underlying granular motions and continually reconnect with the overlying loops. **Figure 2** shows some examples of loops with horizontal dimensions of ~3-4 Mm that are embedded inside "unipolar" plage areas. The inverted-Y structures are jets formed when the moss loops reconnect with the overlying active region loops.

Using SOHO/MDI magnetograms, Hagenaar et al. (2008) found that the emergence rate of ephemeral regions, defined as having fluxes $\lesssim 10^{20}$ Mx, was at least 3 times lower inside unipolar areas than in the quiet Sun. If we instead take the emergence or churning rate of small-scale loops to be the same everywhere on the Sun and assume that the energy released by reconnection with the large-scale field scales as the local field strength B (Wang 2020), we find that the energy flux density is of order 10^7 erg cm^{-2} s^{-1} when B ~ 300 G. This would be sufficient to heat the active region corona (Withbroe and Noyes 1977). Similarly, reconnection between small loops and the open flux inside coronal holes would give an energy flux of ~ 3×10^5 erg cm^{-2} s^{-1} (for B ~ 10 G), enough to drive the solar wind. In both cases, the energy input would be in the form of Ohmic heating, jets, and MHD waves. Heat would be conducted downward from the reconnection site into the transition region, leading to chromospheric evaporation, which would fill the overlying active region loops with hot, dense material (as in solar flares) and drive the solar wind mass flux along open field lines. In closed loops, the Alfvén waves generated by reconnection events near the coronal base would be reflected by the steep gradients in the transition region at the opposite end, and the trapped waves would interact with each other and undergo turbulent decay. In models where the Alfvén waves are excited by photospheric footpoint motions

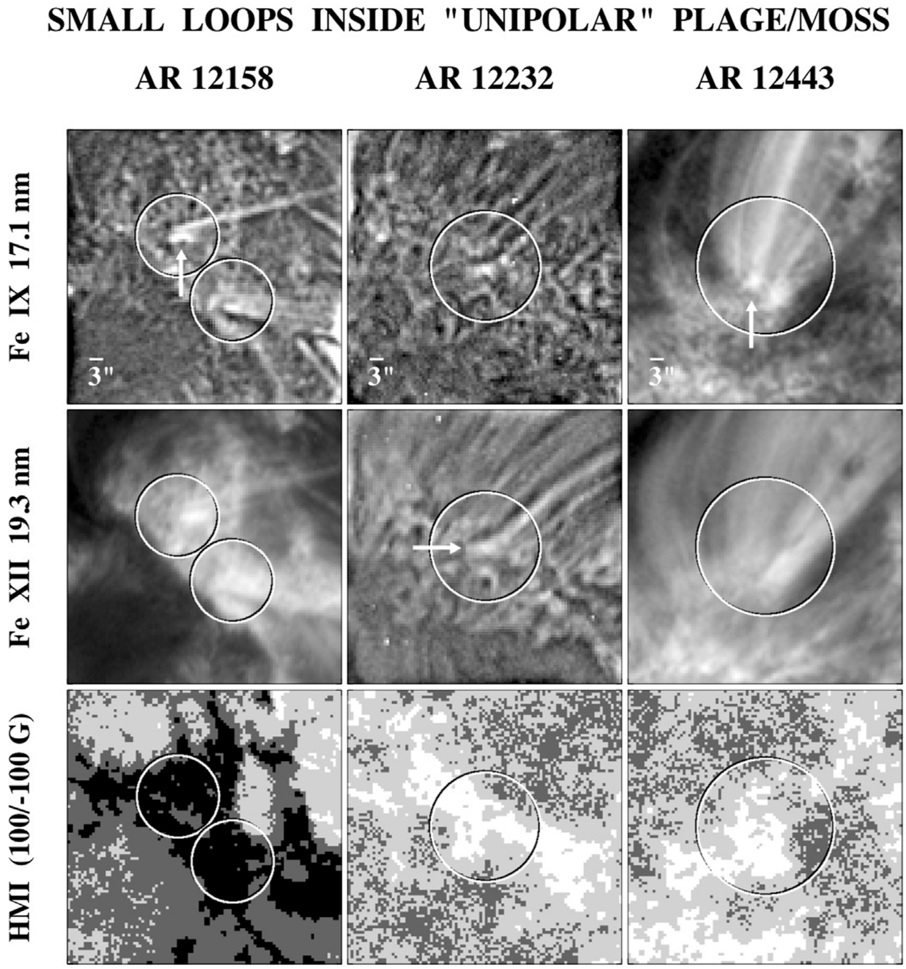

FIGURE 2 | Examples of small loops embedded in active region plages that are purely unipolar according to simultaneous magnetograms. The EUV images and magnetograms are from SDO. Gray scale for the magnetograms is as follows: white (B_{los} >100 G); light gray (0 G < B_{los} < 100 G); dark gray (–100 G < B_{los} < 0 G); black (B_{los} <–100 G). The inverted-Y structures represent jet-like outflows along long coronal loops that are reconnecting with the underlying small-scale loops. Our hypothesis is that the reticulated EUV "moss" covering the plage areas consists mainly of small loops that continually reconnect with the active region loops and are the main source of the coronal heating in active regions.

(e.g., van Ballegooijen et al., 2011), only a small fraction of the available energy leaks into the corona.

5 CONCLUDING REMARKS

Perhaps the most important thing I've learned from my astrophysical career is that it pays to look closely at the available observations, and theoretical models that are not based on a close examination of observations are likely to be wrong. An example of this precept is provided by mean field dynamo models, which have long ignored magnetograph observations showing the essential role of surface flux transport in the formation and evolution of the polar fields. I'm also puzzled as to why specialists in coronal heating and coronal loops continue to ignore the fact (or even possibility) that magnetograms do not show most of the minority-polarity flux present inside plages. As in the case of the mean field dynamo theorists, perhaps it's due to inertia ("the inherent property of a body that makes it oppose any force that would cause a change in its motion").

Conversely, I think that those who are mainly focused on observations should be familiar enough with basic physical concepts to properly (critically) relate their data to theoretical models. Physical interpretation of data doesn't necessarily require elaborate 3D MHD simulations. For example, when studying the formation and evolution of filaments/prominences, it is important to recognize that flux cancellation at a polarity inversion line acts to decrease the transverse component of the field but not the parallel component. This may lead one to be more skeptical of models where filaments are formed by strong footpoint shearing (which may be easier to include in numerical simulations than flux cancellation/submergence). In any case, observations should be used to evaluate the relative importance of flux cancellation and shearing motions around the polarity inversion line.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

REFERENCES

Aschwanden, M. J., Winebarger, A., Tsiklauri, D., and Peter, H. (2007). The Coronal Heating Paradox. *ApJ* 659, 1673–1681. doi:10.1086/513070

Babcock, H. W. (1961). The Topology of the Sun's Magnetic Field and the 22-YEAR Cycle. *ApJ* 133, 572–589. doi:10.1086/147060

Berger, T. E., De Pontieu, B., Schrijver, C. J., and Title, A. M. (1999). High-resolution Imaging of the Solar Chromosphere/Corona Transition Region. *ApJL* 519, L97–L100. doi:10.1086/312088

Charbonneau, P. (2010). Dynamo Models of the Solar Cycle. *Living Rev. Sol. Phys.* 7, 3. doi:10.12942/lrsp-2010-3

Choudhuri, A. R., Schüssler, M., and Dikpati, M. (1995). The Solar Dynamo with Meridional Circulation. *A&A* 303, L29–L32.

Dahlburg, R. B., Einaudi, G., and Taylor, B. D. (2016). Observational Signatures of Coronal Loop Heating and Cooling Driven by Footpoint Shuffling. *ApJ* 817, 47. doi:10.3847/0004-637X/817/1/47

De Pontieu, B., Tarbell, T., and Erdélyi, R. (2003). Correlations on Arcsecond Scales between Chromospheric and Transition Region Emission in Active Regions. *ApJ* 590, 502–518. doi:10.1086/374928

Hagenaar, H. J., DeRosa, M. L., and Schrijver, C. J. (2008). The Dependence of Ephemeral Region Emergence on Local Flux Imbalance. *ApJ* 678, 541–548. doi:10.1086/533497

Leighton, R. B. (1969). A Magneto-Kinematic Model of the Solar Cycle. *ApJ* 156, 1–26. doi:10.1086/149943

Parker, E. N. (1988). Nanoflares and the Solar X-Ray Corona. *ApJ* 330, 474–479. doi:10.1086/166485

Pontin, D. I., and Hornig, G. (2020). The Parker Problem: Existence of Smooth Force-free Fields and Coronal Heating. *LRSP* 17, 5. doi:10.1007/s41116-020-00026-5

Sheeley, N. R., Jr., Nash, A. G., and Wang, Y.-M. (1987). The Origin of Rigidly Rotating Magnetic Field Patterns on the Sun. *ApJ* 319, 481–502. doi:10.1086/165472

Stenflo, J. O. (1989). Differential Rotation of the Sun's Magnetic Field Pattern. *A&A* 210, 403–409.

van Ballegooijen, A. A., Asgari-Targhi, M., Cranmer, S. R., and DeLuca, E. E. (2011). Heating of the Solar Chromosphere and Corona by Alfvén Wave Turbulence. *ApJ* 736, 3. doi:10.1088/0004-637X/736/1/3

Wang, Y.-M., Sheeley, N. R., Jr., and Nash, A. G. (1991). A New Solar Cycle Model Including Meridional Circulation. *ApJ* 383, 431–442. doi:10.1086/170800

Wang, Y.-M., and Sheeley, N. R., Jr. (1994). The Rotation of Photospheric Magnetic Fields: A Random Walk Transport Model. *ApJ* 430, 399–412. doi:10.1086/174415

Wang, Y.-M., and Sheeley, N. R., Jr. (1993). Understanding the Rotation of Coronal Holes. *ApJ* 414, 916–927. doi:10.1086/173135

Wang, Y.-M., Sheeley, N. R., Jr., Walters, J. H., Brueckner, G. E., Howard, R. A., Michels, D. J., et al. (1998). Origin of Streamer Material in the Outer Corona. *ApJL* 498, L165–L168. doi:10.1086/311321

Wang, Y.-M. (2020). Small-scale Flux Emergence, Coronal Hole Heating, and Flux-Tube Expansion: A Hybrid Solar Wind Model. *ApJ* 904, 199. doi:10.3847/1538-4357/abbda6

Wang, Y.-M. (2016). The Ubiquitous Presence of Looplike Fine Structure inside Solar Active Regions. *ApJL* 820, L13. doi:10.3847/2041-8205/820/1/L13

Wang, Y.-M., Ugarte-Urra, I., and Reep, J. W. (2019). Further Evidence for Looplike Fine Structure inside "Unipolar" Active Region Plages. *ApJ* 885, 34. doi:10.3847/1538-4357/ab45f6

Wang, Y.-M., Warren, H. P., and Muglach, K. (2016). Converging Supergranular Flows and the Formation of Coronal Plumes. *ApJ* 818, 203. doi:10.3847:0004-637X/818/2/20310.3847/0004-637x/818/2/203

Withbroe, G. L., and Noyes, R. W. (1977). Mass and Energy Flow in the Solar Chromosphere and Corona. *ARA&A* 15, 363–387. doi:10.1146/annurev.aa.15.090177.002051

Withbroe, G. L., and Wang, Y.-M. (1972). A Model for the Polar Transition Layer and Corona for November 1967. *SoPh* 27, 394–401. doi:10.1007/BF00153110

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by the Office of Naval Research.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 17 June 2022
doi: 10.3389/fspas.2022.937742

Space Physics: The Need for a Wider Perspective

Mats André *

Swedish Institute of Space Physics, Uppsala, Sweden

We argue that many studies in space physics would benefit from putting a detailed investigation into a wider perspective. Three examples of theoretical and observational studies are given. We argue that space physics should aim to be less of an isolated branch of science. Rather, by putting the scientific space results into a wider perspective these results will become more interesting and important than ever. We argue that diversity in a team often is favourable for work on complicated problems and helps to present the results in a wider perspective.

Keywords: diversity, dispersion surfaces, ion energization, ionospheric outflow, low-energy ions, spacecraft charging

1 INTRODUCTION

To be useful now and in the future, space research must be performed in the greatest possible detail, and must be presented in the widest possible content. This is obvious advice for many areas of science. But space can in the short run be more spectacular than other areas of research. Thus, it is important to realize that results from your own narrow area of research in the long run should not be presented only in a detailed and limited way. A limited presentation is not enough to be useful to your fellow scientists, to society, and is not enough to attract attention and funding.

Space physics can remain as a separate science in terms of some special techniques, including launchers, spacecraft technology and scientific instruments for extreme conditions. Space science should aim to be less of an isolated academic research topic (and so should several other topics now treated as individual subjects). Space results should be presented in a way that make them as useful as possible for other basic science disciplines such as astrophysics, astrobiology and laboratory plasma physics and for applied sciences such as space weather, spacecraft interaction with the surrounding environment and thermonuclear fusion. This can then lead to applications in areas we have not even thought of today.

As a start, space research should aim at results with a wide impact within this area of research. Keeping this wider perspective during each study makes the ongoing research more interesting, and the results will be more useful to a wider community. Below I give examples of three studies where I performed detailed science, and tried to present the results in a wide perspective.

2 EXAMPLES OF STUDIES WITH A PERSPECTIVE

To qualify as a study putting results into perspective, an investigation should include new results and new understanding, should put things together in a new way, and should still be interesting 10 years after publication. The last requirement is somewhat arbitrary but has the advantage that it can be tested by checking if other scientists are using a relevant publication as a reference after several years.

Edited by:
Jorge Luis Chau,
Leibniz Institute of Atmospheric Physics (LG), Germany

Reviewed by:
Huishan Fu,
Beihang University, China

***Correspondence:**
Mats André
mats.andre@irfu.se

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *06 May 2022*
Accepted: *27 May 2022*
Published: *17 June 2022*

Citation:
André M (2022) Space Physics: The Need for a Wider Perspective. Front. Astron. Space Sci. 9:937742. doi: 10.3389/fspas.2022.937742

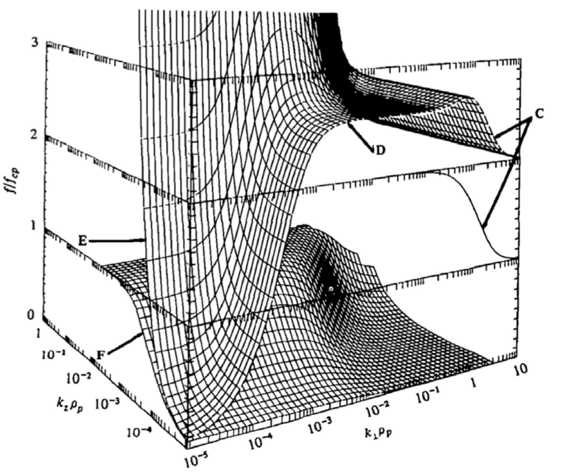

FIGURE 1 | Dispersion surfaces, wave frequency shown as a function of wavevector components. Here f and f_{cp} are the wave frequency and the proton gyrofrequency, $k_{\perp}$ and $k_{\|}$ are the wavevector components perpendicular and parallel to the background magnetic field and ρ_p is the proton gyroradius. Wave modes include: C) Ion Bernstein waves (also called Ion Cyclotron Harmonic, ICH, waves), D) Lower hybrid waves, E) Right circularly polarized whistler waves (fast magnetosonic or compressional Alfvén waves), F) Left circularly polarized shear Alfvén waves (electromagnetic ion cyclotron, EMIC, mode). From André (1985), where details of the plasma model are given.

Rather than attempting a stricter definition of studies with a perspective I give three examples of investigations I considered nice at the time, and which I am still proud of.

2.1 Dispersion Surfaces

The first example concerns the theory of plasma waves in a non-relativistic, collisionless, homogeneous and magnetized plasma. This is a good first approximation for the study of many plasmas in space physics and astrophysics. In collisionless plasmas energy is not transferred between charged particles via Coloumb collsions. Rather, the charged particles interact via electromagnetic waves. Often there is a need to identify which wave mode is observed. The relevant equations are complicated but can be found in textbooks, e.g., Chen (2016). Books and many research papers also give plots illustrating special cases such as a limited range of frequencies and wave-vectors k only parallel or perpendicular to the background magnetic field. As a student, my supervisor developed a computer code to study instabilities of waves in a homogeneous, anisotropic and magnetized plasma, WHAMP (Rönnnmark, 1982; Rönnmark, 1983), while I did a lot of testing and debugging. As part of my studies, I wanted to sort out which (not heavily damped) waves exist for a certain combination of plasma parameters.

Dispersion surfaces, plots of frequency as a function of wave vector for all directions of k (Oakes et al., 1979), are a useful way to display the numerical results. A systematic presentation of dispersion surfaces (André, 1985) was part of my thesis. **Figure 1** shows an example of surfaces, covering frequencies up to three times the proton gyrofrequency. Wave modes labeled with capital letters are often treated as individual and isolated approximations. Here it is clear that they are smoothly connected as k is gradually changed. These surfaces have turned out to be very useful and are included in overviews and books (e.g., Benz (2002); Koskinen (2011)) and recent PhD theses (e.g., Allison (2019)) The point is not so much an individual wave mode represented by a small part of a surface and well described by an analytical approximation in a textbook. The point is more the overall picture of all possible wave modes, showing which wave modes should be considered. Putting the numerical solutions together in this way gives a new and useful perspective.

2.2 Ion Energization

The second example concerns the energization of ions leaving the ionosphere. A large fraction of the plasma in the magnetosphere originate in the ionosphere, with initial energies of less than one electron volt. Ions such as H^+ and O^+ leaving the auroral regions can then be energized to several keV. In addition to upward acceleration of positive ions parallel to the geomagnetic field by the potential drops accelerating electrons downward to cause auroras, ions are often energized in the perpendicular direction. When the ions have left the lower ionosphere dominated by collisions, this heating can be caused by various electromagnetic waves, ultimately powered by energy from the solar wind. These ions may then move adiabatically up the field lines of the inhomogeneous terrestrial magnetic field and form so-called conics in ion velocity space. There are many observations by sounding rockets and satellites of ion conics and associated waves from altitudes of a few hundred km out to several Earth radii. Possible perpendicular energization mechanisms range from nearly static electric field structures, waves below and around the ion gyrofrequency, and waves near the lower hybrid frequency. Many studies have tried to identify the important mechanism. It was clear that more than one process may be acting, but at least for specific regions in space the search was usually for the (only) mechanism.

Having the opportunity to look at a lot of data from the Swedish Freja satellite launched 1992 to investigate the auroral region, I realized that another question would be more interesting. What is the relative importance of different mechanisms (wave modes)? Together with my PhD student Patrik Norqvist and other colleagues, we performed a statistical study using Freja observations close to the apogee of 1700 km, including more than 200 events of simultaneously observed ion conics and waves (André et al., 1998). We found that O^+ energization was mainly caused by broadband waves, where the frequencies around the gyrofrequency were relevant for resonant heating. Sometimes waves around half the proton gyryfrequency (EMIC emissions) or waves around the lower hybrid frequency were more important. We used test particle calculations to verify that the observed wave amplitudes were high enough to explain the observed ion energies. We also considered and disregarded other possible mechanisms of perpendicular ion energization. We found that many previous studies were correct in that they had identified a plausible mechanism for a specific event. Our study included many events and showed that broadband waves overall cause most of the perpendicular ion energization in the auroral region, while other waves sometimes can be important. Considering several wave heating mechanisms in one single study gave a new and useful perspective.

2.3 Outflow of Ionospheric Low-Energy Ions

The third example concerns ions leaving the ionosphere without being much energized. Plasma in the magnetosphere can originate in either the solar wind or the ionosphere, and during decades it gradually became clear that the ionosphere can be an important source also at high altitudes (Chappell et al., 1980; Chappell, 2015). A problem in large regions of the magnetopshere is that a spacecraft in sunlight in a low-density plasma will be positively charged to tens of volts due to the emission of photoelectrons. Hence, positive ions from the ionosphere with eV energies will not reach the spacecraft. These ions often have a drift energy that is higher than the velocity distributions thermal energy. The upward flowing ions will be scattered by a large electrostatic structure around the charged satellite. This will cause an enhanced ion wake behind the spacecraft. For similar electron and ion temperatures the low mass electrons have much higher thermal velocities and immediately fill this wake, causing a region behind the spacecraft with an excess of electrons. The result is a local electric field caused by the combination of drifting low-energy ionospheric ions and the charged spacecraft. As PI with main responsibility for the Electric Field and Wave instruments on the four ESA Cluster satellites launched 2000, I first considered this local field a problem. The EFW instrument includes two pairs of probes on wire booms in the spin plane of each satellite. Each pair has a probe-to-probe separation of 88 m and the electric field is obtained from the potential difference between the probes.

Once we understood that the ion wake is the cause of the local electric field, Anders Eriksson and his PhD student Erik Engwall started to develop a method to use this local field to estimate the flux of outflowing ions (Engwall et al., 2006a; Engwall et al., 2006b; Engwall et al., 2009b). The trick is to use also the Electron Drift Instrument on each spacecraft, measuring the drift of artificially emitted electrons as they gyrate back to the spacecraft under the influence of the geophysical magnetic field measured by the FluxGate Magnetometer FGM. These electrons with an energy of about a keV have large gyroradii and are mot much affected by the local wake. Drifting low-energy ions can then be inferred by detecting a wake electric field, obtained as a large enough difference between the quasi-static electric fields obtained by the EFW (total electric field) and EDI (geophysical electric field) instruments. The direction of the wake gives the direction of the ion outflow. Since the perpendicular E × B drift velocity is known (from EDI and FGM) the parallel velocity can be inferred (Engwall et al., 2009a; André et al., 2015). The density can be obtained by calibrating observations of the spacecraft potential, in practise obtained as the potential difference between the EFW probes (nearly at the plasma potential) and the spacecraft (Lybekk et al., 2012; Haaland et al., 2017). The ion flux can then be

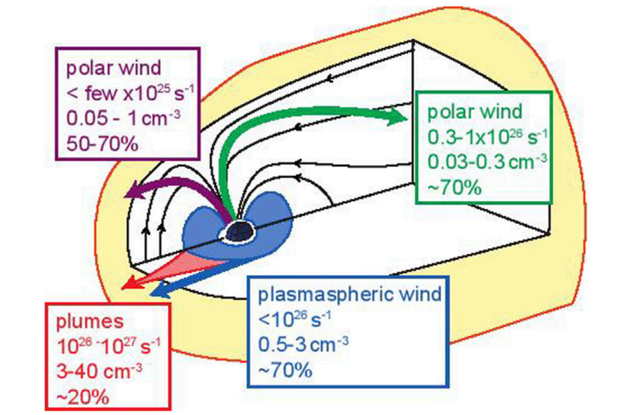

FIGURE 2 | Overview of ion outflow from Earth. Regions dominated by low-energy (eV) plasma of ionospheric origin are indicated. Order of magnitude outflow rates, densities and percentage of time low-energy ions dominate the density are given. From André and Cully (2012).

obtained from the drift velocity and the density. The first important perspective in this example is to seriously try to understand observations which at first seem peculiar or just wrong. In this case two instruments observing the same parameter, the electric field, gave different results. Rather than deciding that at least one observation must be wrong, understanding of the situation shows that both are correct, one showing the local field around the charged spacecraft (EFW) and the other the geophysical field present in large regions (EDI). This wider perspective then gives the possibility to estimate a parameter thought not to be possible to measure, the flux of ionospheric low-energy ions.

The Cluster wake method to estimate the flux of outflowing ionospheric ions has then been used to improve the map of plasma in the magnetosphere. Together with Chris Cully I made a study resulting in the overview of low-energy ions in **Figure 2** (André and Cully, 2012). The night-side polar lobe results are from the Engwall et al. (2009a) investigation using the wake method. For the dayside we used this method together with other estimates. Sometimes ions with low thermal velocity have a large enough E × B drift to be detected by an onboard ion instrument also on a charged spacecraft. We also compared total density obtained from wave observations with particle instrument observations, to estimate how much of the ion population was low-energy and hidden from direct detection. This overall picture has been confirmed by a larger statistical study using 10 years of data and the wake method in the polar lobes (André et al., 2015). Also, this method has been validated in different ways such as using a similar method in the solar wind and comparing with data from particle instruments at lower altitude where spacecraft charging is less of a problem (André et al., 2021b,a). Comparing with a review of many observations, the overview in **Figure 2** is still useful (Toledo-Redondo et al., 2021). Previously, many of the low-energy ions of ionospheric origin could not be detected. Putting observations by several instruments and methods together in an overview such as the one in **Figure 2** gives a new and useful perspective on the magnetosphere.

For all three examples above, some theory or data have been added to previous knowledge. But the most important part is the new overall picture. How different wave modes with different names are related, concluding that several waves can energize ions in the auroral region but one type is typically more important, and realizing that low energy ions of ionospheric origin are not just sporadically detected at high altitudes but are a very common component of the magnetospheric plasma.

3 THE SCIENTIST WITH A PERSPECTIVE

The successful space scientist is intelligent and ambitious, in some wide meaning of these words. But also the surrounding family, society and culture have significant impact on who becomes a great scientist (Gladwell, 2009). Extrovert entrepreneurs are often sought after to lead science projects, while it is clear that introvert and quiet people can be equally good scientists and leaders (Cain, 2013). For complicated (science) problems that takes time to solve, it seems that a group of people with mixed skills and backgrounds is often successful. Different skills can include, for example, expertise in instrument design, data analysis and numerical simulations. Members of the group can differ, for example, concerning gender, ethnic background and age. Diversity in a team is not without problems but I think the positive aspects dominate (Carter and Phillips, 2017; Stangor, 2017; Peters, 2021). This is one reason to achieve roughly equal numbers of female and male scientists at all academic ranks (Coe et al., 2019; Popp et al., 2019). A diverse team is more likely to keep a wider perspective.

4 CONCLUSION

The work on a complicated problem usually becomes more interesting if a wider perspective can be kept during the work. The result becomes more useful to fellow scientists and to society when it is presented in a wider perspective. Diversity in the team can help to keep this wider perspective.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

MA was supported by the Swedish National Space Agency contract SNSA 2020-00058.

ACKNOWLEDGMENTS

Useful discussions concerning the main ideas of this article with Susanne Höfner are acknowledged. Helpful discussions concerning references and related questions with Gabriella Stenberg Wieser and Kristina Andréen are acknowledged.

REFERENCES

Allison, H. J. (2019). *Understanding How Low Energy Electrons Control the Variability of the Earth's Electron Radiation Belts* (UK: University of Cambridge). Ph.D. thesis.

André, M., and Cully, C. M. (2012). Low-energy Ions: A Previously Hidden Solar System Particle Population. *Geophys. Res. Lett.* 39, a–n. doi:10.1029/2011GL050242

André, M. (1985). Dispersion Surfaces. *J. Plasma Phys.* 33, 1–19. doi:10.1017/S0022377800002270

André, M., Eriksson, A. I., Khotyaintsev, Y. V., and Toledo-Redondo, S. (2021a). The Spacecraft Wake: Interference with Electric Field Observations and a Possibility to Detect Cold Ions. *JGR Space Phys.* 126, e29493. doi:10.1029/2021JA029493

André, M., Li, K., and Eriksson, A. I. (2015). Outflow of Low-Energy Ions and the Solar Cycle. *J. Geophys. Res. Space Phys.* 120, 1072–1085. doi:10.1002/2014JA020714

André, M., Norqvist, P., Andersson, L., Eliasson, L., Eriksson, A. I., Blomberg, L., et al. (1998). Ion Energization Mechanisms at 1700 Km in the Auroral Region. *J. Geophys. Res.* 103, 4199–4222. doi:10.1029/97JA00855

André, M., Toledo-Redondo, S., and Yau, A. W. (2021b). "Cold Ionospheric Ions in the Magnetosphere," in *Magnetospheres in the Solar System, Geophysical Monograph 259*. Editors R. Maggiolo, N. André, H. Hasegawa, and D. T. Welling (Washington, DC: John Wiley & Sons), 219–228. doi:10.1002/978111950751210.1002/9781119815624.ch15

Benz, A. (2002). *Plasma Astrophysics, Kinetic Processes in Solar and Stellar Coronae*. second edition, 279. New York: Kluwer Academic Publishers. doi:10.1007/978-0-306-47719-5

Cain, S. (2013). *Quiet: The Power of Introverts in a World that Can't Stop Talking*. New York: Random House. isbn: 978-0-307-35215-6.

Carter, A. B., and Phillips, K. W. (2017). The Double-Edged Sword of Diversity: Toward a Dual Pathway Model. *Soc. Personal. Psychol. Compass* 11, e12313. doi:10.1111/spc3.12313

Chappell, C. R., Baugher, C. R., and Horwitz, J. L. (1980). New Advances in Thermal Plasma Research. *Rev. Geophys.* 18, 853–861. doi:10.1029/RG018i004p00853

Chappell, C. R. (2015). The Role of the Ionosphere in Providing Plasma to the Terrestrial Magnetosphere - an Historical Overview. *Space Sci. Rev.* 192, 5–25. doi:10.1007/s11214-015-0168-5

Chen, F. F. (2016). *Introduction to Plasma Physics and Controlled Fusion*. New York: Springer. doi:10.1007/978-3-319-22309-4

Coe, I. R., Wiley, R., and Bekker, L.-G. (2019). Organisational Best Practices towards Gender Equality in Science and Medicine. *Lancet* 393, 587–593. doi:10.1016/S0140-6736(18)33188-X

Engwall, E., Eriksson, A. I., André, M., Dandouras, I., Paschmann, G., Quinn, J., et al. (2006a). Low-energy (Order 10 eV) Ion Flow in the Magnetotail Lobes Inferred from Spacecraft Wake Observations. *Geophys. Res. Lett.* 33, L06110. doi:10.1029/2005GL025179

Engwall, E., Eriksson, A. I., Cully, C. M., André, M., Puhl-Quinn, P. A., Vaith, H., et al. (2009a). Survey of Cold Ionospheric Outflows in the Magnetotail. *Ann. Geophys.* 27, 3185–3201. doi:10.5194/angeo-27-3185-2009

Engwall, E., Eriksson, A. I., Cully, C. M., André, M., Torbert, R., and Vaith, H. (2009b). Earth's Ionospheric Outflow Dominated by Hidden Cold Plasma. *Nat. Geosci.* 2, 24–27. doi:10.1038/ngeo387

Engwall, E., Eriksson, A. I., and Forest, J. (2006b). Wake Formation behind Positively Charged Spacecraft in Flowing Tenuous Plasmas. *Phys. Plasmas* 13, 062904062904. doi:10.1063/1.2199207

Gladwell, M. (2009). *Outliers: The Story of Success*. New York: Penguin Books. isbn: 978-0-141-04302-9.

Haaland, S., Lybekk, B., Maes, L., Laundal, K., Pedersen, A., Tenfjord, P., et al. (2017). North-south Asymmetries in Cold Plasma Density in the Magnetotail Lobes: Cluster Observations. *J. Geophys. Res.Space Phys.* 122, 136–149. doi:10.1002/2016JA023404

Koskinen, H. E. J. (2011). *Physics of Space Storms: From the Solar Surface to the Earth*. New York: Springer. doi:10.1007/978-3-642-00319-6

Lybekk, B., Pedersen, A., Haaland, S., Svenes, K., Fazakerley, A. N., Masson, A., et al. (2012). Solar Cycle Variations of the Cluster Spacecraft Potential and its Use for Electron Density Estimations. *J. Geophys. Res. (Space Phys.* 117, A01217. doi:10.1029/2011JA016969

Oakes, M. E., Michie, R. B., Tsui, K. H., and Copeland, J. E. (1979). Cold Plasma Dispersion Surfaces. *J. Plasma Phys.* 21, 205–224. doi:10.1017/S0022377800021784

Peters, U. (2021). Hidden Figures: Epistemic Costs and Benefits of Detecting (Invisible) Diversity in Science. *Eur. J. Philosophy Sci.* 11, 33. doi:10.1007/s13194-021-00349-6

Popp, A. L., Lutz, S. R., Khatami, S., van Emmerik, T. H. M., and Knoben, W. J. M. (2019). A Global Survey on the Perceptions and Impacts of Gender Inequality in the Earth and Space Sciences. *Earth Space Sci.* 6, 1460–1468. doi:10.1029/2019EA000706

Rönnmark, K. (1983). Research Note: Computation of the Dielectric Tensor of a Maxwellian Plasma. *Plasma Phys.* 25, 699–701. doi:10.1088/0032-1028/25/6/007

Rönnnmark, K. (1982). *WHAMP - Waves in Homogeneous, Anisotropic, Multicomponent Plasmas*, 179. Kiruna: Kiruna Geophysical Institute. Tech. Rep.

Stangor, C. (2017). *Group Processes*. Oxford: Oxford University Press. doi:10.1093/acrefore/9780190236557.013.255

Toledo-Redondo, S., André, M., Aunai, N., Chappell, C. R., Dargent, J., Fuselier, S. A., et al. (2021). Impacts of Ionospheric Ions on Magnetic Reconnection and Earth's Magnetosphere Dynamics. *Rev. Geophys.* 59, e2020RG000707. n/aE2020RG000707 2020RG000707. doi:10.1029/2020RG000707

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 21 June 2022
doi: 10.3389/fspas.2022.952000

Space Physics Career in a Developing Country: Opportunities and Challenges

J. Américo González-Esparza*

Laboratorio Nacional de Clima Espacial, Instituto de Geofísica, Universidad Nacional Autónoma de México, Morelia, Mexico

The development of an infrastructure of observational instruments is a great challenge for any young scientist especially working in a country where the infrastructure addresses a new field of scientific knowledge and the funds are limited. However, I argue that although it can be questionable in terms of all the difficulties that you might face, the observational infrastructure is a crucial aspect of building a local scientific community. Therefore, it should be pursued as a national priority.

Keywords: space weather, observational networks, solar storms, real-time data, radiotelescope

OPEN ACCESS

Edited by:
Joseph E. Borovsky, Space Science Institute, United States

Reviewed by:
David Gary Sibeck, National Aeronautics and Space Administration, United States
Daniel Okoh, National Space Research and Development Agency, Nigeria

****Correspondence:***
J. Américo González-Esparza
americo@igeofisica.unam.mx

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *24 May 2022*
Accepted: *30 May 2022*
Published: *21 June 2022*

Citation:
González-Esparza JA (2022) Space Physics Career in a Developing Country: Opportunities and Challenges. Front. Astron. Space Sci. 9:952000. doi: 10.3389/fspas.2022.952000

1 INTRODUCTION

I am from Mexico. My country is located in North America and shares a border with the United States of about 3,152 km. There are significant differences concerning wealth and development between the two countries. Although that Mexico has the 15th largest economy in terms of its Gross Domestic Product (GDP), there are many aspects where the country requires more progress. One of them is its investment in science and technology. Whereas the US has a GDP some 20 times larger than that of Mexico, NASA has a budget that is about six orders of magnitude greater than the Mexican space agency. Therefore, to advance a space program, Mexico requires a significant increment in funding by the government.

I did my bachelor's science degree at the National Autonomous University of Mexico (UNAM in Spanish) with a thesis on interplanetary shock waves (1985–1991). My advisor was Dr. Silvia Bravo. Then I was fortunate to obtain a scholarship for my Ph.D. at Imperial College, University of London (1991–1995). My doctoral thesis examined magnetic field data from Ulysses, the most exciting heliospheric mission at that moment. Ulysses was the first spacecraft measure the out-of-ecliptic solar wind *in situ*. My focus was on the analysis of the heliospheric shock waves detected during the first part of the mission (1990–1994) (Balogh et al., 1995). My adviser was Professor André Balogh. When I finished my doctorate, I got a postdoctoral fellowship at the Jet Propulsion Laboratory NASA-CalTech (1995–1997), where I continued my research on interplanetary shocks. My adviser was Dr. Ed Smith, and I also had the opportunity to collaborate with Dr. Marcia Neugebauer, the head of the group at that moment (González-Esparza et al., 1996). I had accomplished a good early career as a young scientist in heliospheric physics, but eventually I had to decide whether I return to my country or continue working abroad. I was happy and proud to work at JPL-NASA, but I missed my family and culture. I wanted to proceed with a career as a professional scientist, and, in that respect, staying in the United States would probably be one of the best options. However, the essential thing in my decision was that I had received a Ph.D. grant from Mexico, and I had the moral commitment to return to my alma mater.

In 1997, I returned to Mexico and incorporated as a researcher at UNAM. Only one group in Mexico studied space physics. It consisted of six members each of whom pursued his or her area

FIGURE 1 | Site of the Mexican Array Radio Telescope (MEXART), main observatory of the Space Weather National Laboratory (LANCE acronym in Spanish).

of research within space physics. Unfortunately, there were personal issues within the working group (a common problem in the academic environment everywhere), and the working atmosphere was not the best when I arrived. I began collaborating with my former advisor. Dr. Silvia Bravo was pursuing a project to build an Interplanetary Scintillation (IPS) array in Mexico. The idea was to construct a radio telescope similar to that of Dr. Antony Hughes's group during the late sixties-seventies in Cambridge, United Kingdom. This group discovered the IPS phenomenon. By using IPS observations from a ground-based radiotelescope, it is possible to infer some solar wind properties (speed and density fluctuations). The IPS technique provides a source of solar wind information, and it becomes a good option when no other instruments are available.

Dr. Silvia Bravo passed away due to cancer in 2000. From that moment, I had the full responsibility to accomplish the project. This IPS project was difficult for us in many respects: I had a poor expertise in managing scientific projects, we lacked an experienced technical workgroup in radio astronomy, and the funding was partial and limited. Learning basic radioastronomy, the principles of the instrument, and struggling to get support and funding in Mexico, did not seem very promising for a good career in Frontier science in space physics at that moment. Rather the straightforward strategy would be to maintain my research collaborations with the Ulysses' scientific team. How many years would we need to invest in getting some valuable IPS scientific data? If I knew it would take us 20 years to do it, I might have taken the former straightforward option. Fortunately, I did not know.

There was also a national pride in the project; we would have our solar wind data. However, the "establishment" of the scientific community in Mexico did not necessarily agree with this idea of developing an observational infrastructure to study solar wind parameters. The "mainstream" philosophy was that the scientific research at UNAM should be based on state-of-the-art international instruments, whose data had become more accessible via internet and the open-access policies for data of international instruments including the important NASA missions. With lack of the human resources and technical expertise, did it make sense to spend time, funding, and effort developing our incipient instrument?

In Mexico, the academic production of a researcher was evaluated mainly by his or her participation in a number of published papers in the indexed journals. The university pays the basic salary, but a substantial part of the income comes from "stimuli" by the government. The "stimuli" are rated and evaluated every 3 years by academic commissions. These evaluations also count in teaching and popularizing science activities, but the main factor is number of articles published. The development of scientific infrastructure counted as well, but as a secondary activity. These ratings are questionable in terms of the country's priorities, and they have been reviewed recently. Developing the country's observational infrastructure should be one of the most significant achievements.

We began construction of the Mexican Array RadioTelescope (MEXART) in the state of Michoacan (about 400 km from Mexico City) (**Figure 1**) (González-Esparza et al., 2004). It was a difficult beginning. We made several technical errors that we could have avoided with more knowledge and experience. However, this was the learning path. The partial and limited funding obliged us to work dividing the project in stages and extended the duration of the array's construction for several years. It was not easy to get technical assistance from the United States without funding, so we had to look for collaborative help from elsewhere. We are very grateful to Professor Shri

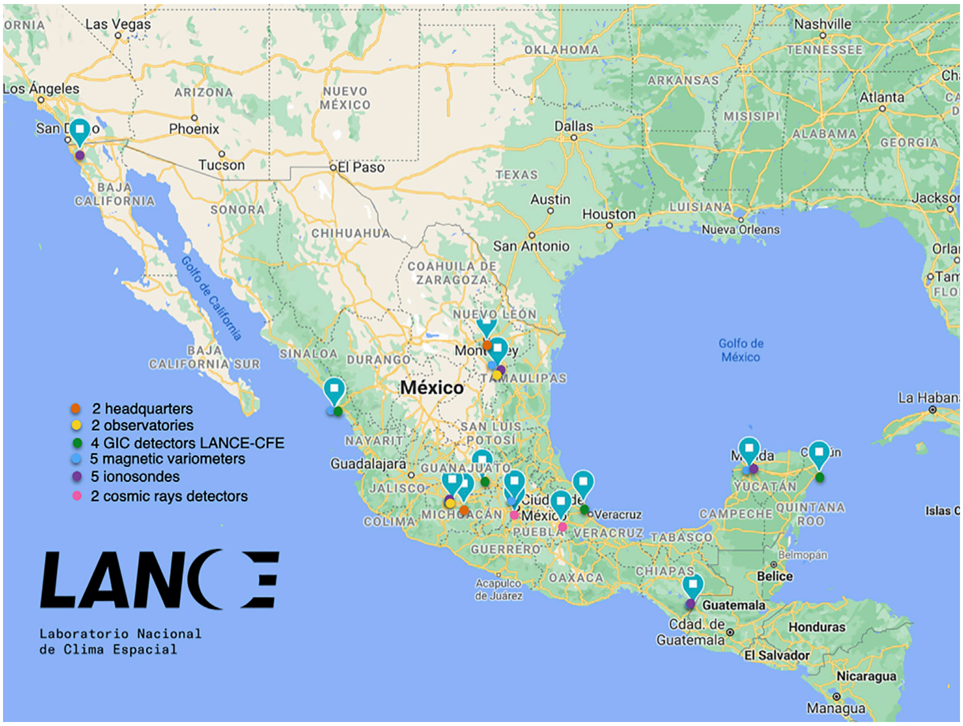

FIGURE 2 | LANCE Space Weather observational network.

Ananthakrishnan and the engineers from his working group sent to Mexico from India's National Centre for Radio Astrophysics, Tata Institute of Fundamental Research. Their expertise in radioastronomy techniques, and their generous support, were crucial for the beginning of the project. We also established a collaboration with the Institute of Geophysics and Astronomy of Cuba whose engineers assisted in construction and testing the array. One of the lessons learned was that we could get assistance from the international community if we looked and asked for it.

Soon it became clear that I could not manage the MEXART project remotely from Mexico City. Therefore, in 2006 my family and I moved to live near the site. In addition, I initiated a new life on a new UNAM campus in Morelia to form a new space physics group.

In 2014 the legislators modified the General Protection Law in Mexico and included Space Weather events in the list of natural hazards that the National Civil Protection System must attend. At that moment, we saw an opportunity for our group in Morelia to grow and begin a new research area, assisting the authorities in accomplishing the law's mandate concerning solar and geomagnetic storms. We created the Mexican Space Weather Service (SCIESMEX acronym in Spanish) at the Institute of Geophysics-UNAM. We began to interact with the Civil Protection System providing an early warning of Space Weather events (Gonzalez-Esparza et al., 2017).

The experience and expertise we got from the construction of MEXART would let us propose a project aiming to develop a set of new observational networks that would include solar, interplanetary, geomagnetic, ionospheric, and cosmic ray phenomena. These networks of instruments are to provide the regional data on Space Weather conditions in Mexico (**Figure 2**). Considering the urgent need for these data and its analysis, we established in 2016 the National Space Weather Laboratory (LANCE by its acronym in Spanish). One of the main issues that we claimed in our proposal was that Mexico did not have an observational infrastructure to study regional effects of Space Weather. We lacked data on geomagnetic field and ionospheric phenomena; nor did we have data to analyze the national electric system's vulnerability to geomagnetic storms. In addition, there are no historical records to study benchmarks of Space Weather phenomena in Mexico. We saw in this problem an opportunity and began to submit projects to our research council to cover this lack of essential infrastructure.

It has become an increasing necessity for the scientific community to justify funding from public funds in recent years. In addition, governments and legislators are demanding that the scientific community be involved in national priorities. This subject can be controversial within our community. We must defend the importance of fundamental research. However, the Space Weather investigations provide an opportunity to work on fundamental problems in space physics and several applications for society. The recent broad interest in the Space Weather by the international community, not only academia but also the governments and industry, triggered opportunities for the space physics community. This is excellent news in the countries where there is still a lack of Space Weather studies. The state entities in different countries realize the importance of these investigations and might open support possibilities. In addition, the promotion of collaborative efforts in Space Weather by United Nations organizations such as UNOOSA, WMO, ICAO, UNDRR, Etc., have bridged excellent opportunities to obtain local funding.

Based on my experience, I would give the following advice for the successful development the scientific instrument infrastructure in the countries where it is still lacking.

1.1 Opportunities

- Each country needs to develop strategic areas in scientific research.
- The governments need to support the establishment of observatories and instrumental networks to obtain regional information and manage its data. It is about national sovereignty.
- Space Weather needs provide an excellent opportunity to begin a project involving basic scientific research and influence areas of national security. It should be emphasized that the subject necessarily requires the development of observational facilities and interdisciplinary working groups.
- Look for international assistance. People in science are united by scientific interest. It is a community in which one is usually open to shaking hands with another. Researchers make strong friendships with each other, even between researchers from different countries.

In order to build up a working group, it is important to consider the following issues:

- Learn about academic and professional management.
- Understand human relationships, challenges of separating personal and professional aspects.
- Provide funding and infrastructure to form a working group.
- Establish a work atmosphere based on respect and support for the individual projects of the group's researchers. Again, personal respect is the key.
- Protect students from personal issues that their supervisor has with other colleagues. His/her students should not inherit the supervisor's emotional problems.
- Ensure that the students have financial support and a good work environment.
- Avoid socially confusing situations. To provide an example, let us imagine a student who presents her research at an internal academic seminar, but there is an awkward atmosphere in the room. The student does not understand anything, and she believes that her study failed and that her classmates are not interested in the subject. The student does not know at the time that the problem is actually between her supervisor have with other senior peers. She is getting a bad reception for things unrelated to her investigation. Unfortunately, in many cases, it takes a few years for the student to understand what the problem was at the time.

2 FINAL REMARKS. LEGACY

In retrospect, I wonder if the decision that I took 25 years ago to focus my scientific career in Mexico on developing a Space Weather observational network was correct? Now, I can answer myself that yes, it was right. I have had the privilege to lead an enthusiastic and professional working group. The group learned to do collaborative and multidisciplinary work. During these years, we built the confidence and self-esteem that comes with hands-on experience and solving problems. We are assembling a comprehensive observational infrastructure covering solar, interplanetary, geomagnetic, ionospheric, and cosmic rays. I feel deep respect and gratitude for all my LANCE colleagues. The data that we will be able to provide in a few years will be crucial to understanding the country's vulnerability toward Space Weather hazards. Keep in mind that global Space Weather parameters, characteristics, or indices may not replace Space Weather regional conditions measurements. These data sets are also important in terms of national sovereignty. This infrastructure and the historical records will provide huge opportunities to future generations for a deeper understanding of Solar impact on the Earth's environment.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

Author acknowledges partial financial support from CONACyT-AEM Grant 2017-01-292684 and LN 315829.

ACKNOWLEDGMENTS

Looking back my student and early research carrier, I owe so much to Silvia Bravo (1945–2000) at UNAM, André Balogh at the Imperial College, and Ed Smith at the Jet Propulsion Laboratory. I owe also to all my colleagues of the Laboratorio Nacional de Clima Espacial at UNAM.

REFERENCES

Balogh, A., Gonzalez-Esparza, J. A., Forsyth, R. J., Burton, M. E., Goldstein, B. E., Smith, E. J., et al. (1995). Interplanetary Shock Waves: Ulysses Observations in and Out of the Ecliptic Plane. *Space Sci. Rev.* 72, 171–180. doi:10.1007/BF00768774

González-Esparza, J. A., Balogh, A., Forsyth, R. J., Neugebauer, M., Smith, E. J., and Phillips, J. L. (1996). Interplanetary Shock Waves and Large-Scale Structures: Ulysses' Observations in and Out of the Ecliptic Plane. *J. Geophys. Res.* 101, 17057–17071. doi:10.1029/96JA00685

González-Esparza, J. A., Carrillo, A., Andrade, E., Enríquez-Pérez, R., and Kurtz, S. (2004). The Mexart Interplanetary Scintillation Array in mexico. *Geofísica Int.* 43, 61–73. doi:10.22201/igeof.00167169p.2004.43.1.215

Gonzalez-Esparza, J. A., De la Luz, V., Corona-Romero, P., Mejia-Ambriz, J. C., Gonzalez, L. X., Sergeeva, M. A., et al. (2017). Mexican Space Weather Service (Sciesmex). *Space Weather* 15, 3–11. doi:10.1002/2016SW001496

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 08 July 2022
doi: 10.3389/fspas.2022.956062

Perspective Article for the Research Topic "Generation-to-Generation Communications in Space Physics" Find Fun and Joy in what You Do

C. Robert Clauer *

Center for Space Science and Engineering Research, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Reflections on career choices and career highlights, including scientific discovery, innovation, and friendships. Some advice to younger generations is offered.

Keywords: magnetosphere (magnetosphere-ionosphere interactions; plasma convection), ionosphere (2487 wave propagation), polar instrumentation, magnetospheric su, magnetic storm and substorm

OPEN ACCESS

Edited by:
Joseph E Borovsky, Space Science Institute, United States

Reviewed by:
Gordon Rostoker, University of Alberta, Canada

***Correspondence:**
C. Robert Clauer
rclauer@vt.edu

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *29 May 2022*
Accepted: *06 June 2022*
Published: *08 July 2022*

Citation:
Clauer CR (2022) Perspective Article for the Research Topic "Generation-to-Generation Communications in Space Physics" Find Fun and Joy in what You Do.
Front. Astron. Space Sci. 9:956062.
doi: 10.3389/fspas.2022.956062

1 WHAT PATH TO TAKE?

As I think back, some of the first and best advice that I received was from my high school self as I was thinking about what I would do with my life. The advice was to pick something for a vocation that you really like, because you will be spending a large part of your life doing that. Along the same lines I remember talking only a few years ago with my colleague and good friend Eigil Friis-Christensen about our work and he said something like. ". . . .of course, why would you work on something if it was not fun". Little did my high school self foresee my fortunate choices and resulting career in space physics. My path was not a planned route. Opportunities that were interesting presented themselves and I followed. I have enjoyed it and it has been fun—mostly. Of course, all paths are fraught with adversity somewhere along the way, but the wonderful friendships, fun and excitement in discovery have outweighed all of the difficulties.

As an undergraduate in physics, I was too distracted by the math that I did not really comprehend the physics as well as I should, I now believe in looking back. I was focused on learning equations rather than principles. I am not good at memorization, yet I tried to memorize equations for the types of problems, rather than building an equation to specify the conditions of the physics problem under consideration. It was not until graduate school that I matured enough to begin to learn the physics, and I believe that it was due to the style of teaching that I encountered at UCLA that enabled me to do this. Ferd Coroniti, for example would present problems, then work out an estimate of the details based on "back of the envelope" calculations. Each of these exercises presented the essential physics without obscuring it with complex equations. Paul Coleman's course "Coleman Club" generally consisted of having each student go to the board and then try to work out a problem that he presented. His Socratic questioning and insights from the other students in the class were foundational in developing a deeper understanding of the physics of the systems under consideration. I enjoyed space science enormously because it was essentially 19^{th} century physics (mostly electricity and magnetism) applied to new and interesting environments. Certainly there was the need for relativistic physics for some of the high energy environments, but it was not required for most of the basic magnetospheric physics that I was learning. My advisor, Bob McPherron was perhaps the most influential person in my education. His clear presentations of physics developed my intuitive understanding and then with time, my math abilities followed -- at least sufficient math ability to have a successful career as an experimental space physicist and data analyst—and to pass the exams. I also developed an appreciation for the breadth of science and it is my opinion that everyone has specific talents that they can apply to advance science. I am not a mathematical or theoretical genius, but I do have skills that can advance our understanding of magnetospheric physics, as do the many students that I have had the pleasure to mentor. Science is large and everyone can have a place and enjoy the excitement of scientific discovery.

At UCLA we had to pass three oral exams. The first was the field oral where a committee essentially asked questions until they found something that you did not know, then watched you try to figure it out. It was Coleman Club on steroids. That was the main hurdle. Following that was the qualifying exam where you presented your thesis topic and some preliminary work as well as a plan to complete the work, and then finally the dissertation defense. My preparation for the field oral was aided by working with other students who were also preparing, Howard Singer in particular. I also owe a great deal of appreciation to Chris Russell who generally came back to the lab after dinner to work late into the evening. He would often come down to where we were studying and begin to quiz us at the board with various problems. It was extremely good practice and also, for me, another learning experience where I developed a deeper understanding of the physics of the magnetosphere. Chris was a great mentor, and I have observed that his deep insight usually brought him to the successful side of resolving the various controversies that abounded in magnetospheric physics at the time. I also have to mention Ray Walker who also became a good friend and mentor.

It was an exciting time to be at UCLA during 1970—1980. Most of the discoveries (solar wind, radiation belts, plasmasphere, magnetotail, etc.,) had been made, and now was the period of trying to understand the dynamics of the systems. Several fundamental concepts were confirmed while I was a student and it was tremendously exciting to be at the cutting edge of this knowledge development. There is also joy in observing something or recognizing for the first time, “Oh, this is something new. I wonder what that is all about?” Of course, the magnetospheric substorm and substorm growth phase was central to being McPherron’s student and being at UCLA. It was there that I observed that scientific argument does not always remain objective, but can become personal. That is a sad and unfortunate aspect of human behavior. It does not need to be this way, and I observed quite the opposite at the University of Michigan where different opinions on ring current development were appreciated and it was a happy and exciting effort to explore and determine the solution that best fit the observations. This was also the case for the development of the BATS-R-US space plasma simulation code which was a newly developed adaptive MHD code built from the ground up to work on massively parallel computers. BATS-R-US was the new kid on the block, and there were many challenges and criticisms leveled at the code. Each was taken seriously and explored and the results reported at the next scientific meeting. The result of this scrutiny is that the code is now considered one of the best and most flexible codes in the community.

2 THE PATH CHOSEN AND THE FUN FOLLOWS

It was a very exciting time to be at UCLA during the seventies. Large new ideas were developing through the close cooperative activities between great experimentalists like Bob McPherron and Chris Russell and theoreticians like Margaret Kivelson, Ferd Coroniti, and George Siscoe. In addition, the list of people who visited UCLA for extended periods was also impressive. A tight bond seemed to develop between UCLA and Imperial College and Jim Dungey and David Southwood were frequent visitors. The idea of magnetic merging between the interplanetary magnetic field and magnetosphere was being examined now closely since Dungey’s 1961 paper described the cycle of energy flow. McPherron’s growth phase ideas were developing along these lines with the southward turning of the interplanetary magnetic field (IMF) leading to dayside reconnection and the transport of flux to the tail and then the sudden release of the accumulated magnetic energy in the stretched field of the tail lobes during the expansion phase of the substorm initiated by a new near-Earth magnetic merging site. It was particularly exciting when an event study was completed that showed an inbound satellite on the dayside during a time when the IMF turned southward and the magnetopause was observed over and over as the satellite moved inward toward the Earth. Simultaneously during this observed magnetopause erosion, a satellite in the tail lobes observed an increase in the lobe field and high latitude ground observations showed an enhanced quiet time ionospheric convection system that supported the hypothesis of increased flow of energy from the dayside to the night side during this “growth phase” period (McPherron et al., 1973).

I remember another new exciting discovery made by a visiting scientist, Torbjorn Pytte. At that time, the idea of a magnetic storm was Sydney Chapman’s view that a storm consisted of a continuous sequence of closely spaced substorms. Pytte studied several storm time sequences that were characterized by extended long periods of southward IMF, strong continuous auroral and magnetic activity, but seemed to lack the characteristic signatures of individual substorm expansions (Pytte et al., 1978). This research indicated that during a long sustained period of southward IMF the tail rate of merging could adjust to match the enhanced dayside merging rate and the entire magnetosphere would operate at an enhanced rate of convection without flux accumulation in the tail. These periods were called convection bays after the observed high latitude magnetic signatures from ground auroral zone magnetic observatories.

3 DISCOVERIES

It is a remarkable feeling to be part of a team that discovers something new. As a young scientist at my first job at Stanford, following graduate school I worked with John Wilcox and the solar group investigating sun-weather relationships. This was entirely new territory for me and I met some extremely interesting people engaged in this speculation that a statistical link could be found between solar variability and weather or climate variability. Among these was Jack Eddy, who is known for his studies of sun spots and the identification of the Maunder Minimum, a 70-year period of low auroral and geomagnetic activity associated with no sunspot activity. In 1987 Eddy was awarded the Arctowski Medal by the National Academy of Sciences for "studies in solar physics and solar-terrestrial relationships and specifically for his demonstration of the existence and nature of solar variations of long term and the consequences of these changes for climate and mankind."

I worked with Wilcox for about 2 years investigating sun-weather relationships, but joined a new research group formed by Peter Banks when he moved to Stanford. With Peter, I was able to initiate a new research program focused on the day side of the Earth to examine the solar wind interaction with the Earth's magnetic field more directly, and I was much more interested in this line of investigation. There, I was able to meet new friends, make new measurements for the first time at the intersection of the day-side magnetic field with the ionosphere and discover new phenomena. The focus on the day side of the Earth where the solar wind first encounters the magnetosphere was extremely exciting and learning to utilize the incoherent scatter radar recently relocated to Sondre Stromfjord, Greenland was like a fresh breeze in my sails. My previous graduate school research had been directed toward substorms and to the development of the ring current, where there are various intervening processes that occur between the solar wind coupling on the dayside and the transport of energy and momentum through the magnetosphere to the substorm or ring current phenomena being investigated. The observations in the dayside ionosphere were quite different because the reaction to a change in the solar wind IMF is seen immediately in the variability of the dayside ionospheric electric field and resulting ionospheric plasma convection (Banks et al., 1984; Clauer et al., 1984; Jorgensen et al., 1984; Clauer and Banks, 1986).

I enjoyed many years of research centered around the Sondrestrom radar, investigating the high latitude electric field and current systems. The west coast Greenland chain of magnetometer measurements provides a powerful tool when combined with the electric field measurements from the radar to examine the high latitude electrodynamics. The large scale DPY and DPZ current systems were examined as they developed in response to changes in the IMF By and Bz components (Banks et al., 1984; Clauer and Banks, 1986). The convection reversal boundary was examined and observed to have both stable and unstable wave-like motions that also were associated with magnetic waves (McHenry et al., 1990; Clauer 2003). It appears that as the dayside ionospheric flows increase, particularly in response to stronger IMF By component, the ionospheric convection reversal boundary shows wave-like behavior that is speculated to be the signature of a flow instability like the Kelvin-Helmholtz instability (Clauer and Ridley, 1995; Ridley and Clauer, 1996; Clauer et al., 1997). However, this flow instability is being generated within the magnetosphere between the tailward and return convection flow. I think that this is a new and exciting discovery that has not yet been explored fully.

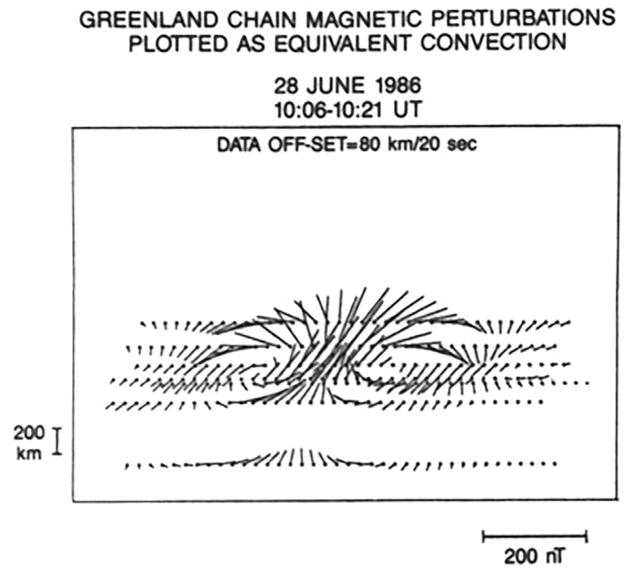

FIGURE 1 | Horizontal magnetic perturbation vectors rotated counter-clockwise and offset by a distance equivalent to 80 km/20 s shown on a geographic spatial coordinate system. (Friis-Christensen et al., 1988).

At the time that my student Mark McHenry and I were investigating the convection reversal boundary, Eigil Friis-Christensen came to Stanford to work with me for about 9 months with the idea to really examine some unusual magnetic impulse events observed in the Greenland magnetometer data. A great deal of attention has been devoted to these phenomena with the initial speculation that they might be the ionospheric signatures of flux transfer events described by Russell and Elphic (1979). Looking at the horizontal magnetic perturbation vectors, they appeared to point toward or away from a point that seemed to move across the Greenland magnetometer chain. Using the more extended array including the east coast stations showed that they were moving east or west roughly away from local noon. This did not support the flux transfer hypothesis in which the disturbance should move poleward or northward rather than in an east-west direction.

It was with the discovery of these magnetic impulse events that I learned the power and art of displaying the data in effective and creative ways. The idea to create a new display of the Greenland data was given to Eigil by Karl-Heinz Glassmeier during a boat ride at the Vancouver IUGG meeting. The horizontal vectors were rotated to be in the direction of the ionospheric F-region plasma convection (opposite to the Hall current direction), and then plotted on a single plot, but each set of measurements offset by a distance determined by the velocity at each measurement interval. The result shown in **Figure 1** was dramatic and made the cover of Geophysical Research Letters when we published (Friis-

Christensen et al., 1988). I was particularly amazed with near perfect organization of the display because it was produced by data and not the result of a model output.

The figure was an epiphany. At the center of each vortex must be a magnetic field-aligned current, in this case, downward in the left vortex and upward in the right (or leading) vortex with a corresponding horizontal ionospheric electric field outward or inward to the vortex center driving a corresponding circular Hall current that produced the magnetic perturbations. These field-aligned currents would map to the outer magnetosphere and were produced by waves on or near the magnetospheric boundary caused by sudden solar wind pressure pulses. These ideas were all later verified by further investigations by many investigators. The deformations in the magnetopause produced by the pressure changes propagated tailward and this was electrodynamically coupled to the ionosphere by field-aligned currents. What an exciting discovery of this direct electrodynamic linkage. These studies further enhanced our investigation of the ionospheric convection reversal boundary because when the reversal boundary became unstable with waves, a similar display of the magnetometer data showed a series of vortices produced by field-aligned currents inward and outward mapping the waves generated at the velocity shear in the outer magnetosphere to the ionosphere.

Since these were propagating structures across Greenland they were occasionally observed in the west coast magnetometers and then later in the east coast magnetometers, but not always. The impulses evolved and changed as they moved across Greenland and this was the motivation to deploy a temporary array of autonomous magnetometer stations near the center of Greenland on the Greenland Ice cap. This was really my start in developing and operating remote autonomous magnetometer arrays and it has been a fun and rewarding activity. It was exciting living in a tent at the Greenland summit and using snow mobile traverses to travel across the ice cap to set up the stations. One of my intense memories is returning to the summit station from a traverse and seeing my first ever mirage. The station was clear as ever but inverted upside down above the horizon where the station was actually sitting. I do not quite understand the optical atmospheric conditions to produce this, but it was remarkable. However, over the years, the array was tedious to maintain because the stations had to be visited each year to download data which was stored in local memory. After a sufficient period of operation, we closed the project and removed the stations from Greenland.

When I moved to Virginia Tech I had the opportunity to develop a new generation of remote measurement system that could be deployed in the Antarctic at extremely remote locations. These systems had to be able to operate unattended for many years and therefore required satellite communication links that were now available through the Iridium satellite network. This led to a new, more robust system that utilized instruments that met lower power requirements (Clauer et al., 2014). The fluxgate magnetometer was developed by Valery Korepanov at the Lviv Center of the Institute for Space Research in the Ukraine. It is an excellent low power instrument. We also added a low power induction magnetometer built by Marc Lessard at the University of New Hampshire, a new and innovative dual frequency GPS receiver developed by Geoff Crowley and his ASTRA research enterprise. The new system was also improved to allow installation without removing gloves. Nuts and bolts are no longer used to attach sections of the tower. Special push pins replaced the bolts, and the battery harness to connect the 16 batteries in parallel was built with snap connectors that could be installed simply and only in the correct way. I am quite proud of this project and achievement. At the time of this writing, the first station installed on the East Antarctic Plateau has been operating successfully for 14 years. The chain established on the East Antarctic Plateau is magnetically conjugate to the chain of magnetometers along the west coast of Greenland and enables the simultaneous measurement of high latitude phenomena in both hemispheres. The Antarctic stations are the next step taken to improve our understanding of the complete coupled system. Data from the Greenland and Antarctic chains are being utilized to investig and the impacts of the seasonal and field asymmetries that exist between hemispheres as they couple with the solar wind and each other.

What a wonderful group of people with whom I have been able to share my existence. All of us watching, examining and thinking about a particular aspect of the world around us, excited when we find some new feature or behavior and getting together to report and discuss our fascination and improve our understanding. It has been rewarding and exciting to be near the formation of new technologies that could be utilized to extend and improve our measurements and understanding.

I enjoy watching the world around me, particularly nature and animal behavior. I have spent many happy hours sitting with my wife Susan on a cliff on Santa Barbara Island off of Los Angeles, watching sea lions and elephant seals. I have enjoyed sitting on the porch of a hut next to a water hole on an African game reserve vacation watching all of the animal activity. And I have enjoyed a career of watching the solar wind interact with the Earth's magnetic field to produce some of the most wonderous phenomena like the aurora, geomagnetic variations, radiation belts and the ring current. Science has given me the tools to organize the observations to develop a deeper understanding and a community who delight and enjoy the discussion and debate over the meaning of our observations and ideas. Indeed, I did find a vocation that I liked and my advice and hope for everyone is that they, also, can find such a happy and fulfilling path in their life.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://mist.nianet.org.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Banks, P. M., Araki, T., Clauer, C. R., St. Maurice, J. P., and Foster, J. C. (1984). The Interplanetary Electric Field, Cleft Currents and Plasma Convection in the Polar Cap. *Planet Space Sci.* 32, 1551–1557. doi:10.1016/0032-0633(84)90024-2

Clauer, C. R., and Banks, P. M. (1986). Relationship of the Interplanetary Electric Field to the High-Latitude Ionospheric Electric Field and Currents: Observations and Model Simulation. *J. Geophys. Res.* 91, 6959. doi:10.1029/JA091iA06p06959

Clauer, C. R., Banks, P. M., Smith, A. Q., Jorgensen, T. S., Friis-Christensen, E., Vennerstrom, S., et al. (1984). Observation of Interplanetary Magnetic Field and of Ionospheric Plasma Convection in the Vicinity of the Dayside Polar Cleft. *Geophys. Res. Lett.* 11, 891–894. doi:10.1029/GL011i009p00891

Clauer, C. R. (2003). Ionospheric Observations of Waves at the Inner Edge of the Low Latitude Boundary Layer. *Geophys. Monogr. Ser.* 133, 297–309. doi:10.1029/133GM30

Clauer, C. R., Kim, H., Deshpande, K., Xu, Z., Weimer, D., Musko, S., et al. (2014). An Autonomous Adaptive Low-Power Instrument Platform (AAL-PIP) for Remote High-Latitude Geospace Data Collection. *Geosci. Instrum. Method. Data Syst.* 3, 211–227. doi:10.5194/gi-3-211-2014

Clauer, C. R., and Ridley, A. J. (1995). Ionospheric Observations of Magnetospheric Low-Latitude Boundary Layer Waves on August 4, 1991. *J. Geophys. Res.* 100, 21873–21884. doi:10.1029/95JA00678

Clauer, C. R., Ridley, A. J., Sitar, R. J., Singer, H. J., Rodger, A. S., Friis-Christensen, E., et al. (1997). Field Line Resonant Pulsations Associated with a Strong Dayside Ionospheric Shear Convection Flow Reversal. *J. Geophys. Res.* 102, 4585–4596. doi:10.1029/96JA02929

Dungey, J. W. (1961). Interplanetary Magnetic Field and the Auroral Zones. *Phys. Rev. Lett.* 6, 47–48. doi:10.1103/PhysRevLett.6.47

Friis-Christensen, E., McHenry, M. A., Clauer, C. R., and Vennerstrøm, S. (1988). Ionospheric Traveling Convection Vortices Observed Near the Polar Cleft: A Triggered Response to Sudden Changes in the Solar Wind. *Geophys. Res. Lett.* 15, 253–256. doi:10.1029/GL015i003p00253

Jorgensen, T. S., Friis-Christensen, E., Wickwar, V. B., Kelly, J. D., Clauer, C. R., and Banks, P. M. (1984). On the Reversal from 'sunward' to 'antisunward' Plasma Convection in the Dayside High Latitude Ionosphere. *Geophys. Res. Lett.* 11, 887–890. doi:10.1029/GL011i009p0087

McHenry, M. A., Clauer, C. R., Friis-Christensen, E., Newell, P. T., and Kelly, J. D. (1990). Ground Observations of Magnetospheric Boundary Layer Phenomena. *J. Geophys. Res.* 95, 14995. doi:10.1029/JA095iA09p14995

McPherron, R. L., Russell, C. T., and Aubry, M. P. (1973). Satellite Studies of Magnetospheric Substorms on August 15, 1968:9. Phenomenological Model for Substorms. *J. Geophys. Res.* 78, 3131–3149. doi:10.1029/ja078i016p03131

Pytte, T., McPherron, R. L., Hones, E. W., and West, H. I., Jr. (1978). Multiple-satellite Studies of Magnetospheric Substorms—Distinction between Polar Magnetic Substorms and Convection-Driven Negative Bays. *J. Geophys. Res.* 83, 663–679. doi:10.1029/JA083iA02p0063

Ridley, A. J., and Clauer, C. R. (1996). Characterization of the Dynamic Variations of the Dayside High Latitude Ionospheric Convection Reversal Boundary and Relationship to Interplanetary Magnetic Field Orientation. *J. Geophys. Res.* 101, 10919–10938. doi:10.1029/ja101ia05p10919

Russell, C. T., and Elphic, R. C. (1979). ISEE Observations of Flux Transfer Events at the Dayside Magnetopause. *Geophys. Res. Lett.* 6, 33–36. doi:10.1029/GL006i001p00033

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 14 July 2022
doi: 10.3389/fspas.2022.943401

Reflections by Bengt Ulf Östen Sonnerup

Bengt Sonnerup *

Dartmouth College, Hanover, NH, United States

First, I will tell readers about memories of my graduate student days at Cornell. I will highlight some of my experiences there, and what I learned, and didn't learn from them. Then I will then discuss a few of the research topics I have been working on over the years, including data interpretation and the tools for it, with special emphasis on the magnetopause. Aspects of MHD shocks and other structures, including boundary layers, are among those topics. I will mention a few people with whom I have worked closely and a few famous individuals, who influenced me in a significant way. The presentation contains material of potential interest to new, as well as more seasoned workers in the field. It will not always be in time order.

Keywords: reconnection, upstream facing waves, shock waves, Hall effect, magnetopause

OPEN ACCESS

Edited by:
Thomas Earle Moore,
Third Rock Research, United States

Reviewed by:
James A. Slavin,
University of Michigan, United States

****Correspondence:***
Bengt Sonnerup
bengt.u.o.sonnerup@dartmouth.edu

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *13 May 2022*
Accepted: *08 June 2022*
Published: *14 July 2022*

Citation:
Sonnerup B (2022) Reflections by Bengt Ulf Östen Sonnerup. Front. Astron. Space Sci. 9:943401. doi: 10.3389/fspas.2022.943401

1 AT CORNELL

My studies at the Graduate School of Aerospace Engineering at Cornell began in 1958. It was there that I first encountered magnetohydrodynamics and space plasma physics. The Aerospace faculty at that time, including my advisor W. R. Sears, had become intensely interested in MHD and were active in building the mathematical and experimental foundations of that discipline. I was swept up by their enthusiasm and have remained active in the field ever since, working on a variety of problems, a particular emphasis of mine being the outer boundary of Earth's magnetic field, the magnetopause. Over many years, I have tried, with variable success, to use basic plasma physics to explain spacecraft observations of the magnetopause, as well as of interplanetary boundaries and other structures.

As a graduate student, I was mostly working on terrestrial applications such as MHD power generation, because, at the time, hardly any *in-situ* spacecraft measurements were available. At a conference somewhere in Japan, Jim McCune, one of my teachers in the Aerospace School, was asked where the Cornell group's work on MHD wing theory could find applications. His reported response was: "Oh somewhere in outer space." A good prediction, it turns out.

One of the things I came to fully appreciate and love was the sublime beauty and tight intertwining of mathematics and physics.

A stark memory from graduate school was the PhD qualifying exam. At Cornell it was entirely oral. My committee consisted of my advisor, Bill Sears together with mathematics professor R. Agnew, and famous astrophysicist, Edwin E. Salpeter. Things went wrong right from the start. As always, Agnew brought along his dog, a large Collie. Salpeter commented that the dog must be very smart after attending all those math classes. To this Sears responded that he knew about dogs and that they were stupid. He quickly realized that Agnew became insulted and tried to improve the situation by saying "now I have insulted his dog and he will flunk my candidate." It probably did not help, and it augmented my sense of doom. The next disaster followed in short order. Salpeter started to ask me questions about the Mössbauer effect, for which the Nobel prize had just been awarded. But he led me through the discussion in a friendly manner. Afterwards, I asked him why he had asked such a question of a low-level mechanical engineer. His response was: “I had already decided to pass you so I thought I could have a bit of fun.” But I myself had difficulty appreciating the fun of it. Since then, I have come to feel that faculty should challenge their students but be careful to not torment them.

Here is another memory: Theodore von Kàrmàn, of vortex street fame, was the mentor of Bill Sears at Cal Tech. This world-famous Hungarian had been invited to spend the Spring term of 1960 at the Aero-School and we, the graduate students were lined up to be introduced to him, one by one. Von Kàrmàn was 79 at the time and rather deaf. When it was my turn, I said something respectful and polite. He paid no attention but said to Sears in a loud voice: “What language does he speak?” The response was quick: “Same as you, broken English.” (At Caltech von Kàrmàn had confounded his students by using the European pronunciation of chaos, which sounds like cows.)

In my last year at Cornell, I was a postdoc with Austrian physicist Tommy Gold, who had been hired as director of the new Arecibo Observatory, the brainchild of EE professor William Gordon. One of the EE graduate students at Cornell, who helped design the Arecibo antenna feed, was a friend of mine, Thomas Laaspere, who then arranged for me to come to Dartmouth, where he, and Millett Morgan (of whistler fame) served as my mentors.

Tommy Gold was fun and friendly, and a great alpine skier who, quite erroneously, claimed that cross country skiing was a dying sport. But he was right in his very early prediction of Earth’s bow shock. And I am very grateful that, in 1962, he arranged a job for me at Hannes Alfvén’s laboratory at KTH in Stockholm.

As it turned out, I left KTH for Dartmouth after only a couple of years. Hannes and I had irreconcilable disagreements over the reality and importance of magnetic reconnection. On my end, I had become convinced that Tommy Gold’s comment was correct: “Reconnection must happen, and at a substantial rate, otherwise the interplanetary magnetic field would become hopelessly entangled.” I don’t think Hannes ever came to agree with that.

It was at Cornell that I first met Ian Axford and developed awe for his work with Colin Axford and Hines (1961) on the global plasma circulation and currents in the magnetosphere. I did not meet Carl Sagan, who was almost never on campus, and not Vladimir Nabokov, who gave lectures about the writing of Lolita, to overflow audiences.

My MS thesis at Cornell was concerned with Hall effects in MHD flow past a wavy wall. It became my first publication (Sonnerup, 1961). To my knowledge it has seldom, if ever, been cited but was a precursor to my later work on the Hall effect in magnetic reconnection (Sonnerup, 1979).

There were memorable papers by Sears and coworkers (1961, 1964), in which they demonstrated the possibility of upstream-facing standing waves in compressible MHD. Many features of such flows remain unexplored, both theoretically and in spacecraft observations. With the high precision and time resolution provided by MMS, the search for such forward facing waves could be a rewarding one. More details will be given in **Section 4**.

But first, I want to make a brief jump back in time to my high school days in the forties in Malmö, Sweden. There, a very special mathematics teacher taught his students something important, in addition to mathematics. He would ask us to solve geometry problems on the blackboard. If you just stood there doing nothing, he would get upset, where he stood right behind you. At the peril of a sharp rap on the posterior from his pointer, he would urge you to do something, for example, draw a help line. It might not solve the problem but thinking about why it didn’t was, he knew, an effective way to figure out a useful next move. His approach has been a great help to me, not only in geometry but in all sorts of problems encountered in life. The message from him was clear: Get on with that first step.

2 MHD SHOCK WAVES

My interest in shock waves was spurred by two facts. The first of these was the paper by Petschek (1964), in which the concept of standing waves associated with reconnection at the magnetopause was first proposed. It was followed by a, now mostly forgotten, but perfectly beautiful, comprehensive, article on MHD waves by Kantrowitz and Petschek (1966). There was also the seminal work of Levy et al. (1964) on magnetopause reconnection in the limit of zero magnetospheric plasma pressure.

The second fact was the arrival in 1988 of Lin-Ni Hau as my postdoctoral coworker. Her studies of MHD shock structures, including Hall effects (Hau and Sonnerup, 1989, 1990), are beautiful, even unique. In her work, ordinary gas-dynamic shocks allow abrupt transfer of the MHD shock structure from a supersonic upstream portion, located on one Riemann sheet, to subsonic conditions on a downstream sheet. Samples of her shock work can be found in **Figure 1B**. She is now a renowned professor at the National Central University in Taiwan.

3 DATA INTERPRETATION

My first encounter with spacecraft data, occurred in collaboration with Larry Cahill, who was then at the UNH in Durham. We used his magnetic field data from Explorer 12 to try to determine the sign and magnitude of the field component perpendicular to the magnetopause. The use of model normal vectors failed completely, which led us to seek for a direction in which that component was as constant as the data would permit. The resulting analysis process became known as magnetic variance analysis or MVAB (Sonnerup et al., 2010). This method is still in common use. But it has proved very difficult to obtain reliable results, a curse caused by eigenvalue degeneracy in combination with the very small value of the field component along the normal.

What my work with Cahill did was to get me deeply involved in developing and using data analysis methods. By far, my most extensive collaboration has been with Goetz Paschmann at the Max-Planck Institute for Extraterrestrial Physics (MPE) in Garching. Our collaboration tended to work this way: Goetz would look at a large data base and identify features that were mysterious and promised to give new insights. I would suggest

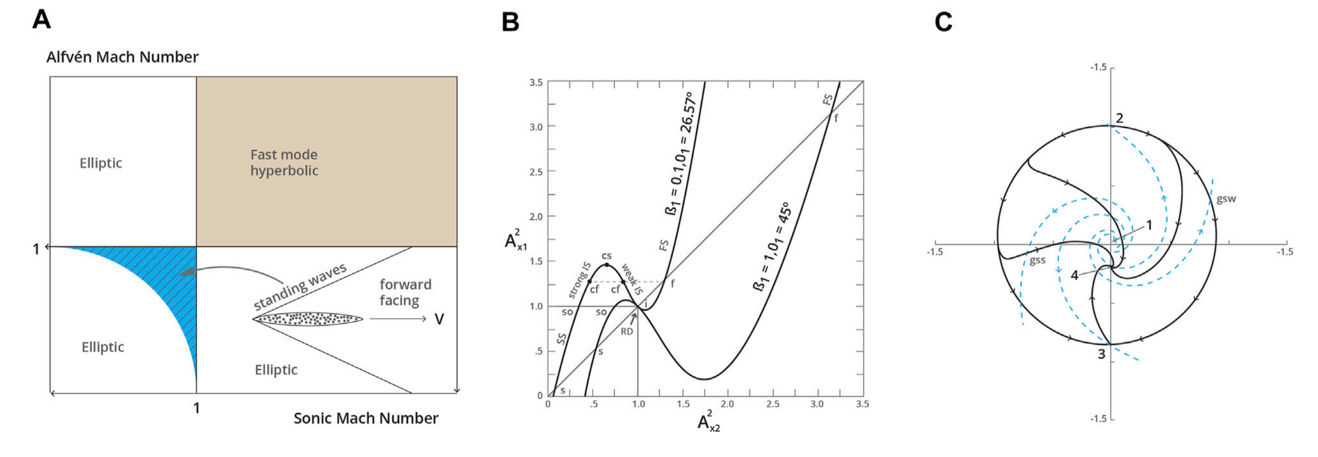

FIGURE 1 | Standing waves and shocks in MHD: (A) Standing upstream-facing slow mode waves in blue shaded region; (B) Slow, intermediate, and fast shock wave jump properties from upstream (1) to downstream (2) conditions; angle θ_1, is between shock normal and upstream magnetic field (C) Example of predicted magnetic hodogram for a shock structure.

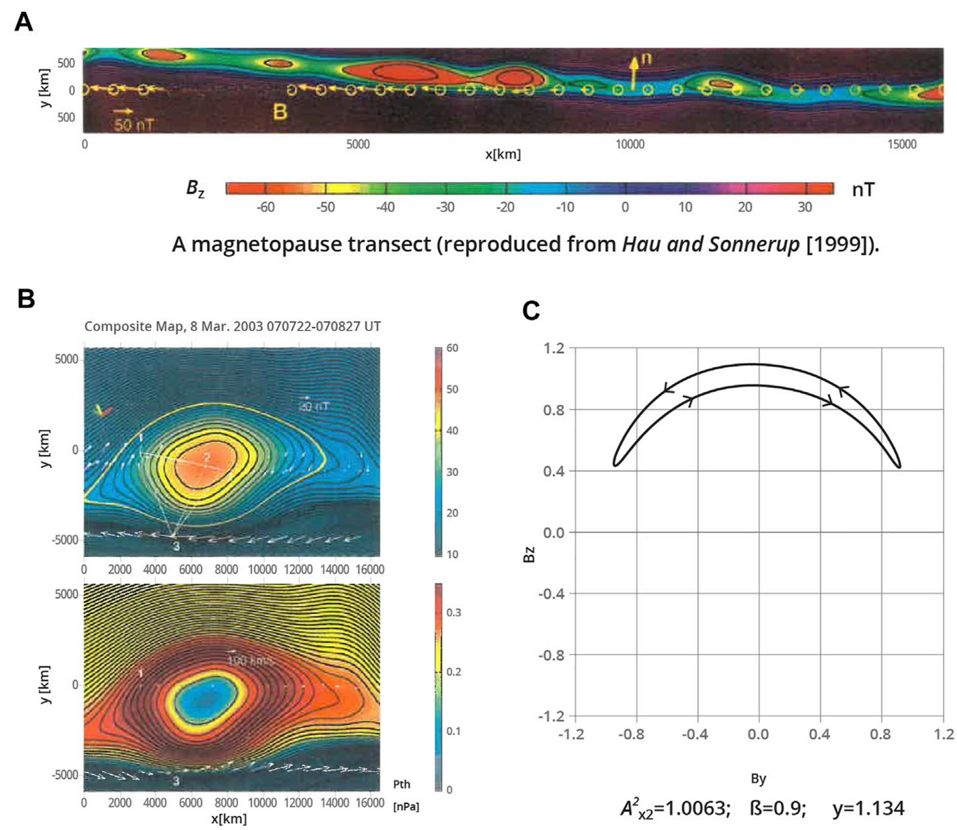

FIGURE 2 | Reconstruction: (A) Nearly tangential discontinuity behavior seen by AMPTE at the magnetopause; (B) Field and plasma pressure in a "Flux transfer" event. (C) Back and forth motion of the field in arc-polarized waves.

various interpretations and then try to do some mathematical analysis to quantify them. Our cooperation started in 1979 and has been extremely fruitful until the present, with some allowance for my old age failings. What I want to highlight is that having a skillful coworker is a tremendous boon. I have been in great luck about that, with students as well as seasoned scientists.

I only had a small number of graduate students here at Dartmouth; among them were Tai Phan, Sasha Khrabrov, Dave Walthour, and Qiang Hu. All have done theoretically based data interpretation work of exceptional quality.

My most influential collaboration here at Dartmouth was with Lin-Ni Hau. One of the most successful projects was the development and use of software to reconstruct field structures, observed during the flight pass of a spacecraft through them or near them. The method was originally based on the Grad Shafranov equation, as discussed by Sonnerup and Guo (1996). Examples of such reconstructions at the magnetopause are shown in **Figures 2A,B**. It has been greatly generalized since then and now involves direct integration of various versions of the ion and electron equations of motion. Applications involving electron dynamics are given by Sonnerup et al. (2016) and by Hasegawa et al. (2017).

In joint work (Sonnerup et al., 2010), Stein, Goetz, and I presented theoretical analysis of ark-polarized structures in the solar wind, such as observed by Bruce Tsurutani, and others. Our analysis incorporated plasma compressibility as well as electron and ion inertia. As shown in **Figure 2C**, it accounts for the back-and-forth motion of the field seen in such structures.

At MPE, I also had wonderful collaboration with Iannis Papamastorakis, leading to new findings about the convection electric field. With Stein Haaland and a group of others, I also participated in a comparison of results from single- and multi-spacecraft measurements concerning magnetopause orientation, motion, and thickness.

4 THE WAVY WALL PROBLEM

The first problem I was exposed to at Cornell was the analysis of waves generated by MHD flow past an impenetrable wavy wall. Such flows remain incompletely studied, both theoretically and observationally. **Figure 1A** shows a diagram of the Alfvén-Mach number M_A, here denoted simply by A, versus ordinary sonic Mach number, $M_s = M$ for such flows, developed by Sears and coworkers [see for example, the papers by McCune and Resler (1960) and Sears (1961)]. The region of interest is shaded blue in **Figure 1A**. In this diagram of A versus M, it is bounded by the two lines $A = 1$ and $M = 1$, and a circular arc from $A = 1, M = 0$ to $A = 0, M = 1$. In that small region upstream-facing, rather than downstream-facing, standing waves are predicted. To my knowledge, such remarkable and unexpected wave orientations have yet to be observed. With the high precision and time resolution provided by MMS, the hunt for them should be a rewarding one. Old age slows us all down, but I am still tempted to join the hunt.

5 MORE COWORKERS

Over the years, I have had good and useful working relations with many researchers. Included are some of my former and present colleagues at Dartmouth, especially Richard Denton. He taught me the extreme importance of asking persistent probing questions. My engineering colleague, Bill Lotko, whose understanding of, and ability to mathematically describe, the entire dynamic magnetospheric system were indispensable in our collaboration. In the sixties, I also worked on magnetic field annihilation with solar physicist Eric Priest.

I have worked with a group of scientists at the Mission Research Corporation, with a branch located in Nashua, NH and headed by Willard W. White. The group included George Siscoe, Nelson Maynard, Keith Siebert, Dan Weimer and others. George Siscoe, who passed away in April of 2022, was an inspiring teacher and mentor of students. He was a soft-spoken intellectual leader, whose work will have lasting impact on our field. He was a gentle soul and a dear friend.

At MPE, I primarily worked with Goetz Paschmann. But I also collaborated, with and befriended, Norbert Schkopke, Wolfgang Baumjohann and Rumi Nakamura, Chuck Carlson (visiting from Berkeley), and many others. At ISSI, it was great to interact with its founder, Johannes Geiss, and co-director, Rudi von Steiger, and with many science visitors to the Institute, some of them well-established or famous, like Rudolf Treumann, others being in earlier stages of their careers.

Among the many people I met at MPE, there were two individuals in addition to Paschmann, who had a strong, albeit more indirect, influence on my development as a scientist.

The first was Reimar Lüst, the founder of MPE. He went on to become president of the entire Max-Planck Society, then director general of ESA. He was the founder and president of the private Jakobs Universität in Bremen. He later became president of the Alexander von Humboldt Foundation, which funded my 9-month stay in Garching and in Bern in 2001–2002.

He second was Gerhard Haerendel. He was a director at MPE and enabled my many visits to the institute. Later, he served as the Dean of Faculty at the Jakobs Universität, before returning to Garching and retirement. Gerhard and I shared a love of Mozart's opera The Magic Flute.

Among the scientists I interacted with were many who are still active, working with data from Cluster, MMS, and other missions. One notable among those individuals is Chris Russell. Here is an old but acute memory I have of him. He was a coauthor on the MPE paper (Paschmann et al., 1979) about the first *in-situ* observations, "the smoking gun evidence", of reconnection at the magnetopause. When we sent him a draft, his prompt, salt-of-the-earth, response was: "This paper starts with a roar and ends with a whimper." Of course, we then did work to remove the whimper part. A bit later on, the MPE based team pursued the reconnection topic in further detail (Sonnerup et al., 1981).

6 CLOSING REMARK

I have had good relations with most people I interacted with in science, even during my period as JGR editor (1981–84). But at that time, as well as both earlier and later, I also encountered people exhibiting what I refer to as the "barracuda syndrome," ranging from simple passive-reactive to outright aggressive behavior, which included an unsuccessful effort to get me fired. But being editor also has many attractive features: It is a great service to the research community, and the editor learns a great deal about the wide activities and personalities of workers in the field. If the opportunity arises, my advice is to give it serious thought.

This account of my thoughts and reflections about space physics would be incomplete without mention of Vytenis Vasyliunas, whose awe-inspiring insights into global heliospheric physics remain difficult to match, and whose organ concerts in church were high points at many science meetings.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The author did all the research and wrote the paper. Preparation of the paper was supported by NASA grant 80NSSC19K0254 (Richard Denton, PI) to Dartmouth College.

REFERENCES

Axford, W. I., and Hines, C. O. (1961). A Unifying Theory of High-Latitude Geophysical Phenomena and Geomagnetic Storms. *Can. J. Phys.* 39, 1433–1464. doi:10.1139/p61-172

Hasegawa, H., Sonnerup, B. U. Ö., Denton, R. E., Phan, T. D., Nakamura, T. K. M., Giles, B. L., et al. (2017). Reconstruction of the Electron Diffusion Region Observed by the Magnetospheric Multiscale Spacecraft: First Results. *Geophys. Res. Lett.* 44 (4), 566–574. doi:10.1002/2017GL073163

Hau, L.-N., and Sonnerup, B. U. Ö. (1989). On the Structure of Resistive MHD Intermediate Shocks. *J. Geophys. Res.* 94 (A66), 539–551. doi:10.1002/2017gl073163

Hau, L.-N., and Sonnerup, B. U. Ö. (1990). The Structure of Resistive-Dispersive Intermediate Shocks. *J. Geophys. Res.* 95 (A1118), 791–818. doi:10.1029/ja095ia11p18791

Kantrowitz, A., and Petschek, H. E. (1966). "MHD Characteristics and Shock Waves," in *Plasma Physics In Theory And Application* (New York: McGraw-Hill Book Co.), 147–206.

Levy, R. H., Petschek, H. E., and Siscoe, G. L. (1964). Aerodynamic Aspects of the Magnetospheric Flow. *AIAA J.* 2, 2065–2076. doi:10.2514/3.2745

McCune, J. E., and Resler, E. L. (1960). Compressibility Effects in Magnetoaerodynamic Flows Past Thin Bodies. *J. Aerosp. Sci.* 27, 493–503. doi:10.2514/8.8616

Petschek, H. E. (1964). "Magnetic Field Annihilation," in *AAS-NASA Symposium on the Physics of Solar Flares*. NASA Special Publication SP-50. Editors W. N. Hess, 425–439.

Paschmann, G., Sonnerup, B. U. Ö., and Papamastorakis, I. (1979). Plasma Acceleration at the Earth's Magnetopause: Evidence for Reconnection. *Nature* 282, 243–246.

Sears, W. R. (1961). Sub-Alfvénic Flow in Magnetoaerodynamics. *J. Aerosp. Sci.* 28 (8), 249–250. doi:10.2514/8.8940

Sonnerup, B. (1961). Some Effects of Tensor Conductivity in Magnetohydrodynamics. *J. Aerosp. Sci.* 28 (8), 612–621. doi:10.2514/8.9115

Sonnerup, B. U. Ö., Haaland, S. E., and Paschmann, G. (2010). On Arc-Polarized Structures in the Solar Wind. *Ann. Geophys.* 28, 1229–1248. doi:10.5194/angeo-28-1229-2010

Sonnerup, B. U. Ö., Paschmann, G., Papamastorakis, I., Sckopke, N., Haerendel, G., Asbridge, R., et al. (1981). Evidence for Magnetic Field Reconnection at the Earth's Magnetopause. *J. Geophys. Res.* 86, 10049–10067. doi:10.1029/ja086ia12p10049

Sonnerup, B. U. Ö., and Guo, M. (1996). Magnetopause Transects. *Geophys. Res. Lett.* 23 (No 253), 679–682. doi:10.1029/96gl03573

Sonnerup, B. U. Ö., Hasegawa, H., Denton, R. E., and Nakamura, T. K. M. (2016). Reconstruction of the Electron Diffusion Region. *J. Geophys. Res. Space Phys.* 121 (4), 279–290. doi:10.1002/2016ja022430

Sonnerup, B. U. Ö. (1979). "Magnetic Field Reconnection," in *Solar System Plasma Physics Vol III*. Editors L. J. Lanzerotti, C. F. Kennel, and E. N. Parker (New York, NY: Elsevier North-Holland Publishing Company), 45–108.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 22 July 2022
DOI 10.3389/fspas.2022.952226

OPEN ACCESS

EDITED BY
Joseph E. Borovsky,
Space Science Institute, United States

REVIEWED BY
Tamas I. Gombosi,
University of Michigan, United States
Michael Denton,
Space Science Institute, United States

*CORRESPONDENCE
P. G. Richards,
pgrichds@gmail.com

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 24 May 2022
ACCEPTED 30 June 2022
PUBLISHED 22 July 2022

CITATION
Richards PG (2022), Ionospheric photoelectrons: A lateral thinking approach.
Front. Astron. Space Sci. 9:952226.
doi: 10.3389/fspas.2022.952226

Ionospheric photoelectrons: A lateral thinking approach

P. G. Richards*

The University of Alabama in Huntsville, Huntsville, AL, United States

This paper presents a personal account of developments in the solution of two related problems that limited the accuracy of ionosphere calculations in the 1980s. The first problem concerns the model-data discrepancies in the ionosphere photoelectron flux spectrum. Early comparisons between measured and modeled data revealed discrepancies in magnitude and shape. A lateral thinking approach revealed that there were problems with two key photoelectron model inputs: namely, the electron impact cross-sections and the solar extreme ultraviolet irradiance. This work led to the development of a widely used EUVAC solar irradiance model for ionosphere electron density calculations. The second problem relates to the neutral winds that are crucial for modeling variations in the ionosphere ion and electron densities. There is a lack of thermosphere neutral wind data because they are difficult to measure. The winds determine the altitude at which the electron density peaks. An accurate solution to this problem was the development of an algorithm that assimilates the altitude of the peak electron density into ionosphere models. This technique works because the altitude of the peak density is very sensitive to variations in the neutral wind.

KEYWORDS

ionosphere, thermosphere, photoelectrons, solar EUV, cross-sections

Introduction

A common approach to solving scientific problems is to 1) build a model, 2) input the best parameters, 3) compare the model output with data, and 4) publish. If there is a model-data discrepancy, see if the model formulation and its input parameters can be improved and repeat steps 3 and 4.

Another approach is to simply do a parameter study, in which step 3 is omitted and the model is run with different input parameters just to see how the model output changes. This approach is of limited value since even if the model output looks reasonable, it may not represent anything physical.

A problem with the first approach is what to do if the model output does not match the data after careful checking of the formulation and the input. That is, given that the coding is correct, which of several input parameters might be responsible for the discrepancy. Input parameter refers to parameters that may be hard-coded or input from a file.

For some complex problems such as weather forecasting, the model may deviate from normal over time despite the best available modeling. In such a case, it may be possible to improve the model performance by assimilating measurements as it steps in time. Such models can be used to forecast future behavior based on current behavior.

Ionosphere photoelectron fluxes

The specific problem addressed in this article relates to the modeling of the ionosphere photoelectron flux. The Earth's atmosphere above about 100 km altitude contains a substantial number of charged particles (ions and electrons produced in equal numbers), which are primarily created by the photoionization of atomic oxygen (O) and molecular nitrogen (N_2) by extreme ultraviolet (EUV) light from the sun. Studying the ionosphere electron distribution is important because of its effect on radio waves that pass through or are refracted by the electrons. Prior to the mid-1990s, most ionosphere model calculations used the F74113 solar irradiance spectrum, which was based on rocket measurements from the 1960s and 1970s (Hinteregger, 1981).

In the photoionization process, some of the photon energy is stored in the resulting positive ion and the rest appears as translational energy of energetic photoelectrons. Unlike the photoelectrons, the ions gain little translational energy. The photoelectrons lose their excess energy in a cascade process through multiple collisions with the neutral gases, before ultimately joining the ambient low-energy thermal electron population. This cascade process resembles a mountain avalanche with most of the debris (low energy electrons) ending up in a pile near the bottom. Some of the photoelectrons lose energy in collisions with neutrals that result in the excitation of internal energy states of the neutral particles, while other collisions result in the creation of new ions in which the impacting (primary) electron loses energy, and an additional (secondary) electron is created. In fact, between about 100 and 170 km altitude, photoelectrons create more ions than are created by the initial photoionization by the solar EUV photons.

If photoelectron transport and energy cascade were not important, the flux for a specific electron energy at each altitude would be determined by the number of ions produced initially by solar EUV and lost due to energy-sapping collisions with the ambient thermosphere particles. The energy cascade process greatly complicates the calculation because there are many different types of collisions that can result in a wide energy spectrum of degraded primary electrons and secondary electrons.

In calculating the photoelectron flux spectrum, the process begins at the highest energy with just the primary electrons and proceeds to lower energies. Secondary electrons and degraded primary electrons are added to the primary electrons at lower energies. There is a lot of bookkeeping because there are many possible energy losses depending on which electronic states are created from which neutral particle.

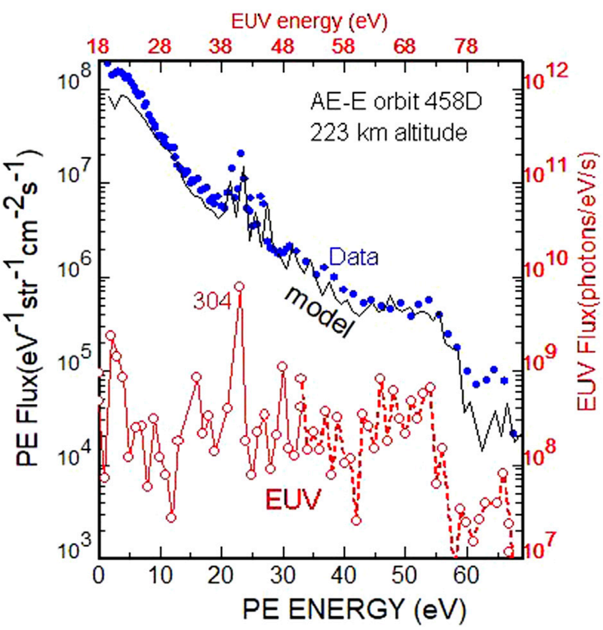

FIGURE 1
Measured (dots) and modeled (solid black line) photoelectron flux at 223 km. The red solid and dashed lines show the F74113 solar EUV reference irradiance as a function of energy, which has been shifted down by 18 eV (top axis) to approximately line up the EUV photons with the primary photoelectrons that they produce. The dashed line irradiances are double those of the standard F74113 irradiances.

At low altitudes, most photoelectrons lose their energy locally, but transport begins to have a significant effect on the energetic electron spectrum above about 300 km at solar minimum and at higher altitudes at solar maximum. Transport adds another layer of complexity to the photoelectron calculation. Photoelectrons can escape upwards along the magnetic field into the plasmasphere where they heat the thermal electron population and deposit energy in the opposite hemisphere. Another modeling complication, when transport is considered, is that the photoelectrons are created with multiple pitch angles. The simplest photoelectron transport model is the two-stream model developed by Nagy and Banks, 1970. It treats the photoelectrons as a single upward flux and a single downward flux. Even this simplified method is unsuitable for large global calculations because of the many ways photoelectrons can be degraded. The computation can be greatly reduced if production frequencies are pre-calculated for each photoelectron energy and for each neutral species. The photoelectron fluxes can then be efficiently recreated by simply folding the production frequencies with the appropriate neutral densities as necessary (see Richards and Peterson, 2008).

Figure 1 shows an example photoelectron flux that was measured by the AE-E satellite (dots) along with a model calculation (solid black line). The exponential increase in flux with decreasing energy reflects, to a large extent, the result of electron energy cascade. Below ~20 eV, the photoelectron flux is overwhelmingly a result of the cascade from higher energies. That means that most of the overall photoelectron population is generated by photons with wavelengths less than about 400Å. The prominent flux peaks between 20 and 30 eV are the result of the photoionization to several different energy states of O and N_2 by the strong 304Å solar irradiance. The red solid and dashed lines with circles show the solar EUV irradiance as a function of energy (top axis). The EUV spectrum has been shifted down by 18 eV to emphasize that the 20 and 30 eV peaks are caused by this strong 304Å (41 eV) solar line.

Ideally, a model calculation should match the photoelectron flux magnitude well at the peaks and also at most other energies up to 60 eV. Beyond 60 eV the low instrument count rate introduces a lot of measurement uncertainty. A model should also reproduce the altitude variation of the photoelectron flux well as the dominant species changes from N_2 to O. If the model does not match the overall flux magnitude, it should at least match the shape of the spectrum.

A key insight of Richards and Torr (1981) was that the 20–30 eV peaks have only a small contribution from the cascade process. That insight greatly simplifies the flux calculation below ~250 km where transport is negligible. So, the prominent peaks can be used to validate the complex full photoelectron calculations at these energies. The calculation using the observed F74113 304Å solar irradiance revealed that the best fit to the peaks is obtained with electron impact cross-sections that are approximately a factor of 2 smaller than most used previously. The problem could be solved if the measured photoelectron fluxes were a factor of 2 too high or the 304Å solar irradiance was a factor of 2 too small. The general opinion amongst modelers in the early 1980s was that the measured photoelectron fluxes were a factor of 2 too high. This was not borne out by later measurements and modeling.

In addition to validating the full flux calculation, the 1981 study revealed that the electron impact cross-sections that were used in prior calculations would create photoelectron flux spectra that did not match the shape of the measured spectra. With this observation, Richards and Torr (1984) decided to try to determine the energy variation of the electron impact cross-sections that would be needed to reproduce the shape and magnitude of the observed photoelectron fluxes. As revealed below, these results published in 1984 identified a problem with the F74113 standard solar EUV irradiance for wavelengths below 250Å (see dashed line in Figure 1).

The technique was to recast the photoelectron problem by using the measured solar EUV irradiance and photoelectron fluxes to determine the total electron impact cross-sections for O and N_2 that would be compatible with those measurements. The O cross-section was determined from ionosphere data around 250 km where O is the dominant species and the N_2 cross-section was determined below ~200 km where N_2 is the dominant species. Both the O and N_2 cross-sections so obtained were only about half of the other sets of cross-sections that were being used at the time in photoelectron models.

A more concerning problem was that there was a distinct discontinuity near 35 eV in both the O and N_2 cross-sections that were calculated from the ionosphere data. The curve labeled AE in Figure 2 shows the discontinuity for the N_2 electron impact cross-section. The oxygen electron impact cross-section had a similar discontinuity. The calculated cross-sections increase smoothly as expected from the threshold near 15 eV to about 35 eV. Then both cross-sections drop to only half of their prior value. This behavior is contrary to the well-established shape of the laboratory cross-sections in this energy range (solid line). Since the discontinuity occurred in both the O and N_2 cross-sections, Richards and Torr (1984) suggested that the F74113 solar EUV irradiance was a factor of 2 too small below 250Å. Private consultations with Hans Hinteregger confirmed that the solar irradiances below 250Å were the least reliable in the F74113 solar EUV spectrum. So, it was decided to double the solar EUV irradiance below 250Å (dashed line in Figure 1) for all subsequent calculations of ionosphere densities and temperatures. Although there was a problem with the magnitude of the F74113 irradiance below 250Å, the variation with changing solar activity proved to be satisfactory, which allowed accurate scaling of all wavelengths with solar activity.

When the cross-section results were submitted for publication, the unconventional technique resulted in a good deal of pushback from two reviewers and a third reviewer was consulted. He was a laboratory scientist involved with N_2 cross-section measurements and was concerned that we might be casting aspersions on the laboratory measurements. After a phone conversation, he became convinced that this result was important because it identified a serious problem with solar EUV irradiance, and the paper was finally published. It turned out to be one of the two most important papers that we published. From this experience, I adopted a policy of never reviewing journal articles anonymously. It has led to some productive and enjoyable interactions with authors.

After the paper was published, the modified F74113 solar irradiances were used in all our ionosphere calculations of densities and temperatures for the next 10 years until 1993 when a reviewer challenged their use in a paper concerned with the comparison of our calculated ion densities and temperatures with measurements. That criticism prompted the development and publication of the EUVAC solar irradiance model that has been widely used (Richards et al., 1994). The EUVAC model is based on the F74113 solar EUV irradiances with the doubling of the solar irradiance short ward of 250Å. It was not until almost 20 years after its initial discovery that the problem with the F74113 solar irradiance was confirmed by new measurements from the Student Nitric Oxide Explorer (SNOE) satellite (Bailey et al., 2000). Further confirmation came from the

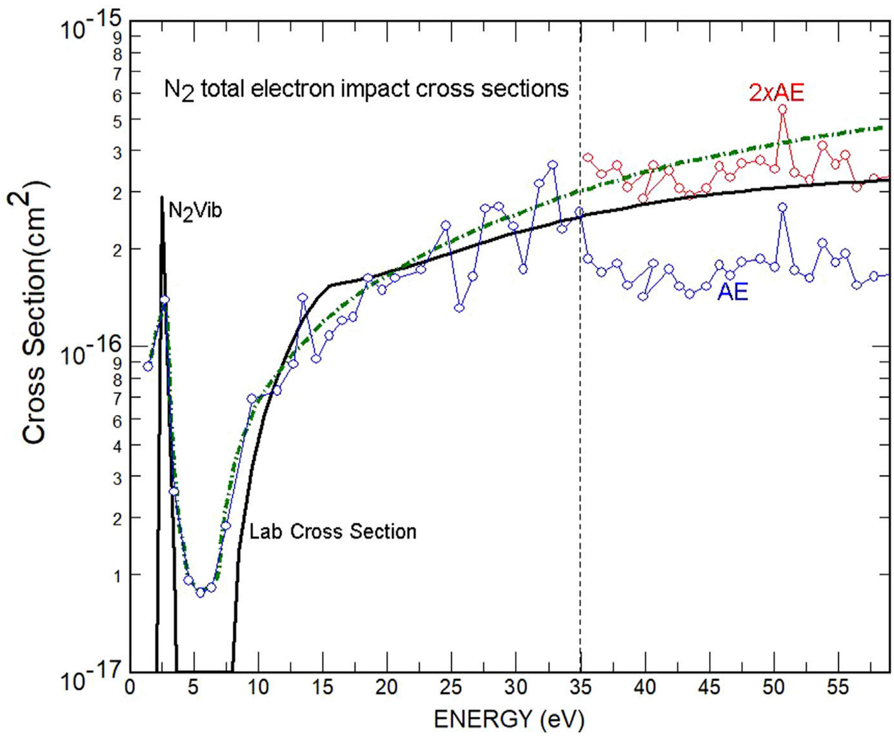

FIGURE 2
Comparison of N_2 cross-sections derived from Atmosphere Explorer solar EUV irradiance and photoelectron flux (AE) with Laboratory cross-section (solid line). The dash-dot line is a weighted least-squares fit to the 1.0xAE data points below 35 eV. The points labeled 2xAE were obtained by doubling the AE values. The label N_2Vib identifies the cross-section peak near 2.5 eV that results from vibrational excitation of N_2.

measurements of the SEE experiment on the TIMED satellite (Woods and Eparvier, 2006).

Despite these corroborating measurements, skepticism continued in the general aeronomy community. When I took up a position as a program scientist at NASA headquarters in 2003, I worked with fellow program scientist, Bill Peterson, who expressed skepticism about the Richards and Torr photoelectron work. To settle the issue, we decided to compare model calculations with his photoelectron data from the FAST satellite. The model-data agreement was highly satisfactory, and our subsequent collaboration led to 11 journal articles on various aspects of photoelectron behavior (e.g., Richards and Peterson, 2008; Peterson 2021). These papers show that the 2-stream photoelectron model not only produces good agreement with photoelectron data in the local ionosphere but also in the plasmasphere including backscattered fluxes from the conjugate ionosphere.

Although great progress has been made in photoelectron flux and ionosphere theory, there are still unresolved problems. None of the current solar irradiance models adequately capture the photoelectron flux variations (Peterson et al., 2012). To date, photoelectron modeling has not been tested with a truly comprehensive set of coincident high-resolution measurements of both photoelectron and solar EUV spectra.

Inadequate knowledge of the photoelectron and EUV spectra is likely the reason that ionosphere models routinely underestimate the E-region peak electron density by > 30% (Solomon et al., 2001). Photoelectrons are the major source of ions and electrons in the E-region. Because the chemical loss rate is a square function of the electron density, a 30% model-data difference could correspond to a factor of 2 underestimate of the photoelectron source. Resolution of this problem would require coincident high-resolution rocket measurements of the photoelectron spectra and high-resolution measurements of the solar EUV X-ray spectrum, along with ion and neutral densities.

Epilogue

The collaboration with Bill Peterson that began at NASA HQ illustrates the scientific importance of taking full advantage of being in the right place at the right time. My career has benefited from several other fortuitous circumstances that have led to a deeper understanding of the space environment. Photoelectron modeling played a central role in the research examples below.

Among these happenstances was a decision in 1971 to get a degree at LaTrobe University. There, I was adopted as a graduate student by the well-known pioneering space physicist professor Keith Cole who was very supportive. His reputation helped me to obtain a post-doc at the University of Michigan with Doug Torr in May 1978 where Andy Nagy introduced me to his 2-stream photoelectron code. For some reason, Doug had great faith in my ability and set challenging research goals that greatly expanded my scientific horizons. At the University of Michigan Space Physics Research Laboratory in those days, it was common practice to submit a meeting abstract before the investigation had even started, possibly as a motivator for post-docs to work long hours. Fortunately, we usually managed to avoid embarrassing performances at meetings.

One such example was the submittal of an abstract in August 1979 for a meeting in Canberra, Australia to be held in December 1979. The abstract described the reevaluation of the solar EUV heating efficiency, which entailed the daunting work of updating the model to include a photoelectron calculation, comprehensive photochemistry, and additional dynamics of minor species. The solar EUV heating efficiency was important for efficient neutral gas heating calculations in global thermosphere models. These heating efficiency calculations were CPU intensive and had to be done remotely on the CRAY-1 computer located in Boulder CO. Just before getting on the plane to Australia the results were output on line-printer paper and the EUV heating efficiency was evaluated using a calculator on the flight to Australia. Our EUV heating efficiencies were different from previous ones in important ways but were later confirmed by others.

On another occasion, chance discussions in a hallway at the 1990 fall American Geophysical Union meeting led to a solution to a vexing problem for ionosphere modeling. By then, the basic ion chemistry had been established from laboratory and Atmosphere Explorer satellite measurements. However, there was great uncertainty in the ion dynamics because of a lack of knowledge of the thermosphere neutral winds that affect the modeled electron density and temperature. It is not possible to accurately model the electron density profile without first accurately reproducing the observed hmF2. The neutral wind determines the height of the ionosphere peak electron density (hmF2). Without an accurate thermal electron density profile, the heating by photoelectrons was not accurate and therefore the electron and ion temperature could not be modeled accurately either. The key insight was to have an ionosphere model assimilate hmF2 as a proxy for the neutral wind (Richards, 1991). The peak height and peak density (NmF2) have been well-measured globally for many decades using ground-based ionosondes. An algorithm was developed to modify the neutral winds to cause the model to closely follow the observed hmF2 automatically as it stepped in time. Just as with the early photoelectron work, this procedure was greeted with much skepticism but has now become widely accepted. Together with the earlier EUV and photoelectron work, this algorithm enabled more accurate studies of the electron density and temperature as a function of altitude. The neutral winds produced by the algorithm are also a useful by-product of the algorithm, greatly increasing the amount of neutral wind data available for other studies.

Another fortuitous collaboration that connected all these ideas together occurred in 1996 during a 3-months sabbatical visit to LaTrobe University. Peter Dyson's group had only just finished extracting high-quality hmF2 and NmF2 data from their ionosonde. They also had accurate optical measurements of the neutral wind and temperature. These data enabled a detailed test of the photoelectron and EUVAC models and the neutral wind algorithm (Dyson et al., 1997; Richards et al., 1998).

There remain some vexing problems in ionosphere modeling. While the modeling of the quiet midlatitude F-region ionosphere is now reasonably mature, more dynamic conditions at high and equatorial latitudes and during geomagnetic storms still present a major challenge primarily due to the difficulty in accounting for the rapid changes in the key inputs such as neutral densities and winds, and electric fields. It is likely that data assimilation will be necessary for further improvement of the modeling of these more dynamic conditions and that will require improved data availability on smaller scales.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: No restrictions. The data are available from the author. Requests to access these datasets should be directed to Phil Richards, pgrichds@gmail.com.

Author contributions

PR performed all research on this perspective and wrote the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Acknowledgments

The author thanks the colleagues mentioned in this article as well as the many colleagues in the space science community who supported and critiqued my research. Special thanks to Andy Nagy who generously shared his photoelectron model and expertise and to Bill Peterson for being a great collaborator and for reviewing this article prior to submission.

References

Bailey, S. M., Woods, T. N., Barth, C. A., Solomon, S. C., Canfield, L. R., Korde, R., et al. (2000). Measurements of the solar soft X-ray irradiance by the Student Nitric Oxide Explorer: First analysis and underflight calibrations. *J. Geophys. Res.* 105 (27), 27179–27193. doi:10.1029/2000ja000188

Dyson, P. L., Davies, T. P., Richards, P. G., Parkinson, M. L., Reeves, A. J., Fairchild, C. E., et al. (1997). Thermospheric neutral winds at southern mid-latitudes: A comparison of optical and ionosonde$h_mF_2$methods. *J. Geophys. Res.* 102, 27189–27196. doi:10.1029/97ja02138

Hinteregger, H. E. (1981). Representations of solar EUV fluxes for aeronomical applications. *Adv. Space Res.* 1, 39–52. doi:10.1016/0273-1177(81)90416-6

Nagy, A. F., and Banks, P. M. (1970). Photoelectron fluxes in the ionosphere. *J. Geophys. Res.* 75, 6260–6270. doi:10.1029/ja075i031p06260

Peterson, W. K. (2021). Perspective on energetic and thermal atmospheric photoelectrons. *Front. Astron. Space Sci.* 8. doi:10.3389/fspas.2021.655309

Peterson, W. K., Woods, T. N., Fontenla, J. M., Richards, P. G., Chamberlin, P. C., Solomon, S. C., et al. (2012). Solar EUV and XUV energy input to thermosphere on solar rotation time scales derived from photoelectron observations. *J. Geophys. Res.* 117, A05320. doi:10.1029/2011JA017382

Richards, P. G. (1991). An improved algorithm for determining neutral winds from the height of the F2 peak electron density. *J. Geophys. Res.* 96, 17839. doi:10.1029/91ja01467

Richards, P. G., Dyson, P. L., Davies, T. P., Parkinson, M. L., and Reeves, A. J. (1998). Behavior of the ionosphere and thermosphere at a southern midlatitude station during magnetic storms in early March 1995. *J. Geophys. Res.* 103 (26), 26421–26432. doi:10.1029/97ja03342

Richards, P. G., Fennelly, J. A., and Torr, D. G. (1994). EUVAC: A solar EUV flux model for aeronomic calculations. *J. Geophys. Res.* 99, 8981. doi:10.1029/94ja00518

Richards, P. G., and Peterson, W. K. (2008). Measured and modeled backscatter of ionospheric photoelectron fluxes. *J. Geophys. Res.* 113, A08321. doi:10.1029/2008JA013092

Richards, P. G., and Torr, D. G. (1981). A formula for calculating the theoretical photoelectron fluxes resulting from the He^+ 304 Å solar spectral line. *Geophys. Res. Lett.* 8, 995–998. doi:10.1029/gl008i009p00995

Richards, P. G., and Torr, D. G. (1984). An investigation of the consistency of the ionospheric measurements of the photoelectron flux and solar EUV flux. *J. Geophys. Res.* 89, 5625. doi:10.1029/ja089ia07p05625

Solomon, S. C., Bailey, S. M., and Woods, T. N. (2001). Effect of solar soft X-rays on the lower ionosphere. *Geophys. Res. Lett.* 28, 2149–2152. doi:10.1029/2001gl012866

Woods, T. N., and Eparvier, F. G. (2006). Solar ultraviolet variability during the TIMED mission. *Adv. Space Res.* 37 (2), 219–224. doi:10.1016/j.asr.2004.10.006

frontiers | Frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 08 August 2022
doi: 10.3389/fspas.2022.957776

A Story of Developing the Idea of Plasma-Sheet Flow Braking

Kazuo Shiokawa *

Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Japan

This paper reports a story of developing the idea of Earthward ion flow braking in the near-Earth plasma sheet and its relationship with substorm onset processes. This idea and the data that support it are the basis for today's two competing models for substorms: the near-Earth neutral-line model and the current disruption model. The idea was developed when the author was staying at the Max Planck Institute for Extraterrestrial Physics (MPE) from July 1996 to June 1997. The story addresses the colleagues and mentors who had contributed to the development of this idea. The lessons learned from this story are also summarized for students and early-career scientists for their development of new scientific ideas.

Keywords: development of new idea, flow braking, plasma sheet, reconnection, magnetotail

Edited by:
Elena E. Grigorenko, Space Research Institute (RAS), Russia

Reviewed by:
Evgeny Panov, Austrian Academy of Sciences, Austria

Correspondence:
Kazuo Shiokawa
shiokawa@nagoya-u.jp

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *31 May 2022*
Accepted: *13 June 2022*
Published: *08 August 2022*

Citation:
Shiokawa K (2022) A Story of Developing the Idea of Plasma-Sheet Flow Braking. Front. Astron. Space Sci. 9:957776. doi: 10.3389/fspas.2022.957776

INTRODUCTION

The study on the braking of high-speed ion flow in the near-Earth plasma sheet (Shiokawa et al., 1997) and a subsequent study on the relationship of flow braking with substorm onset processes (Shiokawa et al., 1998a) are one of the several important steps to understand plasma processes during substorms in the Earth's magnetosphere. These two works were carried out when the author was staying at the Max Planck Institute for Extraterrestrial Physics (MPE) from July 1996 to June 1997 as an overseas researcher supported by the Ministry of Education, Science, Sports, and Culture, Japan. In this short article, we would like to introduce a story of when we developed the idea of flow braking in MPE, in order to clarify how this idea was developed and to address the colleagues and mentors who had contributed to this idea. We hope this story is helpful for students and early-career scientists for their development of new scientific ideas.

MY BACKGROUND BEFORE STARTING THE MAGNETOSPHERE STUDY

I was graduated with a bachelor course (March 1988) and a graduate (master) course (March 1990) from Tohoku University, Japan, under the supervision of Prof. Hiroshi Fukunishi. After that, I joined the Solar-Terrestrial Environment Laboratory (STEL), Nagoya University, in April 1990, as a research assistant, working mainly with Associate Prof. Kiyohumi Yumoto, for optical and magnetic field measurements at ground stations during the Solar-Terrestrial Energy Program (STEP, 1990–1997) operated by the Scientific Committee on Solar-Terrestrial Physics (SCOSTEP). I obtained a PhD in 1994 by combining the works in Tohoku University and STEL, on the topic of auroral electrons and ions using data from Antarctic rocket experiments and Defense Meteorological Satellite Program (DMSP) satellites (Shiokawa et al., 1990a; Shiokawa and Fukunishi, 1991; Shiokawa and Yumoto, 1993). In these works, we estimated the density and temperature of auroral electrons in the source magnetosphere by fitting the accelerated Maxwellian distribution function to the observed auroral electron spectra. I also developed a two-stream transportation code of auroral electrons to calculate electron spectra and auroral emissions in the thermosphere and ionosphere from input of precipitating electron distribution (Shiokawa and Fukunishi, 1990).

In 1996, I received an opportunity for an overseas researcher from the Ministry. In this opportunity, I could choose any overseas institution to stay for 1 year. There were two choices: one was to extend my research to the auroral energy dissipation in the thermosphere by using the electron transport code with Dr. Stan Solomon of the University of Colorado because I learned a lot from his study (Solomon et al., 1988) when I developed my auroral electron code. The other choice was to extend my research to the magnetosphere because my past research was focused on estimating the density and temperature of magnetospheric electrons using data from the ionosphere. We can directly compare the estimated density and temperature with those from direct measurements by magnetospheric satellites and possibly identify the source of auroras in the magnetosphere. I consulted Prof. Yosuke Kamide in STEL about these two possibilities, who had many experiences in international collaboration. Prof. Kamide introduced me to Dr. Wolfgang Baumjohann of MPE, as a possible host researcher on the magnetospheric study. Hence, I decided to stay at MPE with Dr. Baumjohann.

DEVELOPMENT OF THE FLOW BRAKING IDEA

I joined the MPE in July 1996. The magnetospheric satellite data I used for the analysis were the data from the Active Magnetospheric Particle Tracer Explorers/Ion Release Module (AMPTE/IRM) satellite. At the beginning of the study, Dr. Baumjohann suggested me to look into the three-dimensional distribution function of Earthward high-speed ion flow in the plasma sheet, in order to identify the evidence of magnetic reconnection in the magnetotail. This topic was extensively studied later by the Japanese Geotail satellite and made significant progress in understanding the magnetic reconnection processes (e.g., Nagai et al., 1998; Hoshino et al., 2001). Thus, Dr. Baumjohann had an excellent perspective on this direction prior to these extensive studies. However, I hesitated to move forward with this suggestion. I had been a co-investigator of the Geotail mission since 1990, and I knew that there were many excellent scientists who had studied magnetic reconnection. I thought that I should not take the same research direction with these smart scientists. Hence, I asked Dr. Baumjohann to give me time to look into the AMPTE/IRM dataset in detail as I am a beginner for magnetospheric physics. Dr. Baumjohann had already developed a well-organized database of the AMPTE/IRM particle and magnetic field and a Fortran code set to plot them (Baumjohann et al., 1988; Baumjohann et al. 1989; Baumjohann et al. 1990). This database and the code set were really easy for me to prepare various plots of magnetospheric plasma and field features and helped me to develop the idea of what happens in the magnetotail plasma sheet.

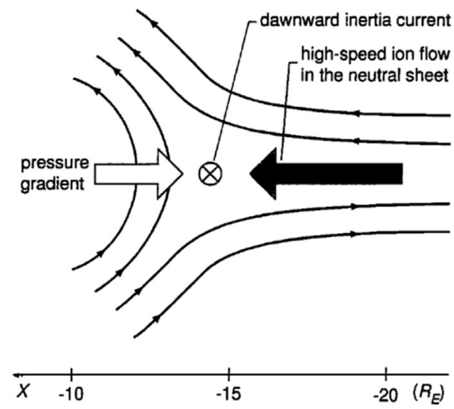

FIGURE 1 | Schematic picture of the proposed magnetic field configuration during Earthward high-speed ion flow (looking from dusk to dawn). The stopping point of the flow is the boundary between dipolar and tail-like magnetic fields which correspond to the inner edge of the neutral sheet. The boundary is formed by the downward inertia current that is caused by braking of the Earthward ion flow and by the pileup of northward magnetic flux carried by the flow (Shiokawa et al., 1997).

Looking into the AMPTE/IRM data, the occurrence rate of earthward flow decreased from ~4 to ~1% as the satellite moved closer to the Earth from 20 Re to 10 Re (**Figure 1A** of Shiokawa et al., 1997). Hence, the question as to how high-speed earthward ion flows stop arose. This question might arise because I was trying to avoid the reconnection topic. Also, I was not a good student of Tohoku University where lectures on magneto-hydro-dynamics (MHD) were made. Then, I started re-learning about MHD using a textbook by Nicholson (1983) and eventually understood that the flow must be stopped by tailward pressure gradient forces, because the plasma and magnetic pressures increase as the flow gets closer to the Earth. But the interesting point was that when the flow stops, there will be a downward inertia current that creates a clear boundary of dipole-like and tail-like magnetic field configuration, as shown in **Figure 1**. This boundary idea suddenly came up to me in the morning in bed, and I said to my wife, “I got a good idea” at 4 a.m. in the morning. This flow-braking process is similar to the process where solar wind hits Earth’s magnetosphere. Then, a magnetopause is formed with a clear boundary of magnetic field intensity due to the magnetopause current (inertia current). Using the AMPTE/IRM data, I confirmed that Earthward ion flows cannot propagate more than a few Re under the average tailward pressure gradient force in the plasma sheet (Figure 3A of Shiokawa et al., 1997).

I was really glad to obtain this idea and discussed it with the senior scientists in MPE, that is, Drs. Götz Paschmann, Manfred Scholer, Nova Scopke, and Rudolf A. Treumann (Dr. Baumjohann was on travel at that time). Dr. Treumann suggested me the possibility that the flow can diverge to dawn/duskward or north/southward, like a river water flow hitting a rock. So, I checked dawn/dusk and north/south velocities (Vy and Vz) in the AMPTE/IRM data and did not find any particular enhancement. But this suggestion helped my

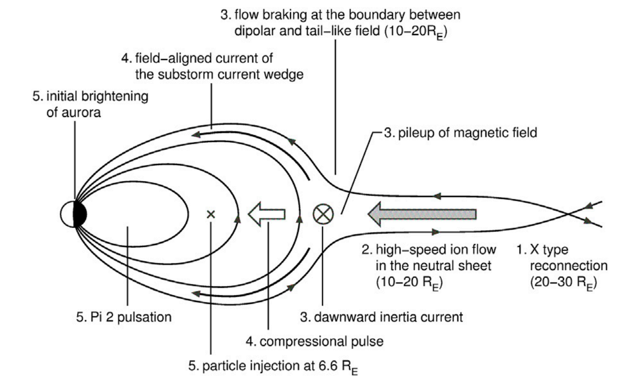

FIGURE 2 | Model proposed by Shiokawa et al. (1998a) for the current wedge formation during the initial stage of the substorm expansion phase. The number preceding each comment indicates the order of occurrence. The braking point of the Earthward flow moves tailward due to the pileup of magnetic field carried by the flow. This motion probably corresponds to the poleward expansion of aurora. The transition from tail-like field to dipolar field at the braking point is due to the downward inertia current. The substorm current wedge is formed at the braking point by the inertia current and the current due to flow shear. The fluctuation of field-aligned currents and the compressional pulses generated at the braking point can be the cause of the Pi2 pulsation in the inner magnetosphere.

understanding of the fluid behavior of magnetospheric plasma. Finally, I discussed the idea with Prof. Gerhard Haerendel, the director of MPE. After explaining my idea, he smiled at me and said to me, "Did you read my paper published in 1992?" Of course, I was lazy and did not search related past literature. He gave me Haerendel (1992) and said to me, "I wrote this paper when I got a heart attack. So this is my heart-attack paper." Haerendel (1992) theoretically discussed the processes of flow braking and its generation of inertia current. My data analysis eventually proved his idea using the AMPTE/IRM data. Haerendel (1992) also pointed out that this process can be a course of substorm current wedge. This paper led me to connect the flow braking idea with the substorm topic.

CONNECTION BETWEEN FLOW BRAKING AND SUBSTORM PROCESSES

After completing the flow-braking article (Shiokawa et al., 1997) in October 1996, it was rather straightforward for me to investigate the timing difference of Earthward flow and substorm onset processes. In that idea development, an excellent review of substorm controversy by McPherron (1995) was beneficial to me. Chapter 13.7 of this review article pointed out several difficulties of pre-existing substorm models to explain observation facts of substorm. For example, in the near-Earth neutral line model, the reconnection (flow reversal) was observed at 20–30 Re in the tail. On the other hand, the auroral breakup at the substorm onset occurs at low latitudes that map to the tail current sheet just outside of 6.6 Re. The flow braking idea can explain this discrepancy by providing an additional point between the reconnection site (20–30 Re) and Earth (1 Re). The flow braking point at ~10 Re can be a magnetospheric source of the auroral breakup at the substorm onset, as shown in **Figure 2** (Shiokawa et al., 1998a). The downward inertia current driven by flow braking can drive the substorm current wedge, as discussed by Haerendel (1992).

The Earthward ion flow with a speed of 400 km/s takes a few minutes to travel from the reconnection region (20–30 Re) to the flow braking region (10–20 Re). If we investigate the timing difference between the flow and auroral breakup, we can identify whether the flow (and reconnection) occurs before or after the auroral breakup. Thus, I collected substorm onset signatures in the ground and satellite data for a substorm-associated flow event on 1 March 1985, as observed by AMPTE/IRM. For this particular event, the onset of Earthward high-speed flow was observed 3 min before the onset of the global current wedge formation. From this observation, we concluded that the substorm current wedge was caused at the braking point of the Earthward high-speed flow during the initial stage of the substorm expansion phase and drew a schematic figure of the substorm onset sequence as shown in **Figure 2** (Shiokawa et al., 1998a, submitted on February 1997).

After this proposal of the substorm model by Shiokawa et al. (1998a), the onset mechanism models seemed to converge into the two major models, that is, the near-Earth neutral line (NENL) model (outside-in model, for e.g., Baker et al., 1996; Shiokawa et al., 1998a) and the current disruption model (inside-out model) (e.g., Lui, 2001). Then, the Time History of Events and Macroscale Interactions during Substorms (THEMIS) mission was proposed to identify the controversy of these models (Angelopoulos, 2008). The current understanding is that both models can work at a substorm onset, although which one is more significant is still unresolved (See Lui, 2015 versus Nagai and Shinohara, 2021 for the present state of the controversy). But I could not contribute much to these developments of new missions and subsequent substorm discussion. One of the reasons is that during my stay in MPE, STEL had obtained a new big budget from the Ministry to construct the Optical Mesosphere Thermosphere Imagers (OMTIs, Shiokawa et al., 1999) to measure airglow and aurora using multiple cameras and interferometers. I was responsible for this project, so I became really busy for ground-based multi-point measurements. The other reason may be that I tried to avoid the scientific topic that many smart scientists were studying.

After these two works in MPE, we (Shiokawa, Haerendel, and Baumjohann) also published one more article on azimuthal pressure gradient during substorms (Shiokawa et al., 1998b) to complete the substorm current budget because flow braking processes were clearly not sufficient to drive the total amount of field-aligned currents during substorms (Angelopoulos et al., 1994). In this work, I again tried to prove one of the many theoretical ideas of Haerendel (1990) using the AMPTE/IRM satellite data. My initial motivation to stay at MPE (magnetosphere-ionosphere coupling study) was finally published as Shiokawa et al. (2000) to compare the electron density and temperature estimated from ionospheric DMSP satellites with those directly measured by AMPTE/IRM and as Shiokawa et al. (2003) to show bi-directional field-aligned electrons observed by AMPTE/IRM possibly coming from the ionosphere and/or generated by magnetospheric Fermi acceleration.

DISCUSSION AND CONCLUSION

There were several lessons learned from my one-year stay at MPE which may be helpful for students and early-career scientists.

1. It is better to avoid taking the same research direction as other smart scientists.
2. It is better to take your class lectures more seriously when you are a student.
3. But, you can learn any time when you are interested in a topic. Maybe that is the best time to learn it.
4. It is better to choose an excellent team/institution when you go abroad for collaborative research. The team/institution should be slightly different from what you are currently doing to extend your research to wider fields.

The MPE team was the best team for me to start studying on the magnetospheric physics, that is, a well-organized database of AMPTE/IRM developed by Dr. Baumjohann and the AMPTE/IRM team and an excellent mentor Prof. Haerendel, who provided me background physics and indicated a new research direction. Actually, Dr. Vassilis Angelopoulos, who has been the principal investigator of the THEMIS mission, had also obtained his PhD in the study of bursty bulk flow in collaboration with MPE (e.g., Angelopoulos et al., 1992; Angelopoulos et al., 1994). These works were carried out just before my stay at MPE. MPE was an excellent institute where many active scientists from various countries joined and interacted with each other.

I sometimes remember the other possible choice to stay at the University of Colorado with Dr. Stan Solomon because my current research on ground-based measurements of airglow and aurora requires comparison with thermospheric modeling which Dr. Solomon developed. If I had had another chance to stay for 1 year abroad, I would have stayed at the University of Colorado to work with Dr. Solomon and his colleagues.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

KS contributed to the conception, design, and writing of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

The author's research at MPE was supported by the Overseas Researcher Scholar (8-Wakate-70) of the Ministry of Education, Science, Sports and Culture, Japan.

ACKNOWLEDGMENTS

The author is grateful to the editorial board of the perspective article for the research topic "Generation-to-Generation Communications in Space Physics" for providing the opportunity to write the article. His work at MPE was supported by the Overseas Researcher Fund (Monbusho Zaigai Kenkyuin) of the Ministry of Education, Science, Sports and Culture, Japan. He also thanks all the MPE colleagues who supported him and made his stay in MPE enjoyable. Not only making the scientific research, but he also enjoyed playing soccer every Wednesday and Friday and sometimes playing tennis with Dr. Baumjohann. He joined a city table tennis club for which Thomas Bauer, his MPE colleague, introduced him, and made a lot of trips to Germany and other European countries every weekend with his wife, Miwa. The stay in Germany was really a memorable 1 year in his life.

REFERENCES

Angelopoulos, V., Baumjohann, W., Kennel, C. F., Coroniti, F. V., Kivelson, M. G., Pellat, R., et al. (1992). Bursty Bulk Flows in the Inner Central Plasma Sheet. *J. Geophys. Res.* 97, 4027–4039. doi:10.1029/91ja02701

Angelopoulos, V., Kennel, C. F., Coroniti, F. V., Pellat, R., Kivelson, M. G., Walker, R. J., et al. (1994). Statistical Characteristics of Bursty Bulk Flow Events. *J. Geophys. Res.* 99 (21), 21257–21280. doi:10.1029/94ja01263

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

Baker, D. N., Pulkkinen, T. I., Angelopoulos, V., Baumjohann, W., and McPherron, R. L. (1996). Neural Line Model of Substorms: Past Results and Present View. *J. Geophys. Res.* 101 (12), 12,975–13,010. doi:10.1029/95ja03753

Baumjohann, W., Paschmann, G., and Cattell, C. A. (1989). Average Plasma Properties in the Central Plasma Sheet. *J. Geophys. Res.* 94, 6597–6606. doi:10.1029/ja094ia06p06597

Baumjohann, W., Paschmann, G., and Lühr, H. (1990). Characteristics of High-Speed Ion Flows in the Plasma Sheet. *J. Geophys. Res.* 95, 3801–3809. doi:10.1029/ja095ia04p03801

Baumjohann, W., Paschmann, G., Sckopke, N., Cattell, C. A., and Carlson, C. W. (1988). Average Ion Moments in the Plasma Sheet Boundary Layer. *J. Geophys. Res.* 93 (11), 11507–11520. doi:10.1029/ja093ia10p11507

Haerendel, G. (1992). "Disruption, Ballooning or Auroral Avalanche - on the Cause of Substorms," in *Proc. Of the First International Conference on Substorms (ICS-1)* (Kiruna, Sweden: ESA), 417–420.

Haerendel, G. (1990). "Field-aligned Currents in the Earth's Magnetosphere," in *Physics of Magnetic Flux Ropes*. Editors C. T. Russell, E. R. Priest, and L. C. Lee (Washington, D. C: AGU), 58, 539–553. Geophys. Monogr.~Ser. doi:10.1029/gm058p0539

Hoshino, M., Mukai, T., Terasawa, T., and Shinohara, I. (2001). Suprathermal Electron Acceleration in Magnetic Reconnection. *J. Geophys. Res.* 106 (A11), 25979–25997. doi:10.1029/2001ja900052

Lui, A. T. Y. (2001). A Multiscale Model for Substorms. *Space Sci. Rev.* 95, 325–345. doi:10.1023/a:1005217304749

Lui, A. T. Y. (2015). "Magnetospheric Substorm Onset by Current Disruption Processes," in *Auroral Dynamics and Space Weather: Understanding and Applications*. Editors Y. Zhang and L. J. Paxton (Washington, DC: American Geophysical Union). doi:10.1002/9781118978719.ch12

McPherron, R. L. (1995). "Magnetospheric Dynamics," in *Introduction to Space Physics*. Editors M. G. Kivelson and C. T. Russell (Cambridge University Press), 400–458. doi:10.1017/9781139878296.014

Nagai, T., Fujimoto, M., Saito, Y., Machida, S., Terasawa, T., Nakamura, R., et al. (1998). Structure and Dynamics of Magnetic Reconnection for Substorm Onsets with Geotail Observations. *J. Geophys. Res.* 103 (97), 4419–4440. doi:10.1029/97ja02190

Nagai, T., and Shinohara, I. (2021). Dawn-dusk Confinement of Magnetotail Reconnection Site in the Near-Earth Magnetotail and its Implication for Depolarization and Substorm Current System. *J. Geophys. Res. Space Phys.* 126, e2021JA026691. doi:10.1029/2021ja029691

Nicholson, D. R. (1983). *Introduction to Plasma Theory*. New York: John Wiley & Sons.

Shiokawa, K., Baumjohann, W., and Haerendel, G. (1997). Braking of High-Speed Flows in the Near-Earth Tail. *Geophys. Res. Lett.* 24, 1179–1182. doi:10.1029/97gl01062

Shiokawa, K., Baumjohann, W., Haerendel, G., and Fukunishi, H. (2000). High- and Low-Altitude Observations of Adiabatic Parameters Associated with Auroral Electron Acceleration. *J. Geophys. Res.* 105, 2541–2550. doi:10.1029/1999ja900458

Shiokawa, K., Baumjohann, W., Haerendel, G., Paschmann, G., Fennell, J. F., Friis-Christensen, E., et al. (1998a). High-speed Ion Flow, Substorm Current Wedge, and Multiple Pi 2 Pulsations. *J. Geophys. Res.* 103, 4491–4507. doi:10.1029/97ja01680

Shiokawa, K., Baumjohann, W., and Paschmann, G. (2003). Bi-directional Electrons in the Near-Earth Plasma Sheet. *Ann. Geophys.* 21, 1497–1507. doi:10.5194/angeo-21-1497-2003

Shiokawa, K., and Fukunishi, H. (1990). “Dependences of Auroral 5577-A and 6300-A Emission Rates on Thermospheric Density Variations,” in *Proc. NIPR Symposium on Coordinated Observations of the Ionosphere and the Magnetosphere in the Polar Regions* (Tokyo: National Institute of Polar Research NIPR), 3, 24–31.

Shiokawa, K., and Fukunishi, H. (1991). Global Characteristics of Field-Aligned Acceleration Processes Associated with Auroral Arcs. *J. Geomagn. geoelec* 43, 691–719. doi:10.5636/jgg.43.691

Shiokawa, K., Fukunishi, H., Yamagishi, H., Miyaoka, H., Fujii, R., and Tohyama, F. (1990a). Rocket Observation of the Magnetosphere-Ionosphere Coupling Processes in Quiet and Active Arcs. *J. Geophys. Res.* 95 (10), 10,679–10,686. doi:10.1029/ja095ia07p10679

Shiokawa, K., Haerendel, G., and Baumjohann, W. (1998b). Azimuthal Pressure Gradient as Driving Force of Substorm Currents. *Geophys. Res. Lett.* 25, 959–962. doi:10.1029/98gl00540

Shiokawa, K., and Yumoto, K. (1993). Global Characteristics of Particle Precipitation and Field-Aligned Electron Acceleration during Isolated Substorms. *J. Geophys. Res.* 98, 1359–1375. (Correction, J. Geophys. Res., 98, 9357-9362, 1993). doi:10.1029/92ja01092

Shiokawa, K., Katoh, Y., Satoh, M., Ejiri, M. K., Ogawa, T., Nakamura, T., et al. (1999). Development of Optical Mesosphere Thermosphere Imagers (OMTI). *Earth, Planets, and Space* 51, 887–896.

Solomon, S. C., Hays, P. B., and Abreu, V. J. (1988). The Auroral 6300 Å Emission: Observations and Modeling. *J. Geophys. Res.* 93 (A9), 9867–9882. doi:10.1029/JA093iA09p09867

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 15 August 2022
DOI 10.3389/fspas.2022.957007

Check for updates

OPEN ACCESS

EDITED BY
Jorge Luis Chau,
Leibniz Institute of Atmospheric Physics (LG), Germany

REVIEWED BY
Erdal Yiğit,
George Mason University, United States

*CORRESPONDENCE
N. M. Pedatella,
nickp@ucar.edu

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 30 May 2022
ACCEPTED 18 July 2022
PUBLISHED 15 August 2022

CITATION
Pedatella NM (2022), Scientific motivations and future directions of whole atmosphere modeling.
Front. Astron. Space Sci. 9:957007.
doi: 10.3389/fspas.2022.957007

Scientific motivations and future directions of whole atmosphere modeling

N. M. Pedatella[1,2]*

[1]High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO, United States, [2]COSMIC Program Office, University Corporation for Atmospheric Research Boulder, Boulder, CO, United States

The recent development of whole atmosphere models that extend from the surface to the upper thermosphere represents a significant advance in modeling capabilities of the ionosphere-thermosphere. Whole atmosphere models have had an especially important influence on understanding the role of terrestrial weather on generating variability in the ionosphere-thermosphere. This paper provides an overview of the scientific motivations and contributions made by whole atmosphere modeling. This is followed by a discussion of future directions in whole atmosphere modeling and the science that they will enable.

KEYWORDS

whole atmosphere model, ionosphere, thermosphere, atmosphere coupling, space weather

1 Introduction

The importance of terrestrial weather on the ionosphere-thermosphere began to be widely recognized in the past few decades. Although a number of researchers had previously explored the role of the lower atmosphere on generating ionosphere-thermosphere variability [e.g., (Chen, 1992; Stening et al., 1996; Forbes et al., 2000; Rishbeth and Mendillo, 2001)], it is only more recently that this coupling is widely understood to be an important source of variability in the ionosphere-thermosphere. The considerable progress that has been made in this area can be found in a number of recent reviews (England, 2012; Pancheva and Mukhtarov, 2012; Liu, 2016; Yiğit et al., 2016; Sassi et al., 2019; Goncharenko et al., 2021; Ward et al., 2021). The increased recognition of the lower atmosphere effects on the ionosphere-thermosphere served as an important motivator for the development of whole atmosphere models, herein considered to be those that seamlessly span altitudes from the surface to the upper thermosphere (~500 km). In addition to observational investigations [e.g., (Immel et al., 2006; Chau et al., 2009; Goncharenko et al., 2010; Gasperini et al., 2020; Goncharenko et al., 2022)], the development of whole atmosphere models played a crucial role in understanding the physical mechanisms by which terrestrial weather impacts the ionosphere-thermosphere, and further confirmed the importance of the lower atmosphere on generating ionosphere-thermosphere variability.

This Perspective discusses the role of whole atmosphere models in ionosphere-thermosphere research and the future directions of whole atmosphere modeling. The focus is primarily on their role in understanding the impact of terrestrial weather on the ionosphere-thermosphere. Following a brief background on the initial development of whole atmosphere models, recent scientific progress enabled by whole atmosphere models is discussed. This is followed by a personal vision for the future of whole atmosphere model development and the science that these developments will enable.

2 Development of whole atmosphere models

A brief historical overview of the development of whole atmosphere models is first warranted in order to provide context for both recent advances and future developments. For a more detailed discussion, including what is involved in the development of a whole atmosphere model, the reader is referred to Akmaev (2011). Roble (2000) significantly advanced the concept of a whole atmosphere model by coupling together a model developed for the lower atmosphere (NCAR Community Climate Model, CCM3) with one developed for the upper atmosphere (Thermosphere-Ionosphere-Mesosphere-Electrodynamics General Circulation Model, TIME-GCM). Though viewed by Roble (2000) as a "feasibility study to determine just how processes in the lower atmosphere affect the upper atmosphere", the exploratory model proved to be highly valuable with regards to understanding coupling processes between the lower and upper atmospheres [e.g., (Liu and Roble, 2002; Mendillo et al., 2002)]. This success led to the subsequent development of several stand-alone whole atmosphere models by researchers around the world, including the Ground-to-topside model of Atmosphere and Ionosphere for Aeronomy [GAIA, (Jin et al., 2011),], HI Altitude Mechanistic general Circulation Model [HIAMCM, (Becker and Vadas, 2020)], Whole Atmosphere Model [WAM, (Akmaev et al., 2008)], and the Whole Atmosphere Community Climate Model with thermosphere-ionosphere extension [WACCM-X, (Liu et al., 2018)]. Note that in the present context, whole atmosphere models are considered those that are comprised of a single model that seamlessly extends from the surface to the upper thermosphere. Though not the focus of the present paper, it is important to recognize that a variety of other models that are not considered whole atmosphere models by this definition have also contributed significantly to the understanding of how variability in the lower atmosphere is imprinted on the middle and upper atmospheres. This includes models with upper boundaries in the lower thermosphere (~100–250 km) and models that extend into the ionosphere-thermosphere but not all the way down to the surface, as well as one- and two-dimensional models [e.g., (Hagan and Forbes, 2003; Hagan et al., 2007; Hickey et al., 2009; Qian et al., 2009; Vadas and Liu, 2009; Yiğit et al., 2009)].

One aspect of the coupled CCM3-TIMEGCM that is important to highlight is that it leveraged decades of historical model developments in what could be considered to be disparate communities. Specifically, it could not have been realized without the prior (generally separate) developments that occurred in climate and ionosphere-thermosphere modeling. Middle atmosphere models, those extending up to the mesosphere-lower thermosphere, were also of fundamental importance. Whole atmosphere models thus require expertise across a range of disciplines. Future developments will continue to require wide-ranging expertise, including atmospheric scientists, software engineers, and space physicists. While this can, at times, present a challenge, I have personally found that it makes working with whole atmosphere models full of opportunities to broaden one's perspective and learn significantly from those with a variety of areas of expertise.

3 Science enabled by whole atmosphere models

Despite their recent development, whole atmosphere models have already made significant contributions to ionosphere-thermosphere research. Some of the areas where whole atmosphere models have led to new scientific understanding are discussed below. Note that what follows is focused on the role of the lower atmosphere on generating ionosphere-thermosphere variability and it is not intended to be an exhaustive list of the scientific applications of whole atmosphere models. Other scientific topics of relevance to whole atmosphere models include long-term trends (Solomon et al., 2019; Cnossen and Maute, 2020; Liu et al., 2020) and the solar influence on chemistry and climate. All of these areas remain active areas of research and will continue to see progress with the continued development of whole atmosphere models.

The influence of SSWs on the ionosphere-thermosphere was one of the first scientific applications of stand-alone whole atmosphere models (Wang et al., 2011; Fang et al., 2012; Jin et al., 2012; Pedatella et al., 2012). Detailed discussion of the contributions of whole atmosphere models in the understanding of the coupling mechanisms between SSWs and the ionosphere-thermosphere can be found in Goncharenko et al. (2021). Notably, whole atmosphere model simulations advanced understanding of the variability of different solar and lunar tides during SSWs and their role in generating ionosphere-thermosphere variability. Another important contribution was the finding that using data assimilation systems to initialize whole atmosphere model forecasts could lead to forecasting the SSW effects on the ionosphere ~10 days in advance (Wang et al., 2014; Pedatella et al., 2018). This demonstrates the potential increased space weather forecast skill that may be obtained by incorporating lower atmospheric effects, especially during periods of strong lower atmospheric forcing.

Observational studies have long shown that the lower atmosphere contributes a significant fraction of the day-to-day ionosphere variability. The advent of whole atmosphere models has helped to quantify the variability in the ionosphere-thermosphere that is driven by the lower atmosphere. Model simulations by Fang et al. (2018) have shown that the lower atmosphere contributes ~10–20% of the ionosphere variability. This is generally consistent with prior observational estimates, demonstrating that whole atmosphere models can reasonably represent the day-to-day variability of the ionosphere. Additional modeling studies have shown that there exists large day-to-day variability in atmospheric tides and planetary waves and that these are likely to be the source of the persistent day-to-day variability in the ionosphere [e.g., (Jin et al., 2011; McDonald et al., 2018; Gasperini et al., 2020; Liu, 2020)].

Though whole atmosphere models are typically run at a relatively coarse resolution (1-2°), the development of models with resolutions of 0.25–0.50° has enabled investigation into smaller scale variability. High resolution whole atmosphere model simulations have led to new understanding of the pathways by which gravity waves reach the thermosphere where they can imprint themselves on the ionosphere by generating traveling ionospheric disturbances (TIDs) at middle latitudes and plasma instabilities in the equatorial region. Though previous research investigated the impacts of gravity waves on the thermosphere [e.g., (Vadas and Fritts, 2004; Vadas and Liu, 2009; Yiğit et al., 2009; Yiğit and Medvedev, 2009)], high resolution whole atmosphere model simulations by Vadas and Becker (2019) and Becker and Vadas (2020) provided new insight into the gravity waves reaching the thermosphere, which include an important contribution from secondary and higher-order waves that are generated by the momentum deposition that results from wave breaking. Complete understanding of this process for gravity waves to reach the thermosphere would be difficult without high resolution whole atmosphere modeling owing to the difficulty in observing gravity waves throughout their full altitude range. The capability to simulate small-scale waves in the thermosphere enabled by high resolution models further enables simulations of small-scale structures in the ionosphere, such as TIDs and equatorial instabilities (Miyoshi et al., 2018; Huba and Liu, 2020).

An important feature of whole atmosphere models is that they can simultaneously capture ionosphere-thermosphere variability that is driven by the lower atmosphere as well as variability due to solar and geomagnetic activity. This is critical as solar and geomagnetically driven variability occurs on top of the background state of the ionosphere-thermosphere, which is in-part controlled by waves propagating upwards from the lower atmosphere. Previous studies (Hagan et al., 2015; Pedatella, 2016) found that incorporation of lower atmospheric effects can significantly alter the simulated response to a geomagnetic storm. This was confirmed in the context of a whole atmosphere model by Pedatella and Liu (2018), who found that regional differences in the ionosphere-thermosphere response to a geomagnetic storm can reach 50–100% due to lower atmospheric effects.

4 Future of whole atmosphere modeling

Whole atmosphere models will continue to play a critical role in enabling scientific understanding of the ionosphere-thermosphere system. It is likely that they will also have an increasing role operationally, as evidenced by the recent implementation of the NOAA WAM for operational space weather forecasting (https://www.swpc.noaa.gov/products/wam-ipe). Here I outline a number of areas for advances in model development along with how they will facilitate advances in ionosphere-thermosphere research and space weather operations.

Fully capturing the range of spatial scales that influence the ionosphere-thermosphere requires high-resolution whole atmosphere models. Initial high-resolution whole atmosphere model simulations with horizontal resolutions of ~0.25–0.50° have shown the profound influence of small-scale waves on the thermosphere-ionosphere, including the generation of equatorial ionosphere instabilities (Huba and Liu, 2020). Such high-resolution capabilities have only been developed in the past several years and have yet to be fully exploited in terms of understanding the influence of atmospheric waves of various scale sizes on the ionosphere-thermosphere. At the same time, it is also crucial to continue advancing the development of high-resolution modeling capabilities. Current models rely on hydrostatic dynamical cores, which inherently limits their ability to simulate the full extent of the waves that influence the ionosphere-thermosphere. This can be addressed by adopting non-hydrostatic dynamical cores, though incorporating a non-hydrostatic dynamical core is nontrivial owing in-part to the need to control dynamical instabilities (Griffin and Thuburn, 2018). Minimizing unphysical noise, for example through hyperdiffusion or hyperviscosity (Dennis et al., 2012; Ullrich et al., 2018), is also critical to separate real wave variability from unphysical noise. The development of new dynamical cores allows for regionally refined grids, enabling extremely high resolutions [O (5–10 km)] over specific areas within a coarser resolution global grid. Regionally refined grids have yet to be employed in whole atmosphere models, though they are likely the only feasible approach to obtain resolutions on the order of 5–10 km within the context of a global model in the foreseeable future. Important scientific questions that can be addressed through the continued development and application of high-resolution whole atmosphere models include cross-scale wave coupling processes and the mechanisms responsible for the day-to-day variability of small-scale ionospheric structures, such as TIDs and equatorial irregularities.

High-resolution simulations will continue to be inhibited by their computational demands, restricting their applications to simulation lengths on the order of years. There will thus continue to be a need for whole atmosphere model configurations with coarser resolutions (~1-2° degrees) for certain applications (e.g., long-term trends, multi-year climatological studies, etc.). These resolutions necessitate parameterization of the atmospheric gravity waves that influence the middle and upper atmospheres. Though critical for reproducing the mean state of the middle and upper atmosphere, gravity wave parameterizations remain a significant source of uncertainty in whole atmosphere models (Pedatella et al., 2014). This is partly due to the fact that many existing gravity wave parameterizations rely on a number of assumptions, such as strictly vertical and instantaneous propagation, that are known to be incorrect. Updated gravity wave parameterization schemes may alleviate some of the uncertainty due to gravity wave parameterizations [e.g., (Yiğit et al., 2008; Bölöni et al., 2021)]. They additionally neglect secondary and higher-order waves that are now thought to have an increasingly important role at higher altitudes (Becker and Vadas, 2020). It is important to note that even high-resolution whole atmosphere models will continue to rely on parameterized processes for the near future. This is due to the fact that certain processes, such as convective generation of gravity waves, wave dissipation, and mixing, will continue to be on sub-grid scales. Development of improved parameterization schemes for both high- and low-resolution whole atmosphere models will therefore be necessary to address existing uncertainties in whole atmosphere models.

While a number of data assimilation systems have been developed that extend into the lower thermosphere [e.g., (Wang et al., 2011; Eckermann et al., 2018; Koshin et al., 2020)], a true whole atmosphere data assimilation system that assimilates observations from the surface to the ionosphere-thermosphere has only recently been realized (Pedatella et al., 2020). There thus remains considerable room for improvement in current data assimilation capabilities for whole atmosphere models. Data assimilation systems have been extensively used for numerical weather prediction (NWP), again providing the opportunity to leverage the extensive prior developments in a different discipline. However, data assimilation systems need to be tailored to the specific demands of whole atmosphere models owing to differences between the troposphere-stratosphere and ionosphere-thermosphere. Important differences that influence the data assimilation system include different spatial and temporal scales of the dynamical variability, greater influence of external driving in the ionosphere-thermosphere compared to the troposphere, less understanding of model error characteristics in the ionosphere-thermosphere, and the relative sparsity of observations compared to the troposphere. Dealing with unbalanced adjustments, which can generate spurious waves, will also be critical due to the large wave growth with altitude. The development of high-quality whole atmosphere data assimilation systems will provide the opportunity to advance a wide-range of scientific areas of interest to the space physics community, much in the way that atmospheric reanalysis products (e.g., ERA-5, MERRA2) are widely used across the atmospheric science research community. Furthermore, whole atmosphere data assimilation systems can provide initial conditions for space weather forecasting, enabling the study of the predictability and forecast skill of the ionosphere-thermosphere, an area that is vastly understudied in the authors opinion. Development of whole atmosphere data assimilation systems is also critical for operational space weather forecasting, especially for forecasting the day-to-day variability of the ionosphere-thermosphere during periods of quiet solar and geomagnetic activity.

Though slightly outside the primary focus of the present article, it should be noted that improvements to the specification of high-latitude forcing in whole atmosphere models are also required. Whole atmosphere models currently typically rely on empirical specifications of the high-latitude electric potential and auroral precipitation that are known to be deficient. Improvements in the high-latitude forcing may be realized through data-driven approaches, such as the Assimilative Mapping of Ionosphere Electrodynamics [AMIE, (Richmond and Kamide, 1988)] and Assimilative Mapping of Geospace Observations [AMGeO, (Matsuo, 2020)]. Coupling with a magnetospheric model is an alternative approach, and has proven to be beneficial for improving the high-latitude forcing specification in ionosphere-thermosphere simulations (Wang et al., 2004; Pham et al., 2022). Additionally, the current capability of whole atmosphere models to simulate the effects of particle precipitation on the chemistry of the middle atmosphere is inhibited by large uncertainties in the particle precipitation [e.g., (Nesse Tyssøy et al., 2022; Sinnhuber et al., 2022)]. Improved specifications of particle precipitation will enable better representation of solar influences on chemistry and climate.

It is important to recognize that although the above advances in whole atmosphere modeling and data assimilation capabilities will themselves enable new understanding of the ionosphere-thermosphere, it remains important to continually assess the fidelity of model simulations. This entails both confronting the model with observations as well as performing detailed inter-model comparisons. Such comparisons provide crucial insight into model shortcomings and can help identify areas for future development. Observational verification of whole atmosphere models is especially critical; however, it is inhibited by the deficiency of observations, especially in the thermosphere. A robust observing system is thus essential for ensuring the continued advancement of whole atmosphere models.

5 Conclusion

The development of whole atmosphere models have significantly advanced our understanding of the influence of the lower atmosphere on the ionosphere-thermosphere across a range of temporal and spatial scales. The advances outlined above will serve to advance our existing modeling capabilities, leading to new understanding of the processes that generate ionosphere-thermosphere variability. Some of the important scientific topics that can be addressed with advanced whole atmosphere modeling capabilities include: 1) the influence of terrestrial weather on the day-to-day variability of the ionosphere, including TIDs and equatorial irregularities; 2) cross-scale coupling between small and large scale waves; 3) long-term trends; 4) predictability of the ionosphere-thermosphere; 5) interaction between lower atmosphere and solar/geomagnetic driven variability; and 6) solar influences on atmospheric chemistry and climate. Advances in whole atmosphere modeling will thus enable new understanding across a range scientific areas, demonstrating the need to continue advancing current modeling capabilities. They may additionally serve to improve operational space weather forecasts.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the United States National Science Foundation under Cooperative Agreement 1852977.

Acknowledgments

I thank the many colleagues and collaborators—too many to name—for many discussions related to atmospheric coupling and whole atmosphere modeling. These discussions significantly influenced my views on the subject of whole atmosphere modeling.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor JC declared a past co-authorship with the author NP.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akmaev, R. A., Fuller-Rowell, T. J., Wu, F., Forbes, J. M., Zhang, X., Anghel, A. F., et al. (2008). Tidal variability in the lower thermosphere: Comparison of Whole Atmosphere Model (WAM) simulations with observations from TIMED. *Geophys. Res. Lett.* 35, L03810. doi:10.1029/2007GL032584

Akmaev, R. A. (2011). Whole atmosphere modeling: Connecting terrestrial and space weather. *Rev. Geophys.* 49, RG4004. doi:10.1029/2011RG000364

Becker, E., and Vadas, S. L. (2020). Explicit global simulation of gravity waves in the thermosphere. *J. Geophys. Res. Space Phys.* 125. doi:10.1029/2020JA028034

Bölöni, G., Kim, Y.-H., Borchert, S., and Achatz, U. (2021). Toward transient subgrid-scale gravity wave representation in atmospheric models. Part I: Propagation model including nondissipative wave–mean-flow interactions. *J. Atmos. Sci.* 78, 1317–1338. doi:10.1175/JAS-D-20-0065.1

Chau, J. L., Fejer, B. G., and Goncharenko, L. P. (2009). Quiet variability of equatorial E × B drifts during a sudden stratospheric warming event. *Geophys. Res. Lett.* 36, L05101. doi:10.1029/2008GL036785

Chen, P.-R. (1992). Two-day oscillation of the equatorial ionization anomaly. *J. Geophys. Res.* 97, 6343–6357. doi:10.1029/91JA02445

Cnossen, I., and Maute, A. (2020). Simulated trends in ionosphere-thermosphere climate due to predicted main magnetic field changes from 2015 to 2065. *J. Geophys. Res. Space Phys.* 125, e2019JA027738. doi:10.1029/2019JA027738

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., et al. (2012). CAM-SE: A scalable spectral element dynamical core for the community atmosphere model. *Int. J. High. Perform. Comput. Appl.* 26, 74–89. doi:10.1177/1094342011428142

Eckermann, S. D., Ma, J., Hoppel, K. W., Kuhl, D. D., Allen, D. R., Doyle, J. A., et al. (2018). High-altitude (0–100 km) global atmospheric reanalysis system: Description and application to the 2014 austral winter of the deep propagating gravity wave experiment (DEEPWAVE). *Mon. Weather Rev.* 146, 2639–2666. doi:10.1175/MWR-D-17-0386.1

England, S. L. (2012). A review of the effects of non-migrating atmospheric tides on the earth's low-latitude ionosphere. *Space Sci. Rev.* 168, 211–236. doi:10.1007/s11214-011-9842-4

Fang, T.-W., Fuller-Rowell, T., Akmaev, R., Wu, F., Wang, H., and Anderson, D. (2012). Longitudinal variation of ionospheric vertical drifts during the 2009 sudden stratospheric warming. *J. Geophys. Res.* 117. doi:10.1029/2011JA017348

Fang, T.-W., Fuller-Rowell, T., Yudin, V., Matsuo, T., and Viereck, R. (2018). Quantifying the sources of ionosphere day-to-day variability. *J. Geophys. Res. Space Phys.* 123, 9682–9696. doi:10.1029/2018JA025525

Forbes, J. M., Palo, S. E., and Zhang, X. (2000). Variability of the ionosphere. *J. Atmos. Sol. Terr. Phys.* 62, 685–693. doi:10.1016/S1364-6826(00)00029-8

Gasperini, F., Liu, H., and McInerney, J. (2020). Preliminary evidence of madden-julian oscillation effects on ultrafast tropical waves in the thermosphere. *J. Geophys. Res. Space Phys.* 125, e2019JA027649. doi:10.1029/2019JA027649

Goncharenko, L. P., Chau, J. L., Liu, H.-L., and Coster, A. J. (2010). Unexpected connections between the stratosphere and ionosphere. *Geophys. Res. Lett.* 37. doi:10.1029/2010GL043125

Goncharenko, L. P., Harvey, V. L., Liu, H., and Pedatella, N. M. (2021). *Sudden stratospheric warming impacts on the ionosphere–thermosphere system*. Washington, DC: American Geophysical Union (AGU, 369–400. chap. 16. doi:10.1002/9781119815617.ch16

Goncharenko, L. P., Harvey, V. L., Randall, C. E., Coster, A. J., Zhang, S.-R., Zalizovski, A., et al. (2022). Observations of Pole-to-Pole, stratosphere-to-ionosphere connection. *Front. Astron. Space Sci.* 8. doi:10.3389/fspas.2021.768629

Griffin, D. J., and Thuburn, J. (2018). Numerical effects on vertical wave propagation in deep-atmosphere models. *Q. J. R. Meteorol. Soc.* 144, 567–580. doi:10.1002/qj.3229

Hagan, M. E., and Forbes, J. M. (2003). Migrating and nonmigrating semidiurnal tides in the upper atmosphere excited by tropospheric latent heat release. *J. Geophys. Res.* 108. doi:10.1029/2002JA009466

Hagan, M. E., Häusler, K., Lu, G., Forbes, J. M., and Zhang, X. (2015). Upper thermospheric responses to forcing from above and below during 1–10 April 2010: Results from an ensemble of numerical simulations. *J. Geophys. Res. Space Phys.* 120, 3160–3174. doi:10.1002/2014JA020706

Hagan, M. E., Maute, A., Roble, R. G., Richmond, A. D., Immel, T. J., and England, S. L. (2007). Connections between deep tropical clouds and the Earth's ionosphere. *Geophys. Res. Lett.* 34, L20109. doi:10.1029/2007GL030142

Hickey, M. P., Schubert, G., and Walterscheid, R. L. (2009). Propagation of tsunami-driven gravity waves into the thermosphere and ionosphere. *J. Geophys. Res.* 114. doi:10.1029/2009JA014105

Huba, J. D., and Liu, H.-L. (2020). Global modeling of equatorial spread F with Sami3/WACCM-X. *Geophys. Res. Lett.* 47, e2020GL088258. doi:10.1029/2020GL088258

Immel, T. J., Sagawa, E., England, S. L., Henderson, S. B., Hagan, M. E., Mende, S. B., et al. (2006). Control of equatorial ionospheric morphology by atmospheric tides. *Geophys. Res. Lett.* 33, L15108. doi:10.1029/2006GL026161

Jin, H., Miyoshi, Y., Fujiwara, H., Shinagawa, H., Terada, K., Terada, N., et al. (2011). Vertical connection from the tropospheric activities to the ionospheric longitudinal structure simulated by a new Earth's whole atmosphere-ionosphere coupled model. *J. Geophys. Res.* 116. doi:10.1029/2010JA015925

Jin, H., Miyoshi, Y., Pancheva, D., Mukhtarov, P., Fujiwara, H., and Shinagawa, H. (2012). Response of migrating tides to the stratospheric sudden warming in 2009 and their effects on the ionosphere studied by a whole atmosphere-ionosphere model GAIA with COSMIC and TIMED/SABER observations. *J. Geophys. Res.* 117. doi:10.1029/2012JA017650

Koshin, D., Sato, K., Miyazaki, K., and Watanabe, S. (2020). An ensemble Kalman filter data assimilation system for the whole neutral atmosphere. *Geosci. Model. Dev.* 13, 3145–3177. doi:10.5194/gmd-13-3145-2020

Liu, H.-L., Bardeen, C. G., Foster, B. T., Lauritzen, P., Liu, J., Lu, G., et al. (2018). Development and validation of the whole atmosphere community climate model with thermosphere and ionosphere extension (WACCM-X 2.0). *J. Adv. Model. Earth Syst.* 10, 381–402. doi:10.1002/2017MS001232

Liu, H.-L. (2020). Day-to-Day variability of prereversal enhancement in the vertical ion drift in response to large-scale forcing from the lower atmosphere. *Space weather.* 18, e2019SW002334. doi:10.1029/2019SW002334

Liu, H.-L., and Roble, R. G. (2002). A study of a self-generated stratospheric sudden warming and its mesospheric–lower thermospheric impacts using the coupled TIME-GCM/CCM3. *J. -Geophys. -Res.* 107 (D23), 4695. doi:10.1029/2001JD001533

Liu, H.-L. (2016). Variability and predictability of the space environment as related to lower atmosphere forcing. *Space weather.* 14, 634–658. doi:10.1002/2016SW001450

Liu, H., Tao, C., Jin, H., and Nakamoto, Y. (2020). Circulation and tides in a cooler upper atmosphere: Dynamical effects of CO2 doubling. *Geophys. Res. Lett.* 47, e2020GL087413. doi:10.1029/2020GL087413

Matsuo, T. (2020). *Recent progress on inverse and data assimilation procedure for high-latitude ionospheric Electrodynamics*. Cham: Springer International Publishing, 219–232. doi:10.1007/978-3-030-26732-2_10

McDonald, S. E., Sassi, F., Tate, J., McCormack, J., Kuhl, D. D., Drob, D. P., et al. (2018). Impact of non-migrating tides on the low latitude ionosphere during a sudden stratospheric warming event in January 2010. *J. Atmos. Sol. Terr. Phys.* 171, 188–200. doi:10.1016/j.jastp.2017.09.012

Mendillo, M., Rishbeth, H., Roble, R. G., and Wroten, J. (2002). Modelling F2-layer seasonal trends and day-to-day variability driven by coupling with the lower atmosphere. *J. Atmos. Sol. Terr. Phys.* 64, 1911–1931. doi:10.1016/S1364-6826(02)00193-1

Miyoshi, Y., Jin, H., Fujiwara, H., and Shinagawa, H. (2018). Numerical study of traveling ionospheric disturbances generated by an upward propagating gravity wave. *J. Geophys. Res. Space Phys.* 123, 2141–2155. doi:10.1002/2017JA025110

Nesse Tyssøy, H., Sinnhuber, M., Asikainen, T., Bender, S., Clilverd, M. A., Funke, B., et al. (2022). HEPPA III intercomparison experiment on electron precipitation impacts: 1. Estimated ionization rates during a geomagnetic active period in april 2010. *JGR. Space Phys.* 127, e2021JA029128. doi:10.1029/2021JA029128

Pancheva, D., and Mukhtarov, P. (2012). Global response of the ionosphere to atmospheric tides forced from below: Recent progress based on satellite measurements. *Space Sci. Rev.* 168, 175–209. doi:10.1007/s11214-011-9837-1

Pedatella, N. M., Anderson, J. L., Chen, C. H., Raeder, K., Liu, J., Liu, H.-L., et al. (2020). Assimilation of ionosphere observations in the whole atmosphere community climate model with thermosphere-ionosphere EXtension (WACCMX). *J. Geophys. Res. Space Phys.* 125, e2020JA028251. doi:10.1029/2020JA028251

Pedatella, N. M., Fuller-Rowell, T., Wang, H., Jin, H., Miyoshi, Y., Fujiwara, H., et al. (2014). The neutral dynamics during the 2009 sudden stratosphere warming simulated by different whole atmosphere models. *J. Geophys. Res. Space Phys.* 119, 1306–1324. doi:10.1002/2013JA019421

Pedatella, N. M. (2016). Impact of the lower atmosphere on the ionosphere response to a geomagnetic superstorm. *Geophys. Res. Lett.* 43, 9383–9389. doi:10.1002/2016GL070592

Pedatella, N. M., Liu, H.-L., Marsh, D. R., Raeder, K., Anderson, J. L., Chau, J. L., et al. (2018). Analysis and hindcast experiments of the 2009 sudden stratospheric warming in WACCMX+DART. *J. Geophys. Res. Space Phys.* 123, 3131–3153. doi:10.1002/2017JA025107

Pedatella, N. M., Liu, H.-L., Richmond, A. D., Maute, A., and Fang, T.-W. (2012). Simulations of solar and lunar tidal variability in the mesosphere and lower thermosphere during sudden stratosphere warmings and their influence on the low-latitude ionosphere. *J. Geophys. Res.* 117. doi:10.1029/2012JA017858

Pedatella, N. M., and Liu, H.-L. (2018). The influence of internal atmospheric variability on the ionosphere response to a geomagnetic storm. *Geophys. Res. Lett.* 45, 4578–4585. doi:10.1029/2018GL077867

Pham, K. H., Zhang, B., Sorathia, K., Dang, T., Wang, W., Merkin, V., et al. (2022). Thermospheric density perturbations produced by traveling atmospheric disturbances during august 2005 storm. *JGR. Space Phys.* 127, e2021JA030071. doi:10.1029/2021JA030071

Qian, L., Solomon, S. C., and Kane, T. J. (2009). Seasonal variation of thermospheric density and composition. *J. Geophys. Res.* 114. doi:10.1029/2008JA013643

Richmond, A. D., and Kamide, Y. (1988). Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique. *J. Geophys. Res.* 93, 5741–5759. doi:10.1029/JA093iA06p05741

Rishbeth, H., and Mendillo, M. (2001). Patterns of F2-layer variability. *J. Atmos. Sol. Terr. Phys.* 63, 1661–1680. doi:10.1016/S1364-6826(01)00036-0

Roble, R. G. (2000). *On the feasibility of developing a global atmospheric model extending from the Ground to the exosphere*. Washington, DC: American Geophysical Union (AGU, 53–67. doi:10.1029/GM123p0053

Sassi, F., McCormack, J. P., and McDonald, S. E. (2019). Whole atmosphere coupling on intraseasonal and interseasonal time scales: A potential source of increased predictive capability. *Radio Sci.* 54, 913–933. doi:10.1029/2019RS006847

Sinnhuber, M., Nesse Tyssøy, H., Asikainen, T., Bender, S., Funke, B., Hendrickx, K., et al. (2022). Heppa III intercomparison experiment on electron precipitation impacts: 2. Model-measurement intercomparison of nitric oxide (NO) during a geomagnetic storm in april 2010. *JGR. Space Phys.* 127, e2021JA029466. doi:10.1029/2021JA029466

Solomon, S. C., Liu, H.-L., Marsh, D. R., McInerney, J. M., Qian, L., and Vitt, F. M. (2019). Whole atmosphere climate change: Dependence on solar activity. *J. Geophys. Res. Space Phys.* 124, 3799–3809. doi:10.1029/2019JA026678

Stening, R. J., Meek, C. E., and Manson, A. H. (1996). Upper atmosphere wind systems during reverse equatorial electrojet events. *Geophys. Res. Lett.* 23, 3243–3246. doi:10.1029/96GL02611

Ullrich, P. A., Reynolds, D. R., Guerra, J. E., and Taylor, M. A. (2018). Impact and importance of hyperdiffusion on the spectral element method: A linear dispersion analysis. *J. Comput. Phys.* 375, 427–446. doi:10.1016/j.jcp.2018.06.035

Vadas, S. L., and Becker, E. (2019). Numerical modeling of the generation of tertiary gravity waves in the mesosphere and thermosphere during strong mountain wave events over the southern andes. *J. Geophys. Res. Space Phys.* 124, 7687–7718. doi:10.1029/2019JA026694

Vadas, S. L., and Fritts, D. C. (2004). Thermospheric responses to gravity waves arising from mesoscale convective complexes. *J. Atmos. Sol. Terr. Phys.* 66, 781–804. doi:10.1016/j.jastp.2004.01.025

Vadas, S. L., and Liu, H. l. (2009). Generation of large-scale gravity waves and neutral winds in the thermosphere from the dissipation of convectively generated gravity waves. *J. Geophys. Res.* 114. doi:10.1029/2009JA014108

Wang, H., Akmaev, R. A., Fang, T.-W., Fuller-Rowell, T. J., Wu, F., Maruyama, N., et al. (2014). First forecast of a sudden stratospheric warming with a coupled whole-atmosphere/ionosphere model IDEA. *J. Geophys. Res. Space Phys.* 119, 2079–2089. doi:10.1002/2013JA019481

Wang, H., Fuller-Rowell, T. J., Akmaev, R. A., Hu, M., Kleist, D. T., and Iredell, M. D. (2011). First simulations with a whole atmosphere data assimilation and forecast system: The January 2009 major sudden stratospheric warming. *J. Geophys. Res.* 116. doi:10.1029/2011JA017081

Wang, W., Wiltberger, M., Burns, A. G., Solomon, S. C., Killeen, T. L., Maruyama, N., et al. (2004). Initial results from the coupled magnetosphere–ionosphere–thermosphere model: Thermosphere–ionosphere responses. *J. Atmos. Sol. Terr. Phys.* 66, 1425–1441. doi:10.1016/j.jastp.2004.04.008

Ward, W., Seppälä, A., Yiğit, E., Nakamura, T., Stolle, C., Laštovička, J., et al. (2021). Role of the sun and the middle atmosphere/thermosphere/ionosphere in climate (ROSMIC): A retrospective and prospective view. *Prog. Earth Planet. Sci.* 8, 47. doi:10.1186/s40645-021-00433-8

Yiğit, E., Aylward, A. D., and Medvedev, A. S. (2008). Parameterization of the effects of vertically propagating gravity waves for thermosphere general circulation models: Sensitivity study. *J. Geophys. Res.* 113, D19106. doi:10.1029/2008JD010135

Yiğit, E., Koucká Knížová, P., Georgieva, K., and Ward, W. (2016). A review of vertical coupling in the Atmosphere–Ionosphere system: Effects of waves, sudden stratospheric warmings, space weather, and of solar activity. *J. Atmos. Sol. Terr. Phys.* 141, 1–12. doi:10.1016/j.jastp.2016.02.011

Yiğit, E., Medvedev, A. S., Aylward, A. D., Hartogh, P., and Harris, M. J. (2009). Modeling the effects of gravity wave momentum deposition on the general circulation above the turbopause. *J. Geophys. Res.* 114, D07101. doi:10.1029/2008JD011132

Yiğit, E., and Medvedev, A. S. (2009). Heating and cooling of the thermosphere by internal gravity waves. *Geophys. Res. Lett.* 36, L14807. doi:10.1029/2009GL038507

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 16 August 2022
DOI 10.3389/fspas.2022.934216

OPEN ACCESS

EDITED BY
Joseph E. Borovsky,
Space Science Institute, United States

REVIEWED BY
Michael Schulz,
Lockheed Martin Solar and Astrophysics Laboratory (LMSAL), United States
Robert McPherron,
California State University, Los Angeles, United States

*CORRESPONDENCE
Nikolai A. Tsyganenko,
nikolai.tsyganenko@gmail.com

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 02 May 2022
ACCEPTED 05 July 2022
PUBLISHED 16 August 2022

CITATION
Tsyganenko NA (2022), A lifetime with models, or toils and thrills of number crunching.
Front. Astron. Space Sci. 9:934216.
doi: 10.3389/fspas.2022.934216

A lifetime with models, or toils and thrills of number crunching

Nikolai A. Tsyganenko *

Laboratory of Magnetospheric Physics, Department of Physics, Saint-Petersburg State University, Saint-Petersburg, Russia

Modeling the magnetosphere from satellite data has some analogy to the construction of a person's 'identikit' from patchy information, collected by different people at different times under different lighting/viewing conditions. This article is a brief overview of the author's half-century-long efforts in building Earth's magnetosphere models, with some feats, setbacks, and lessons learned.

KEYWORDS

magnetosphere, modeling, satellite data, magnetopause, geomagnetic storms

1 Early pursuits, first steps and missteps

In the spring of 1968, being then a fourth-year physics student, I considered changing my major from quantum mechanics to something more earthly. One day, while hanging around our department I came across a poster for a habilitation thesis defense by a certain Mikhail Pudovkin. The title, "Morphology of Polar Storms", stirred associations with blizzards, howling winds, icebergs, and Klondike stories by Jack London. As I learned at the event, the dissertation subject was not atmospheric, but geomagnetic storms. It was just there that I first heard the magic terms 'magnetosphere', 'auroral precipitation', 'pitch angle', etc. All that stuff sounded exciting, and Pudovkin himself turned out to be quite a pleasant person, which finally put an end to my vacillations. He agreed to supervise my work and suggested to explore the role of plasma isotropization in the substorm triggering. First calculations with a pancake distribution function gave a wrong polarity of the tail current; to understand the paradox I resorted to tracing particle orbits, then dug into the plasma–field equilibrium issues. However naive those early efforts seem now, they helped me get initial experience and resulted in a publication with my mentor (Pudovkin and Tsyganenko, 1973).

For my PhD work, Pudovkin suggested to calculate magnetospheric configurations using the then existing information of the distant field and magnetopause size/shape. It was just then that he uttered his fateful motto which shaped my future as a geophysicist: "Everybody is drawing the magnetosphere, but nobody knows where the field lines go". While paying due tribute to the ground-based experiment, he was well aware of the growing role of *in-situ* observations. As the space age just started, lack of data had to be offset by theoretical estimates; nevertheless the situation gradually improved. By mid-1970s, our laboratory grew in size and became an internationally recognized team. Even though most of its staff could not travel beyond the Iron Curtain, we were often visited by scientists from all over the world. In the fall of 1968, Syun-Ichi Akasofu gave a talk at our seminar on conjugate aurora observations under the Northern and Southern ovals. In the following years, the lab was visited by Alex Dessler, Keith Cole, Alv Egeland, Takesi Nagata, Leif Svalgaard, and other distinguished scientists. Such informal contacts helped exchange ideas and, sometimes, get the needed data. In mid-1970s, being impressed by a paper of Mead and Fairfield (Mead and Fairfield, 1975), I sent them a request for IMP magnetometer data and, to my delight, shortly received a pair of hefty IBM card decks. The favorable reply was a pleasant surprise and, much more important, a good lesson in what was coined later in the AGU's slogan "unselfish cooperation in science".

As time went on, it became clear that new data called for a new mathematical framework: realistic and flexible, on the one hand, and computationally simple/fast, on the other. A host of problems emerged; one of the hurdles was to properly represent the magnetotail field. First attempts to find closed solutions failed, while the numerical approach was unacceptable as computationally expensive. Eventually a simple idea emerged: instead of deriving the field from the current, start from the outset with a simplest field from a diffuse wire. Its integration along the tail with variable intensity factors yielded the long-sought solution, embodied in the TU82 model (Tsyganenko and Usmanov, 1982) and published with my then PhD student and now lifetime friend Arcadi Usmanov. In the next T87 model (Tsyganenko, 1987), the validity region was extended through the Moon's orbit owing to newly added distant IMP data. Providentially, one of reviewers of that paper was David Stern of GSFC, who happened to play an outstanding role in my career a few years later.

In terms of mistakes and lessons, an instructive episode is worth mentioning. In 1981, I got interested in revisiting the magnetotail equilibrium: the idea was to numerically trace proton orbits and use the obtained scattering matrix to derive steady-state pitch-angle distributions. They turned out isotropic, except for strongly non-adiabatic regimes. The latter result lingered in the back of my mind as counterintuitive, but I ventured to publish the work anyway (Tsyganenko, 1982). Three years later, a paper appeared (Wright, 1985) under almost the same title, with an elegant proof that the stationary scattering must always provide strictly isotropic distributions. Somewhat embarrassed, I went back to my tracing code and found the culprit of the paradox in using single precision instead of double. On a positive side, however, that work had far-reaching ramifications: the sharp boundary between the particle motion regimes was realized as a powerful means to probe the magnetospheric configuration and resulted in a joint paper (Sergeev and Tsyganenko, 1982) with my lifelong colleague Victor Sergeev, followed by whole series of his seminal works on the subject (e.g., (Sergeev et al., 1983), (Sergeev et al., 1993), (Sergeev et al., 2015) and refs. therein).

2 Further developments

The next step forward was made in the end of 1980s, which also deserves a brief recounting. The TU82/T87 models shared a common shortfall: the straight tail current with rectilinear inner edge resulted in artificial "pockets" of depressed field near the dawn/dusk magnetopause. In reality, the innermost tail current gradually curves and smoothly joins the ring current. That suggested to explore axisymmetric vector potentials with azimuthal component A_ϕ satisfying Poisson's equation in cylindrical coordinates $\{\rho, \phi, z\}$. A particular solution had already been derived for a Jovian disk model (Connerney and Ness, 1981) as a Fourier-Bessel convolution of an ad hoc current profile. However, models of that kind required lengthy numerical integrations and, hence, were out of the question. An effective workaround was eventually discovered, based on required profile of the magnetic field depression $B_z(\rho, a)$, peaking at origin and falling off to zero at infinity with a spatial scale a. The corresponding 3D magnetic field was obtained by deriving a convolution kernel function and then integrating that kernel with the adopted $B_z(\rho, a)$. To my delight, the integrations yielded an incredibly simple analytic vector potential $A_\phi(\rho, a)$. Moreover, its successive derivatives $\partial^i A_\phi/\partial a^i$ provided a whole family of independent potentials with progressively steeper radial fall-off. Their simplicity was amazing, and I still do not know any easier way to derive them, other than to scramble through the direct and inverse integral transformations. The T89 model (Tsyganenko, 1989a) based on those results is still the fastest and widely used in studies where the computation speed is critical, e.g., tracing high-energy solar protons. No wonder that the T89 citation count (now 1,370) was reported among the highest (Abt, 2000); looking back through many years, that work brought me the most satisfaction ever. In this regard, I cannot but gratefully remember another person whose advice stuck in my mind since my freshman/sophomore years. That was late Vladimir Buslaev, then our mathematical physics professor who taught us students to savor the beauty of equations, not to fear of getting lost in long calculations, not to give up too early, and to carefully check/recheck everything with no stones left unturned. Here I would like to add my own more specific advice: in all calculations, try to invent and apply various tests based on fundamental laws. In particular, an unfailing bug-hunting weapon is to make sure that the model $\nabla \cdot \mathbf{B}$ is zero.

Another important task was to consistently combine the internal field sources with magnetopause currents, such that the total field be confined (shielded) within the boundary. In particular, a conceptual hurdle persisted with the magnetotail, where the shielding currents also had to at least partially contribute to the plasma sheet closure currents. Still in early theoretical speculations, the tail currents were believed to close at high latitudes via theta-shaped dual vortex loops. However, 'wire' models based on that notion (Olson and Pfitzer, 1974) could not accurately confine the field inside the magnetopause. An alternative solution was just to extend the equatorial current beyond the magnetosphere and simulate the magnetic effect of closure currents by adding ad hoc curl-free fields. In spite of lack of explicitly defined magnetopause, such models (TU82, T87, T89) still provided fairly realistic configurations, owing to their flexibility and reliance on large amounts of spacecraft data. The magnetopause appeared in those models as a *de facto* comet-shaped surface separating the field lines with and without topological connection to Earth.

3 Goddard years (1992–2007)

At the end of 1980s, the advent of perestroika brought major upheavals, both on the large societal scale and in my personal life and scientific activities. In the spring of 1989, a joint US–USSR Space Science workshop was held in Moscow, and it was there that my first personal meeting with David Stern took place. Owing to his efforts, I was offered an NRC fellow position at GSFC, which started in February 1992 and opened a new 15-years long page in my personal life and magnetospheric modeling studies.

By that time, enough data of direct magnetopause crossings had been collected, which allowed to create first empirical models of the boundary ((Sibeck et al., 1991), (Roelof and Sibeck, 1993)) and explicitly introduce the magnetopause in the data-based models. The main task was to find a shielding method, not limited to specific boundary shapes. Previous models used simple surfaces that allowed to expand the shielding potential into series of eigenfunctions. Only a few such surfaces satisfied that criterion, in particular, a paraboloid ((Alexeev and Shabansky, 1972), (Stern, 1985)) or an ellipsoid (Tsyganenko, 1989b); a composite model was also developed (Voigt, 1981) combining a cylinder with a hemisphere. All such models shared a common deficiency: lack of flexibility. A breakthrough idea was conceived by Schulz and McNab (Schulz and McNab, 1987): instead of seeking an exact solution of Neumann's problem, they suggested to derive the shielding field by minimizing the residual flux across the 'source-surface' boundary. An advantage of that method was that, instead of a limited number of analytic boundaries, it allowed to represent the magnetopause by any suitable surface and tap the immense variety of curl-free shielding fields. In the data-based modeling, that method was first implemented in the T96 model ((Tsyganenko, 1995), (Tsyganenko et al., 1996)) and serves since then as the unfailing workhorse.

Another major complication stemmed from that the geodipole axis is tilted by $\approx 10°$ to the planet's spin axis, which is in turn inclined by 23.44° to the ecliptic polar axis. This results in diurnal/seasonal deformations of the magnetosphere, which had to be somehow replicated by the models. A general method was eventally developed and implemented (Tsyganenko, 1998), based on a powerful field deformation technique, first introduced in an earlier seminal paper by Stern (Stern, 1987). Its essence is to properly modify the original coordinates and accordingly transform the field vector, keeping it divergence-free. Adding a few variable parameters in the deformation makes it flexible, which allows to extract the dipole tilt effects from data.

Incidentally, the observed tilt-related effects raise an interesting physics question. Still in early studies (Russell and Brody, 1967), it was found that the deformed tail current assumes a trough-like shape, elevated at midnight but depressed at flanks or vice versa. Much later (Arridge et al., 2008), a similar effect was detected in Saturn's magnetosphere; the most interesting feature was the bowl-shaped deformation of the equatorial current, shifted away from the magnetic equator in the same direction at all longitudes. At first sight, that might appear as a direct "blowing-off" by the solar wind; however, since the solar wind cannot penetrate inside the magnetosphere, some other factors should be at work. A simple explanation was proposed in a study made with my student Varya Andreeva (Tsyganenko and Andreeva, 2014), based on calculating the inverse mirror ratio $\varepsilon = B_{\min}/B$ in a vacuum model with fully shielded tilted dipole. Even though the model did not include any internal currents, the obtained ε distributions were found to be strikingly similar to the observed bowl-shaped current. In the limit of a "pole-on" magnetosphere with 90° tilt angle, the ε pattern turned into an axisymmetric surface resembling the simulated current [e.g. (Eggington et al., 2020)]. Therefore, the bowl-shaped deformation is just a combined effect of the North-South asymmetry induced by the dipole tilt and the day-night asymmetry due to the solar wind flow.

The principal goal of the empirical modeling is to reconstruct the observed or expected state of a system using the entire body of experimental information. Because of the complexity of Sun-Earth interactions, the main problem is to optimally relate the model parameters to the state of the magnetosphere and its external drivers. Due to lack of the solar wind data, early models were binned into consecutive intervals of the Kp index. After mid-1990s, a more or less continuous monitoring of the solar wind began, which allowed to parameterize the models using both interplanetary and ground-based input. The original approach was to represent the field as a sum of terms (modules) associated with basic current systems, relate their coefficients to external drivers and/or ground indices and fit the model to the entire database. It was shortly realized that the magnetospheric inertia is important, such that not only the driving intensity, but also its previous history was a factor. A simple analogy with the Dst index prediction (Burton et al., 1975) prompted to represent the individual module magnitudes as a result of continuous competition between the external pumping and internal dissipation. That concept was implemented in our joint work with Misha Sitnov (Tsyganenko and Sitnov, 2005) and resulted in a popular storm-time TS05 model.

My following collaboration with Misha resulted in two important developments, of which the first one was to remove the assumption of axial symmetry in the T89 tail disk current and, instead of seeking simple analytical solutions, represent the convolution integrals by discrete series over a set of radial wave numbers and azimuthal harmonics. Because of lack of axial symmetry, the problem could no longer be solved easily by using a vector potential. Instead, we had to first define two separate scalar potentials northward and southward from the current sheet, and then convert them into a single vector potential using a method discovered by luck in a paper by Stern (Stern, 1987). In its final formulation (Tsyganenko and Sitnov, 2007) the new TS07 model represented the field by double Fourier-Bessel series with free coefficients. This opened the way to construct new-generation models with (theoretically) unlimited resolution in the radial distance and longitude angle.

The second upgrade area (Sitnov et al., 2008) dealt with the model parameterization and resulted in a complete revision of the previous approach. In essence, the new method reincarnated the original data binning used in early models, but raised it to a sophisticated level rooted in the formal systems theory. Namely, the system's evolution was modeled by following its trajectory in the state space and selecting data from a limited volume around the moment of interest. As a result, the magnetosphere dynamics was reconstructed by creating a sequence of 'instantaneous' submodels, each fitted to relatively small data subsets, mined from the whole database by means of the 'nearest-neighbor' (NN) technique. In tandem with the high-resolution TS07 framework, the NN data mining (DM) revealed many important features previously beyond the reach of empirical models. That marked a genuine leap forward (if not a revolution) in the empirical modeling of the magnetosphere, which already produced a number of interesting results (e.g. (Sitnov et al., 2020), (Tsyganenko et al., 2021a), and references therein).

4 Quo vadis?

In the summer of 2007, after 15 years at Goddard I moved back to my alma mater to accept a teaching position and continue my modeling studies. In that year, five Themis satellites were launched, adding to the flurry of data the Cluster tetrahedron has already been contributing since its launch in 2001. The fast buildup of magnetospheric data archives opened new prospects for the modeling and called for even more powerful methods. To that end, an idea emerged to construct a high-resolution model from finite elements covering not only the equatorial plane as in the TS07, but extending into the entire 3D space. Another innovation was to make the building blocks local, which allowed to easily shrink the model's field of view to a limited area of interest. A feasible solution was eventually found, based on representing the magnetic field as a sum of toroidal and poloidal parts and expanding their generating functions into sums of radial/cylindrical basis functions (RBF/CBF), each centered around a 3D system of nodes. As a realization of that idea, a new model (Tsyganenko and Andreeva, 2016) was constructed, followed by several local studies focused on isolated regions around dayside cusps (Tsyganenko and Andreeva, 2018) or low-latitude magnetosphere (Tsyganenko and Andreeva, 2020). A potentially promising approach (Tsyganenko and Andreeva, 2017), conceived in parallel with the purely RBF/CBF-based models, was to combine the latter with old modular frameworks within hybrid architectures, in which the modular part is fitted to data first, and then the RBF/CBF part is determined, playing thus a role of higher-order correction term.

The page limits do not allow for a more detailed coverage of the subject. An important topic with many interesting aspects that I had to completely leave outside this memoir, is the spacecraft data and methods of their ingestion into the empirical models. A more or less detailed coverage of that subject can be found in a recent review (Tsyganenko et al., 2021b).

In summary (and keeping in mind primary goals of this Perspective issue), I would like to briefly formulate my view of the future prospects of the data-based modeling. Unlike in meteorology, the basic curse of space weather forecasting is lack and irregularity of data from very limited number of satellites probing the magnetosphere at a given moment. As already convincingly demonstrated ((Sitnov et al., 2020), (Stephens et al., 2020)), that shortage can be effectively solved using the DM and deep-learning methods, based on tapping huge resources of archived data. First-principle simulations on the other hand, while rapidly improving in their realism, often provide a quantitatively biased view of what is actually going on in the magnetosphere. A highly promising solution would be to synergistically combine the two approaches, using the DM-based empirical models as a source of virtual assimilation data.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

The author's studies in the field of magnetospheric modeling have been supported by the Russian Foundation for Basic Research grants 17-05-00415 and 20-05-00218.

Acknowledgments

The author acknowledges support of RFBR grant 20–05-00218.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abt, H. A. (2000). The most frequently cited astronomical papers published during the past decade. *Bull. Am. Astronomical Soc.* 32, 937. https://adsabs.harvard.edu/full/2000BAAS. . .32..937A.

Alexeev, I. I., and Shabansky, V. P. (1972). A model of a magnetic field in the geomagnetosphere. *Planet. Space Sci.* 20 (1), 117–133. doi:10.1016/0032-0633(72)90146-8

Arridge, C. S., Russell, C. T., Khurana, K. K., Achilleos, N., Cowley, S. W. H., Dougherty, M. K., et al. (2008). Saturn's magnetodisc current sheet. *J. Geophys. Res.* 113 (8). doi:10.1029/2007ja012540

Burton, R. K., McPherron, R. L., and Russell, C. T. (1975). An empirical relationship between interplanetary conditions and Dst. *J. Geophys. Res.* 80 (31), 4204–4214. doi:10.1029/JA080i031p04204

Connerney, J. E. P., and Ness, N. F. (1981). AcuaModeling the jovian current sheet and inner magnetosphere. *J. Geophys. Res.* 86 (A10), 8370–8384. doi:10.1029/JA086iA10p08370

Eggington, J. W. B., Eastwood, J. P., Mejnertsen, L., Desai, R. T., and Chittenden, J. P. (2020). Dipole tilt effect on magnetopause reconnection and the steady-state magnetosphere-ionosphere system: global MHD simulations. *J. Geophys. Res. Space Phys.* 125, e2019JA027510. doi:10.1029/2019JA027510

Mead, G. D., and Fairfield, D. H. (1975). A quantitative magnetospheric model derived from spacecraft magnetometer data. *J. Geophys. Res.* 80 (4), 523–534. doi:10.1029/JA080i004p00523

Olson, W. P., and Pfitzer, K. A. (1974). A quantitative model of the magnetospheric magnetic field. *J. Geophys. Res.* 79 (25), 3739–3748. doi:10.1029/JA079i025p03739

Pudovkin, M. I., and Tsyganenko, N. A. (1973). Particle motion and currents in the neutral sheet of the magnetospheric tail. *Planet. Space Sci.* 21, 2027–2037. doi:10.1016/0032-0633(73)90131-1

Roelof, E. C., and Sibeck, D. G. (1993). Magnetopause shape as a bivariate function of interplanetary magnetic field$B_z$and solar wind dynamic pressure. *J. Geophys. Res.* 98 (12), 21421450–21450. doi:10.1029/93ja02362

Russell, C. T., and Brody, K. I. (1967). Some remarks on the position and shape of the neutral sheet. *J. Geophys. Res.* 72 (23), 6104–6106. doi:10.1029/jz072i023p06104

Schulz, M., and McNab, M. C. (1987). Source-surface model of the magnetosphere. *Geophys. Res. Lett.* 14 (3), 182–185. doi:10.1029/gl014i003p00182

Sergeev, V. A., Chernyaeva, S. A., Apatenkov, S. V., Ganushkina, N. Y., and Dubyagin, S. V. (2015). Energy-latitude dispersion patterns near the isotropy boundaries of energetic protons. *Ann. Geophys.* 33, 1059–1070. doi:10.5194/angeo-33-1059-2015

Sergeev, V. A., Malkov, M., and Mursula, K. (1993). Testing the isotropic boundary algorithm method to evaluate the magnetic field configuration in the tail. *J. Geophys. Res.* 98 (5), 7609–7620. doi:10.1029/92JA02587

Sergeev, V. A., Sazhina, E. M., Tsyganenko, N. A., Lundblad, J. A., and Soraas, F. (1983). Pitch angle scattering of energetic protons in the magnetotail current sheet as the dominant source of their isotropic precipitation into the nightside ionosphere. *Planet. Space Sci.* 31, 1147–1155. doi:10.1016/0032-0633(83)90103-4

Sergeev, V. A., and Tsyganenko, N. A. (1982). Energetic particle losses and trapping boundaries as deduced from calculations with a realistic magnetic field model. *Planet. Space Sci.* 30 (10), 999–1006. doi:10.1016/0032-0633(82)90149-0

Sibeck, D. G., Lopez, R. E., and Roelof, E. C. (1991). Solar wind control of the magnetopause shape, location, and motion. *J. Geophys. Res.* 96 (4), 5489. doi:10.1029/90ja02464

Sitnov, M. I., Stephens, G. K., Tsyganenko, N. A., Korth, H., Roelof, E. C., Brandt, P. C., et al. (2020). Reconstruction of extreme geomagnetic storms: breaking the data paucity curse. *Space weather.* 18, e2020SW002561. doi:10.1029/2020SW002561

Sitnov, M. I., Tsyganenko, N. A., Ukhorskiy, A. Y., and Brandt, P. C. (2008). Dynamical data-based modeling of the stormtime geomagnetic field with enhanced spatial resolution. *J. Geophys. Res.* 113, A07218. doi:10.1029/2007JA013003

Stephens, G. K., Bingham, S. T., Sitnov, M. I., Gkioulidou, M., Merkin, V. G., Korth, H., et al. (2020). Storm time plasma pressure inferred from multimission measurements and its validation using Van Allen Probes particle data. *Space weather.* 18, e2020SW002583. doi:10.1029/2020SW002583

Stern, D. P. (1985). Parabolic harmonics in magnetospheric modeling: the main dipole and the ring current. *J. Geophys. Res.* 90 (11), 863. doi:10.1029/ja090ia11p10851

Stern, D. P. (1987). Tail modeling in a stretched magnetosphere: 1. methods and transformations. *J. Geophys. Res.* 92 (5), 4437. doi:10.1029/JA092iA05p04437

Tsyganenko, N. A. (1989a). A magnetospheric magnetic field model with a warped tail current sheet. *Planet. Space Sci.* 37 (1), 5–20. doi:10.1016/0032-0633(89)90066-4

Tsyganenko, N. A. (1989b). A solution of the chapman-ferraro problem for an ellipsoidal magnetopause. *Planet. Space Sci.* 37 (9), 1037–1046. doi:10.1016/0032-0633(89)90076-7

Tsyganenko, N. A., and Andreeva, V. A. (2017). A hybrid approach to empirical magnetosphere modeling. *J. Geophys. Res. Space Phys.* 122, 8198–8213. doi:10.1002/2017JA024359

Tsyganenko, N. A., and Andreeva, V. A. (2016). An empirical RBF model of the magnetosphere parameterized by interplanetary and ground-based drivers. *J. Geophys. Res. Space Phys.* 121 (10), 10786–10802. doi:10.1002/2016JA023217

Tsyganenko, N. A., and Andreeva, V. A. (2018). Empirical modeling of dayside magnetic structures associated with polar cusps. *J. Geophys. Res. Space Phys.* 123, 9078–9092. doi:10.1029/2018JA025881

Tsyganenko, N. A., Andreeva, V. A., Kubyshkina, M. V., Sitnov, M. I., and Stephens, G. K. (2021). "Data-based modeling of the Earth's magnetic field," in Space Physics and aeronomy collection volume 2: Magnetospheres in the solar system*, geophysical monograph 259Romain maggiolo, nicolas andre, hiroshi hasegawa, and daniel T. Welling. American geophysical union*. First Edition (Hoboken: Wiley & Sons, Inc.). doi:10.1002/9781119815624.ch39

Tsyganenko, N. A., and Andreeva, V. A. (2020). Magnetospheric "penetration" of IMF viewed through the lens of an empirical RBF modeling. *JGR. Space Phys.* 125, e2019JA027439. doi:10.1029/2019JA027439

Tsyganenko, N. A., and Andreeva, V. A. (2014). On the 'bowl-shaped' deformation of planetary equatorial current sheets. *Geophys. Res. Lett.* 41 (4), 1079–1084. doi:10.1002/2014gl059295

Tsyganenko, N. A., Andreeva, V. A., Sitnov, M. I., Stephens, G. K., Gjerloev, J. W., Chu, X., et al. (2021). Reconstructing substorms via historical data mining: Is it really feasible? *JGR. Space Phys.* 126, e2021JA029604. doi:10.1029/2021JA029604

Tsyganenko, N. A. (1996). "Effects of the solar wind conditions on the global magnetospheric configuration as deduced from data-based models," in *International conference on substorms, proceedings of the 3rd international conference*. Editors E. J. Rolfe, and B. Kaldeich (Paris: Publication ESA SP-389, European Space Agency), 181–185.

Tsyganenko, N. A. (1987). Global quantitative models of the geomagnetic field in the cislunar magnetosphere for different disturbance levels. *Planet. Space Sci.* 35 (11), 1347–1358. doi:10.1016/0032-0633(87)90046-8

Tsyganenko, N. A. (1998). Modeling of twisted/warped magnetospheric configurations using the general deformation method. *J. Geophys. Res.* 103 (10), 23551–23563. doi:10.1029/98JA02292

Tsyganenko, N. A. (1995). Modeling the Earth's magnetospheric magnetic field confined within a realistic magnetopause. *J. Geophys. Res.* 100 (4), 5599. doi:10.1029/94JA03193

Tsyganenko, N. A. (1982). Pitch-angle scattering of energetic particles in the current sheet of the magnetospheric tail and stationary distribution functions. *Planet. Space Sci.* 30 (5), 433–437. doi:10.1016/0032-0633(82)90052-6

Tsyganenko, N. A., and Sitnov, M. I. (2007). Magnetospheric configurations from a high-resolution data-based magnetic field model. *J. Geophys. Res.* 112, A06225. doi:10.1029/2007JA012260

Tsyganenko, N. A., and Sitnov, M. I. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *J. Geophys. Res.* 110, A03208. doi:10.1029/2004JA010798

Tsyganenko, N. A., and Usmanov, A. V. (1982). Determination of the magnetospheric current system parameters and development of experimental geomagnetic field models based on data from IMP and HEOS satellites. *Planet. Space Sci.* 30 (10), 985–998. doi:10.1016/0032-0633(82)90148-9

Voigt, G.-H. (1981). A mathematical magnetospheric field model with independent physical parameters. *Planet. Space Sci.* 29 (1), 1–20. doi:10.1016/0032-0633(81)90134-3

Wright, A. (1985). Stationary distribution function for non-adiabatic particle motion in a one-dimensional current sheet. *Planet. Space Sci.* 33 (7), 847–852. doi:10.1016/0032-0633(85)90039-X

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 17 August 2022
DOI 10.3389/fspas.2022.912036

Check for updates

OPEN ACCESS

EDITED BY
Joseph E. Borovsky,
Space Science Institute, United States

REVIEWED BY
Christine Gabrielse,
The Aerospace Corporation,
United States
Robert McPherron,
California State University, Los Angeles,
United States

*CORRESPONDENCE
Gordon Rostoker,
gr@ualberta.ca

SPECIALTY SECTION
This article was submitted
to Space Physics,
a section of the journal
Frontiers in Astronomy and Space
Sciences

RECEIVED 03 April 2022
ACCEPTED 26 July 2022
PUBLISHED 17 August 2022

CITATION
Rostoker G (2022), Lessons from a
career in space physics.
Front. Astron. Space Sci. 9:912036.
doi: 10.3389/fspas.2022.912036

Lessons from a career in space physics

Gordon Rostoker*

Department of Physics, University of Alberta, Edmonton, AB, Canada

Frameworks for describing magnetospheric substorms are as good as the data on which they are based. While disturbances in the ionosphere are well described by existing data bases, the plasma and field data acquired by satellites in the magnetosphere are sparse and unable to lend an element of uniqueness to any model for substorm activity. This paper describes the voyage of discovery experienced by the author from the leadup to his career to the time of his retirement. Perhaps it will provide young scientists with a sense of how space physics developed and what needs to be done before the solar wind-magnetosphere-ionosphere interaction becomes fully understood.

KEYWORDS
substorms, aurora, magnetic field, interplanetary magnetic field, magnetotail

A bit of my history and how I fell into space science

I grew up in Toronto starting my schooling in the mid 1940s. I was not a particularly good student, and I was probably more interested in playing competitive chess than in the subjects I was taught in school. Nonetheless, I believe what I learned playing chess probably was at least as important as what was taught in school. In my work in space physics, the ability to visualize complex scenes in three dimensions and to think ahead several moves was a distinct asset. When I arrived in university, I chose to enter MPC (Math, Physics and Chemistry) at the University of Toronto. It seemed a reasonable thing to try, since my three older brothers all had gone through university obtaining their doctorates in the physical sciences. I was, at best, a mediocre student for the first 3 years. Then a stroke of luck changed everything. I got to work in a scientific lab for the first time during the summer between my third and fourth year, building equipment and getting involved with one of the first brain scanning machines. I finished my last year with a flourish that got me into Graduate School. For my M. Sc., I set up an array of magnetometers to study the subsurface conductivity across southern Ontario. I was supposed to be studying the induced electric currents in the Earth, but I wondered what was inducing those currents. I spent many hours at the Agincourt Magnetic Observatory looking at magnetograms, and after a while I began to see patterns. I was actually looking at the magnetic signatures of what are now called substorms—in those days they were called magnetic bays. When I was persuaded to go the University of British Columbia for my doctorate studying perturbations in the Earth's magnetic field, I was already hooked. That is how I fell into Space Physics as a career.

After completing my Ph.D. in 1966 studying geomagnetic pulsations, I spent a postdoctoral period at the Royal Institute of Technology in Stockholm, Sweden attracted there by the presence of Hannes Alfvén. There I had the opportunity to work with some of the first interplanetary electric field data, establish a great working relationship with Rolf Boström, and learn another language (Swedish). Little did I know, but the market for university academics was closing down rapidly as I was doing my postdoctoral studies. I was fortunate to be recruited by Professor Jack Jacobs (my department head at UBC who had moved to University of Alberta to assume an endowed Chair). And that is how, in 1968, I started my academic career.

I knew exactly what I wanted to do at that point in time—to set up a line of magnetometers spanning the average position of the auroral oval and monitor the development of the auroral electrojets. What those data did for me, other than allowing me to learn a lot about geomagnetic disturbances, was to provide an entry into the world of scientists who were studying those same disturbances using satellite data. The balance of my career was spent cooperating with colleagues using my ground-based data to help them know what was happening in the magnetosphere when they were looking at their data.

There are two papers that I published during my career which, in retrospect, are a source of pride. The first was co-authored with Rolf Boström (cf. Rostoker and Boström, 1976) stemming from work done during my first sabbatical at the Royal Institute of Technology. It came as a consequence of me sitting in a nearly empty office for some weeks contemplating the three-dimensional structure of the magnetotail as it pertained to the electric field configuration. I was responsible for the geometry of the situation and Rolf Boström was responsible for the mathematical formalisms. One of the referees commented that he didn't believe it, but he couldn't find anything wrong with it. You can't ask for more than that! The second paper was a review of substorm phenomenology along with my framework for the processes that are responsible for substorms (Rostoker, 1996). In Figure 8 of that paper, I showed the development of an auroral substorm as projected on the ionosphere. This differed from the original Akasofu picture (Akasofu, 1964) in that it showed the full development of a substorm in the context of the changes in the interplanetary magnetic field. It emphasized the presence of two distinct regions of activity—the equatorward portion of the auroral oval where substorm activity initiates and the poleward portion where auroral surge forms develop (see Figure 1). It turns out that a substorm expansion phase onset on the equatorward branch of the midnight sector auroral oval looks very much like the development of a surge structure on the poleward branch of the oval in terms of their magnetic signature on the ground—a sharp negative perturbation in the north-south component of the magnetic field. It is all too easy to take the magnetic signature of a surge and misinterpret it as an expansion phase onset.

What I learned during my career

After entering the field of space physics, I watched battles develop in which more than one possible explanation presented itself for observed phenomena. In all these cases, the available data were inadequate to allow one to distinguish between competing models. Over the course of my career, I saw three of these dilemmas resolved when key data became available.

The first of these was a battle between Sidney Chapman (cf. Vestine and Chapman, 1938), who believed that magnetic field disturbances detected at Earth's surface were caused by ionospheric currents alone, and Hannes Alfvén (cf. Alfven, 1940) who believed that those magnetic field disturbances were caused by a combination of field-aligned and ionospheric currents after the idea of Kristian Birkeland (c.f. Birkeland, 1913). Chapman based his claim on the fact that, in the collisionless plasma above the ionosphere, the conductivity along field lines was infinite and hence any electric field would drive an infinite current. When Alfred Zmuda and others (c.f. Zmuda et al., 1966) flew a magnetometer aboard a polar orbiting satellite hundreds of kilometers above the ionosphere, it detected the unmistakable magnetic signatures of field-aligned currents and Alfvén was proved to have been correct thanks to the right data becoming available.

A second such dilemma arose when the first satellite to sample the interplanetary medium (IMP-1) in the late 1960s detected magnetic field pointing towards the Sun for some days, then away from the Sun for some days, then towards the Sun for some days and then away from the Sun for some days. The pattern had an ~ 27- day periodicity, so it was suggested by John Wilcox (cf. Wilcox and Ness, 1965) that there were sectors on the Sun in which the field pointed alternately away from or towards the solar surface. He called this the sector structure and initially this was believed to be the explanation of the observations. A few years later Michael Schulz (cf. Schulz, 1973) proposed that the observations could be explained if there was a wavy current sheet near the ecliptic plane, and sometimes the satellite was above the current sheet and sometimes it was below. The Wilcox viewpoint prevailed until serendipity intervened when the Pioneer 11 satellite, after reaching Jupiter, was slung out of the ecliptic plane on its way to Saturn. When Pioneer got to about 16° above the ecliptic plane, the sector structure disappeared and until the satellite moved back towards the ecliptic plane, the magnetic field pointed in one direction (cf. Smith et al., 1978). It was immediately apparent that Schulz had been correct, but it took the right data to distinguish between the two explanations.

A third such dilemma centered on the cause of geomagnetic storms. From the time of the observation of a giant solar flare in 1869 by Carrington, it was believed that geomagnetic storms were caused by solar flares. Until the early 1990s, efforts were made to clearly establish this causal relationship. At NOAA, Jo Anne Joselyn and her group worked intensely on this problem but were frustrated by the fact that sometimes geomagnetic

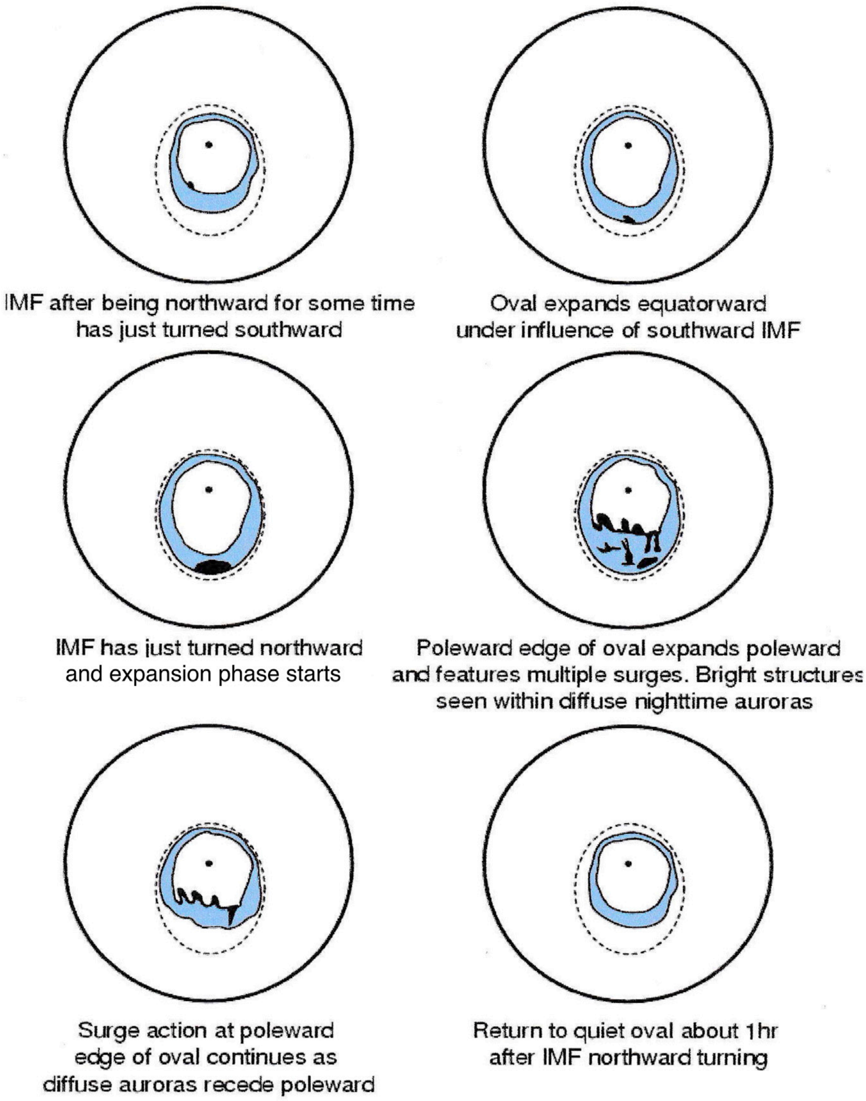

FIGURE 1
Evolution of the auroral signatures during a magnetospheric substorm. Note that this differs from the classic Akasofu auroral substorm in that his substorm features equatorward drifting auroral arcs during his recovery phase whereas this picture features a poleward contracting oval during recovery. Note also that one can have localized auroral forms develop on the poleward branch of the oval during quiet times and pseudo breakups develop on the equatorward branch of the oval in advance of the expansion phase onset. The poleward expansion of the region of discrete auroral arcs during the development of the expansion phase occurs in steps rather than smoothly (Kisabeth and Rostoker, 1974). [modified after Rostoker, 2007].

storms occurred with no apparent flare to blame, and sometimes there were major flares but no ensuing geomagnetic storms (cf. Joselyn, 1995). It was not until Jack Gosling at Los Alamos National Laboratory used Skylab data to reveal that coronal mass ejections (CME's) were to blame for geomagnetic storms that flares were abandoned as the causal agent (cf. Gosling, 1993). Until that time, the presence of coronal mass ejections could not easily be established because the observations from the ground were simply not up to the task of detecting CME's.

My career was dominated over the years by the study of magnetospheric substorms, which involved working with ground-based magnetometer data on which the definition of the substorm was based in part. To understand the physical processes responsible for substorms, it was necessary to have observations in space of magnetic fields, plasmas and electric fields. While such measurements were essential to follow the development of substorms in space, the number of observation points at any one time was woefully inadequate to provide unique solutions to the physical phenomena involved with the substorm process. In the

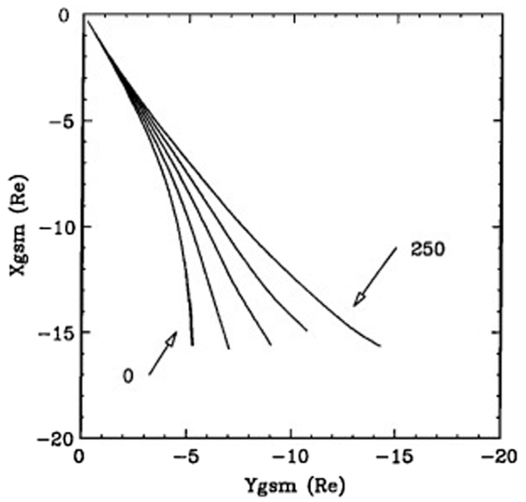

FIGURE 2
Projections in the XYgsm plane of field lines traced using the T87 Kp = 2 model modified to include the magnetic effects of field-aligned and closure currents. The volume filling field-aligned currents stretch across 5 degrees of latitude with downward field aligned current equatorward of 69.45° and upward field-aligned current poleward of that latitude. The five field lines shown here were traced with amounts of field-aligned current added that would produce eastward perturbations of the magnetic field at 800 km altitude of 0 nT, 62.5 nT, 125 nT, 187.5 and 250 nT (0 nT corresponds to no field-aligned current added.) The field lines all have an ionospheric footprint 2 hours from local midnight midway between the current sheets and terminate in the equatorial plane (after Donovan, 1993).

early stages of my career, space-based data were sparse. By the time I retired, the vastly increased amount of data was surprisingly still inadequate in terms of a multipoint set of observations that would provide a clear physical insight to the problem. Two issues were paramount. The first involved the ability to know whether a disturbance was temporal in nature or reflected the passage of plasma and field structures past the observation points. To this day, researchers assume they are dealing with temporal changes despite the fact that evidence dating back to the beginning of the 1970s indicated that plasma structures could sweep past a spacecraft leading to measurements that could incorrectly be attributed to a temporal change. The second issue centered around being able to map along magnetic field lines from the ionosphere to points in the magnetosphere; this involves being able to model volume filling electric currents whose perturbation magnetic fields add to the background geomagnetic field (Donovan, 1993). Between the antiparallel current sheets, the magnetic field lines are skewed towards the flanks of the magnetosphere (see Figure 2). While ionospheric disturbances nowadays can often have their locations precisely defined, if one does not know where in the magnetosphere these disturbances (e.g., auroral forms) map to, one cannot be confident about the physics of the processes that led to these disturbances. It remains for future space scientists to address these issues and to not accept, without question, existing paradigms in this field of research.

What you can look forward to during a career in space science (or any science, for that matter)

While I was spending 2 years on Faculty at the Solar-Terrestrial Environment Laboratory of Nagoya University in Japan a couple of years after my retirement from the University of Alberta in 1997, I was asked to give a talk on what it takes to become a successful scientist. What I told them can be summarized in three points:

1 Successful scientists are persistent in doing what they want to do regardless of what others think.
2 Successful scientists must be creative and believe in themselves. What they discover will likely disagree with what is the current belief, and a scientist must be prepared to disagree publicly with famous people with gray hair who have written the “bible” in their area of research.
3 A successful scientist must be able to communicate very effectively in both the spoken and written language. Some researchers find it hard to write up what they have done, although they love doing the research. Some researchers give talks in a manner which puts the audience to sleep in a manner of minutes. Successful scientists are usually very animated speakers who know just how much information they can provide in the time allotted.

Some final thoughts

I was fortunate to have the opportunity to work in a real laboratory prior to my final year of undergraduate studies. That provided the motivation to propel me into graduate school. That I ended up in space science was a matter of luck—it was either biophysics or geophysics and, as it turned out, the head of Geophysics in Toronto (J. Tuzo Wilson) admitted me to graduate studies. It was good fortune that put me in front of years of magnetograms from Agincourt Magnetic Observatory; that allowed me to recognize patterns in disturbances which turned out to be the signatures of the substorms that I ended up studying for the rest of my career. If you are a young scientist reading this, keep in mind how much luck is involved in finding your final area of research. The rest is hard work, although it can be a lot of fun and very rewarding!

I learned during my career that, to provide a unique solution to aspects of the physical processes that take place in the solar wind-magnetosphere-ionosphere system, one must have adequate data. The human mind is very creative and can think of more than one mechanism that satisfies the available data. As I have pointed out earlier, even during my time in space science, I could identify three instances in which more than one explanation existed to explain a set of observations which was resolved when new data became available that allowed one to identify the correct explanation.

Despite the vast amount of data that has been accumulated *in situ* observations of the solar wind-magnetosphere-ionosphere system, it has still been grossly inadequate to resolve issues that are still outstanding. In fact, the present paradigm for explaining how magnetic field line reconnection in the magnetotail leads to auroras and electric current flows in the ionosphere and magnetosphere is still very fragile and does not account for the fact that there are two separate regions of space in which auroral brightenings take place—the equatorward edge of the midnight sector auroral oval where substorm expansion phases are initiated and the poleward edge where auroral surge phenomena are initiated. When mapping from the ionosphere to the magnetotail along magnetic field lines during significant levels of activity becomes possible, that will be an important step towards answering many of the questions that were left unanswered when I retired from the field.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author performed all research on this perspective and wrote the manuscript.

Acknowledgments

The author is forever grateful to W. Kerr who gave me an opportunity to work in a science lab as an undergraduate student, to Professor Jack Jacobs who found a faculty position for me at the University of Alberta and helped me launch my career, and to the graduate students and postdoctoral associates who helped me in my exploration of magnetic pulsations and magnetospheric substorms. The author is grateful to the Natural Sciences and Engineering Research Council of Canada for support of his research over the years.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akasofu, S.-I. (1964). The development of the auroral substorm. *Planet. Space Sci.* 12, 273–282. doi:10.1016/0032-0633(64)90151-5

Alfven, H. (1940). *A theory of magnetic storms and of the aurorae; 2. The aurorae; 3. The magnetic disturbances*. Stockholm: Almqvist & Wiksell. Kungl. Sv. Vetensk.-Akad. Handlingar III, 18, No. 9, Stockholm.

Birkeland, K. (1913). *The Norwegian aurora polaris expedition, 1902-1903*, 1. Christiania: H. Aschelhoug & Company. Second Section.

Donovan, E. (1993). Modeling the magnetic effects of field-aligned currents. *J. Geophys. Res.* 98 (13), 13529–13543. doi:10.1029/93ja00603

Gosling, J. T. (1993). The solar flare myth. *J. Geophys. Res.* 98 (18), 18937–18949. doi:10.1029/93ja01896

Joselyn, Jo Ann (1995). Geomagnetic activity forecasting: The state of the art. *Rev. Geophys.* 33 (3), 383. doi:10.1029/95rg01304

Kisabeth, J. L., and Rostoker, G. (1974). The expansive phase of magnetospheric substorms: 1. Development of the auroral electrojets and auroral arc configuration during a substorm. *J. Geophys. Res.* 79, 972–984. doi:10.1029/ja079i007p00972

Rostoker, G., and Boström, R. (1976). A mechanism for driving the gross Birkeland current configuration in the auroral oval. *J. Geophys. Res.* 81, 235–244. doi:10.1029/ja081i001p00235

Rostoker, G. (1996). Phenomenology and physics of magnetospheric substorms. *J. Geophys. Res.* 101 (12), 12955–12973. doi:10.1029/96ja00127

Rostoker, G. (2007). "Substorms," in *Handbook of the solar-terrestrial environment*. Editors By Y. Kamide and A. Chian (Springer-Verlag Berlin Heidelberg), 375–395.

Schulz, M. (1973). Interplanetary sector structure and the helliomagnetic equator. *Astrophys. Space Sci.* 24, 371–383. doi:10.1007/bf02637162

Smith, E. J., T Tsurutani, B., and Rosenberg, R. L. (1978). Observations of the interplanetary sector structure up to heliographic latitudes of 16°: Pioneer 11. *J. Geophys. Res.* 83, 717. doi:10.1029/ja083ia02p00717

Vestine, E. H., and Chapman, S. (1938). The electric current system of geomagnetic disturbance. *J. Geophys. Res.* 43, 351. doi:10.1029/te043i004p00351

Wilcox, J. M., and Ness, N. F. (1965). Quasi-stationary corotating structure in the interplanetary medium. *J. Geophys. Res.* 70, 5793–5805. doi:10.1029/jz070i023p05793

Zmuda, A. J., Martin, J. H., and Heuring, F. T. (1966). Transverse magnetic disturbances at 1100 Kilometers in the auroral region. *J. Geophys. Res.* 71, 5033–5045. doi:10.1029/jz071i021p05033

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 29 August 2022
DOI 10.3389/fspas.2022.952248

Check for updates

Daring to think of the impossible: The story of Vlasiator

Minna Palmroth[1,2]*

[1]Department of Physics, University of Helsinki, Helsinki, Finland, [2]Space and Earth Observation Centre, Finnish Meteorological Institute, Helsinki, Finland

OPEN ACCESS

EDITED BY
Elena E. Grigorenko,
Space Research Institute (RAS), Russia

REVIEWED BY
Gian Luca Delzanno,
Los Alamos National Laboratory (DOE),
United States

*CORRESPONDENCE
Minna Palmroth,
minna.palmroth@helsinki.fi

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 24 May 2022
ACCEPTED 18 July 2022
PUBLISHED 29 August 2022

CITATION
Palmroth M (2022), Daring to think of the impossible: The story of Vlasiator.
Front. Astron. Space Sci. 9:952248.
doi: 10.3389/fspas.2022.952248

Vlasiator is the world's first global Eulerian hybrid-Vlasov simulation code, going beyond magnetohydrodynamics in the solar wind—magnetosphere—ionosphere system. This paper gives the story of Vlasiator. An important enabler of Vlasiator is the rapid increase of computational resources over the last decade, but also the open-minded, courageous forerunners, who have embraced this new opportunity, both as developers but also as co-authors of our papers. Typically, when starting a new coding project, people think about the presently available resources. But when the development continues for multiple years, the resources change. If instead, one targets to upcoming resources, one is always in possession of a code which does not contain large legacy parts that are not able to utilize latest resources. It will be interesting to see how many modelling groups will take the opportunity to benefit from the current high-performance computing trends, and where are we in the next 10 years. In the following, a simulation that directly handles and manipulates the phase space density f(**r**,**v**,t) is referred to as a *Vlasov approach*, whereas a simulation system that traces phase space samples by their kinetic characteristics of motion is a *Particle-in-Cell approach*. This terminology is consistent with its use in the magnetospheric simulation community.

KEYWORDS

space physics, modelling, ion-kinetic physics, magnetohydrodynamics, space weather, numerical simulations

Life at the turn of the millennium

In 2004, as a young postdoc, I was doing my postdoctoral period in Boulder, USA. This time was generally marked by high hopes and positive expectations for the future of space physics. The first constellation space physics mission *Cluster* (Escoubet et al., 2001) had just been launched, introducing an opportunity to distinguish spatial effects from temporal variations for the first time. Simultaneously, we still had many of the International Solar Terrestrial Physics (ISTP) satellites in operation, like *Polar* and *Geotail*. I had just finished the methodology to assess magnetospheric global energy circulation using magnetohydrodynamic (MHD) simulations (Palmroth et al., 2003; Palmroth et al., 2004). It felt like anything would be possible, and we could, for example, explain magnetospheric substorms within no time. In fact, I remember wondering what to do once we understand the magnetosphere.

I had conflicting thoughts about the Cluster mission, though. On one hand, this European leadership mission would surely solve all our scientific questions. On the other hand, for me personally, Cluster posed a difficult problem. As a four-point tetrahedron mission, it would provide observations of *ion-kinetic* physics. This was an intimidating prospect, as I had just written my PhD thesis concerning the fluid physics, using a global MHD simulation GUMICS-4 (Janhunen et al., 2012). During those times, simulations were thought primarily as context to data, not really experiments on their own. It seemed that even though we had built eye-opening methods based on MHD, they would be left behind of the development. With Cluster, observations took a giant leap forward, and it seemed that MHD simulations would soon become obsolete.

The modelling community was also thinking about how to go forward. Code coupling and improving grid resolution were frequent topics of conversations. One of the most challenging places to do MHD is the inner magnetosphere, where most of the societally critical spacecraft traversed. The inner magnetosphere is characterized by co-located multi-temperature plasmas of the cold plasmasphere, the semi-energetic ring current, and the hot radiation belts. It is a source region of the Region-2 field-aligned current system that closes through the resistive ionospheric medium providing Joule heating that can bring spacecraft down (Hapgood et al., 2022). MHD fails in the inner magnetosphere because it represents the multi-temperature plasmas by a Maxwellian approximation of the temperature (e.g., Janhunen and Palmroth, 2001). Therefore, it does not reproduce the Region-2 current system (Juusola et al., 2014) and is possibly off by orders of magnitude in estimating Joule heating (Palmroth et al., 2005). Hence, many researchers were relying on code coupling to improve the representation of the inner magnetosphere (e.g., Huang et al., 2006).

During the postdoctoral period, I visited the Grand Canyon. While taking pictures, I anticipated that my old camera would not convey the truth about the place. It struck me that this is like me, using an MHD simulation to reproduce our great magnetosphere. MHD was the best we had, and it was very useful in some respects—but it did not really describe the near-Earth space like Cluster would in the coming years. One-way code coupling, like coupling an MHD magnetosphere to non-MHD inner magnetosphere would not yield a better representation of the global description because that would still be represented as a fluid. Besides, I had doubts towards code coupling (and still do): It would be more about coding than physics, and I was not really interested in that. I wanted to understand how the Cluster measurements would fall into context. But that meant that one would have to change the physics in the simulation.

Beyond MHD?

I remember watching Nick Omidi's work about the formation of the foreshock (Omidi et al., 2004). They had developed a 2-dimensional (2D) hybrid particle-in-cell (hybrid-PIC) code, in which protons were macroparticles describing ion-kinetic physics, and electrons were a charge-neutralizing fluid. Now we are getting somewhere, I thought. However, even though I was amazed of their new capabilities, in comparison to satellite observations the results seemed rather hard to interpret in terms of foreshock wave characteristics, like amplitudes and frequencies. The physics in a hybrid-PIC simulation depends on the ion velocity distribution function (VDF) constructed from the macroparticle statistics, and since they were not able to launch very many particles due to computational restrictions, the outcome was noisy. The other option to simulate ion-kinetic physics was the Vlasov approach (e.g., Elkina and Büchner, 2006), which did not launch macroparticles, but modified the VDF itself in time. Many Vlasov solvers were called *spectral*, i.e., they used the property of the distribution function being constant along the characteristic curves according to the Liouville theorem. The benefit was that there was no noise. However, their problem was filamentation. Formulated as a differential equation on $f(\mathbf{r},\mathbf{v},t)$, nothing prevents the phase space from forming smaller and smaller structures *ad infinitum*. The spectral Vlasov simulations need to address this issue through filtering.

Simply mimicking someone else's approach did not seem very appealing, and a noiseless representation of the same physics would give nice complementarity, I thought, and started to think how a global hybrid-Vlasov simulation would become possible. Let's take the number of GUMICS-4 cells in a refined state, this is 300,000 cells in the **r**-space. Then, let's set the VDF into each **r**-space grid cell to form the **v**-space, use a Eulerian method for propagation in time to get rid of the filamentation, and assume 3,000 grid cells, yielding about 10^9 phase space cells in total. These numbers turned out to be about 100–1,000 times too small in the end, but they were my starting point. Any code development takes time, so let's look at where the computational resources are in 5 years. If one would start to develop a global hybrid-Vlasov code that currently does not fit into any machine, global runs would eventually become possible if the Moore's law continued to increase available resources. This led to two strategic factors: First, the code would need to be always portable to the best available machine irrespective of the architecture, indicating that the latest parallelisation technologies should be used. On the other hand, it meant that I would need to have sustained funding to develop the code for at least during the time at which the computational resources are increasing. So—I thought—my only problem is to get a five-year grant with which I could hire a team to develop the code.

In 2004–2005 the plan did not seem plausible, but in 2007 an opportunity presented itself. The European Union established the European Research Council (ERC). ERC's motto became "*Excellence is the sole criterion*," as they wanted to fund frontier paradigm-changing research, bottom up, from all fields. They had a two-stage call to which I submitted an improved version of the old plan. Based on the then available resources it seemed that in a few years even the Finnish Meteorological Institute would have machines that could be utilised. The first stage proposal

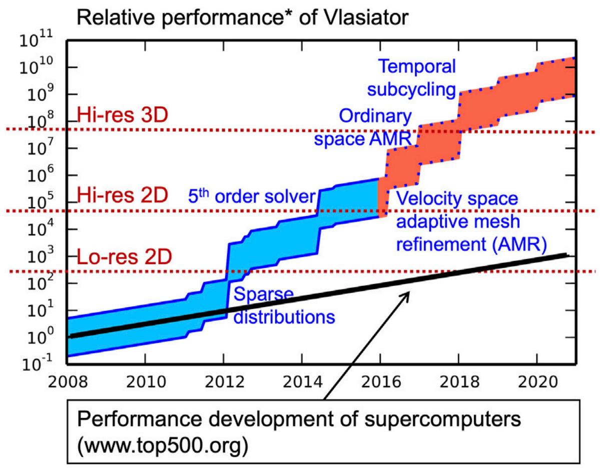

FIGURE 1
Relative performance of Vlasiator normalized to the performance in the beginning of the project. The target simulation is global, so incorporating parts of the solar wind and a large part of the magnetosphere. The black line gives the performance development of the world's 500 top supercomputers. The blue and red areas give the relative performance of Vlasiator depending on whether we used a local machine or a top European supercomputer available through Partnership for Advanced Computing in Europe (PRACE). The plot was made in 2016, explaining the discontinuity in colors (blue is past and red is future).

FIGURE 2
First global view of the 6D Vlasiator, depicting plasma density. Solar wind comes to the simulation box from the right, and a real-size dipole sits in the origin. Every ordinary space grid cell includes a velocity space, where the ion VDF can have an arbitrary form. The solution allows to investigate ion-kinetic phenomena within the global magnetosphere, including also physics that is beyond the MHD description, like Hall fields in reconnection, drifts, instabilities, co-located multi-temperature plasmas, field-aligned and field-perpendicular velocities, just to name a few.

deadline was also my own deadline, as it was the same day as my child was due. In the summer of 2007, I received notice that my proposal was accepted to the second stage. The deadline of the full proposal was in the fall, and the interview in Brussels was at my child's 6-months birthday. I wrote the second stage proposal in 1-h slots when the baby was sleeping, and to make my time more efficient, I utilised a "power-hour" concept established by Finnish explorers. If you concentrate all your daily courage into 1 h, you can do whatever during that hour. So, I called around and talked to different researchers about Vlasov solvers and managed to submit in time. The interview went beyond all expectations. In May 2008, I received information that from the ~10,000 submitted proposals, 300 were funded, my proposal among the successful ones.

Vlasiator

In the beginning, there were many obstacles. Not many people were willing to come to Finland, because the challenge seemed enormous even with knowledge of the Moore's law. Further, the community did not seem to believe in the project. I remember participating in the International Symposium of Space Simulations 2009, where the first ever Vlasiator poster was presented (Daldorff et al., 2009). Most people thought developing Vlasiator is impossible because there would not be enough supercomputer capacity to realize the simulation in the global scale. Figure 1 displays the relative performance of Vlasiator in time, showing that indeed in 2009, we did not have anything concrete yet. Several optimizations enabled our first 3D test-Vlasov simulation (Palmroth et al., 2013). The first real breakthrough in Vlasiator development was the use of a sparse velocity space. As a large part of the VDF is negligible, we decided not to store nor propagate those VDF grid cells that were below a certain threshold in phase space density. Sparsity enabled the first 2D paper of the foreshock waves (Pokhotelov et al., 2013). The second breakthrough was when we replaced the diffusive finite volume solver with a semi-Lagrangian approach (see details in Palmroth et al., 2018). At the end of my ERC Starting grant in 2013 we were able to carry out similar studies as Omidi showed in 2004, but without noise in the solution.

Throughout the project, we developed and utilized new parallelization schemes, and one critical success factor was, and is, the collaboration with the Finnish supercomputing center CSC. Even though the scientific community doubted Vlasiator, CSC believed in it, and helped us at every step of the way. They thought it was remarkable that someone is not thinking about the current resources but is aiming for the coming resources. It made sense to them, too, because they had seen how many codes are obsolete and not able to utilize the newest resources efficiently; supercomputing architectures are experiencing a constant change. CSC is tasked to look ahead in high-performance computing (HPC), and by helping us they proofed many coming HPC architectures and technologies. When a supercomputer is installed, it needs to be piloted, i.e., executed under a heavy load to understand how the system behaves. Few codes can scale to hundreds of thousands of compute cores linearly, and so we have piloted many supercomputer installations providing information both to vendors as well as supercomputing centers. The collaboration with the HPC professionals is very rewarding.

Then, in 2015, I won an ERC Consolidator grant, this time to make Vlasiator 3D and to couple it with the ionosphere. Long story short, at the end of my second ERC grant, this is where we are now. The team increased from 3 to over 20, and we have held international hackathons, and invited guest first authors to utilize the results. We have presented the first 6D results using a global 3D ordinary space, which includes a 3D velocity space in every grid cell. We have coupled the code with an ionospheric solution. Even though we only have submitted papers of the new capabilities, I can say that the first results are so breath-takingly and utterly beautiful, worth all these years of blood, sweat, tears, and trying to convince people who doubted the project. Figure 2 shows preliminary results. We are finally able to address many of the great mysteries of our field, like what is the interplay and spatiotemporal variability of ion-kinetic instabilities and reconnection in the substorm onset (Palmroth et al., 2021a). We have made a paradigm shift and finally, for the first time, see how the global magnetosphere looks like in a Eulerian hybrid-Vlasov simulation. We have gone beyond global MHD. We can also compare to the complementary hybrid-PIC simulations that have also recently extended their approach to 3D (e.g., Lin et al., 2021).

Vlasiator has two strengths compared to the complementary hybrid-PIC approach; the noiseless representation of the physics, and the fact that we can give the results in non-scaled SI units that are not factors of the ion scale lengths but are directly comparable with spacecraft observations. We have also weaknesses compared to the hybrid-PIC: We can only follow the particle trajectories by particle tracing (at run-time we only see how the VDFs evolve), and we are possibly using more computational resources. The latter weakness is not certain, though, because to be able to represent also the tenuous parts of the magnetosphere as accurately as in a Eulerian hybrid-Vlasov scheme, the hybrid-PIC simulations would need to use so many macroparticles that the computational resources would possibly be of the same order of magnitude as we use. On the other hand, the Vlasiator performance is also increasing in time as seen in Figure 1. In 2022, we can carry out 6D global simulations at the cost of about 15 million core hours, the number we used to carry out our first 5D simulation.

Closing remarks

One of the most important lessons I've learnt is that if the motivation to make a paradigm shift comes from the matter itself—in my case the need to understand how our system behaves beyond MHD—the seeds of success have been planted. This is a great shield against the inevitable misfortunes and drawbacks, which are an integral part of any success. If I was foremost interested in appreciation or approval, I would not have pulled this project through. Another lesson concerns recruiting. When doing something that has never been done before, prior skills and knowledge are of lesser importance. In fact, it may even hurt to have too much prior knowledge due to a psychological concept called confirmation bias[1], a tendency to favor information that confirms or supports prior beliefs. When developing Vlasiator, a confirmation bias would have magnified the notion of a shortage of available computational resources, leading to misguided thinking about what is or will become possible in the future. If the will to understand is strong enough, resources will come. Right now, there are more HPC resources than ever before—but—they are again changing. In the beginning of 2010s, the direction was towards compute processing unit (CPU) parallelism, now it is in heterogeneous architectures including graphics processing units (GPUs). To meet this challenge, of course we have an active development branch concentrating on how to harness GPUs in improving Vlasiator performance.

It feels that this age has more frustration than 20 years ago, perhaps because it is more difficult to get new spacecraft missions. Our field possibly suffers from some sort of a first in—first out challenge. It was our field, which first started using spacecraft to understand how the near-Earth space behaves (van Allen and Frank, 1959). Increasingly in the past decades, other fields have started to use space as well, and therefore there is more competition in getting new missions. Since we have studied the near-Earth space using spacecraft already from 1960s, the other space-faring fields may have the (wrong) impression that we already know how our system behaves. Right now, there is more need for our field than ever before due to the increasing economical use of space (Palmroth et al., 2021b), and also because the society is critically depending on space. If we do not understand the physical space environment, it is like sending an increasing number of ships to wreck in the Cape of Good Hope.

The new 6D Vlasiator results may improve our chances in convincing the selection boards that new missions are needed. Indeed, it seems that the tides have turned: The Cluster mission was one of the major reasons to develop Vlasiator in the first place. However, now it is the models which lead the search for new physics. For example, we have found that magnetopause

1 https://en.wikipedia.org/wiki/Confirmation_bias

reconnection can launch bow waves that deform the bow shock shape upstream and influence the particle reflection conditions in the foreshock (Pfau-Kempf et al., 2016). In fact, Vlasiator results make the whole concept of scales obsolete: If we can look at ion-kinetic physics globally, there is no fluid-scale anymore, there is only scale-coupling—how small-scale physics affects at global scales and vice versa. For example, we can see how small variability like reconnection finally emerges as large-scale changes, like eruptions of plasmoids and brightening of the aurora. We will not know the answers with the current missions, but with ion-kinetic models we can build a picture that guides future mission development.

My advice for the next generation? First, follow your own nose, and take only projects that are worth carrying out at least for 10 years. Otherwise, you are completing someone else's dream, and contributing incrementally to the present state-of-the-art. Confirmation bias means that our field does not renew fast enough if we do not dare to think of the impossible. When you do follow your own nose, do not expect that people see or understand what you are working towards. Strategies are very difficult to discern from the outside. When your strategy succeeds, your accomplishments seem easily earned, even if they required an enormous effort. This leads to another problem: only you know how hard it was! When your hard-earned results are showcasing new science, you may get nods of approval and positive feedback, followed by requests to run parameter studies which are feasible with previous-generation tools. For Vlasiator, developing it took over 10 years and several millions of euros, while running it at state-of-the-art scale requires competitive funding and computational proposals.

Our field also needs holistic thinking, unity, and a sense of community. We are sandwiched between the large Earth system sciences and astrophysics. They are often the gatekeepers in deciding who gets funding, new missions, and high-impact papers. Our internal quarrels are interpreted as a weakness, and a sign that investing into us may be wasting resources. Hence, every time someone succeeds, we should celebrate because as our field progresses, so do we on the surf. Every time someone struggles, we should help because at the same time, we help ourselves. We also need to improve the general work-wellbeing to be able to lure new people to our field. Our work forms such a large part of our identity that we might as well enjoy the road and invest in a great, forward-going atmosphere.

Data availability statement

The raw data supporting the conclusion of this article will be made available as per Vlasiator rules of the road (see: http://helsinki.fi/vlasiator).

Author contributions

The author confirms being the sole contributor of this article and has approved it for publication.

Funding

Work was supported by European Research Council Consolidator grant 682068-PRESTISSIMO and Academy of Finland grants 1336805, 1312351, 1339327, 1335554, 1347795, and 1345701.

Acknowledgments

I have been blessed in finding great people in the Vlasiator team. I would like to thank each past and current member equally, as working with you has made the journey enjoyable. I am also extremely grateful to the co-authors of our papers, and the numerous Vlasiator-enthusiasts we have encountered over the years: Your support was crucial to continue. Vlasiator (http://helsinki.fi/vlasiator) was developed with the European Research Council Starting grant 200141-QuESpace, and Consolidator grant 682068—PRESTISSIMO. The CSC—IT Center for Science in Finland and the Partnership for Advanced Computing in Europe (PRACE) Tier-0 supercomputer infrastructure in HLRS Stuttgart are acknowledged for their constant support and absolutely brilliant advice on HPC (grant numbers 2012061111, 2014112573, 2016153521, and 2019204998). Academy of Finland has been a great contributor to Vlasiator work as well, especially through grants 138599, 267144, 1312351, 1336805, 1309937, 1345701, 1347795, 1335554, and 1339327.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Daldorff, L. K. S., Yando, K. B., Palmroth, M., Hietala, H., Janhunen, P., and Lucek, E. (2009). "Preliminary studies on the development of a global hybrid-Vlasov simulation," in poster presentation in 9th International Symposium for Space Simulations, July 3–10, Paris France.

Elkina, N. V., and Büchner, J. (2006). A new conservative unsplit method for the solution of the Vlasov equation. *J. Comput. Phys.* 213, 862–875. doi:10.1016/j.jcp.2005.09.023

Escoubet, C. P., Fehringer, M., and Goldstein, M. (2001). Introduction the cluster mission. *Ann. Geophys.* 19, 1197–1200. doi:10.5194/angeo-19-1197-2001

Hapgood, M., Liu, H., and Lugaz, N. (2022). SpaceX—sailing close to the space weather? *Space weather.* 20, e2022SW003074. doi:10.1029/2022SW003074

Huang, C.-L., Spence, H. E., Lyon, J. G., Toffoletto, F. R., Singer, H. J., and Sazykin, S. (2006). Storm-time configuration of the inner magnetosphere: Lyon-Fedder-Mobarry MHD code, Tsyganenko model, and GOES observations. *J. Geophys. Res.* 111, A11S16. doi:10.1029/2006JA011626

Janhunen, P., and Palmroth, M. (2001). Some observational phenomena are well reproduced by our global MHD while others are not: Remarks on what, why and how. *Adv. Space Res.* 28, 1685–1691. doi:10.1016/S0273-1177(01)00533-6

Janhunen, P., Palmroth, M., Laitinen, T. V., Honkonen, I., Juusola, L., Facskó, G., et al. (2012). The GUMICS-4 global MHD magnetosphere – ionosphere coupling simulation. *J. Atmos. Sol. Terr. Phys.* 80, 48–59. doi:10.1016/j.jastp.2012.03.006

Juusola, L., Facsko, G., Honkonen, I., Tanskanen, E., Viljanen, A., Vanhamäki, H., et al. (2014). Statistical comparison of seasonal variations in the GUMCIS-4 global model ionosphere and measurements. *Space weather.* 11, 582–600. doi:10.1002/2014SW001082

Lin, Y., Wang, X. Y., Fok, M. C., Buzulukova, N., Perez, J. D., Cheng, L., et al. (2021). Magnetotail-inner mag- netosphere transport associated with fast flows based on combined global-hybrid and CIMI simulation. *JGR. Space Phys.* 126, e2020JA028405. doi:10.1029/2020JA028405

Omidi, N., Sibeck, D., and Karimabadi, H. (2004). "Global hybrid simulations of solar wind and bow shock related discontinuities interacting with the magnetosphere," in Proceeding AGU Fall Meeting Abstracts, SM21B-0, December 2004, Available at: https://ui.adsabs.harvard.edu/abs/2004AGUFMSM21B.01O.

Palmroth, M., Pulkkinen, T. I., Janhunen, P., and Wu, C.-C. (2003). Stormtime energy transfer in global MHD simulation, *J. Geophys. Res.* 108 (A1), 1048. doi:10.1029/2002JA009446

Palmroth, M., Janhunen, P., Pulkkinen, T. I., and Koskinen, H. E. J. (2004). Ionospheric energy input as a function of solar wind parameters: global MHD simulation results. *Ann. Geophys.* 22, 549–566. doi:10.5194/angeo-22-549-2004

Palmroth, M., Janhunen, P., Pulkkinen, T. I., Aksnes, A., Lu, G., Ostgaard, N., et al. (2005). Assessment of ionospheric joule heating by GUMICS-4 MHD simulation, AMIE, and satellite-based statistics: towards a synthesis. *Ann. Geophys.* 23, 2051–2068. doi:10.5194/angeo-23-2051-2005

Palmroth, M., Honkonen, I., Sandroos, A., Kempf, Y., von Alfthan, S., and Pokhotelov, D. (2013). Preliminary testing of global hybrid-Vlasov simulation: magnetosheath and cusps under northward interplanetary magnetic field. *J. Atmos. Sol. Terr. Phys.* 99, 41–46. doi:10.1016/j.jastp.2012.09.013

Palmroth, M., Ganse, U., Pfau-Kempf, Y., Battarbee, M., Turc, L., Brito, T., et al. (2018). Vlasov methods in space physics and astrophysics. *Living Rev. comput. Astrophys.* 4, 1. doi:10.1007/s41115-018-0003-2

Palmroth, M., Ganse, U., Alho, M., Suni, J., Grandin, M., Pfau-Kempf, Y., et al. (2021a). "First 6D hybrid-Vlasov modelling of the entire magnetosphere: Substorm onset and current sheet flapping," in AGU Fall Meeting 2021. New Orleans, 13–17.

Palmroth, M., Tapio, J., Soucek, A., Perrels, A., Jah, M., Lönnqvist, M., et al. (2021b). Toward sustainable use of space: Economic, technological, and legal perspectives. *Space Policy* 57, 101428. doi:10.1016/j.spacepol.2021.101428

Pfau-Kempf, Y., Hietala, H., Milan, S. E., Juusola, L., Hoilijoki, S., Ganse, U., et al. (2016). Evidence for transient, local ion foreshocks caused by dayside magnetopause reconnection. *Ann. Geophys.* 36, 943–959. doi:10.5194/angeo-34-943-2016,

Pokhotelov, D., von Alfthan, S., Kempf, Y., Vainio, R., Koskinen, H. E. J., and Palmroth, M. (2013). Ion distributions upstream and downstream of the earth's bow shock: first results from vlasiator. *Ann. Geophys.* 31, 2207–2212. doi:10.5194/angeo-31-2207-2013

van Allen, J., and Frank, L. (1959). Radiation around the Earth to a radial distance of 107, 400 km. *Nature* 183, 430–434. doi:10.1038/183430a0

TYPE Perspective
PUBLISHED 01 September 2022
DOI 10.3389/fspas.2022.951411

The evolution of heliophysics: Complexity, community, and open science

Ryan M. McGranaghan*

Orion Space Solutions, LLC, Louisville, CO, United States

Check for updates

OPEN ACCESS

EDITED BY
Joseph E. Borovsky,
Space Science Institute, United States

REVIEWED BY
Lev Zelenyi,
Space Research Institute (RAS), Russia

*CORRESPONDENCE
Ryan M. McGranaghan,
ryan.mcgranaghan@gmail.com

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 24 May 2022
ACCEPTED 12 August 2022
PUBLISHED 01 September 2022

CITATION
McGranaghan RM (2022), The evolution of heliophysics: Complexity, community, and open science.
Front. Astron. Space Sci. 9:951411.
doi: 10.3389/fspas.2022.951411

Responding to the grand challenges that confront the Earth and Space Sciences requires an embrace of methods from the field of complexity and systems science that can adapt our thinking and our science to be more inter- and cross-disciplinary and enable broader connection across individuals, teams, communities, and sciences. Culturally, as scientifically, broader disciplinary approaches are imperative. The cultural challenge is the disconnect that exists between groups. These disconnects preclude plurality in discussions, harm creativity and innovation, and give rise to a palpable malaise, especially at the early career stage. Together, the scientific and cultural grand challenges we describe point to a need for a new set of literacies and curriculum that the advent of open science supports–increased cross-disciplinarity, team science that generates community connections, plurality and inclusion in our science and in how we connect.

KEYWORDS

data science, information science, complexity, heliophysics, epistemology, philosophy of science, open science, commons

Introduction

At 2 a.m. on February 15, the Electric Reliability Council of Texas (ERCOT) declared an Energy Alert Level 3 and utilities began rotating outages due to high consumer demand. The heightened alert was the result of unusual, though perhaps not unexpected, cold temperatures and unpredictable power consumption behavior of individuals and businesses–together causing ERCOT officials to nervously watch the frequency of the electric power grid drop outside of the narrow 60 Hz band, a number affected by innumerable moving pieces and dynamics from the weather to the operation of the power grid to the user demand on the system. The events of February 2021 reawakened the world to the precarity of the power grid, a massively complex and integrated system whose resilience in the face of the variability of the natural and human world is anything but guaranteed. The way we see and attempt to control the grid is like trying to know everything about a room we are not standing in when all we have is a temperature reading from a thermometer within it.

The grid is at the whim of the natural world and the vicissitudes of human behavior, a truly complex system (Meadows and Wright, 2008). The interconnected power grid is merely an especially visible example of the complexity of the world that we attempt to

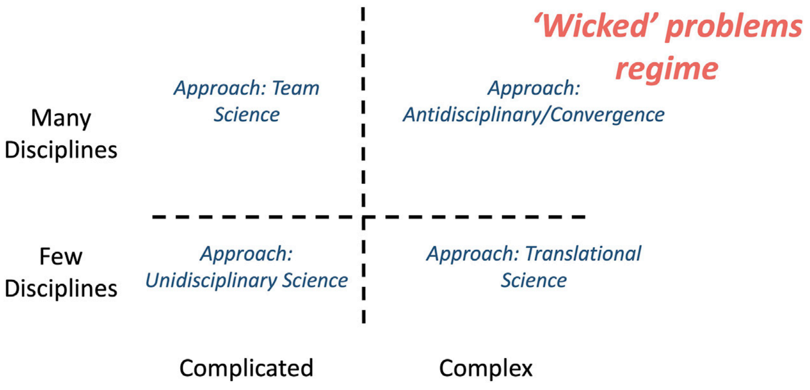

FIGURE 1
An illlustration of the problemscape and approaches of the Earth and Space Sciences.

understand. We realize that we face a new challenge in the Earth and Space Sciences. While the paradigm of reducing problems to separate sub-disciplines to study as distinct constituent parts has produced remarkable insights, all scientists are confronted with interconnected problems of increasing existential importance and yet have been obstinate to progress.

Our world is interconnected. John Muir wrote, "When we try to pick out anything by itself, we find it hitched to everything else in the Universe." (Muir and Gleason, 1911) As these interconnections become more important to the problems we are trying to solve, the whole becomes more than the sum of the parts–or in other words, the system is complex and exhibits emergent behavior. Figure 1 illustrates the situation. A given problem can be graded over four quadrants detailing the number of disciplines and whether the problem is complicated or complex (Kurtz and Snowden, 2003; Kurtz and Snowden, 2007). We use complicated to refer to hard problems that can be addressed by reduction to rules or processes. Alternatively, complexity refers to phenomena that emerge from a collection of interacting objects. The term 'emergence' is important in that it describes how the phenomena is not present in any of the interacting parts alone and cannot be reduced in the same manner as complicated problems (more below in *Introduction*). For problems involving few disciplines and complicated behavior (lower left quadrant), unidisciplinary approaches are effective. As more disciplines are required where the system behavior remains complicated rather than complex (upper left quadrant), the approach is one of team science (Council, 2015) where the most important advance required is improved collaboration, coordination, and communication. In the lower right quadrant, behavior is complex but perhaps can be addressed with a relatively few disciplines the approach is translational, or amenable to borrowing methods of complexity science across those few disciplines. The upper right quadrant requires many disciplines and the system exhibits complex behavior. This is the 'wicked problems regime.' The approach to wicked problems must be *antidisciplinary* or convergent, where we must merge innovative ideas, approaches, and technologies from a wide and diverse range of sectors and expertise. Note here that I define antidisciplinary as increased plurality of thought and transdisciplinary connections. It does *not* mean against disciplines. A better metaphor might be that antimatter as a partner to ordinary matter. This is the problem landscape of the Earth and Space Sciences. New approaches are needed when we cross the thresholds into the wicked problems regime. An increasing number of our problems land in this upper right regime.

Improving the resiliency of the power grid in the face of compounding human-natural forces is one of these wicked problems and a part of a class of them in our society (e.g., global pandemics, climate change). Indeed, problems are often elevated to this regime when human behavior is tied in. I suggest that to respond to the wicked problems in the Earth and Space Sciences requires methods from the field of complexity science, expanded or new literacies that we need to develop as individuals and incentivize as a community, and open science as an emerging framework for these changes and to make them sustainable and scalable.

> We have reached a stage where the pace of discovery and the nature of shared knowledge bring the whole venerable exercise of disciplinary fads into question . . . The cost of [disciplinarity] is that it restricts the scope of our inquiries and causes us to lose sight of the numerous extradisciplinary ideas and methods that have contributed to (and will be required to further) our progress through the thorny branches of science. -David Krakauer (Krakauer, 2019)

The cultural reach

Coupled to the scientific challenges are cultural ones. There is a growing disconnect between individuals, whether within Heliophysics or between Heliophysicists and Earth Scientists (or even between Heliophysicists and artists and designers). That disconnect is linked to a malaise among the communities of scientists and engineers in the Earth and Space Sciences. This lack of fulfillment has been attributed to a lack of connection, diversity, inclusivity, and a general feeling of fatigue (McGranaghan et al., 2020). Though no cross-sections of Earth and Space Science are immune to the effects, it is perhaps the early career community that experiences it most acutely (Evans et al., 2018; Bankston, 2021) as growing connections is particularly important to their work and lives.

In the scientific as in the cultural, perhaps the solution is through interconnections. I propose, based on professional experiences and many personal conversations and observations, that much of this cultural malaise is because of the now-undeniable recognition that our disciplinary approaches, the strictures of our thinking, no longer describe the reality that we are faced with–in the scientific sense as in the cultural. Our disciplinary silos cannot describe the scientific problems we witness, which depend on the interconnections, just as in our communities the ways we segment and separate ourselves deny richer interactions and relationality. In the early career community, perhaps, this feeling is particularly acute.

Our challenge may be indicative of a broader cultural problem: a disconnect between scientists and the public. Perhaps born of different expectations of literacies between the scientist and the citizen (e.g., critical skepticism and hypothesis testing), the disconnect can lead to distrust of science. Below we outline new literacies we need to address issues across the Heliophysics community, but we should also consider those that will allow us to reach across the scientific community to the public.

The complex response

We need to revitalize our vision for the field. The goal of this piece is to clarify and reveal the nature of the problems facing the Earth and Space Sciences to enable conversations about solutions to them. I have chosen to ground it in the concept of complexity because of its ability to deal with interconnections. We define complexity and relate it specifically to the field of Heliophysics, revitalizing it in the process. We also describe the implications of complexity–the degree of collaboration that it requires and the philosophy that will underlie our efforts to reach it. Though it is contextualized in Heliophysics, the discourse is relevant to all areas of science.

Our development will lead us to a new vision and a remaking of Heliophysics, one with a broader scope and more open and cohesive community. We describe the practices we need to adopt and the literacies/capacities we need to create in our workforce to achieve frontier scientific discovery at the pace and complexity that society needs. We suggest new metrics we need to consider that drive resources and policy for a more flourishing community and science. We use the concept of *Open Science* (Vicente-Saez and Martinez-Fuentes, 2018) as a portmanteau to encompass the cultural implications of this shift in philosophy.

This commentary comes from an early career perspective. It is an introduction to a new way of thinking about Heliophysics that can connect the sub-disciplines in our science, our science to other sciences, and the society of Earth and Space Scientists–creating a healthier community. It is also an attempt to let that perspective create a map to tangible, actionable recommendations and be a basis for new comprehensive solutions. In T*he complex response* below I offer suggestions that may help our community focus in new ways: new metrics, literacies and capacities we need to value and build, and a curriculum that encompasses them.

Complexity and heliophysics

> The [21st] century will be the century of complexity. -Stephen Hawking

Complexity science is the study of phenomena that emerge from a collection of interacting objects. To understand a complex system requires a plurality of frameworks and we must be able to move between levels (e.g., micro and macro). As such, complexity science spans numerous dimensions. In the context of Heliophysics, complexity science is the study of a star, interplanetary environment, magnetosphere, upper and terrestrial atmospheres, and planetary surface as interacting subsystems. Each of these subsystems can be further broken down into regions (e.g., the auroral region of the upper atmosphere) and all the way down to more elementary components such as electrons and protons. Complexity science is a paradigm that suggests ways of reconciling the micro and macro scales. It is the collection of methods to understand a system across scales, the smaller scale behavior in connection with the larger-scale phenomena that emerge from it. The complex systems paradigm transcends the concepts of scale and discipline, providing methods to connect across them (Thayer, 2011). To evolve toward a complexity paradigm in Heliophysics requires understanding and adopting the methods of complexity. In the process of envisioning this transformation, other fields provide examples and inspiration: biology (Kauffman and Kauffman, 1993), ecology (Wilson, 1999), cognitive science (Varela et al., 1992), to name a few.

The methods of complexity

There are numerous methods that undergird complexity science. We only highlight a select few that have basis in

Heliophysics research already and can be foundations on which to build. The research cited below is not comprehensive, but meant to be a way into these topics for members of our community and those of other communities to find commonality.

Self-organization, emergence, and scaling theory Emergence is the term used to describe phenomena that are 'more than the sum of their parts' (Rosas et al., 2020). Emergence is observed in virtually all areas of inquiry, such as how large numbers of individual fish are able to behave dynamically as a school when threatened by a predator (Parrish et al., 2002). In terms of scale, emergence is the occurrence of actions at one scale giving rise to a phenomenon on another level. The idea that order at some higher dimension, or coarse-grained level of a system, is organized by a number of interacting sub-systems is called self-organization. Self-organization is a powerful toehold in complexity science because it reveals that emergence is observable in statistical characteristics of the system. If there are underlying driving mechanisms that are identical at all scales, a statistical signature is created that is consistent across scales–a power law (West, 2017). Emergence is a way that order is extracted from many interacting parts and power laws describe that order statistically. Self-organization and power law relationships have produced cross-system Heliophysics understanding for decades (e.g., (Consolini, 2002; Chang et al., 2003; Aschwanden et al., 2014; Budaev et al., 2015)). Though self-organization and scaling laws may be relatively new terms for many, the concept of developing an effective theory from coarse-grained principles is well understood to all scientists. Temperature is an average of all of the particles' motions in a gas, and is a better predictor of the system's future at a certain macro-level. Coarse-graining is how we model the behavior of a complex system without specifying every underlying cause and component that lead to system-level changes.

Information theory To analyze order mathematically, the driving principle of the complexity paradigm, one must begin with information and its counterpart, entropy. Information quantifies the amount of dependency or connection between a random variable and itself at a different time or with other variables at the same or different times. Entropy quantifies the amount of micro-states involved in the value of a random variable. Information theory provides rigorous mathematical formalisms to study the nonlinear relationships and feedbacks that characterize complex systems (Thayer, 2011), especially because they can go beyond linear correlational analyses, capture nonlinear relationships, and establish causalities. Entropy-based information theory is already a valuable tool in Heliophysics to determine the information flow among cross-system parameters, infer potential causalities, untangle the drivers, and provide observational constraints that can help guide the development of the theories and physics-based models (Wing and Johnson, 2019).

Network Science If the complexity science paradigm is about understanding the emergence of patterns from the interactions of their parts, then networks are its specimens and network science its toolkit. A network is simply a collection of entities, or nodes, and their relationships, or edges. For example, in a social network the nodes are people and the edges are the relationships with one another. As the network structure is remarkably representative of the natural world (Kauffman and Kauffman, 1993), thinking of a system in this way can lead to new and useful insights for Heliophysics (Dods et al., 2015; McGranaghan et al., 2017b; Hughes et al., 2022).

Resilience framework New frameworks are required to handle uncertainty and embody the complexity paradigm. A framework of resilience acknowledges complexity, taking into account the holistic system, and the probabilistic nature of complexity science. In this framework, a system is treated as complex and can be defined by whether or not it can accommodate changes and reorganize itself while maintaining the crucial attributes that give the system its unique characteristics (Scheffer et al., 2001). In Heliophysics a resilience framework involves two important principles: 1) considering the Sun-to-society system; and 2) quantifying uncertainty that arises from coarse-graining and statistical simplification. Heliophysics, with its societal implications (Schrijver et al., 2015), requires a resilience framework in order to translate the science of Heliophysics into actionable knowledge for space weather. Resilience offers a way that decisions can be made based on complex systems understanding.

Literacies, curriculum, and metrics for Complexity Heliophysics

Complexity Heliophysics requires our community to develop new literacies and the curriculum that encompasses them.

New capacities and literacies

The literacies are both technical and cultural. The methods of complexity science listed above reveal important technical competencies: scaling relationships, information theory, and network science (and the computational techniques required by them). Several others are less explicit in the development so far.

Data science Data science refers to scalable architectural approaches, techniques, software and algorithms which alter the paradigm by which data are collected, managed and analyzed and communicated. For years, our understanding of complex systems has benefited from taking advantage of comprehensive data-intensive approaches (McGranaghan et al., 2017a). Those skills for state-of-the-art data-driven sciences and technologies are even more important in light of the need to synthesize more encompassing disciplinary information for Heliophysics-related science. Knowledge engineering, or the skill of building the technologies that represent our knowledge, is emerging as an important sub-component of a data science literacy. Building better knowledge representation systems is a cornerstone of any approach to identify where information asymmetries and bottlenecks exist, recognize the whole of an individual's, group's, or project's contributions, and create information needed to design more productive incentive structures for our community.

Gathering and organizing The dramatic increase in the scale and complexity of scientific research required to address wicked problems must be reflected in the scale and diversity of our collaborations. A seldom-recognized skill in managing collaboration and fostering knowledge generation that now becomes incumbent upon all researchers is the ability to effectively bring larger groups together and cultivate effective connections across them (Council, 2015). We do not appropriately recognize and value the challenge and importance of the ability to create a cohesive and broad gathering. The importance of a leader's gathering, organization, and facilitation skills will be elevated. Indeed, these skills are central to improving diversity, equity, inclusivity, and accessibility (DEIA).

Resilience Martin Scheffer (Scheffer et al., 2001) defines a resilient system as one that can accommodate changes and reorganize itself while maintaining the crucial attributes that give the system its unique characteristics. It is the ability of an entity–e.g., asset, organization, community, region–to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance. It is clear that the researcher of today must adapt to more rapidly changing conditions, as the late Buckminster Fuller termed the pace of the appearance of ideas–accelerating acceleration. The resilient researcher is one prepared and equipped to respond to quickly changing conditions and capable of continual learning and reinvention of their frameworks.

Trans-media communication Knowledge is created and consumed in myriad new ways in the 21st century. While technical journal publication remains a primary outlet for the dissemination of scientific knowledge, to reach broader audiences and widen the impact of science in society researchers must embrace and become skilled in communicating across mediums, including blogs and newsletters, audio (e.g., podcasts), video (e.g., YouTube), and interactive data visualization.

It will be the job of all Heliophysicists to figure out how to develop these literacies. The list of jobs and responsibilities for Heliophysicists seems overly burdensome and it is not difficult to see why early career researchers feel overwhelmed. Perhaps not all skills need to be tackled by all Heliophysicists. Instead, infrastructure of various types (e.g., in the ability to construct more capable teams) can assist in meeting 21st century needs. We need to offset the burdensome nature of too much expectation on the individual and facilitate more collective activity and intelligence, which we describe in *New capacities and literacies* below.

Metrics of the future

To cultivate new literacies requires rethinking the visible quantities that our community uses to drive resources, particularly our most precious one: attention. These visible quantities are our metrics. Like exploring an unlit room with a flashlight, our understanding and the ways we choose to move depend on what we shine the light on. We need to rethink our metrics to incentivize the complexity paradigm and the more connected, healthier community it can create. This has been written about in the context of evaluating our models (Liemohn et al., 2018; Hietala et al., 2020; Morley, 2020; McGranaghan et al., 2021b)and those comments are important, but we also need new metrics that describe healthier community.

The call to our whole community is to think about what might be metrics for a future Heliophysics community. For all of our metrics, we must more carefully define what is being illuminated and what is being neglected. These new metrics should be matched to the literacies listed above, which are in turn derived from the tenets of complexity science. For instance, to incentivize the skill of gathering, we should value conference and workshop organization and better assess the success of such events. Complexity science suggest that the use of network measures can be used to assess the density of collaborative networks and diffusion of ideas and techniques and therefore to provide insight into the success of gathering. For trans-media communication we can value dissemination beyond just technical publications. Already we are beginning to recognize open source software contributions. Elevating that emphasis will promote better knowledge engineering capacities. Overall, these new metrics can become measures of success and feedback tools to understand and improve connectivity, communication, and collaboration in our community.

A curriculum for complexity heliophysics and a more healthy community: open science

These literacies can coalesce into a new curriculum–a more information-literate Heliophysics community. This must be done together through co-creation; I believe that open science is an ethos under which we can join these literacies. There are many defintions of open science, but the one that I think best captures our needs is:

> Open science is transparent and accessible knowledge that is shared and developed through collaborative networks. (Vicente-Saez and Martinez-Fuentes, 2018)

There are two distinct components of the definition, pointing to substantive directions that our field needs to progress. The first (*transparent and accessible knowledge*) alludes to the need for intelligent and accessible data infrastructure and the platform to use it. The second (*collaborative networks*) subtly identifies a grand challenge that confronts our field and our community: the need to imagine and construct participatory ecosystems of knowledge sharing, governance, and trust. Together these two components indicate a *Knowledge Commons* (McGranaghan et al., 2021a).

> I argue that the commons is what we need to build Complexity Heliophysics and a healthier community.

Knowledge commons

A knowledge commons is a combination of intelligent information representation and the openness, governance, and trust required to create a participatory ecosystem whereby the whole community maintains and evolves this shared information space (McGranaghan et al., 2021a). It is one structure for bringing together transdisciplinary knowledge, both explicit in the form of datasets and publications and tacit in the form of the knowledge held by individuals. The commons elevate data to knowledge through the FAIR data principles (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al., 2016). Co-creation and maintenance also promotes a healthier more connected community in much the same way that a community functions around shared farmland or pastures.

A knowledge commons is a solution to another problem that we must quickly address: the haphazard or irresponsible use of AI/ML. Semantically enriching our data facilitates integrating data and tracing the provenance of data. Making open and accessible the process followed to create and train the model and the source code of the subsequently produced model would improve our ability to interrogate it and broaden participation in the evaluation. The result would be more robust AI/ML models that are able to be built on rather than isolated and opaque that remain only within the researcher's mind (and sometimes pass even from there). Thus, the KCs support robustness and resiliency of our research artifacts (over brittleness) and collective improvement over silo'ed individual development.

The KCs have the potential to democratize access to information, knowledge, and one another in the Earth and Space Science community.

The implications for early career researchers are immense. The KCs raise awareness about the resources (people, capabilities, assets, contents, data, models) available–awareness that newcomers to the field, inordinately, are working to develop. Further, the KCs are a place for dissemination of research artifacts that may not fit into the traditional publishing model (e.g., data analysis pipelines), leading to greater visibility and credit across the full research process. Finally, the commons are a place for richer engagement, providing opportunities for connection outside of in-person conferences and events, which may not be well-suited to all researchers and, worse, can sometimes reinforce existing cliques to the detriment of inclusivity. Asymmetries in knowledge lead to unhealthy communities, and the knowledge commons offer a framework to overcome the asymmetries.

Conclusion

Our world is increasingly interconnected. Society's most pressing problems dictate a new ability to contend with the interconnections. We suggest that a complexity science philosophy–the approaches for understanding phenomena that emerge from a collection of interacting objects–is required. We have outlined the methods of *Complexity Heliophysics* and discussed the implications of the complexity mindset. Those implications include a healthier and more flourishing community that is better connected, a set of literacies that our community should cultivate, and new metrics that those driving resources might adopt. We coalesced these ideas under the emerging approaches of Open Science and the knowledge commons. Ultimately, we suggest that it may be time for our project groups, departments, and research institutes to embrace the full implications of a shift in philosophy toward complexity and to incentivize new literacies through a redesign of curricula and adoption of open science principles that might create a more flourishing science and community.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The author extends deep thanks to Jeffrey Thayer, Seebany Datta-Barua, and Joshua Semeter for conversation extending over many years on the topics of complexity and community and for thoughtful input on drafts of this piece. The author gratefully acknowledges the support of the NASA Early Career

Investigator Program (ECIP) Program (NASA Grant Number: 80NSSC21K0622) for the resources to research, pursue, and write articles on the philosophy of Heliophysics science and how to make breakthroughs in our epistemology such as this one. Additionally, the author is deeply appreciative of the NASA Center for HelioAnalytics, funded by NASA ISFM, for supporting this work and creating a community in which conversations like these occur.

Conflict of interest

R. McGranaghan was employed by Orion Space Solutions LLC.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aschwanden, M. J., Crosby, N. B., Dimitropoulou, M., Georgoulis, M. K., Hergarten, S., McAteer, J., et al. (2014). 25 years of self-organized criticality: Solar and astrophysics. *Space Sci. Rev.* 198, 47–166. doi:10.1007/s11214-014-0054-6

Bankston, A. (2021). Great science begins with nurturing early-career researchers. *Issues Sci. Technol.*

Budaev, V., Zelenyi, L. M., and Savin, S. (2015). Generalized self-similarity of intermittent plasma turbulence in space and laboratory plasmas. *J. Plasma Phys.* 81, 395810602. doi:10.1017/s0022377815001099

Chang, T., Tam, S. W., Wu, C.-C., and Consolini, G. (2003). Self-organized criticality: An explanation of the 1/f noise. *Space Sci. Rev.* 107, 425–445. doi:10.1023/A:1025502023494

Consolini, G. (2002). Self-organized criticality: A new paradigm for the magnetotail dynamics. *Fractals* 10, 275–283. doi:10.1142/S0218348X02001397

Council, N. R. (2015). *Enhancing the effectiveness of team science*. Washington, DC: The National Academies Press. doi:10.17226/19007

Dods, J., Chapman, S. C., and Gjerloev, J. W. (2015). Network analysis of geomagnetic substorms using the supermag database of ground-based magnetometer stations. *J. Geophys. Res. Space Phys.* 120, 7774–7784. doi:10.1002/2015JA021456

Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T., and Vanderford, N. L. (2018). Evidence for a mental health crisis in graduate education. *Nat. Biotechnol.* 36, 282–284. doi:10.1038/nbt.4089

Hietala, H., Dimmock, A. P., Zou, Y., and Garcia-Sage, K. (2020). The challenges and rewards of running a geospace environment modeling challenge. *JGR. Space Phys.* 125, e2019JA027642. doi:10.1029/2019ja027642

Hughes, J., Mcgranaghan, R., Kellerman, A. C., Bortnik, J., Arrit, R. F., Venkataramani, K., et al. (2022). Revealing novel connections between space weather and the power grid: Network analysis of ground-based magnetometer and geomagnetically induced currents (gic) measurements. *Space weather.* 20, e2021SW002727. doi:10.1029/2021sw002727

Kauffman, S., and Kauffman, M. (1993). *The origins of order: Self-organization and selection in evolution. The origins of order: Self-organization and selection in evolution*. Oxford, UK: Oxford University Press.

Krakauer, D. (2019). *Worlds hidden in plain sight*. Santa Fe, NM: SFI Press.

Kurtz, C. F., and Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Syst. J.* 42, 462–483. doi:10.1147/sj.423.0462

Kurtz, C. F., and Snowden, D. J. (2007). A leader's framework for decision making. *Harv Bus. Rev.* 85, 68–149.

Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space weather.* 16, 2079–2102. doi:10.1029/2018SW002067

McGranaghan, R., Klein, S. J., Cameron, A., Young, E., Schonfeld, S., Higginson, A., et al. (2021). The need for a space data knowledge commons. *Struct. Collect. Knowl.* Available at: https://knowledgestructure.pubpub.org/pub/space-knowledge-commons.

McGranaghan, R. M., Bhatt, A., Matsuo, T., Mannucci, A. J., Semeter, J. L., and Datta-Barua, S. (2017). Ushering in a new frontier in geospace through data science. *JGR. Space Phys.* 122 (12), 586–590. doi:10.1002/2017JA024835

McGranaghan, R. M., Mannucci, A. J., Verkhoglyadova, O., and Malik, N. (2017). Finding multiscale connectivity in our geospace observational system: Network analysis of total electron content. *JGR. Space Phys.* 122, 7683–7697. doi:10.1002/2017JA024202

McGranaghan, R. M., Ziegler, J., Bloch, T., Hatch, S., Camporeale, E., Lynch, K., et al. (2021). Toward a next generation particle precipitation model: Mesoscale prediction through machine learning (a case study and framework for progress). *Space weather.* 19, e2020SW002684. doi:10.1029/2020sw002684

[Dataset] McGranaghan, R., Thompson, B., and Halford, A. (2020). The science of team science and inclusivity in Heliophysics. *Zenodo.* doi:10.5281/zenodo.4031888

Meadows, D., and Wright, D. (2008). *Thinking in systems: A primer*. Vermont, USA: Chelsea Green Pub.

Morley, S. K. (2020). Challenges and opportunities in magnetospheric space weather prediction. *Space weather.* 18, e2018SW002108. doi:10.1029/2018sw002108

Muir, J., and Gleason, H. W. (1911). *My first summer in the Sierra*. Cambridge, Mass: Houghton Mifflin Company & Riverside Press.

Parrish, J., Viscido, S., and Grünbaum, D. (2002). Self-organized fish schools: An examination of emergent properties. *Biol. Bull.* 202, 296–305. doi:10.2307/1543482

Rosas, F. E., Mediano, P. A. M., Jensen, H. J., Seth, A. K., Barrett, A. B., Carhart-Harris, R. L., et al. (2020). Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. *PLoS Comput. Biol.* 16, e1008289–22. doi:10.1371/journal.pcbi.1008289

Scheffer, C. S. F. J. E. A., M., Carpenter, S., Foley, J. A., Folke, C., and Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature* 413, 591–596. doi:10.1038/35098000

Schrijver, C. J., Kauristie, K., Aylward, A. D., Denardini, C. M., Gibson, S. E., Glover, A., et al. (2015). Understanding space weather to shield society: A global road map for 2015–2025 commissioned by cospar and ilws. *Adv. Space Res.* 55, 2745–2807. doi:10.1016/j.asr.2015.03.023

Thayer, J. (2011). Coupling, energetics, and dynamics of atmospheric regions (cedar) the new dimension, strategic vision

Varela, F., Rosch, E., and Thompson, E. (1992). *The embodied mind: Cognitive science and human experience*. Cambridge, MA, USA: MIT Press.

Vicente-Saez, R., and Martinez-Fuentes, C. (2018). Open science now: A systematic literature review for an integrated definition. *J. Bus. Res.* 88, 428–436. doi:10.1016/j.jbusres.2017.12.043

West, G. (2017). *Scale: The universal laws of life and death in organisms, cities and companies*. New York City, NY: Orion.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi:10.1038/sdata.2016.18

Wilson, E. (1999). "Consilience: The unity of knowledge," in *A Borzoi book* (New York, NY, USA: Vintage Books).

Wing, S., and Johnson, J. R. (2019). Applications of information theory in solar and space physics. *Entropy* 21, 140. doi:10.3390/e21020140

TYPE Mini Review
PUBLISHED 26 September 2022
DOI 10.3389/fspas.2022.1033542

My dealings with the aurora borealis

Gerhard Haerendel*

Max Planck Institute for Extraterrestrial Physics, Garching, Germany

Check for updates

OPEN ACCESS

EDITED BY
Maria Usanova,
University of Colorado Boulder,
United States

REVIEWED BY
Gordon Rostoker,
University of Alberta, Canada
Robert Lysak,
University of Minnesota Twin Cities,
United States

*CORRESPONDENCE
Gerhard Haerendel,
hae@mpe.mpg.de

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 31 August 2022
ACCEPTED 12 September 2022
PUBLISHED 26 September 2022

CITATION
Haerendel G (2022), My dealings with the aurora borealis.
Front. Astron. Space Sci. 9:1033542.
doi: 10.3389/fspas.2022.1033542

Two important decisions on my career path had the consequence that, after a quarter century of experimentation with barium plasma clouds, I was directing my research more and more towards physics of the auroral borealis. The combination of ground-based optical and radar observations and two national satellite missions were our means to deepen the understanding of the plasma physics background of the aurora, especially of discrete auroral arcs. Our contributions are put in perspective with the global research in this field by a quick run through the major steps in the exploration of the physics of the aurora. Although by the end of the 20th century all key ingredients for the understanding of auroral arcs were available, the present state leaves many open questions, foremost with respect to the true generators processes and the overall flow of momentum and energy. Some of these questions I tried to address during my retirement years.

KEYWORDS

career decisions, auroral plasma physics, brief auroral history, auroral generators, incomplete theories

1 Introduction

There are two interlinked goals for this Generation-to-Generation communications article. The first is that a career decision in the less challenging direction, based on the recognition of one's intellectual, physical, or other limitations, need not have negative effects on one's creativity and professional success. This applies to my way. It led more and more into the physics of the aurora borealis. As my second goal, I briefly review the development of this field during my professional life and conclude with some comments on the present state.

2 Recognize your limitations

Hardly any career has taken place without an event of luck and/or mentorship. This was so in my case. The lucky event was that I had been working on my thesis about the Van Allen Belt, when Germany decided to enter space research. A working group was formed to this end at the Max Planck Institute for Physics and Astrophysics in Munich in October 1961 under the direction of Reimar Lüst. It was almost natural that he invited the graduate student to join his group. The experiments with barium plasma clouds were quickly rewarded with fascinating optical phenomena, primarily in the auroral zone. That needed theoretical work and it turned out that I soon assumed the position of the house theoretician. On one of the first evenings of a rocket campaign in Ft. Churchill on the Hudson Bay, I experienced a fantastic auroral event, lying with my back on a rock and watching for at least half an hour the continuously moving and changing auroral arcs. I fell in love with the aurora borealis. Our work became very popular. For me it meant that I had to give presentations at conferences, was asked to join working groups of the European Space Research Organization, ESRO, and the German Ministry for Science and Technology (BMFT). For instance, I was the scientific German delegate to the Scientific Programme Board, SPB, ESRO, soon to mutate into ESA, and, from late 1972, (komma) member of the Launching Programme Advisory Committee, LPAC, ESRO's highest scientific advisory group. I sensed that the committee work and concomitant responsibilities would increase. At the same time, I directed rocket campaigns in Southern India, in Northern Sweden, and Greenland. Furthermore, in 1974 we had just begun to receive data from the dayside magnetopause taken by the ESRO satellite HEOS-2. We were at the frontiers of space physics. I searched myself and recognized it was difficult to do justice to these different tasks at the same time. I recognized my limitations in dealing with high-level management. Therefore, 1 day in late 1974 during an LPAC meeting, I declared my intention to resign from LPAC and soon after also from the SPB. That caused angry reactions from various sides and disappointment from my mentor Reimar Lüst. However, it set the course for my future. I had decided in favor of my own scientific work.

This work progressed well, in spite of some severe setbacks, like the crash of the Firewheel spacecraft in 1980. A few years later, we had reached our ultimate goal. Producing two artificial comets in the solar wind, we had opened a chapter of hitherto unknown plasma processes. Up to this point, my work had consisted of creating or exploiting opportunities for the application of the plasma cloud technique, and suddenly we noticed that we had largely exhausted its possibilities. What to do next? Already a few years before, I had taken over the leadership of a young group in infrared astronomy at our institute. I liked this new task and saw the possibility of a new frontier for myself. We had established a wonderful cooperation with a Dutch group on ESA's ISO (Infrared Space Observatory) mission and were planning a European infrared flying observatory, the Astroplane. Again, I examined myself. Was my education and technical knowhow sufficient to lead the IR group into a great future? The implicit answer was that I began to look for a promising young IR astronomer. I found him in the person of Reinhard Genzel in Berkeley working with Nobel Laureate Charley Townes. It soon proved that he did what I never would have been able to achieve, namely designing novel instrumentation for a most ambitious research program. The Nobel Prize in Physics of 2020 was a deserved recognition (Genzel, 2022).

3 Towards auroral physics

Until completion of the artificial comet experiments in 1985 (Haerendel et al., 1986; Valenzuela et al., 1986), I had had a wonderful career. As of 1972, I was director at MPE, had exciting work, and enjoyed a long-term support of my research by the Max Planck Society. The freedom I gained with my decision to step down from the high-level committees of ESRO/ESA, I devoted to a full engagement in our plasma clouds experiments and theoretical support. With our international sounding rocket campaigns in the auroral zone we had made some significant contributions to the physics of magnetosphere and aurora borealis, such as the penetration of a barium ion jet through the auroral acceleration region in Greenland in 1975 (Haerendel et al., 1976; Haerendel, 2019). In 1980, I conceived the fracture theory for embedded evening arcs (Haerendel, 1980). Until then our focus was on the application of the barium cloud technique with artificial comets and equatorial spread F being the main goals. When that was completed and the IR astronomy at MPE had been handed over to more competent hands, auroral research began to take the front seat for me.

In the late 1980s the opportunities arose with the participation in the Freja mission (Launch 1992) with Sweden (Lundin and Haerendel, 1993) and in the use of the incoherent radar technique with EISCAT (European Incoherent Scatter Radar). While the Freja mission produced many new insights into the nature of the primary auroral particles and their effects on the ionosphere (Lundin et al., 1994), a special topic could be addressed by EISCAT in combination with our highly developed imaging technique. Determination of plasma motions in the ionospheric F region with the first and tracking the motion of auroral arcs with the latter, we could prove the prediction of my fracture theory that embedded arcs have a proper motion with respect to the ambient plasma (Haerendel et al., 1993). This is essential for a continued energy supply out of the magnetic field stretched by magnetospheric convection to counter friction in the ionosphere. Many other insights into the structure of auroral arcs were obtained helping an increasing understanding of the physical processes behind. In the 1990s my attention was drawn more and more to the processes of the substorm onset and the energy entry into the magnetosphere. [The onset is the beginning. It is a matter of definition whether you call the entry simultaneous or subsequent, since entry is coupled with redistribution of the energy]. For an *in-situ* study of these processes, we had conceived and finally built the Equator-S spacecraft. It was a great pity that this mission ended abruptly in 1996, after 5 months of operation and just 2 months before the orbit had drifted into the midnight sector. Throughout the 1990s, my group and guests engaged in the data reduction and interpretation of the observed auroral phenomena and the theory.

4 Major steps in the development of auroral physics

The work of my group to auroral physics, of course, only represented scattered contributions to the worldwide research of the origin of the fascinating aurora borealis. When I was working on my thesis, I could observe already the first results of the impressive progress in auroral physics made by means of space flight. In 1960, Carl McIlwain (1960) had discovered nearly mono-energetic electrons above an auroral arc indicating electrostatic acceleration. Equally fast expanding ground-based research resulted in the recognition of what Syun-Ichi Akasofu had named the substorm (Akasofu, 1964), and Rolf Boström defined the global current system driving ionospheric convection (Boström, 1964). [Magnetospheric forces drive motions of the hot plasma which are coupled to the ionosphere by the global current systems.] The seventies were the time of great discoveries and setting the theoretical foundations. Already in 1972, Vasyliunas (1972) related the origin of the global currents to pressure gradients in the hot magnetospheric plasma. A most important finding was the inverted-V structure of the auroral electron spectrum, discovered by Frank and Ackerson (1971), and the accompanying interpretation as originating from U-shaped electrostatic potentials by Don Gurnett (1972). Hallinan and Davis (1970) had found that auroral rays were, in reality, moving curls indicating the presence of strong shear flows or transverse electric fields above an arc. This was experimentally proven by the employment of Langmuir double probes to measure electric fields and the identification of electrostatic U-shaped potentials by Mozer et al. (1977). At the same time theorists wondered about the processes able to sustain parallel electric fields. There were two widely different proposals, current driven anomalous resistivity by Kindel and Kennel (1971) and Papadopoulos (1977) and a current-voltage relationship derived by Knight (1973) on the basis of kinetic theory in presence of the mirror effect. [Repositioned and reformulated.] The latter theory was elaborated by Fridman and Lemaire (1980). Lyons (1980) used a field-parallel conductance on the basis of the Knight-relation in combination with the Pedersen conductivity to derive M-I coupling scales, not applicable to auroral arcs. Whereas the author (Haerendel 1980) proposed oblique propagating quasi-static Alfvén waves as energy suppliers to the auroral acceleration region, Goertz (1981) associated kinetic or inertial Alfvén waves with short-lived auroral structures. Realistic scales of auroral arcs followed from the Alfvén wave conductance coupled with the Knight conductance (Lysak 1985). Measurements of the Freja spacecraft (Lundin et al., 1994) showed that, instead of accelerating magnetospheric electrons, kinetic Alfvén waves deliver energy to the cool plasma of the topside ionosphere, accelerating electrons parallel to the magnetic field and the ions transversely. The functioning of this ionospheric erosion process has been studied by Chaston et al. (2006). The Fast (Fast Auroral SnapshoT) mission, launched in 1996 (Carlson et al., 1998), brought an unprecedented enrichment of the physics of auroral arcs. This pertained foremost to the micro-processes excited in the acceleration region, which, on the one hand, play a role in exchanging energy and momentum between the e. m. field and the charged particles, thus contributing to the field-parallel resistivity. On the other hand, they give rise to a host of wave fields serving as diagnostic tools. Further great progress coming from the FAST data applies to the downward currents. The data elucidated and underpinned the "pressure cooker" theory of Gorney et al. (1985).

I think that by the end of the 20th century the ingredients to understand discrete auroral arcs were available. What was needed was to identify the energy sources and respective mechanical forces that drive quasi-steady currents or waves thus transferring energy into the magnetic field. Waves may deliver their energy directly to auroral particles or by interaction with the cool plasma of the topside ionosphere. Quasi-stationary currents are set up by interaction of the driving forces with the frictional ionosphere thereby storing energy in the sheared magnetic field, from where it may be extracted in auroral acceleration regions. Apart from diffuse arcs owed to the precipitation of pitch-angle scattered electrons or ions, I maintain that all discrete aurora is caused by extracting energy from intermediately stored energy in terms of sheared magnetic field components. This is grossly different from the suspicion voiced in the early days of space research that anti-parallel magnetic field reconnection in the tail was a source of auroral arcs (Atkinson 1992). By contrast, the powerful energy conversion processes in the solar corona are generally attributed to reconnection processes. [The point lies in the contrast between sheared and anti-parallel field components]. At this point, I am asking: Why have so far only few of the striking auroral structures been explained along these lines? I will have a brief look at that in the next section.

5 What do we understand?

An impressive outcome of a workshop at the International Space Science Institute, ISSI, in Berne led by Dave Knudsen is the series of comprehensive review papers on auroral research (Knudsen et al., 2021). It is a great thesaurus for the state of the art but acknowledges also the many unsolved questions. In the spirit of the considerations at the end of the preceding section, I will look primarily into the reviews covering quiet discrete arcs, namely (Borovsky et al., 2020; Karlsson et al., 2020; Lysak et al., 2020). I will largely neglect the other eight reviews covering the wide range of other auroral phenomena, such as small-scale and mesoscale auroral forms, or dayside and subauroral auroral forms.

The review by Lysak et al. (2020) is largely devoted to understanding of the setup and maintenance of parallel electric fields or voltages in the presence of quasi-steady currents and propagating and reflected Alfvén waves. This is done with kinetic theory as well as with two-fluid descriptions. The relation between transverse and parallel electric fields is discussed in detail, whereas the potential role of anomalous resistivity finds little coverage. Other topics are the reflection of Alfvén waves, and the production of ion conical distributions, the connection of double layers with discontinuities in the plasma parameters, and the relation of density cavities and downward currents. All these topics are key issues in auroral physics and, apart from some subtleties, largely understood. There is, however, one shortcoming; they are treated in non-moving reference systems, which is typical for most of the literature on auroral arcs. Since the experiments of Wescott et al. (1975), it has been known that auroral arcs can have a proper motion with respect to the ambient plasma. It turns out that acceleration regions propagating with respect to the ambient plasma require new considerations of energy and momentum transport (Haerendel, 2021).

The review by Karlsson et al. (2020) presents the key scales of auroral arcs, widths, height distributions of the acceleration regions, potential distributions and relations to upward and downward currents. Much attention is directed to electrostatic potential structure and the current closure in the ionosphere. A special topic is the unipolar and multipolar nature of arcs, i.e., the absence or presence of return currents in the immediate neighborhood of an arc. It is connected with the spatial distribution of the energy supply. Also this review regards arcs as electrostatic entities. Proper motions are mentioned.

The essence of understanding auroral arcs is the identification of potential generator processes. Since at least discrete arcs seem to be related to acceleration by parallel electric fields embedded in upward field-aligned currents or in propagating Alfvén waves, the identification of the generator of these currents is necessary. This is the subject of the review of Borovsky et al. (2020). The authors distinguish between the classes of generators of high-latitude and low altitude arcs, meaning sources at magnetospheric outer interfaces or inside the magnetosphere, respectively. In the first class, there are plenty of ideas for the generation of Alfvén waves with small perpendicular scales, i.e., kinetic or inertial Alfvén waves. Owing to their parallel electric field component, they can accelerate electrons in flight, but only to limited energies. Conversion of the wave energy in the topside ionosphere, as suggested by the discoveries of the Freja mission (Lundin et al., 1994) is not mentioned. Quiet high-latitude arcs are attributed to the generation of transverse potentials at interfaces of plasmas with different thermodynamic properties. These potentials are meant to propagate along the magnetic field down to the ionosphere, where current closure and Pedersen conductivity are used to determine the local potential. The difference with the high-altitude potential leads to the electron acceleration. I have problems with these theories, because current closure in the ionosphere means momentum dumping, whereas the origin of this momentum remains obscure. This problem does not exist with the second class of generators, which are typically current generators, driven by pressure gradients. One example is the arcs embedded in the magnetospheric convection in the evening auroral oval. However, the review also considers static pressure gradients to be sources for the pre-midnight quiescent arcs without flows associated [e.g., (Stasiewicz, 1985)]. [The low pressure region was a cloud. I refrain from quoting details]. Again, these models lack identifications of the source of momentum. If flows are driven, they constitute voltage generators at first sight. However, they cannot sustain the arc for more than a few Alfvén wave reflection periods (Haerendel 2014) and must be maintained by pressure gradients. In case of embedded arcs, the momentum is supplied from the release of stored magnetic shear stresses. Another viable current generator is flow braking. While the merit of the review of Borovsky et al. (2020) is the extensive listing of potential auroral generators, it also raises many questions. In spite of the existence of many incomplete theories, very few auroral forms have been described in a way allowing quantitative evaluation and comparison with data. Why is that so? The information on any observed auroral form is incomplete. Therefore it needs focused and temporally better resolved observations, intuition, theory, and numerical modeling, which, as the authors write, "... encompass the entire auroral-arc region from the equatorial magnetosphere to the resistive ionosphere, ... ".

6 Conclusion

I return to the beginning of my story. Insights into my limitations had led to two decisions that may seem as downward steps on my career ladder. In some sense they were. Such decisions turn out to be unavoidable, whenever the self-examination is done honestly. Cheating may lead to unpleasant consequences. However, the process of self-examination may lead to sensing one's unexploited abilities and to take the path into a risky future, often with success. My case should be seen as rather atypical. What were the benefits for me? It left me in my established realm of competence, planning experiments, conceiving missions, participating personally in rocket and satellite campaigns, being deeply immersed in interpreting data, and providing theoretical support. It led my way inevitably into the fascinating auroral plasma physics and some ambitious projects. It continued to fill my meanwhile 18 years of retirement with the pleasure of addressing a few of the many unanswered questions. During this period, I was able to publish 25 theoretical or interpretative papers on the aurora as well as in solar physics as sole or first author and a few more as co-author only. Thus, from my entry into space research until today, I had a rich, often exciting, and certainly intellectually challenging life. More can be read on that in my professional autobiography: "My Life in Space Exploration," Springer Biographies, 2022.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akasofu, S.-I. (1964). The development of the auroral substorm. *Planet. Space Sci.* 12, 273–282. doi:10.1016/0032-0633(64)90151-5

Atkinson, G. (1992). Mechanism by which merging at X-lines causes discrete auroral arcs. *J. Geophys. Res.* 97, 1337–1344. doi:10.1029/91ja02443

Borovsky, J. E., Birn, J., Echim, M. E., Fujita, S., Lysak, R. L., Knudsen, D. J., et al. (2020). Quiescent discrete auroral arcs: A review of magnetospheric generator mechanisms. *Space Sci. Rev.* 216, 1. doi:10.1007/s11214-019-0619-5

Boström, R. (1964). A model of the auroral electrojets. *J. Geophys. Res.* 69 (23), 4983–4999. doi:10.1029/jz069i023p04983

Carlson, C. W., Pfaff, R. F., and Watzin, J. G. (1998). The fast auroral SnapshoT (FAST) mission. *Geophys. Res. Lett.* 25 (12), 2013–2016. doi:10.1029/98gl01592

Chaston, C. C., Gernot, V., Bonnell, J. W., Carlson, C. W., McFadden, J. P., Ergun, R. E., et al. (2006). Ionospheric erosion by Alfven waves. *J. Geophys. Res.* 111, A03206. doi:10.1029/2005ja011367

Frank, L. A., and Ackerson, K. L. (1971). Observations of charged particle precipitation into the auroral zone. *J. Geophys. Res.* 76, 3612–3643. doi:10.1029/ja076i016p03612

Fridman, M., and Lemaire, J. (1980). Relationship between auroral electrons fluxes and field aligned electric potential difference. *J. Geophys. Res.* 85, 664–670. doi:10.1029/ja085ia02p00664

Genzel, R. (2022). Nobel lecture 23 march 2022. Available at: www.nobelprize.org/prizes/physics/2020/genzel/lecture/.

Goertz, C. K. (1981). "Discrete breakup arcs and kinetic Alfvén waves," in *Physics of auroral arc formationProceedings of the chapman conference on formation of auroral arcs* (Washington, DC: American Geophysical Union), 451–455.

Gorney, D. J., Chiu, Y. T., and Croley, D. R. (1985). Trapping of ion conics by downward parallel electric fields. *J. Geophys. Res.* 90 (A5), 4205–4210. doi:10.1029/ja090ia05p04205

Gurnett, D. (1972). "Electric field and plasma observation in the magnetosphere," in *Critical problems of magnetospheric physics, proc* (Madrid: COSPAR/IAGA/URSI Symposium), 123–138.

Haerendel, G. (2021). Auroral arcs: The fracture theory revisited. *JGR. Space Phys.* 126, e2020JA028194. doi:10.1029/2020JA028194

Haerendel, G. (1980). "Auroral particle acceleration - an example of a universal plasma process," in Proceedings of the 5th ESA-PAC Symposium on European Rocket & Balloon Programmes & Related Research, Bournemouth Southern England, September, 1980 197–210. (ESA SP-152).

Haerendel, G., Buchert, S., LaHoz, C., Raaf, B., and Rieger, E. (1993). On the proper motion of auroral arcs. *J. Geophys. Res.* 98, 6087–6099. doi:10.1029/92ja02701

Haerendel, G. (2019). Experiments with plasmas artificially injected into near-Earth space. *Front. Astron. Space Sci.* 30. doi:10.3389/fspas.2019.00029

Haerendel, G., Paschmann, G., Baumjohann, W., and Carlson, C. W. (1986). Dynamics of the AMPTE artificial comet. *Nature* 320, 720–723. doi:10.1038/320720a0

Haerendel, G., Rieger, E., Valenzuela, A., Föppl, H., and Stenbaek-Nielsen, H. (1976). "First observation of electrostatic acceleration of barium ions into the magnetosphere," in Proceedings of the Symposium on Present and Future European Sounding Rocket and Balloon Research in the Auroral Zone, Bavaria, Germany, May, 1976 203–211. Schloss Elmau, ESA SP-155.

Hallinan, T. J., and Davis, T. N. (1970). Small-scale auroral arc distortions. *Planet. Space Sci.* 18, 1735–1744. doi:10.1016/0032-0633(70)90007-3

Karlsson, T., Andersson, L., Gillies, D. M., Lynch, K., Marghitu, O., Partamies, N., et al. (2020). Quiet, discrete auroral arcs – Observations. *Space Sci. Rev.* 216, 16. doi:10.1007/s11214-020-0641-7

Kindel, J. M., and Kennel, C. F. (1971). Topside current instabilities. *J. Geophys. Res.* 76, 3055–3078. doi:10.1029/ja076i013p03055

Knight, S. (1973). Parallel electric fields. *Planet. Space Sci.* 21, 741–750. doi:10.1016/0032-0633(73)90093-7

Knudsen, D. J., Borovsky, J. E., Karlsson, T., Kataoka, R., and Partamies, N. (2021). Editorial: Topical collection on auroral physics. *Space Sci. Rev.* 217, 19. doi:10.1007/s11214-021-00798-8

Lundin, R., Eliasson, L., Haerendel, G., Boehm, M., and Holback, B. (1994). Large-scale auroral plasma density cavities observed by Freja. *Geophys. Res. Lett.* 21, 1903–1906. doi:10.1029/94gl00888

Lundin, R., and Haerendel, G. (1993). The Freja satellite mission. *Eos Trans. AGU.* 74, 321–324. doi:10.1029/93eo00394

Lyons, L. R. (1980). Generation of large-scale regions of auroral currents, electric potentials, and precipitation by the divergence of the convection electric field. *J. Geophys. Res.* 85, 17–24. doi:10.1029/ja085ia01p00017

Lysak, R. L. (1985). Auroral electrodynamics with current and voltage generators. *J. Geophys. Res.* 90 (A5), 4178–4190. doi:10.1029/ja090ia05p04178

Lysak, R. L., Echim, M., Karlsson, T., Marghitu, O., Rankin, R., Song, Y., et al. (2020). Quiet, discrete auroral arcs: Acceleration mechanisms. *Space Sci. Rev.* 216, 92. doi:10.1007/s11214-020-00715-5

McIlwain, C. (1960). Direct measurement of particles producing visible auroras. *J. Geophys. Res.* 65, 2727–2747. doi:10.1029/jz065i009p02727

Mozer, F., Carlson, C. W., Hudson, M. K., Torbert, R. B., Parady, B., Yatteau, J., et al. (1977). Observations of paired electrostatic shocks in the polar magnetosphere. *Phys. Rev. Lett.* 38, 292–295. doi:10.1103/physrevlett.38.292

Papadopoulos, K. (1977). A review of anomalous resistivity for the ionosphere. *Rev. Geophys.* 15, 113–127. doi:10.1029/rg015i001p00113

Stasiewicz, K. (1985). Generation of magnetic-field aligned currents, parallel electric fields, and inverted-V structures by plasma pressure inhomogeneities in the magnetosphere. *Planet. Space Sci.* 9, 1037–1048. doi:10.1016/0032-0633(85)90022-4

Valenzuela, A., Haerendel, G., Föppl, H., Melzner, F., Neuss, H., Rieger, E., et al. (1986). The AMPTE artificial comet experiments. *Nature* 320, 700–703. doi:10.1038/320700a0

Vasyliunas, V. M. (1972). "The interrelationship of magnetospheric processes," in *Earth's magnetospheric processes*. Editor B. M. McCormac (Dordrecht-Holland: D. Reidel Publ. Comp.), 29–38.

Wescott, E. M., Stenbaek-Nielsen, H. C., Davis, T. N., Murcray, W. B., Peek, H. M., and Bottoms, P. J. (1975). The $L = 6.6$ Oosik barium plasma injection experiment and magnetic storm of March 7, 1972. *J. Geophys. Res.* 80, 951–967. doi:10.1029/ja080i007p00951

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 27 September 2022
DOI 10.3389/fspas.2022.1008888

Check for updates

OPEN ACCESS

EDITED BY
Noora Partamies,
The University Centre in Svalbard,
Norway

REVIEWED BY
Luke Barnard,
University of Reading, United Kingdom
Nithin Sivadas,
National Aeronautics and Space
Administration, United States

*CORRESPONDENCE
Elena A. Kronberg,
kronberg@geophysik.uni-muenchen.de

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 01 August 2022
ACCEPTED 09 September 2022
PUBLISHED 27 September 2022

CITATION
Kronberg EA (2022), Data analysis in space physics: My experience and lessons learned.
Front. Astron. Space Sci. 9:1008888.
doi: 10.3389/fspas.2022.1008888

Data analysis in space physics: My experience and lessons learned

Elena A. Kronberg*

Department of Earth and Environmental Sciences, Ludwig Maximilian University of Munich, Munich, Germany

The specific area of investigation in this perspective is data analysis in space physics. This paper is intended to be useful for those who start working with observations in space physics, especially with a focus on charged particle measurements. I forward lessons I learned regarding the data analysis such as calibration, statistics and machine learning. I also list practices which I find important in research in general. An outlook on possible future directions in space physics is given.

KEYWORDS

space physics, charged particle observations, data analysis, statistics, machine learning

1 Introduction

A wide spectrum of methods of data analysis in space physics are well presented in, e.g., Paschmann and Daly (1998) and Paschmann and Daly (2008). However, there are common mistakes that are not described in the literature. Because of time pressure and "result orientation", people sometimes do not bother enough to familiarize themselves with the data and the calibrations involved. It is also common to put undue trust in data. In Section 2.1, I describe the lessons I learned about the handling of data and best practices for communicating with data providers. Large data sets can provide global pictures of physical processes in the space environment. Statistical methods in space physics and also their misuse can be found in, e.g., Reiff (1990). However, the processing of large data sets can easily lead to erroneous results if not done carefully. In Section 2.2, common mistakes in the processing of the large data sets are described. Machine learning techniques are popular for dealing with long observation series. Their application in space sciences is highlighted, e.g., in Bortnik and Camporeale (2021). In Section 2.3, common mistakes in the application of these methods are pointed out. In Section 3, general lessons for space physicists, which I learned during my career, are discussed. In Section 4, I will give an outlook on directions in space physics.

2 Lessons: The data analysis

2.1 Get to know your data

I always recommend to students:

- Lesson 1: Use the data carefully, read the metadata, read User Guides and Calibration Reports, contact PIs and Co-investigators about the data

Data are often taken from a data archive. After my graduation, I became a member of the RAPID [the Research with Adaptive Particle Imaging Detector (Wilken et al., 1997)] team at the Max Planck Institute for Solar System Research led by Dr. Patrick Daly, the principal investigator (PI) of this instrument on the Cluster mission (Escoubet et al., 2001) by the European Space Agency. My job was to work on calibration and preparation of the RAPID data for the Cluster Science Archive (CSA, https://www.cosmos.esa.int/web/csa), writing the User Guide (Daly and Kronberg, 2022), Calibration report (Kronberg et al., 2022) and Interface Control Document [ICD, (Daly et al., 2021)]. Before this project, during my doctorate, I was using data from the Galileo mission to analyze the dynamics of the Jovian magnetosphere without considering that the data could have errors and biases. However, working on data calibrations and processing I understood how much data is altered before it is archived. I learned about the problems and corrections of particle data, how many iterations are needed to obtain "ideal" values, and how much work is involved in the preparation of scientific datasets. For example, to convert raw counts measured by the Cluster/RAPID into electron differential fluxes, one has to apply geometry factors, take the time-dependent efficiency of the detectors into account, shift the initial energy threshold of the lowest channel, and remove the solar contamination and the pedestal noise, see more details in Kronberg et al. (2022). I realized the importance of archiving the data and maintaining of data archives such as CSA. I advocate for this work to be recognized by the scientific community by using DOIs for data sets and related documents such as User Guides, Calibration reports and ICDs.

The highlighting characteristic of the CSA is that the data quality is controlled by a dedicated team. The observations from different instruments, spacecrafts and missions are cross-calibrated and, therefore, the data is complemented and improved. The calibrations which were applied on the data are well described in the Calibration Reports and User Guides of the corresponding instruments. The calibrations, are therefore, not a black box and a user can, in principle, take the raw data and apply a calibration procedure to receive a scientific product. Another advantage of the CSA is that the archiving team works closely with scientists. Because of this interaction, the archive offers many useful scientific products and convenient interfaces, e.g., for plotting.

I show several examples of my own work in which the lesson above was crucial to avoid wrong results.

In my work on the origin of energetic ion events measured upstream of the Earth's bow shock by STEREO, Cluster, and Geotail missions (Kronberg et al., 2011), I worked on explaining upstream events observed far away (> 70 R_E) from the Earth. For this I have combined observations from the above-mentioned space missions. I used particle measurements by STEREO which were given to me in the form of an ASCII table without any metadata. Measurements by the Cluster/RAPID instrument are delivered to the archive in keV units for the particle intensities. Being naive, I thought that the same is true for the STEREO data. It was quite striking that the energetic particle intensities measured by STEREO were very strong compare to those measured near the Earth by the Cluster and Geotail. We even had an explanation for such an interesting observation. Luckily, before submitting the manuscript an expert in STEREO data has noted that instead of keV, the intensity units are MeV. This spoiled our initial interpretation of the data (which by the way was very exciting), we needed to rework the interpretation quite a lot but we avoided submitting an incorrect study.

Here is another instructive example. The Van Allen Probe mission has discovered a temporal third radiation belt which was observed for more than 4 weeks (Baker et al., 2013). Generally the data in radiation belts observed by the RAPID instrument were considered to be rather useless due to background contamination. A warning about this issue has been stated in the RAPID User Guide. Still, a manuscript using Cluster/RAPID observations was submitted to the Nature journal, about the discovery of a third radiation belt which is persistent on long time scales, for several months and during several years. This could have been a great discovery. The reviewers have commented that the manuscript can be published if the RAPID experts confirm that this belt is not a contamination of the observations. The RAPID team was already working on simulations of the RAPID/Imaging Electron Spectrometer (IES) in the radiation belt environment. The detector was bombarded with particles at an energy spectrum corresponding those in the radiation belts. Our results have shown that the "third radiation belt" is indeed a contamination (Kronberg et al., 2016). Unfortunately, the manuscript was not published in Nature but we avoided the publication of wrong results. Since then, a novel cleaning technique for background contamination, also described in the RAPID Calibration report, has helped to make the RAPID data in the radiation belts useful for science. This allowed, for example, an extensive statistical study of radiation belts (Smirnov et al., 2019) and the deduction of information on particle anisotropy for the calculation of the wave power of chorus waves (Breuillard et al., 2015). We also created a guide on how to calculate adiabatic invariants using the Cluster/RAPID data (Smirnov et al., 2020a) and the LSTAR product for the CSA.

Eventually, the Galileo/Energetic Particle Detector (EPD, (Williams et al., 1992)) ion observations which I used for my doctorate, never doubting their accuracy, were corrected for radiation background contamination. It did not affect the results of my thesis. However, in my recent study of the ion acceleration in plasmoids (Kronberg et al., 2019), I excluded the formerly included helium observations because after the correction we did not have a sufficient amount of reliable helium data. Thanks to EPD experts!

- Lesson 2: Question "gold standards"

The data are not "static", meaning they may change after many years if a better calibration technique is found. Moreover, calibrations are a form of measurement interpretation. They can be subjective. This can affect older studies. This can also affect "gold standards in observations". For example, the charged particle observations by Combined Release and Radiation Effects Satellite (CRRES) launched in 1990 were considered as a "gold standard". The charged particle observations by Polar and LANL satellites were cross-calibrated with those from the CRRES. I was working on the cross-calibration of protons observed by Cluster between the two instruments: the RAPID and the Cluster Ion Spectrometry [CIS, (Rème et al., 2001)]. The cross-calibrations were relatively good (Kronberg et al., 2010) [they were redone later for both instruments but still having relatively good agreement, see Kronberg et al. (2022)]. However, comparing the RAPID proton observations with those from the Polar mission we found a difference of about one order of magnitude. We were not happy to see this, because the data from the Polar mission were well aligned with the "gold standard". However, the agreement of the measurements by the CIS and the RAPID instruments and later the agreement found with the measurements from the Van Allen Probes (new "gold standard" in the radiation belts) and observations from the Arase mission gave us confidence in our data.

2.2 Statistics

- Lesson 3: A value of zero is also a measurement, do not remove it without a reason

It is generally advisable to plot the data to check the type of the distribution, analyze outliers and clean the data before doing statistics. It often happens in particle observations that zeroes are ignored because they are not suitable for logarithmic plots. Also, the absence of an observation (commonly indicated by "fill values" in the data) is often not distinguished from values of zero. In plots, both are then shown as data gaps. Please remember that a value of zero is also a valid measurement, meaning that there was no particle entering the detector at a specific energy at this time. Slip-ups in post processing are less likely if NaNs (not a number, defined in IEEE 754) are used for missing values.

- Lesson 4: Be careful with interpolations

Another mistake which I often observe is interpolation of data between large data gaps. Such inappropriate interpolations often remain undiscovered in the data. Please make sure that the interpolation is reasonable. For example, the spacecraft should not cross several different plasma regimes during a data gap. I recommend avoiding interpolations or using them only for short data gaps.

- Lesson 5: Be careful with possible solar cycle related biases in statistics

You should use as much data as possible. Different phases of the solar cycle (which is 22 years!) may lead to quite different statistical results depending on the phase the sampling was done. In space observations it is often difficult to avoid biases related to the solar cycle, but you need to be aware of it.

- Lesson 6: Please calculate the uncertainty of your results

I always tell my students: please add error bars. I often see a lack of error bars in manuscripts which I review, and conclusions are made just based on a visible trend or the difference of the color in a spectrogram. It is especially dangerous if a spectrogram is made using a rainbow color map (Borland et al., 2007). The differences often appear less prominent when using perceptually uniform color maps. You should always question the uncertainty of the results and separate signal from noise. Even simple, random uncertainties can create a statistical or systematic bias (Sivadas and Sibeck, 2022). Remember that measurements have (systematic) error. We usually measure only a subset of a population, leading to sampling errors. This is very obvious but often ignored. Also calibrations of the data introduce errors but this is usually not taken into account in most studies and data sets.

In charged particle measurements, individual intensity measurements may have different uncertainties, depending on how many counts were accumulated during the time interval used to derive the intensity. In proper data archives, such as CSA, an uncertainty is provided for each measurement. This is especially important for the estimation of the spectral slope in particle distributions.

There are many methods used by statisticians for problem of separation of signal from noise and making conclusions under uncertainties, see, e.g., Wasserstein et al. (2019). Conclusions in space physics have to be made by taking into account all known uncertainties.

2.3 Machine learning

Applying machine learning techniques to observations in space physics for derivation of prediction and forecasting models can be very useful. My students found the book by Geron (2019) quite useful.

- Lesson 7: Be careful with splitting time series

One common mistake is to apply the Scikit-Learn (Pedregosa et al., 2011) train_test_split procedure on time series and getting excellent predictions that occur because a model just interpolates between adjacent times.

- Lesson 8: Make sure there is no overfitting

One easily gets excited about an excellent performance of the model on training data. However, this is often a sign of overfitting.

Namely, there is a large discrepancy in the performance on the training data and the (unseen) validation data (Ghojogh and Crowley, 2019). In this case the model just remembers the training data. In the ideal case the gap between training and validation errors should be small (Goodfellow et al., 2017).

- Lesson 9: Be careful with the interpretation of feature importances

One should be careful with interpretation of the importance of features (for predictors such as solar wind parameters) for understanding underlying physics. Machine learning models combine individual features to get the best output result and this combination can vary from model to model and also be different from considering one input and one output variable in isolation. Please also consider uncertainties of the importances.

- Lesson 10: Archive your codes, data sets and models

It is great to archive the codes, the data and the models on, for example, zenodo or make them available through GitHub, so that other scientists can build up on it in future studies.

3 Discussion: General lessons for space physicists

- Lesson 11: Do not try to accommodate the data with the expected physical picture: physics is complicated and there can be various reasons for why the data does not fit.

For example, one can expect that the mass loading from the moon Io in the Jovian magnetotail leads to a pressure increase in the magnetodisk (I searched for a long time for such signatures in the data during my PhD). However, it can be that the disk just becomes larger and the plasma pressure equilibrium does not change.

- Lesson 12: Use as many observational points as possible

The physical picture may become more complex and bring more questions but it also helps to make a global picture of a phenomena. For example, in Kronberg et al. (2017a) we used observations from 14 satellites to monitor a substorm event. This gave us an opportunity to simultaneously observe phenomena which are usually studied separately such as current sheet flapping, magnetic field dipolarization, signatures of reconnection in the near-Earth tail, dispersionless and dispersed injections and their propagation, electron acceleration by ultra low frequency waves etc.

- Lesson 13: Do not give up if you believe in your research after your manuscript is rejected: it will become better.

A couple of my now well cited papers were initially rejected. However, it can happen that one has to give up on a manuscript because one realizes that the approach was wrong.

- Lesson 14: Find a mentor

It is great to have a mentor who can give directions and set up goals. For different aspects of a scientific career one may need different mentors. Also one can learn a lot from younger people.

- Lesson 15: Be a part of such a team as at an International Space Science Institute (ISSI) in Bern

One of the best places to conceive scientific ideas is the International Space Science Institute (ISSI) in Bern, Switzerland, which allows to gather teams of experts and make them collaborate closely in an informal way for about 1 week several times. About one third of my first author papers were conceived in this place.

4 Outlook

In summary I outline several directions which in my opinion should be developed in space physics:

- Lesson 16: Combination of data and models

Communication between observers and modelers is difficult, although a lot of effort has been made in this direction. Models should be verified by observations. Observations, on the other hand, can be better understood if they are related to physical models.

- Lesson 17: Combine different energies and species

It is common to separate, for example, in the inner magnetosphere the regions by the energy of electrons: plasmasphere (less than ~3 eV), warm plasma cloak [~10 eV–3 keV, (Chappell et al., 2008)], ring current (~3–100 keV) and the radiation belts (above 50 keV). However, efforts are still needed to understand how the particle populations move along these energy scales. For example, the dynamics of cold ions and electrons is still not well understood (Delzanno et al., 2021). Assessing just bulk energies without considering cold and energetic parts may be misleading (Kronberg et al., 2017c). Measurements of heavy ions are still far from ideal and their influence on the magnetospheric dynamics is not well understood (Kronberg et al., 2014).

- Lesson 18: Look in 3D

In space physics it is common to map the data to the equatorial plane in GSE/GSM coordinates. This is fine. But a lot of new physics is also hidden if one looks at the magnetosphere in 3D. Examples are mysterious north-south hemispheric asymmetries and diamagnetic cusps (from the particle observations point of view).

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This work is supported by the German Research Foundation (DFG) under number KR 4375/2-1 within SPP "Dynamic Earth."

Acknowledgments

I am thankful to Songyan Li, Aljona Blöcker and Patrick W. Daly for helpful advices.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Baker, D. N., Kanekal, S. G., Hoxie, V. C., Henderson, M. G., Li, X., Spence, H. E., et al. (2013). A long-lived relativistic electron storage ring embedded in earth's outer van allen belt. *Science* 340, 186–190. doi:10.1126/science.1233518

Borland, D., Russell, R. T., and Ii, T. (2007). Rainbow color map (still) considered harmful. *IEEE Comput. Graph. Appl.* 27, 14–17. doi:10.1109/mcg.2007.323435

Bortnik, J., and Camporeale, E. (2021). Ten ways to apply machine learning in Earth and space sciences. *EOS* 102. doi:10.1029/2021EO160257

Breuillard, H., Agapitov, O., Artemyev, A., Kronberg, E. A., Haaland, S. E., Daly, P. W., et al. (2015). Field-aligned chorus wave spectral power in Earth's outer radiation belt. *Ann. Geophys.* 33, 583–597. doi:10.5194/angeo-33-583-2015

Chappell, C. R., Huddleston, M. M., Moore, T. E., Giles, B. L., and Delcourt, D. C. (2008). Observations of the warm plasma cloak and an explanation of its formation in the magnetosphere. *J. Geophys. Res. (Space Phys.* 113, A09206. doi:10.1029/2007JA012945

Daly, P. W., and Kronberg, E. A. (2022). *User guide to the RAPID measurements in the Cluster Science Archive (CSA)*. Paris: Tech. Rep. CAA–EST–UG–RAP.

Daly, P. W., Mühlbachler, S., and Kronberg, E. A. (2021). *Cluster Science Archive: Interface Control document for RAPID*. Paris: Tech. Rep. CAA–EST–ICD–RAP.

Delzanno, G. L., Borovsky, J. E., Henderson, M. G., Resendiz Lira, P. A., Roytershteyn, V., and Welling, D. T. (2021). The impact of cold electrons and cold ions in magnetospheric physics. *J. Atmos. Solar-Terrestrial Phys.* 220, 105599. doi:10.1016/j.jastp.2021.105599

Escoubet, C. P., Fehringer, M., and Goldstein, M. (2001). The cluster mission–introduction. *Ann. Geophys.* 19, 1197–1200. doi:10.5194/angeo-19-1197-2001

Geron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and TensorFlow*. Sebastopol: O'Reilly Media, Inc.

Ghojogh, B., and Crowley, M. (2019). *The theory behind overfitting, cross validation, regularization, bagging, and boosting: Tutorial*. Ithaca: arXiv e-prints, arXiv:1905.12787.

Goodfellow, I., Bengio, Y., and Courville, A. (2017). *Machine learning basics*. Ithaca: MIT Press, 98–165.

Kronberg, E. A., Ashour-Abdalla, M., Dandouras, I., Delcourt, D. C., Grigorenko, E. E., Kistler, L. M., et al. (2014). Circulation of heavy ions and their dynamical Effects in the magnetosphere: Recent observations and models. *Space Sci. Rev.* 184, 173–235. doi:10.1007/s11214-014-0104-0

Kronberg, E. A., Bučík, R., Haaland, S., Klecker, B., Keika, K., Desai, M. I., et al. (2011). On the origin of the energetic ion events measured upstream of the Earth's bow shock by STEREO, Cluster, and Geotail. *J. Geophys. Res.* 116, A02210. doi:10.1029/2010JA015561

Kronberg, E. A., Daly, P. W., Dandouras, I., Haaland, S., and Georgescu, E. (2010). Generation and validation of ion energy spectra based on cluster RAPID and CIS measurements. *Clust. Act. Archive, Stud. Earth's Space Plasma Environ.*, 301–306. doi:10.1007/978-90-481-3499-1_20

Kronberg, E. A., Daly, P. W., and Vilenius, E. (2022). *Calibration report of the RAPID measurements in the Cluster Science Archive (CSA)*. Paris: Tech. Rep. CAA–EST–CR–RAP.

Kronberg, E. A., Grigorenko, E. E., Malykhin, A., Kozak, L., Petrenko, B., Vogt, M. F., et al. (2019). Acceleration of ions in jovian plasmoids: Does turbulence play a role? *J. Geophys. Res. (Space Phys.* 124, 5056–5069. doi:10.1029/2019JA026553

Kronberg, E. A., Grigorenko, E. E., Turner, D. L., Daly, P. W., Khotyaintsev, Y., and Kozak, L. (2017a). Comparing and contrasting dispersionless injections at geosynchronous orbit during a substorm event. *J. Geophys. Res.* 122, 3055–3072. doi:10.1002/2016JA023551

Kronberg, E. A., Li, K., Grigorenko, E. E., Maggiolo, R., Haaland, S., Daly, P. W., et al. (2017c). Dawn-dusk asymmetries in the near-earth plasma sheet: Ion observations. *Dawn-Dusk Asymmetries Planet. Plasma Environments,Geophysical Monogr. Ser.* 230, 243–253. doi:10.1002/9781119216346.ch19

Kronberg, E. A., Rashev, M. V., Daly, P. W., Shprits, Y. Y., Turner, D. L., Drozdov, A., 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, 449–462. doi:10.1002/2016SW001369

Paschmann, G., and Daly, P. W. (1998). *Analysis methods for multi-spacecraft data. ISSI scientific reports series SR-001*. ESA/ISSI, Vol. 1. ISBN 1608-280X, 1998. *ISSI Scientific Reports Series* 1.

Paschmann, G., and Daly, P. W. (2008). *Multi-spacecraft analysis methods revisited*

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Reiff, P. H. (1990). The use and misuse of statistics in space physics. *J. Geomagnetism Geoelectr.* 42, 1145–1174. doi:10.5636/jgg.42.1145

Rème, H., Aoustin, C., Bosqued, J. M., Dandouras, I., Lavraud, B., Sauvaud, J. A., et al. (2001). First multispacecraft ion measurements in and near the Earth's magnetosphere with the identical Cluster ion spectrometry (CIS) experiment. *Ann. Geophys.* 19, 1303–1354. doi:10.5194/angeo-19-1303-2001

Sivadas, N., and Sibeck, D. G. (2022). Regression bias in using solar wind measurements. *Front. Astronomy Space Sci.* 9, 924976. doi:10.3389/fspas.2022.924976

Smirnov, A. G., Kronberg, E. A., Daly, P. W., Aseev, N. A., Shprits, Y. Y., and Kellerman, A. C. (2020a). Adiabatic invariants calculations for cluster mission: A long-term product for radiation belts studies. *J. Geophys. Res. (Space Phys.* 125, e27576. doi:10.1029/2019JA027576

Smirnov, A. G., Kronberg, E. A., Latallerie, F., Daly, P. W., Aseev, N., Shprits, Y. Y., et al. (2019). Electron intensity measurements by the cluster/RAPID/IES instrument in Earth's radiation belts and ring current. *Space weather.* 17, 553–566. doi:10.1029/2018SW001989

Wasserstein, R. L., Schirm, A. L., and Lazar, N. A. (2019). Moving to a world beyond "p0.05. *Am. Statistician* 73, 1–19. doi:10.1080/00031305.2019.1583913 <

Wilken, B., Axford, W. I., Daglis, I., Daly, P., Guttler, W., Ip, W. H., et al. (1997). Rapid - the imaging energetic particle spectrometer on Cluster. *Space Sci. Rev.* 79, 399–473. doi:10.1023/A:100499420229610.1007/978-94-011-5666-0_14

Williams, D. J., McEntire, R. W., Jaskulek, S., and Wilken, B. (1992). The Galileo energetic particles detector. *Space Sci. Rev.* 60, 385–412. doi:10.1007/978-94-011-2512-3_16

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 30 September 2022
DOI 10.3389/fspas.2022.1000766

OPEN ACCESS

EDITED BY
Yoshizumi Miyoshi,
Nagoya University, Japan

REVIEWED BY
Nickolay Ivchenko,
Royal Institute of Technology, Sweden
Hitoshi Fujiwara,
Seikei University, Japan

*CORRESPONDENCE
Franz-Josef Lübken,
luebken@iap-kborn.de

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 22 July 2022
ACCEPTED 24 August 2022
PUBLISHED 30 September 2022

CITATION
Lübken F-J (2022), Physics in the mesosphere/lower thermosphere: A personal perspective.
Front. Astron. Space Sci. 9:1000766.
doi: 10.3389/fspas.2022.1000766

Physics in the mesosphere/lower thermosphere: A personal perspective

Franz-Josef Lübken*

Leibniz Institute of Atmospheric Physics, Kühlungsborn, Germany

The scope of this paper is to present some progress being made in the last few decades regarding some aspects of physical processes in the mesosphere/lower thermosphere and to point to some open questions. This summary is presented from a personal perspective, i.e., this is not a review of a certain science topic. Most citations reflect my own work or are representative examples only. They are not meant to be complete or comprehensive.

KEYWORDS

mesosphere, lower thermosphere, summer mesopause, noctilucent clouds, turbulence, gravity waves, mesopause jumps, climate change

1 Introduction

The (upper) mesosphere and lower thermosphere (MLT) is a scientifically exciting region since it comprises the transition between the Earth's atmosphere and space. More specifically, several physical processes change their general characteristics in this region. The mean free path of molecules (and atoms) increases with altitude and approaches macroscopic dimensions in the MLT. This implies that objects entering from space 'feel' the Earth's atmosphere here for the first time. This is relevant for the reentry of satellites (including space shuttle etc.) but also for meteors. This also means that the Knudsen number ($Kn = \lambda/\ell$) is on the order of unity, i.e., the description of flows changes from fluid dynamics to molecular kinetics (λ = mean free path of molecules; ℓ = typical macroscopic dimension, e.g., ℓ = 0.1 m). For insitu-based measurements, e.g., sensors on sounding rockets, this imposes a significant challenge when correcting for ram effects.

We note that parameters describing molecular diffusion processes increase exponentially with altitude because they are roughly proportional to 1/n (n = atmospheric number density). In the MLT, more precisely at the turbopause, the mixing effect by molecular diffusion (D_m) reaches turbulent mixing described by the eddy diffusion coefficient (K): $D_m \approx K$. In the MLT a major part of solar EUV and UV (and particles) gets absorbed which contributes to the fact that the ionosphere is located above approximately 60–70 km. Furthermore, geomagnetically induced variations can be significant.

Since atmospheric density is decreasing with height, the time between collisions of molecules decreases. In the MLT this time gets close to the lifetime of excited states due to spontaneous emissions for some relevant molecular excitations, e.g., in carbon dioxide. This implies that radiation and atmospheric matter are no longer in thermodynamic equilibrium

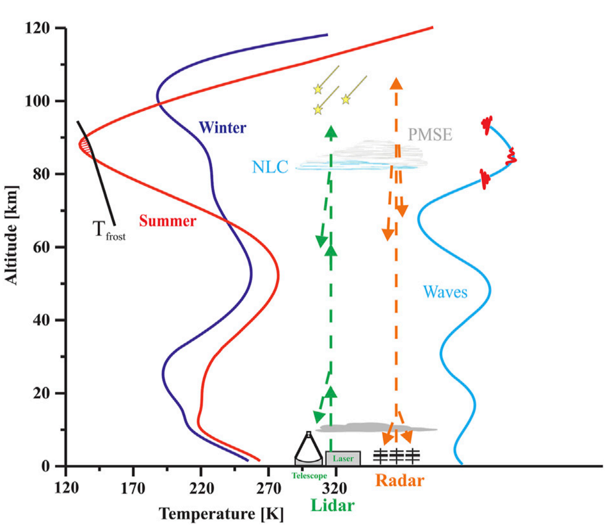

FIGURE 1
Temperature profiles at high latitudes (70°N) for summer (red) and winter (dark blue), respectively. In summer, frost temperatures (black) in the mesopause region are larger than atmospheric temperatures. This allows for the existence of ice particles which are known as noctilucent clouds (NLC), and which are crucial for radar echoes known as polar mesosphere summer echoes (PMSE). The thermal structure in the MLT is mainly caused by a residual circulation driven by the interaction of the mean flow with gravity waves and tides.

for those transitions (non-LTE: non local thermodynamic equilibrium). Therefore, the corresponding radiation cannot readily be taken to derive temperatures which imposes a significant challenge for some satellite retrievals of temperatures.

2 The thermal structure of the MLT

The thermal structure of the MLT is of specific importance since it reflects the energy balance in this atmospheric region and is therefore a good indicator for the physical processes involved. This is particularly relevant for the summer mesopause region at middle and polar latitudes which is the coldest place in the terrestrial atmosphere (T ~130 K), much colder than in winter (see Figure 1). On the other hand, it is rather difficult to measure temperatures in the MLT region because of aerodynamical and non-LTE effects mentioned above. Furthermore, the heat capacity of the atmosphere is too low to alow for direct classical temperature measurements by thermistors etc. One of very few methods to directly derive high altitude resolution temperature profiles quasi-permanently (depending on weather conditions) is to determine the Doppler broadened line-width of atomic transitions by resonance lidars (Fricke and von Zahn, 1985; von Zahn and Höffner, 1996). Lidars typically measure at one location only. However, recent developments promise to cover a substantial horizontal region [VAHCOLI = Vertical And Horizontal COverage by LIdar (Lübken and Höffner, 2021)].

In the late 1990s a climatology of temperatures was deduced from density measurements by falling spheres (FS) which is frequently being used as a ‘standard’ for comparison with other observations and with models (Lübken, 1999). The idea to measure densities and deduce temperature profiles assuming hydrostatic equilibrium is also used by other instruments like the rocket-borne CONE ionization gauge or by ground-based Rayleigh lidars (Strelnikov et al., 2013).

Although temperatures in the summer MLT are difficult to measure, they can easily be checked for consistency due to the presence of ice particles known as NLC (noctilucent clouds) or related phenomena (PMSE, polar mesosphere summer echoes, see below): mean temperatures must be below the frost point temperature of water vapor, T_{frost}, which is on the order of 150–145 K at 82–88 km, respectively [see Figure 1 and reference (Lübken et al., 1996), the abundance of water vapor plays a minor role for T_{frost}; 82 and 88 km are typical heights for the lower and upper edge of NLC/PMSE, respectively]. Typical temperatures from the FS climatology mentioned above are 150 and 135 K at 82 and 88 km, respectively, which are consistent with the presence of ice particles. Note that PMSE are present nearly permanently at polar latitudes, i.e., temperatures must be smaller than T_{frost} nearly permanently and not just sporadically.

Energy budget considerations employing observational campaigns and modelling should be contemplated again, taking into account also non-local processes. A significant contribution to the energy budget comes from observations of turbulence on sounding rockets demonstrating that heating of the atmosphere due to the dissipation of turbulent motions caused by wave breaking can be up to 10–20 K/day which is on the same order of magnitude as other energy budget contributions (Lübken, 1997). Remarkably, this heating is largest around the summer mesopause at polar latitudes, i.e., at the coldest place in the Earth’s atmosphere. Heating in the MLT by turbulent dissipation of parameterized or resolved gravity waves is occasionally taken into account in global models, although often incomplete, inconsistent, or tuned by fudge factors [see discussion in (Becker, 2003)].

A long-standing and still unresolved problem in the MLT relates to non-LTE and its relevance for the energy budget in that region. The coefficient for the deactiviation (and activation) of excited carbon dioxide molecules by atomic oxygen is on the order of $k(CO_2\text{-}O) \sim 1.5\text{–}6 \times 10^{-12}$ cm^3/sec. Unfortunately, values for $k(CO_2\text{-}O)$ as used in satellite retrieval algorithms are different from values obtained from laboratory measurements (Feofilov and Kutepov, 2012). This coefficient has a major influence on radiative heating/cooling and therefore on background temperatures. Non-LTE is therefore of high relevance for MLT science and deserves to be studied further in the future.

In the summer mesopause region dynamical processes drive the thermal state of the atmosphere away from an otherwise

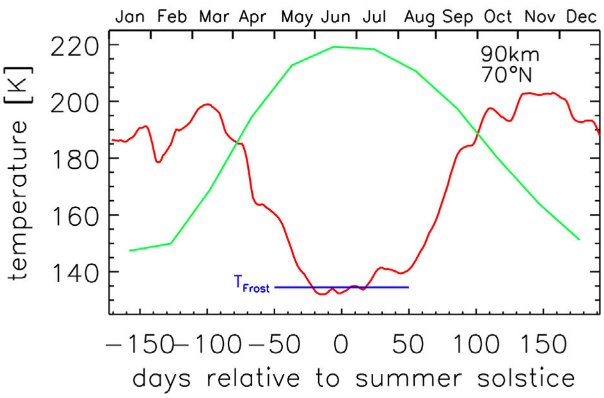

FIGURE 2
Temperatures as function of season at the summer mesopause (90 km), as measured by an Fe-lidar at ALOMAR (red) and from a model (green) where the impact of waves is intentionally eliminated. The blue line indicates the frost point temperature for water ice. The data are provided by Josef Höffner (lidar) and Erich Becker (model).

balanced state by up to 100 K (Figure 2). This is the largest impact of dynamics on the thermal state in the Earth's atmosphere. The forcing comes from wave-induced residual circulation leading to up- and downwelling and associated cooling and heating, respectively. This situation should therefore be used as a benchmark test for models describing this dynamical control, and, more general, of studying the energy balance of the atmosphere.

Winds are also crucial when evaluating the physics of the MLT, including the energy and momentum balance. Measuring background winds in the MLT is realized by radars which since recently cover a substantial horizontal region (Chau et al., 2019). The observational gap of measuring winds on a continuous basis between the (middle) stratosphere and mesosphere is nowadays filled by sophisticated lidars measuring line-of-sight winds (Baumgarten, 2010; Lübken et al., 2016).

3 Layers in the MLT: NLC, PMSE, PMWE

Ice layers in the MLT can be observed by naked eye, by lidars, or from satellites, all of which are here referred to as 'noctilucent clouds', NLC (Russell et al., 2009; Fiedler et al., 2017). Polar mesosphere summer echoes (PMSE) are large radar echoes from the summer MLT which require the presence of (charged) ice particles, but also neutral air turbulence. The main physics aspects of PMSE are well understood, i.e., they can now be used as a diagnostic tool to study the atmospheric background, in particular temperatures and turbulence (Rapp and Lübken, 2003).

When using PMSE as an indicator for neutral air turbulence, fossile turbulence can play a role: Fluctuations in the plasma may still exist although neutral air turbulence has already decayed (Rapp and Lübken, 2003). This effect increases with increasing Schmidt number, Sc = D/ν (D = diffusion coefficient for plasma irregularities, ν = kinematic viscosity), i.e., with increasing mass of the (charged) ice (or dust) particles, m_{ice}. If m_{ice} is small, Sc ~1 and fossile turbulence plays no role. This is the case in the upper altitude range of PMSE but also for 'polar mesosphere winter echoes', PMWE. The latter are weaker compared to PMSE and are observed in the lower and middle mesosphere mainly outside the summer seasons (Lübken et al., 2006; Latteck et al., 2021). They are mainly caused by neutral air turbulence. Although (charged) dust can support the creation of PMWE, their presence is of secondary importance. Note that PMSE and PMWE also occur at middle latitudes like Kühlungsborn (54°N).

PMSE and PMWE should be used to deduce the morphology of neutral air turbulence (occurrence rate, intermittency, etc.) which is important to judge the significance of snapshot observations performed by sounding rockets. Of course, other prerequisites, like ice particles (for PMSE) and sufficient electron number densities (for PMSE and PMWE) must also be fulfilled.

NLC and PMSE also allow to study the microphysics of ice particle formation, growth, and sublimation in a unique environment, including the frostpoint temperature of water vapor over ice at very low temperatures, or the charging of ice and dust particles in a weakly ionized plasma, or the impact of gravity waves and instabilities on the formation of ice (Fritts et al., 2014). NLC are also considered as an indicator for climate change (see Section 5).

4 Waves, turbulence, and transport

4.1 Gravity waves and tides

Gravity wave (GW) and tidal forcing of the background atmosphere is absolutely essential to understand the thermal structure of the MLT. Unfortunately, the physics of gravity wave forcing of the atmosphere is rather complex regarding sources, propagation, generation of higher harmonic waves, dissipation of momentum and energy, all of which in a time-varying background. Mesopause jumps in the southern hemisphere, first observed in 2010 by resonance lidars, are a nice example for a chain of processes of GW forcing of the MLT involving nearly the entire height range from the troposphere to the lower thermosphere (Lübken et al., 2017). They rely on an early breakdown of the winter polar vortex which only takes place in the southern hemisphere. Mesopause jumps should be used as a benchmark test for GW models since they exhibit a distinct and unique signature. Note that in 1998 we performed falling sphere observations of temperatures from Rothera (68°S), the first scientific rocket launchings in Antarctica (Lübken et al., 1999a). The polar stratospheric vortex breakdown in that season occurred rather late, i.e., temperatures in the southern hemisphere MLT were very similar to the northern hemisphere in that season.

Lidars offer the only method to measure GW and tides quasi-continuously (clear sky conditions provided) from the troposphere to the lower thermosphere, both in temperatures (potential energy) and in winds (kinetic energy). Recent developments in lidar technology promise to extend this capability from local to regional scales (Lübken and Höffner, 2021). Note that the morphology of gravity waves can only be treated correctly if the background wind field is known, which is nowadays the case for some lidars (Baumgarten, 2010; Lübken and Höffner, 2021).

Recent lidar observations of GW temperatures and winds have shown that a substantial part of GW propagate downward which confirmed earlier model predictions (Becker and Vadas, 2018; Strelnikova et al., 2020). This mechanism is not taken into account when parameterizing GW in models. Other deficiencies concern non-oblique propagation, unjustifiably assuming an instantaneous impact, ignoring refraction in a time-varying background etc [see, e.g., (Senf and Achatz, 2011)]. Continuous measurements of GW in the stratosphere and lower mesosphere over a period of 10 days showed some unexpected variations on timescales of days to weeks which are presumably caused by the interaction of GW and tides (Baumgarten et al., 2018). In the upper mesosphere such variations of tides are frequently observed by radar (Kero et al., 2019).

Thousends of hours of lidar observations are now available at ALOMAR (69°N) and Kühlungsborn (54°N) which were used to derive a climatology of gravity waves (Strelnikova et al., 2021). The seasonal variation of potential energy (E_{pot}) is more than an order of magnitude smaller compared to a gravity wave resolving model [KMCM, Kühlungsborn Meachnistic Circulation Model (Becker et al., 2022)]. Currently, there is no straight forward explanation of this discrepancy since 1) GW in KMCM are self-generated and cannot easily be 'tuned', and b) any change of GW forcing in KMCM will presumably change the background atmosphere, in particular temperatures at the summer mesopause. This discrepancy between observations and modelling provides a promising field for further studies.

A crucial task for the future will be to describe the complex morphology of GW (and tides) in models in a background which varies in time and space and to correctly model the interaction between waves and the mean flow. It is obvious that the creation/propagation of GW in the troposphere and stratosphere are also very important in this context. Up-to-date observations by ground based instruments (lidars, radars, etc.) and by sounding rockets and satellites should be used to constrain these concepts and to provide hints for an adequate description.

4.2 Turbulence

Our early approach to study turbulence around the turbopause at ~100 km concentrated on the effect of mixing/demixing by turbulence and molecular diffusion, respectively. We evaluated the ratio of Argon to Nitrogen number densities, n_{Ar}/n_{N2}, measured by the BUGATTI mass spectrometer (von Zahn et al., 1990). It turned out that other processes may substantially contribute to a constant mixing ratio (i.e., vertical winds), and that it may take a long time (up to weeks) for demixing. This implies that the mixing ratio n_{Ar}/n_{N2} may not result in an adequat description of turbulence. To study the abundance and vertical transport of water vapor we later developed a tunable diode laser spectrometer on sounding rockets called MASERATI, which was, however, rather complicated and too costly for frequent applications (Lübken et al., 1999b).

It should be noted that the transport of tracers from the thermosphere to the stratosphere is not yet fully understood (Smith et al., 2011), but is an important coupling mechanism between the MLT and lower altitudes. This topic requires more attention in the future.

Realizing the limitations outlined above we decided to use relative neutral air density fluctuations $\Delta n/n$ as a passive tracer for turbulent motions. The magnitude of this effect is on the order of $\Delta n/n \sim (\omega_B \cdot v_{turb})/g$ ~0.1–1% (depending on ω_B and v_{turb}; ω_B = Brunt-Väisälä frequency; v_{turb} = turbulent velocity ~1 m/s,; g = Earth's acceleration). Indeed, we observed spectra of fluctuations consistent with inertial subrange turbulence (ISR). However, deducing turbulent energy dissipation rates, ϵ, is hampered by several uncertain parameters, for example, the dissipation rate of fluctuations (Thrane et al., 1994). A more precise method relies on obtaining the transition between the ISR and the viscous subrange (VSR) in the spectra, which only depends on ϵ, and on the kinematic viscosity, ν, which can easily be derived from atmospheric background parameters. This method imposes a challenge for the experimental method since spectra have to be measured with high temporal resolution (few milli-seconds on a sounding rocket) and high precision (0.1–1%). For this purpose we developed the CONE instrument which was successfully applied in many flights (Lübken, 1992; Lübken et al., 1993). The idea to derive ϵ from the 'break' in the spectrum is nowadays frequently applied on sounding rockets and on balloons, and has also been compared to direct numerical simulations (Lehmacher et al., 2006; Theuerkauf et al., 2011; Triplett et al., 2018; Strelnikov et al., 2022).

A climatology for ϵ at 69°N (Andoya Rocket Range) was derived from CONE observations showing large turbulent heating rates of up to ~20 K/d at the summer mesopause, the coldest place in the Earth's atmosphere (Lübken, 1997). Obviously, other processes have to compensate for this heating.

Turbulence in the MLT can also be derived from the spectral broadening of radar backscattering which, however, is associated with some uncertainties, e.g., regarding non-turbulent spectral broadening processes (Hocking, 1996; Lübken, 2014).

5 Trends and solar cycle

Negative temperature trends in the strato-, meso-, and thermosphere are meanwhile established in models and are partly confirmed by observations (Roble, 1995; Berger and Lübken, 2011). There is one exception, namely the summer mesopause region at middle and polar latitudes, where temperature trends are positive which is partly due to an enhanced radiative coupling of the cold mesopause to the 'warm' stratopause. Unfortunately, this implies that trends are very small in the NLC/PMSE region. On the other hand, ice particles are very sensitive to temperatures and may therefore still be used to study long term trends of temperatures and/or water vapor (Lübken et al., 2021).

Trends in mesospheric ice layers are controversially disputed in the literature (Thomas, 1996; von Zahn, 2003). Model studies by LIMA/MIMAS show that cooling in the stratosphere/lower mesosphere due to an increase of CO_2 reduces the (geometric) heights of NLC because of atmospheric shrinking. The decrease of NLC pressure altitudes is much smaller (Berger and Lübken, 2011; Lübken et al., 2021). On the other hand an increase of H_2O (due to an increase of CH_4 which oxidizes nearly completely to H_2O in the mesosphere) leads to an enhancement of NLC brightness and other related parameters, e.g., ice water content (IWC). Note that trends in dynamics have been ignored in these studies. The results from LIMA/MIMAS are consistent with rather limited observations, for example with phase height measurements and with the NLC record at ALOMAR (Bremer and Berger, 2002; Fiedler et al., 2017). Regarding an unambiguous confirmation of trends in NLC or PMSE, the available observations are still not long and comprehensive enough, and they compete with natural variability. A recent extension of LIMA/MIMAS applying future scenarios of greenhouse gas emissions shows that NLC are likely to absorb a significant part of solar radiation, in particular in the UV. This will presumably affect photochemical processes in the stratosphere and MLT.

The upper atmosphere also affects climate on Earth's surface. Empirical studies indicate a significant modulation of regional temperatures by solar cycle, but the mechanism of the coupling from the MLT (where radiation with largest solar cycle variation is absorbed) to the surface is not yet understood (Seppälä et al., 2009; Gray et al., 2010). Presumably, a complicated chain of processes is involved which includes the transport of photochemically active species, the influence of waves (GW, tides, and planetary waves), and the modification of the background circulation. There is no reasonable doubt that anthropogenic activity is the main reason for global climate change in the troposphere. Still, a climate impact from higher altitudes (and from the Sun) should be considered, at least on regional scales.

Little is known regarding trends in dynamics, including GW (Shepherd, 2014). To repeat: 1) the polar summer mesopause region is the strongest manifestation of GW forcing, and 2) mesopheric ice layers (NLC, PMSE) show rather small long term changes. This imposes a major constraint on any model results or speculations about GW trends. Long term observations of radar winds have shown trends in GW (Hoffmann et al., 2011). It is not clear yet, why this trend has not resulted in an accompanying change of MLT temperatures and NLC/PMSE parameters. This topic requires further attention.

6 Concluding remarks and outlook

The mesosphere/lower thermosphere is a unique and stimulating region for applying various and partly competing physics concepts from rather different fields. More specifically, this region provides an impressive example for an impact of dynamics on the background thermal structure. Some of the involved processes are also relevant for regional climate change.

From my experience I recommend publishing (hopefully) unique observations even without interpretation by modelling, since they can trigger new physical insights. If observations do not agree with models, this should be taken as a motivation to investigate the physical reasons (provided that mistakes in the observations can virtually be excluded) and in the ideal case to modify the models. On the other hand, it is crucial to understand the underlying physics and to consider local/regional processes in the view of global impacts. This can only be acquired by appropriate simulations and modelling. A more recent example of such interaction of observations and modelling comes from our improvements in understanding PMSE/PMWE which were subsequently applied to study phenomena such as interhemispheric coupling and mesopause jumps. I think that science in the MLT can easily stimulate and motivate young scientists, also because they can achieve visibility in the science community more quickly than in many other fields of physics. I'm grateful that such a unique opportunity was provided to me.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at doi: 10.22000/705.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This work was partly supported by the Leibniz SAW project FORMOSA (K227/2019) and by the TIMA project of the BMBF research initiative ROMIC (FZ 01LG 1902A). More details of funding sources used in the last decades are provided in the references.

Acknowledgments

I'm grateful to numerous friends and colleagues who inspired and triggered this research over many years. I also thank various funding organizations such as DLR, BMBF, and DFG who not only supported individual projects, but also initiated larger programs such as CAWSES (DFG) and ROMIC (BMBF) which involved several institutes.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Baumgarten, G. (2010). Doppler Rayleigh/Mie/Raman lidar for wind and temperature measurements in the middle atmosphere up to 80 km. *Atmos. Meas. Tech.* 3, 1509–1518. doi:10.5194/amt-3-1509-2010

Baumgarten, K., Gerding, M., Baumgarten, G., and Lübken, F. J. (2018). Temporal variability of tidal and gravity waves during a record long 10 day continuous lidar sounding. *Atmos. Chem. Phys.* 18, 371–384. doi:10.5194/acp-18-371-2018

Becker, E., Vadas, S. L., Bossert, K., Harvey, V. L., Zülicke, C., and Hoffmann, L. (2022). A high-resolution whole-atmosphere model with resolved gravity waves and specified large-scale dynamics in the troposphere and stratosphere. *JGR. Atmos.* 127. doi:10.1029/2021JD035018

Becker, E., and Vadas, S. L. (2018). Secondary gravity waves in the winter mesosphere: Results from a high-resolution global circulation model. *J. Geophys. Res. Atmos.* 123, 2605–2627. doi:10.1002/2017JD027460

Becker, E. (2003). Frictional heating in global climate models. *Bull. Amer. Meteorol. Soc.* 84, 318.

Berger, U., and Lübken, F. J. (2011). Mesospheric temperature trends at mid-latitudes in summer. *Geophys. Res. Lett.* 38, L22804. doi:10.1029/2011GL049528

Bremer, J., and Berger, U. (2002). Mesospheric temperature trends derived from ground-based LF phase-height observations at mid-latitudes: Comparison with model simulations. *J. Atmos. Sol. Terr. Phys.* 64, 805–816. doi:10.1016/s1364-6826(02)00073-1

Chau, J. L., Urco, J. M., Vierinen, J. P., Volz, R. A., Clahsen, M., Pfeffer, N., et al. (2019). Novel specular meteor radar systems using coherent MIMO techniques to study the mesosphere and lower thermosphere. *Atmos. Meas. Tech.* 12, 2113–2127. doi:10.5194/amt-12-2113-2019

Feofilov, A. G., and Kutepov, A. A. (2012). Infrared radiation in the mesosphere and lower thermosphere: Energetic effects and remote sensing. *Surv. Geophys.* 33, 1231–1280. doi:10.1007/s10712-012-9204-0

Fiedler, J., Baumgarten, G., Berger, U., and Lübken, F. J. (2017). Long-term variations of noctilucent clouds at ALOMAR. *J. Atmos. Sol. Terr. Phys.* 162, 79–89. doi:10.1016/j.jastp.2016.08.006

Fricke, K. H., and von Zahn, U. (1985). Mesopause temperatures derived from probing the hyperfine structure of the D2 resonance line of sodium by lidar. *J. Atmos. Terr. Phys.* 47, 499–512. doi:10.1016/0021-9169(85)90116-3

Fritts, D. C., Baumgarten, G., Wan, K., Werne, J., and Lund, T. (2014). Quantifying kelvin-helmholtz instability dynamics observed in noctilucent clouds: 2. Modeling and interpretation of observations. *J. Geophys. Res. Atmos.* 119, 9359–9375. doi:10.1002/2014JD021833

Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., et al. (2010). Solar influences on climate. *Rev. Geophys.* 48, RG4001. doi:10.1029/2009RG000282

Hocking, W. (1996). An assessment of the capabilities and limitations of radars in measurements of upper atmosphere turbulence. *Adv. Space Res.* 17 (11), 37–47. doi:10.1016/0273-1177(95)00728-w

Hoffmann, P., Rapp, M., Singer, W., and Keuer, D. (2011). Trends of mesospheric gravity waves at northern middle latitudes during summer. *J. Geophys. Res.* 116, D00P08. doi:10.1029/2011JD015717

Kero, J., Campbell-Brown, M. D., Stober, G., Chau, J. L., Mathews, J. D., and Pellinen-Wannberg, A. (2019). "Radar observations of meteors," in *Meteoroids: Sources of meteors on Earth and beyond*. Editors G. O. Ryabova, D. J. Asher, and M. J. Campbell-Brown (Cambridge: Cambridge University Press), Vol. 336, 65–89.

Latteck, R., Renkwitz, T., and Chau, J. L. (2021). Two decades of long-term observations of polar mesospheric echoes at 69°N. *J. Atmos. Sol. Terr. Phys.* 216, 105576. doi:10.1016/j.jastp.2021.105576

Lehmacher, G. A., Croskey, C. L., Mitchell, J. D., Friedrich, M., Lübken, F. J., Rapp, M., et al. (2006). Intense turbulence observed above a mesospheric temperature inversion at equatorial latitude. *Geophys. Res. Lett.* 33, L08808. doi:10.1029/2005GL024345

Lübken, F. J., Baumgarten, G., and Berger, U. (2021). Long term trends of mesopheric ice layers: A model study. *J. Atmos. Sol. Terr. Phys.* 214, 105378. doi:10.1016/j.jastp.2020.105378

Lübken, F. J., Baumgarten, G., Hildebrand, J., and Schmidlin, F. (2016). Simultaneous and co-located wind measurements in the middle atmosphere by lidar and rocket-borne techniques. *Atmos. Meas. Tech.* 9, 3911–3919. doi:10.5194/amt-9-3911-2016

Lübken, F. J., Dingler, F., von Lucke, H., Anders, J., Riedel, W. J., and WolfMaserati, H. (1999). MASERATI: A rocketborne tunable diode laser absorption spectrometer. *Appl. Opt.* 38, 5338–5349. doi:10.1364/ao.38.005338

Lübken, F. J., Fricke, K. H., and Langer, M. (1996). Noctilucent clouds and the thermal structure near the Arctic mesopause in summer. *J. Geophys. Res.* 101, 9489–9508. doi:10.1029/96jd00444

Lübken, F. J., Hillert, W., Lehmacher, G., and von Zahn, U. (1993). Experiments revealing small impact of turbulence on the energy budget of the mesosphere and lower thermosphere. *J. Geophys. Res.* 98, 20, 369–20, 384. doi:10.1029/93jd02055

Lübken, F. J., and Höffner, J. (2021). VAHCOLI, a new concept for lidars: Technical setup, science applications, and first measurements. *Atmos. Meas. Tech.* 14, 3815–3836. doi:10.5194/amt-14-3815-2021

Lübken, F. J., Jarvis, M. J., and Jones, G. O. L. (1999). First *in situ* temperature measurements at the Antarctic summer mesopause. *Geophys. Res. Lett.* 26, 3581–3584. doi:10.1029/1999gl010719

Lübken, F. J., Latteck, R., Becker, E., Höffner, J., and Murphy, D. (2017). Using polar mesosphere summer echoes and stratospheric/mesospheric winds to explain summer mesopause jumps in Antarctica. *J. Atmos. Sol. Terr. Phys.* 162, 106–115. doi:10.1016/j.jastp.2016.06.008

Lübken, F. J. (1992). On the extraction of turbulent parameters from atmospheric density fluctuations. *J. Geophys. Res.* 97, 20, 385–20, 395. doi:10.1029/92jd01916

Lübken, F. J. (1997). Seasonal variation of turbulent energy dissipation rates at high latitudes as determined by insitu measurements of neutral density fluctuations. *J. Geophys. Res.* 102, 13, 441–13, 456. doi:10.1029/97jd00853

Lübken, F. J., Strelnikov, B., Rapp, M., Singer, W., Latteck, R., Brattli, A., et al. (2006). The thermal and dynamical state of the atmosphere during polar mesosphere winter echoes. *Atmos. Chem. Phys.* 6, 13–24. doi:10.5194/acp-6-13-2006

Lübken, F. J. (1999). Thermal structure of the Arctic summer mesosphere. *J. Geophys. Res.* 104, 9135–9149.

Lübken, F. J. (2014). Turbulent scattering for radars: A summary. *J. Atmos. Sol. Terr. Phys.* 107, 1–7. doi:10.1016/j.jastp.2013.10.015

Rapp, M., and Lübken, F. J. (2003). On the nature of PMSE: Electron diffusion in the vicinity of charged particles revisited. *J. Geophys. Res.* 108, 8437. doi:10.1029/2002JD002857

Roble, R. G. (1995). Major grenhouse cooling (yes, cooing): The upper atmosphere response to increased CO_2. *Rev. Geophys.* 33, 539–546. doi:10.1029/95rg00118

Russell, J. M., III, Bailey, S. M., Gordley, L. L., Rusch, D. W., Horányi, M., Hervig, M. E., et al. (2009). The aeronomy of ice in the mesosphere(AIM)mission: Overview and early science results. *J. Atmos. Sol. Terr. Phys.* 71, 289–299. doi:10.1016/j.jastp.2008.08.011

Senf, F., and Achatz, U. (2011). On the impact of middle-atmosphere thermal tides on the propagation and dissipation of gravity waves. *J. Geophys. Res.* 116, D24110. doi:10.1029/2011JD015794

Seppälä, A., Randall, C. E., Clilverd, M. A., Rozanov, E., and Rodger, C. J. (2009). Geomagnetic activity and polar surface air temperature variability. *J. Geophys. Res.* 114. doi:10.1029/2008JA014029

Shepherd, T. G. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* 7, 703–708. doi:10.1038/ngeo2253

Smith, A. K., Garcia, R. R., Marsh, D. R., and Richter, J. H. (2011). WACCM simulations of the mean circulation and trace species transport in the winter mesosphere. *J. Geophys. Res.* 116, D20115. doi:10.1029/2011JD016083

Strelnikov, B., Rapp, M., Fritts, D. C., and Wang, L. (2022). Assessment of the precision of spectral model turbulence analysis techniques using direct numerical simulation data. *J. Geophys. Res.* 127, e2021JD035516. doi:10.1002/2021JD035516

Strelnikov, B., Rapp, M., and Lübken, F. J. (2013). "*In-situ* density measurements in the mesosphere/lower thermosphere region with the TOTAL and CONE instruments," in *An introduction to space instrumentation*. Editor K. Oyama (Moscow, Russia: Terra Publishers), 1–11.

Strelnikova, I., Almowafy, M., Baumgarten, G., Baumgarten, K., Ern, M., Gerding, M., et al. (2021). Seasonal cycle of gravity wave potential energy densities from lidar and satellite observations at 54° and 69°N. *J. Atmos. Sci.* 78, 1359–1386. doi:10.1175/JAS-D-20-0247.1

Strelnikova, I., Baumgarten, G., and Lübken, F. J. (2020). Advanced hodograph-based analysis technique to derive gravity-wave parameters from lidar observations. *Atmos. Meas. Tech.* 13, 479–499. doi:10.5194/amt-13-479-2020

Theuerkauf, A., Gerding, M., and Lübken, F. J. (2011). LITOS – A new balloon-borne instrument for fine-scale turbulence soundings in the stratosphere. *Atmos. Meas. Tech.* 4, 55–66. doi:10.5194/amt-4-55-2011

Thomas, G. (1996). Is the polar mesosphere the miner's canary of global change? *Adv. Space Res.* 18 (3), 149–158. doi:10.1016/0273-1177(95)00855-9

Thrane, E., Blix, T., Hoppe, U. P., Lübken, F. J., Hillert, W., Lehmacher, G., et al. (1994). A study of small-scale waves and turbulence in the mesosphere using simultaneous *in situ* observations of neutral and plasma fluctuations. *J. Atmos. Terr. Phys.* 56, 1797–1808. doi:10.1016/0021-9169(94)90011-6

Triplett, C. C., Li, J., Collins, R. L., Lehmacher, G. A., Barjatya, A., Fritts, D. C., et al. (2018). Observations of reduced turbulence and wave activity in the arctic middle atmosphere following the January 2015 sudden stratospheric warming. *J. Geophys. Res. Atmos.* 123, 13259–13276. doi:10.1029/2018JD028788

von Zahn, U., and Höffner, J. (1996). Mesopause temperature profiling by potassium lidar. *Geophys. Res. Lett.* 23, 141–144. doi:10.1029/95gl03688

von Zahn, U., Lübken, F. J., and Pütz, C. (1990). 'BUGATTI' experiments: Mass spectrometric studies of lower thermosphere eddy mixing and turbulence. *J. Geophys. Res.* 95, 7443–7465. doi:10.1029/jd095id06p07443

von Zahn, U. (2003). Are noctilucent clouds truly a "miner's canary" for global change? *Eos Trans. AGU.* 84 (28), 261–264. doi:10.1029/2003eo280001

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Mini Review
PUBLISHED 14 October 2022
DOI 10.3389/fspas.2022.1025466

Check for updates

OPEN ACCESS

EDITED BY
Tuija I Pulkkinen,
University of Michigan, United States

REVIEWED BY
Hannu Koskinen,
University of Helsinki, Finland

*CORRESPONDENCE
Juan G. Roederer,
juro2115@colorado.edu

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 22 August 2022
ACCEPTED 13 September 2022
PUBLISHED 14 October 2022

CITATION
Roederer JG (2022), First shots of the climate revolution: An untold story.
Front. Astron. Space Sci. 9:1025466.
doi: 10.3389/fspas.2022.1025466

First shots of the climate revolution: An untold story

Juan G. Roederer*

Laboratory of Atmospheric and Space Physics, University of Colorado at Boulder, Boulder, CO, United States

This is an account of the history of the "climate revolution of the eighties", focusing on early discussions and, at times, fierce disputes about what's wrong with Planet Earth and why, and what to do about it through coordinated research.

The paper describes the genesis and initial planning of the Global Change program within the International Council of Scientific Unions, and the early split of participating scientists into two camps: one emphasizing the need for a truly global, widely interdisciplinary and basic-science oriented approach that views the Earth system as one single whole of strongly interacting parts; the other camp defending a much more restricted approach by focusing exclusively on that which has greatest and most immediate impact on society.

As a spin-off from the defeat of the "globalists" came the generation of yet another, ultimately international (and politically far less contested) program, centered on what today is called Space Weather—the study of solar-variability- and human-induced changes in the space environment of Earth, and the ensuing effects on technological and human systems in space, as well as the possible physical downward actions of these space perturbations on our more immediate environment of air, water, land and biota.

KEYWORDS
climate change, history, space weather, earth system, initial planning

Part I: Weather and climate on earth

After a 20-years lull, even a slight decrease, before 1975, the global temperature started rising again with gusto in correlation with the unrelenting increase of atmospheric CO_2 concentration. In the early eighties, concerned scientists started asking the question: If this global warming keeps going on at this rate, what will happen with the world food supply in the longer-term future?

NASA was finding itself under increasing pressure from politicians to do something "of more direct relevance" to voters than landing astronauts on the Moon or building a futuristic space station. At the same time, science in general was feeling the pressure from an increasingly powerful worldwide environmental movement to do something about this global warming threat. So, NASA decided to develop a program that it called Global Habitability (Tilford, 1984). Unfortunately, it was poorly designed from the scientific

FIGURE 1
Cover of the first edition of the book (Malone and Roederer, 1985) with all contributions to the 1984 Ottawa Symposium on the IGBP. The picture of the eruption of Alaska's St. Augustine volcano was chosen on purpose to emphasize the original global interdisciplinary character envisaged for the original proposal.

point of view and did not recognize the need for international scientific cooperation, essential for the success of any such an endeavor.

Scientists of the United States Academy of Sciences/National Research Council (NAS) became quite alarmed and, under the leadership of Herbert Friedman, co-chair of its Commission on Physical Sciences, established a study group which met in summer 1983 in Woods Hole, MA, to craft the proposal for an international cooperative program to be carried out under the leadership of the International Council of Scientific Unions (ICSU) in similar fashion as the tremendously successful International Geophysical Year (IGY; Korsmo, 2007) 30 years earlier.

The President of NAS, Frank Press, took personal interest in this project, and participated in the Woods Hole study. The final product was the proposal for an International Geosphere-Biosphere Program (IGBP) (Friedman et al., 1983). The Foreign Secretary of the NAS (Thomas Malone) and the Foreign Secretary of the American Geophysical Union (me) were designated to act as the initial coordinators of this US initiative. Because of a sudden illness of Tom Malone, I was designated to formally present the Woods Hole study report to the ICSU Council meeting in Warsaw a few weeks after the Woods Hole study.

In this proposal, the goal of the IGBP was defined as:

> *"…to describe and understand the interactive physical, chemical and biological processes that regulate the total Earth System, the unique environment that it provides for life, the changes that are occurring in this system, and the manner in which they are influenced by human activity."*

ICSU approved the proposal and asked Tom Malone and me to lead a 2-year ground-laying study that included the organization of an international, truly interdisciplinary symposium, which took place in Ottawa in September 1984 and was a resounding success. Individual presentations were published speedily in book form (Figure 1; in Malone and Roederer, 1985).

A Planning Group was established by ICSU (Bolin et al., 1986), consisting of five Working Groups: WG1 Terrestrial Ecosystems and Atmospheric Interactions; WG2 Marine Ecosystems and Atmospheric Interactions; WG3 Geological Processes, Past, and Present; WG4 Upper Atmosphere and Near-Space Environment; and WG5 Remote Sensing. These groups worked diligently between the 1984 symposium and the ICSU 1986 General Assembly in Berne, Switzerland. Also, during that time, Malone and I realized that the Peoples Republic of China (PRC), so important for a future IGBP from the environmental point of view, seemed not yet fully integrated in the ICSU family. With the sponsorship of the Ford Foundation we organized a small delegation which visited the PRC in April 1984 to meet with authorities of the Chinese Academy of Science and Technology (CAST).

And during all this time a monumental fight had broken out—between the scientists themselves! There were two mutually warring factions worldwide.

One faction was comprised mainly of broad-minded *geophysicists* with expertise in relevant fundamental physical-chemical-biological processes relevant to planet Earth, who were fully aware of the immense complexities (Waldrop, 1984) and inherent unpredictability of this "terrestrial machine," in which "nothing was proportional to anything." Many were already aware of the new mathematical field of "Catastrophe Theory" (later called Complexity Theory). This group was also aware of how difficult it was to communicate such complexities to the lay public and to convey the lurking long-term danger to impatient, scientifically naïve politicians. At this early stage already, this group called the attention to the possibility of a much faster increase of the frequency of extreme events (not yet calculable in climate models of the time), than the predicted rise of global temperature.

The other faction included mainly traditional synoptic meteorologists, experts in weather and climate forecasting, and in crafting the emerging, still highly simplified and

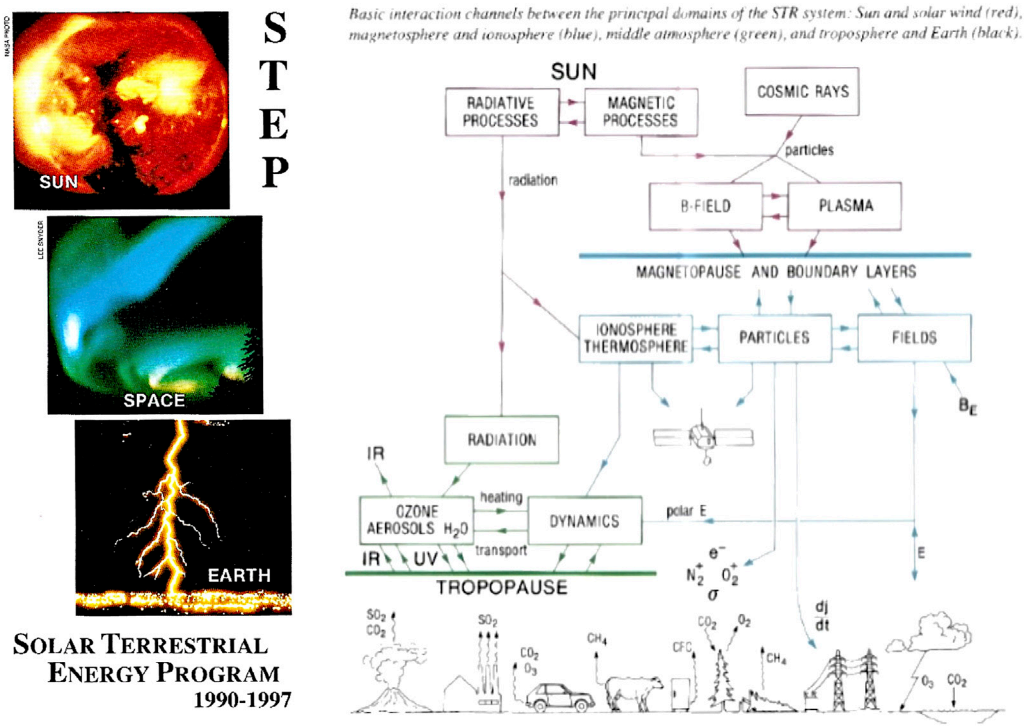

FIGURE 2
Brochure for the ICSU international Solar-Terrestrial Energy Program 1990–1997 (STEP), showing relevant domains and processes in the solar-terrestrial domain, with a sketch of the perturbations determining the "climate and weather in space" and its coupling to the Earth.

coarse-grained computer models. They were joined by environmental biologists, experts in marine, riparian and land ecological systems. All were well versed in relevant economic, hence political impacts of their profession. Environmentalists, a majority of whom were young enthusiasts not specifically trained in science, tended to allay themselves with this second group.

The first group viewed IGBP more as an intellectual challenge of *basic science*; their opponents instead considered it *applied* science focusing on problems of socio-economic impact. The all-encompassing target of the first group was the study of *Global Change* of the *total* earth system including the "new" regions of near-earth space populated by crucial technological systems (and, occasionally, people), whereas the target of the second group was *Global Warming* of the troposphere and its effects on the biosphere, including the anthroposphere.

ICSU had placed the IGBP planning activities under the general direction of the late Swedish meteorologist Bert Bolin who, with manners I have never witnessed in 27 years of participation in international scientific committees, tried to vociferously silence anyone who disagreed with his personal views of the IGBP. In Berne the ICSU formally launched the IGBP (Roederer, 1986a). The basic tenets of the IGBP shown above were preserved, but then came a severe blow to the original concept. Bolin's faction managed to include in the ground-laying document a conditional clause:

> "*Priority in the IGBP will fall on those areas that deal with key interactions and significant change, that most affect the biosphere, that are most susceptible to human perturbations, and that will most likely lead to practical predictive capability.*"

As a result, the entire subject of Working Group 4, near-space and its solar-caused perturbations, was thrown out of IGBP—including the WG itself. Despite having been the chairman of that group, miraculously I still remained on the IGBP Committee. I found it necessary to defend our Working Group's proposals and approach to the IGBP, and published some related articles (Roederer, 1986b, Roederer, 1987). This warfare also propagated right into the US Academy of Sciences, and Herb Friedman, a very mild-mannered person, bitterly complained and withdrew from direct personal involvement in this program, which now was taken over by several meteorologists, atmospheric scientists and environmental biologists who conducted another Woods Hole study to craft a plan for the US contribution to the IGBP (Eddy et al., 1986).

My good friend and colleague Valeria Troitskaya, representing the Soviet Academy of Sciences, and I were sort of standard-bearers of the "opposition" in the ICSU IGBP Committee until the ICSU meeting in Lisbon in 1989. My parting shots were given on the floor in a debate during the meeting (Roederer, 1989):

> "*What would the reaction of impatient politicians be if in a few years' time scientists came to the conclusion that global predictability of a chaotic, turbulent system like the atmosphere is basically impossible on the decadal timescale envisaged? . . . What if the real perturbations caused by greenhouse gas increases manifest themselves first on a much smaller spatial scale, say, as increases in regional variability and turbulence, which are not treatable in any of the current supercomputer global circulation models?*"

That was said in 1989 but, I submit, is still valid today. The rest is history. In 1988 the IGBP Committee had entered in negotiations with the intergovernmental World Meteorological Organization (WMO), and the Intergovernmental Panel on Climate Change (IPCC) was created, headed by—of course—Bert Bolin. It shared the no doubt well-deserved 2007 Nobel *Peace* Prize with former US Vice President Al Gore.

Part II: Weather and climate in space

Have the fields of geoscience tossed out of the IGBP in 1986 suffered? Not at all. One night even argue that this turned out beneficial for space physics! In the late eighties, under the initiative of Donald J. Williams the *international Solar-Terrestrial Energy Program* (STEP—Figure 2) got underway, organized by the ICSU Scientific Committee on Solar-Terrestrial Physics (SCOSTEP). It was based on the original vision of IGBP Working Group 4, and can be summarized as shown in the diagram of Figure 2. In this brochure one can read, for the first time in official print, the expression "*climate and weather in space*."

At the national level in the United States, two initiatives were launched that also developed a few years later into international cooperative programs. Already before the ICSU Assembly in Berne, the brand-new US Arctic Research Commission (ARC), of which I was the Vice Chairman at that time, adopted the original "Earth Systems" approach as one of the *Arctic Research Priorities* (Roederer, 1986c). It should be noted that this was a logical step: near-earth space indeed has its most significant and active coupling with the upper atmosphere in the polar cap regions. In addition, the first steps were taken in 1987 to establish an international Arctic Committee (now the Arctic Council; https://www.arctic-council.org), and on 1 October 1987, Soviet Union President Mikhail Gorbachev opened the vast Soviet Arctic to international research (Roederer, 1988b) calling the community to an international conference in Leningrad. In that conference, which took place December 12–15, 1988, the IGBP program played a fundamental role, but *with* a prominent place for upper atmospheric and near-earth space research, as originally envisaged by IGBP WG4.

The other development was not restricted to Arctic science. As a matter of fact, it led to an enduring world-wide program of basic and applied research in space physics. This story begins right after the 1986 ICSU Assembly in Berne. I had returned home to Alaska deeply concerned about possible consequences of the assault on near-earth space science prior to and during the ICSU meeting, and felt the obligation of doing something at the national level. Taking shameful advantage of now being the Chairman of ARC (a presidential commission with the Director of the NSF an ex-officio member), during its meeting coffee breaks I held conversations with NSF director Erich Bloch, filling him in on this problem. He encouraged me to write him a letter, of course not as Chairman of ARC but as director of the Geophysical Institute of the University of Alaska.

The letter, dated 6 August 1986, begins by defining the problem as: "how to place solar-terrestrial research (STR), a relatively new interdisciplinary field, into the framework of a funding agency such as the NSF." It then points out that STR has now entered a new phase of trying to understand how the solar-terrestrial system works as a single whole, and pointing out that:

> "*. . .near-earth space has become a crucial technological resource . . . yet the medium in which such earth-orbiting resources operate is hostile . . . Prediction of weather and climate in space is rapidly becoming an economic necessity . . .* "

The letter also addresses the fact that the solar activity-controlled outer regions of the geosphere play a role for life—significant in the long term because of their shielding effect from the constant solar-wind flow, but also more subtly in the short term through variations of the ozone layer. Finally, since this letter mainly addresses STR in the domestic arena, I also elaborated on the importance of the predictability of space disturbances for astronautics and the national defense systems.

As a result, NSF invited a small group of STR scientists (Louis Lanzerotti, Stamatios Krimigis, George Reid and myself) to make a formal presentation to NSF Director Bloch and Assistant Director William Merrell, in which it was decided that a proposal for action be submitted to organize a planning workshop. The proposal was approved quickly and the workshop took place at Washington University in August 1987.

And thus, the Geosphere Environment Modeling (GEM) program was born (Roederer, 1988a). A few years later it spun off the international *Space Weather Program*—in close connection with SCOSTEP but outside the ICSU organization (Roederer, 1988b). Today the US space weather central is located at NOAA, and also Europe has an active space weather network distributed among ESA Member States.

Part III: Lessons learned

My personal involvement in the early phase of the "Climate Revolution" and related consequences have taught me (and hopefully others too) some important lessons.

- The crucial importance of international cooperation not just for the advancement of geophysical research per se, but also for the advancement of science in general in developing countries.
- The crucial importance of engaging in interdisciplinary studies, because in the real world, everything is coupled with everything else.
- The importance of learning to communicate with the public and politicians in their language and not that of science.
- The fact that not all scientists are equal, that there are some who can bully and others who are so shy that they withdraw whenever confronted with debate.
- That to understand and propose practical solutions in climate research, whether terrestrial or space, a solid knowledge of physics and mathematics is imperative.

There is one lesson not yet learned enough concerning climate in space: That there is an equivalent of greenhouse gas pollution in the form of ever-increasing orbiting debris from past satellite and rocket missions, posing a fatal threat to technological activity and human habitability in space.

Finally, I would like to state a personal opinion: decreasing the anthropogenic generation of greenhouse gases, while necessary, is not sufficient. We are dealing with the integral Earth System as one single whole—of which a thin veneer, the anthroposphere, is being polluted not just by greenhouse gases but with sheer numbers of people—whether they drive cars or ride on oxen: it is the good old Second Principle of Thermodynamics that counts. Every bit of organization in whatever form out of disorganization costs > (ln2) kT of energy. Ask the bit-coin miners.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The events described above happened during my tenure as professor and institute director at the University of Alaska, Fairbanks. I am most grateful for the unwavering support of its authorities for my organizational activities in programs of international scientific cooperation. I am most indebted to my LASP colleagues Drs. Fran Bagenal and Maria Usanova for their encouragement and assistance in preparing the present article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Bolin, B., Beloussov, V. V., di Castri, F., Eddy, J. A., Ernster, L., Kendrew, J. C., et al. (1986). "The international geosphere-biosphere programme: A study of global change," in Proceedings of the Final report of the ad-hoc planning group, ICSU 21st general assembly, Berne, Switzerland, September 1986, 14–19.

Eddy, J. A., Baker, D. J., Broecker, W. S., Drake, C., Eagleson, P., Fyfe, W. S., et al. (1986). *Global change in the geosphere-biosphere: Initial Priorities for an IGBP*. Washington, DC: National Academy Press.

Friedman, H., Johnson, D., Johnson, F. S., and Malone, T. F. (1983). *Toward an International Geosphere-biosphere Program: A study of Global Change : Report of a National Research Council Workshop*. Woods Hole, MA: National Academy Press.

Korsmo, F. L. (2007). The genesis of the international geophysical year. *Phys. Today* 60 (7), 38–43. doi:10.1063/1.2761801

Malone, T. F., and Roederer, J. G. (1985). *Global change*. Cambridge: Cambridge University Press, 512.

Roederer, J. G. (1986b). Controversy enlivens planning of the international geosphere-biosphere program: One point of view. *Eos Trans. AGU.* 67 (37), 674. doi:10.1029/EO067i037p00674-01

Roederer, J. G. (1988b). GEM: Geospace environment modeling. *Eos Trans. AGU.* 69 (33), 786–787. doi:10.1029/88EO01064

Roederer, J. G. (1988a). Glasnost in the arctic. *Eos Trans. AGU.* 69 (18), 561–570. doi:10.1029/88EO00153

Roederer, J. G. (1986a). ICSU gives green light to IGBP. *Eos Trans. AGU.* 67 (41), 777–781. doi:10.1029/EO067i041p00777-01

Roederer, J. G. (1989). Igbp: Crown jewel or prodigal son? *Eos Trans. AGU.* 70 (50), 1552–1553. doi:10.1029/89EO00388

Roederer, J. G. (1986c). Research priorities in the arctic: U.S. Arctic research commission gets down to business. *Eos Trans. AGU.* 67 (24), 521–522. doi:10.1029/EO067i024p00521

Roederer, J. G. (1987). "The role of solar-terrestrial research in the international program on global change," in Proceedings of the IAF, Int. Astronaut. Congr, 37th, Innsbruck, Austria, October 1986.

Tilford, S. G. (1984). Global habitability and earth Remote sensing. *Philosophical Trans. R. Soc. Lond.* 312 (1519), 115–118.

Waldrop, M. M. (1984). An inquiry into the state of the Earth. *Science* 226 (4670), 33–35. doi:10.1126/science.226.4670.33

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 01 November 2022
DOI 10.3389/fspas.2022.1045689

Check for updates

OPEN ACCESS

EDITED BY
Maria Usanova,
University of Colorado Boulder,
United States

REVIEWED BY
Colin Forsyth,
University College London,
United Kingdom
Joseph E. Borovsky,
Space Science Institute, United States
Lev Zelenyi,
Space Research Institute (RAS), Russia

*CORRESPONDENCE
Elizaveta E. Antonova,
elizaveta.antonova@gmail.com

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 15 September 2022
ACCEPTED 21 October 2022
PUBLISHED 01 November 2022

CITATION
Antonova EE (2022), From physics of polar aurora to changes of the fundamental approaches to the physics of the magnetospheric processes.
Front. Astron. Space Sci. 9:1045689.
doi: 10.3389/fspas.2022.1045689

From physics of polar aurora to changes of the fundamental approaches to the physics of the magnetospheric processes

Elizaveta E. Antonova[1,2]*

[1]Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia, [2]Space Research Institute RAS, Moscow, Russia

One of the main problems of modern magnetospheric physics is the lack of a self-consistent explanation of the main physical processes based on the laws of plasma physics. Among all the traditionally studied phenomena, the polar aurora stands out for being key to our understanding of several magnetospheric processes. In this manuscript, I would like to share with the younger generation my view about main auroral processes which I have developed during my career over the past 50 years.

KEYWORDS

magnetosphere, auroral particle acceleration, alfvénic aurora, magnetospheric turbulence, substorm, outer radiation belt

Introduction

The year 2022 marks the 50th anniversary of my first publication in the field of magnetospheric physics. It was published while I was still a physics student at the Moscow State University. During that time, numerous satellite studies were carried out, and brilliant analytical and numerical models of various magnetospheric processes were developed. However, it was often difficult to achieve a real consensus among groups from different scientific centers or even groups within one center, to the point that until today, some basic magnetospheric processes are still heavily under debate. In this paper, and based on my personal experience and a number of examples, I show how the evolution of our knowledge about the aurora leads to the better understanding of the main magnetospheric phenomena, including the physics of geomagnetic storms and substorms.

The base of the work

My life in space physics began in the Division of Cosmic Rays and Space Physics at the Faculty of Physics of the Lomonosov Moscow State University (MSU) in 1970. I made the decision to attend the MSU when I was a student at a school with advanced teaching of mathematics in 1964, 2 years before my graduation. I had to accomplish a lot to get admitted to MSU. Although the education was free, the admission exams at MSU were hard and extremely competitive. More than eight applicants from all over the country were competing for each place at the Faculty of Physics in the year of my admission. The symbol of the Faculty of Physics was and remains the "root of the factorial" ($\sqrt{!}$), which means "look at the root." This principle served as the foundation for all my subsequent work.

Each member of the Laboratory of Theoretical Space Physics dealt with his/her own problems and independently chose approaches and methods for solving them. My advisor Boris Arkadievich Tverskoy believed that it was necessary only to formulate the problem, and the student should come up with its final solution. Therefore, at the beginning, I did not understand what difficulties I would have to face during the initial stages of research. Tverskoy never explained how the problem could be solved and did not provide references to any work. This meant that we had to spend many hours every day in libraries. In addition, I remember well the following recommendation: with all due respect to the authors, do not read the sections containing the interpretation of the experimental results, look only at the figures. This is the only way to get the objective information. Through my career, I encountered numerous times proof of the validity of this approach. While I faced various difficulties at the early stage of my career associated with the almost complete independence provided (I managed to get B.A. Tverskoy attention no more than 5–10 min per month), that very independence has made me a real scientist. The Internet age has fundamentally changed the availability of scientific publications. Orders of magnitude more scientific articles are published now than at the beginning of my career, but almost each article contains way less new information. Here I fully agree with the assessment made in the recently published article by Borovsky (2022).

My goal during the postgraduate study was more straightforward. It was necessary to understand the nature of auroras and the mechanisms of auroral particle acceleration. I have to say, I keep working on this problem even now.

Since 1975, I have been working as a researcher at the Skobeltsyn Institute of Nuclear Physics (SINP). While working in theory, I always appreciated reliable empirical data and did my best to cultivate collaboration with experimentalists. Brilliant instrument designers E.N. Sosnovets, S.N. Kuznetsov, and many others, and excellent data analyzers, such as O.V. Khorosheva, worked at that time at SINP MSU. Olga Khorosheva (1961) compared simultaneous observations of all-sky cameras, and proved the existence of a continuous glow ring surrounding the pole that shifted to the night side, in contrast to the previously statistically obtained auroral zone. It was called the auroral oval (Feldshtein, 1963). Thus, Olga Khorosheva became my first mentor in the field of polar aurora. The well-known scientists Velior Shabansky and Yuri Galperin became my PhD thesis opponents. At numerous national conferences, I met almost all famous space scientists who worked at that time in the USSR and many of their students. It was possible to have discussions with Valeria Troitskaya, the first president of the International Association of Geomagnetism and Aeronomy (IAGA), Viktor Trakhtengerts, Mikhail Pudovkin and other famous scientists. The beginning of the INTERBALL project brought together many scientists from the USSR and abroad and gave me an opportunity to get acquainted with Albert Galeev and his follower Lev Zelenyi. Since then, I, like many my colleagues, have been working part-time at the Space Research Institute of the Russian Academy of Sciences (IKI RAN). A separate page should be dedicated to the enormous efforts of Lev Zeleny (later a member of the Russian Academy of science) in the 90s to support and save local space physicists in this difficult time after the collapse of the Soviet Union. Since then, I have been one of his friends. Anatoly Petrukovich, who replaced Zelenyi as the director of IKI, continued the effort to unite and support all specialists in our field.

I have traveled a lot to scientific conferences, first within my country, and since the late 80s, internationally as well. This provided me with a unique opportunity to meet the world's leading scientists. Communication with the elite of Space physics community was extremely beneficial to my work. The most fruitful were acquaintances with David Sibeck, Joe Borovsky, Tony Lui, Larry Lyons, Patrick Newell, Maha Ashour-Abdalla and many other well-known scientists in the field of magnetospheric physics. Particularly noteworthy is my acquaintance with sansei Syun-Ichi Akasofu, whose approach (Akasofu, 2022) is very similar to one that I have developed over the years of analyzing magnetospheric processes.

My ideas about the structure of the auroral oval and the persistent patterns of its dynamics changed after I met the unique observer of polar lights Galina Kirillovna Nazarchuk and worked with her for almost a month at the Tiksi geomagnetic observatory of the Institute of Cosmophysical Research and Aeronomy (IKFIA). I have seen with my own eyes what is called a diffuse aurora. In difference to the common view, it consists of thin strips with almost constant thickness of ~2 km. However, these structures can only be seen when the aurora band is at the magnetic zenith. This phenomenon has been barely studied. I discussed it in (Antonova et al., 1999), where the existing evidence of the non-magnetized nature of the motion of auroral electrons in the region at the top of the magnetic field line was collected. Thin discrete auroral draperies were observed almost every night in Tiksi. A curtain consists of thin beams; the thickness of each is more than two orders of magnitude less than its vertical extent. When such a structure passes through the magnetic zenith, a corona-like aurora is observed due to the foreshortening effect. The main thing that we managed to see is the real projection of magnetospheric turbulence on ionospheric altitudes (rapid movements, appearance and disappearance of structures) during the 10–11 March 1979 magnetic storm.

Acceleration of electrons and ions in the auroral oval as the result of medium scale electric field dynamics

My first two papers, published in 1972, were focused on theoretical analysis of the properties of Pc1 and Pi2 micro pulsations. Therefore, from the start of my career, I have been interested in working on the ring current (source of Pc1) and magnetospheric substorms, leading to the excitation of Pi2 at the time of the substorm onset. It was clear that the processes of the appearance of substorms and the formation of a ring current are very closely related (substorm injections form a ring current). However, the situation began to clear up only in the last decade, when we obtained evidence that the main part of the auroral oval is not mapped into the plasma sheet, but into the outer part of the ring current (Antonova et al., 2014; Antonova et al., 2018a).

Field-aligned acceleration of auroral electrons

My PhD thesis "Passage of strong magnetospheric currents through the ionosphere and acceleration of electrons in high-latitude regions", defended in 1975, provides a solution to the problem of the dependence of the field-aligned currents flowing from the ionosphere on the field-aligned potential difference. A simple dependence, which later received the name Knigth relation (Knight, 1973) in the literature, was obtained at the very beginning of my study, but was not immediately published since it was required to describe the mechanisms for creating a field-aligned potential drop. I did not like the widely discussed mechanism of generation of a field-aligned electric field due to the appearance of anomalous resistivity, since the occurrence of finite resistivity in a collisionless magnetospheric plasma meant the occurrence of a very rapid heating of the plasma, i.e., the rapid emergence of a non-equilibrium situation. The heated plasma with the development of anomalous resistivity should be quickly transported away from the heating region. Therefore, it was impossible to explain the existence of a relatively stable auroral structure in which electrons were observed to accelerate to energies much higher than thermal ones. The kinetic description, based on the analysis of the motion of particles at a fixed potential drop along the magnetic field lines, for a given distribution function of electrons and ions in the area at the top of the field line and at ionospheric altitudes, led to a violation of plasma quasi-neutrality. Such violation contradicted the definition of plasma described in the first pages of each plasma textbook. The time of postgraduate studies was coming to an end, yet it was impossible to obtain a quasi-neutral solution of the problem. There was an agonizing period of troubleshooting until an article was found in the journal Cosmic Electrodynamics (Carlqvist, 1972) that described the double diode (Langmuir, 1929) and talked about the possibility of violation of quasi-neutrality in plasma. The resulting relation between the value of the field-aligned current and the field-aligned potential drop was used in my PhD thesis to describe the structure of the ionosphere in the auroral region. Articles (Antonova and Tverskoy, 1975a; Antonova and Tverskoy, 1975b) were published and, despite the fact that publications in the journal Geomagnetism and Aeronomy were not translated into English at that time, were cited in a review (Paschmann et al., 2002). It turned out to be really successful. It was possible to make a change in the definition of plasma, from which the quasi-neutrality disappeared. The large-scale double layer model predicted a sharp drop in the plasma density in the acceleration region compared to neighboring regions. The discovery of the magnetospheric cavity (Calvert, 1981) confirmed this prediction of the model. During this period, many models of double layers of different nature and configuration were published. The problem has not been completely solved to date, but the existence of large potential variations along auroral field lines has been fully proven, which greatly limits the applicability of the frozen-in approximation in describing magnetospheric processes. Later, it became clear that the existence of the field-aligned potential drops was recognized by practically all magnetospheric scientists. Other auroral observations (see Chapter 4 of Paschmann et al., 2002) had different controversial explanations and additionally had difficulties in connecting the aurora with the plasma sheet phenomena which is ordinarily considered as the main source of auroral phenomena.

According to Akasofu (1964) and numerous subsequent works, a substorm expansion phase starts with a brightening of the equatorward auroral arc or with the appearance of a new arc. Therefore, it was very interesting to understand how auroral arcs are formed and the factors that contribute to why the arc becomes brighter at the moment of the onset. The theory developed in (Antonova, 1993) considered the stability of the azimuthal distribution of the magnetospheric plasma pressure and large-scale field-aligned currents during magnetosphere-ionosphere interactions. Arc brightening was considered as the result of the penetration of cold ionospheric plasma into the auroral cavity with field-aligned potential drop at its edge. Different studies (Stepanova et al., 2002; Antonova and Tverskoi, 1979; ets.) supported predictions of this theory including the absence of brightening and large magnetic field distortions before the onset. Many other scenarios of onset were developed in this period. However, the mechanism of brightening represents the real interest to date. In the late 1970s and early 1980s, none of the researchers doubted the validity of such a scenario for the appearance of thin auroral arcs observed at the edges of inverted V structures (Antonova, 1979; McFadden et al., 1986). It described the formation of a beam of cold electrons of ionospheric origin accelerated to energies (1–10 keV) with a density of ~10^2 cm^{-3}, the transverse energy of which is ~several

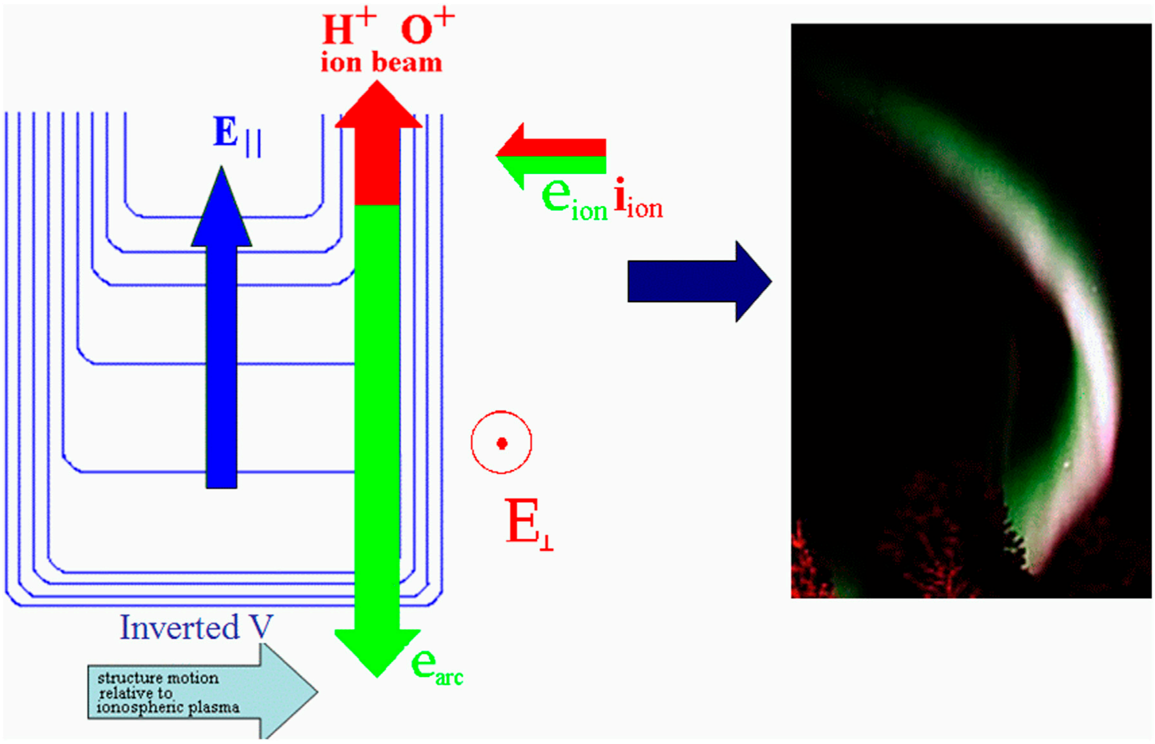

FIGURE 1
Formation of thin bright auroral arc at the boundary of the inverted V/U structure with particle flux J~10^{11} $cm^{-2}s^{-1}$ due to penetration of cold ionospheric plasma inside the plasma cavity ["grader model" in accordance with (Antonova et al., 2012)].

eV, less than the field-aligned energy by orders of magnitude. Figure 1 illustrates the mechanism in action. Such beams were observed on the Fast and DMSP auroral satellites (Mende et al., 2003; Shiokawa et al., 2005). Nonetheless, simultaneous fluctuations of oblique Alfvén waves led to the hypothesis of the wave nature of the acceleration of cold ionospheric electrons although the simulation results required also the action of the mechanism of electron penetration into the region of the field-aligned potential drop (Chaston et al., 2002). It was also obvious that the relaxation of the electron beam carrying the field-aligned current leads to the generation of a wide spectrum of waves, including Alfvén waves. Nevertheless, the term Alfvénic aurora has become established in the literature. Acceleration of powerful field-aligned electron beam was recognized as the main acting mechanism of acceleration. The mechanism of ionospheric plasma penetration in the region of field-aligned potential drop had some difficulties and later practically was not discussed as "a 2D simulation of auroral arc structure, with appropriate wave turbulence may be required to solve this problem (Paschmann et al., 2002, p. 165–166)". Our paper using this mechanism for interpretation of auroral arc splitting and showing that the discussed mechanism difficulties can be illuminated, was rejected. It was only possible to publish an explanation of the generation of multiple arcs by single inverted V structure in our university journal (Antonova et al., 2012). So, in general, the situation should be considered as a failure. However, the situation may be revised in the future with the emergence of new experimental data and a more thorough analysis of existing ones.

Auroral ion acceleration

The acceleration of auroral electrons from the magnetosphere to the ionosphere resulted in the acceleration of ionospheric ions from the ionosphere to the magnetosphere, the formation of ion beams (Antonova, 1979), and the filling of the magnetosphere with ions of ionospheric origin, which has not been studied in detail until now. However, such a process did not allow one to explain the filling of the magnetosphere with heavy ions of ionospheric origin (mainly, O^+ ions were considered) since the formation of ion beams during field-aligned acceleration occurred at sufficiently high altitudes (~ the radius of the Earth). All-sky observations at Tiksi showed storm time fast motions of structures stretched along the magnetic field, which could lead to the action of a two-dimensional analogue of Fermi acceleration of the second order (Antonova, 1983). The developed mechanism described

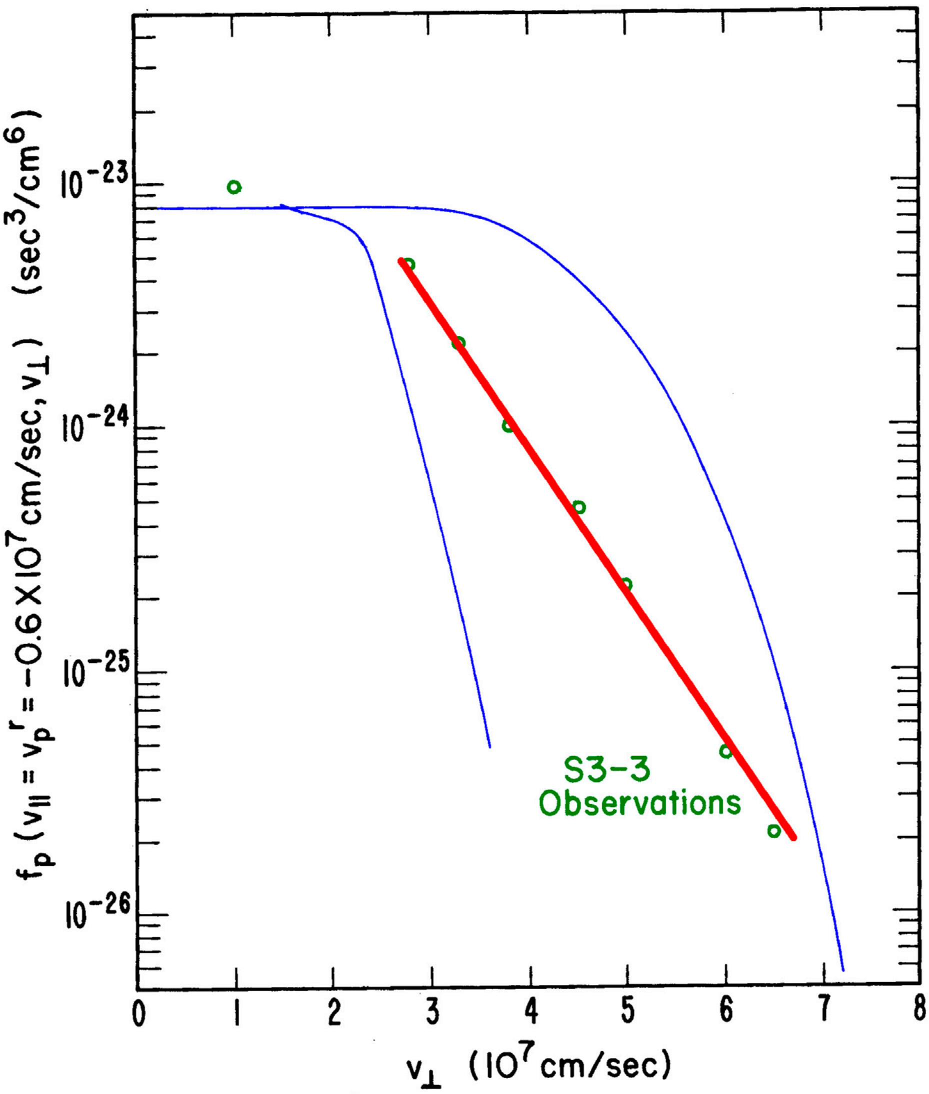

FIGURE 2
Comparison of experimentally measured conic type distribution function on S3-3 satellite on 11 October 1976, at 0046:06 UT (points) with the results of modelling. Blue curve show the results (Dusenbery and Lyons, 1981) model, red curve (Antonova, 1983) model (Antonova, 1987).

the formation of ion "conics" with distribution functions f ~ exp (-V⊥/τ), where τ is the acceleration time. The mechanism explained the predominant acceleration of O^+ ions compared to H^+ ions up to acceleration to the same value of speed, simultaneous observations of ion-cyclotron waves, and well described the shape of the distribution function (Figure 2), but did not become popular. The dominant version was the acceleration due to interaction with cyclotron waves or broadband ELF waves (Paschmann et al., 2002) in spite of the existence of strong transverse electric fields in the auroral ionosphere and difficulties with the explanation of the predominant acceleration of heavy ions. Therefore, the situation was far from consensus. The well-developed mechanism of particle acceleration by simultaneously observed waves was more popular than the ordinary considered in laboratory plasma: if nonequilibrium distribution function is observed simultaneously with waves which can be generated due to its relaxation, it is necessary to search the mechanism of distribution functions formation different from interaction with observed waves.

Relativistic electron acceleration

The same situation occurs with field-aligned electron beams and conics in the auroral plasma appeared with relativistic electrons of the outer radiation belt (ORB). The resonant wave-particle interaction mechanism was considered as the main acceleration mechanism is the very popular quasilinear mechanism of acceleration of by chorus waves. The mechanism could not explain the acceleration time (it required about a day, but in reality, this time was very short (Foster et al., 2017)) and the existence of magnetic storms, during which the storm ORB fluxes are restored to the pre-storm level (Antonova and Stepanova, 2015; Antonova et al., 2018b). Such situation appeared due to absence of the attention to substorm injections at low latitudes in the region of depressed magnetic field. Therefore we again have no consensus.

The possibility of finding more powerful mechanisms of particle acceleration than interaction with waves in the cases mentioned above is associated with the existence of large electrostatic and electromagnetic medium scale electric fields in a collisionless magnetospheric plasma and the development of magnetospheric turbulence. The obvious inapplicability of the frozen-in condition greatly complicates its description. At the same time, the observed balance of pressures at the magnetopause and across the plasma sheet (see the references in (Antonova and Stepanova, 2021)) makes it possible to obtain adequate solutions to existing problems, to obtain consensus and move towards more accurate Space Weather predictions than the current ones.

Conclusion and discussion

The main conclusion of our discussion, which can be useful for students and scientists at the beginning of their careers, is the possibility of revising even the well-known and analyzed processes and finding fundamentally new and more adequate explanations. Such results will lead to a consistent picture of magnetospheric phenomena and facilitate a consensus among various scientific groups. This will obviously require a lot of effort and a lot of work. Therefore, I want to wish the new generation of young scientists, quoting the famous American writer Mark Twain, "find a job you enjoy doing, and you will never have to work a day in your life."

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The author is forever grateful to all teaches and former students, coauthors and scientific opponents for attention and fruitful collaboration.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Akasofu, S.-I. (2022). When I encountered difficult problems. *Front. Astron. Space Sci.* 9, 879290. doi:10.3389/fspas.2022.879290

Akasofu, S. I. (1964). The development of the auroral substorm. *Planet. Space Sci.* 12, 273–282. doi:10.1016/0032-0633(64)90151-5

Antonova, E. E., Kirpichev, I. P., Ryazantseva, M. O., Maryin, B. V., Pulinets, M. S., Znatkovaa, S. S., et al. (2012). Magnetospheric substorms and discrete arcs of the polar aurora. *Mosc. Univ. Phys.* 67 (6), 500–507. doi:10.3103/S0027134912060033

Antonova, E. E. (1987). Magnetospheric turbulence and magnetospheric particle dynamics. *Prepr. SIP MSU* 87-005, 32.

Antonova, E. E. (1979). On the formation of the field-aligned potential difference, isotropic and collimated electron fluxes in the evening sector of the auroral magnetosphere. *Geomagnetism Aeronomy* 19 (6), 1064–1069. Translation.

Antonova, E. E., Stepanova, M., Kirpichev, I. P., Ovchinnikov, I. L., Vorobjev, V. G., Yagodkina, O. I., et al. (2018a). Structure of magnetospheric current systems and mapping of high latitude magnetospheric regions to the ionosphere. *J. Atmos. Sol. Terr. Phys.* 177, 103–114. doi:10.1016/j.jastp.2017.10.013

Antonova, E. E., Stepanova, M. V., Moya, P. S., Pinto, V. A., Vovchenko, V. V., Ovchinnikov, I. L., et al. (2018b). Processes in auroral oval and outer electron radiation belt. *Earth Planets Space* 70, 127. doi:10.1186/s40623-018-0898-1

Antonova, E. E., and Stepanova, M. V. (2021). The Impact of turbulence on physics of the geomagnetic tail. *Front. Astron. Space Sci.* 8, 622570. doi:10.3389/fspas.2021.622570

Antonova, E. E., and Stepanova, M. V. (2015). The problem of the acceleration of electrons of the outer radiation belt and magnetospheric substorms. *Earth, Planets, Space* 67, 148. doi:10.1186/s40623-015-0319-7

Antonova, E. E., Stepanova, M. V., Vikhreva, E. A., Ovchinnikov, I. L., and Teltsov, M. V. (1999). Generation of unmagnetized motion of plasma sheet electrons and its possible causes. *J. Geophys. Res.* 104 (A9), 19941–19953. doi:10.1029/1999JA900036

Antonova, E. E. (1993). The development of initial substorm expansive phase disturbance due to generation of localized electric fields in the region of maximum upward field-aligned current. *Adv. Space Res.* 13 (4), 261–264. doi:10.1016/0273-1177(93)90343-A

Antonova, E. E. (1983). The transverse acceleration of ionospheric ions induced by collisions with moving inhomogeneities. *Geomagnetism Aeronomy* 23 (5), 609–612. Translation.

Antonova, E. E., and Tverskoi, B. A. (1975a). Nature of the electron precipitation band of the 'inverted V' type and of the Harang discontinuity in the evening sector of the auroral ionosphere. *Geomagnetism Aeronomy* 15, 85–90. Translation.

Antonova, E. E., and Tverskoi, B. A. (1975b). On acceleration of auroral electrons by a stationary double layer. *Geomagnetism Aeronomy* 15, 453–454. Translation.

Antonova, E. E., and Tverskoi, B. A. (1979). On the role of the polar ionosphere as source of ions in the magnetosphere. *Geomagnetism Aeronomy* 19, 118–119. Translation.

Antonova, E. E., Vorobjev, V. G., Kirpichev, I. P., and Yagodkina, O. I. (2014). Comparison of the plasma pressure distributions over the equatorial plane and at low altitudes under magnetically quiet conditions. *Geomagn. Aeron.* 54, 278–281. doi:10.1134/S0016793214030025

Borovsky, J. E. (2022). Science: Through the years. *Perspect. Earth Spac.* 3, e2022CN000171. doi:10.1029/2022CN000171

Calvert, W. (1981). The auroral plasma cavity. *Geophys. Res. Lett.* 8 (8), 919–921. doi:10.1029/GL008i008p00919

Carlqvist, C. E. (1972). On the formation of double layers in plasma. *Cosm. Electrodyn.* 3 (3), 377–388.

Chaston, C. C., Bonnell, J. W., Peticolas, L. M., Carlson, C. W., and McFadden, J. P. (2002). Driven alfven waves and electron acceleration: A fast case study. *Geophys. Res. Lett.* 29 (11), 1535. doi:10.1029/2001GL013842

Dusenbery, P. B., and Lyons, R. L. (1981). Generation of ion-conic distribution by upgoing ionospheric electrons. *J. Geophys. Res.* 86 (A9), 7627–7638. doi:10.1029/JA086iA09p07627

Feldstein, Y. (1963). Some problems concerning the morphology of auroras and magnetic disturbances at high latitudes. *Geomagnetism Aeronomy* 3, 183–192. Translation.

Foster, J. C., Erickson, P. J., Omura, Y., Baker, D. N., Kletzing, C. A., and Claudepierre, S. G. (2017). Van Allen probes observations of prompt MeV radiation belt electron acceleration in nonlinear interactions with VLF chorus. *J. Geophys. Res. Space Phys.* 122 (1), 324–339. doi:10.1002/2016JA023429

Khorosheva, O. V. (1961). The space and time distribution of auroras and their relationship with high-latitude geomagnetic disturbances. *Geomagnetism Aeronomy* 1, 615–620. Translation.

Knight, S. (19731973). Parallel electric fields. *Planet. Space Sci.* 21 (5), 741–750. doi:10.1016/0032-0633(73)90093-7

Langmuir, I. (1929). The interaction of electron and positive ion space charges in cathode sheaths. *Phys. Rev.* 33 (6), 954954–954989. doi:10.1103/physrev.33.954

McFadden, J. P., Carlson, C. W., and Boehm, M. H. (1986). Field-aligned electron precipitation at the edge of an arc. *J.* Geophys. Res. 91, 1723–1730. doi:10.1029/JA091iA02p01723

Mende, S. B., Carlson, C. W., Frey, H. U., Peticolas, L. M., and Østgaard, N. (2003). FAST and IMAGE-FUV observations of a substorm onset. *J. Geophys. Res.* 108 (A9), 1344. doi:10.1029/2002JA009787

Paschmann, G., Haaland, S., and Treumann, R. (2002). Auroral plasma physics. *Space Sci. Rev.* 103. doi:10.1023/A:1023030716698

Shiokawa, K., Yago, K., Yumoto, K., Baishev, D. G., Solovyev, S. I., Rich, F. J., et al. (2005). Ground and satellite observations of substorm onset arcs. *J.* Geophys. Res. 110, A12225. doi:10.1029/2005JA011281

Stepanova, M. V., Antonova, E. E., Bosqued, J. M., Kovrazhkin, R. A., and Aubel, K. R. (2002). Asymmetry of auroral electron precipitations and its relationship to the substorm expansion phase onset. *J. Geophys. Res.* 107 (A7), 1134. doi:10.1029/2001JA003503

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 11 January 2023
DOI 10.3389/fspas.2022.1101945

Check for updates

OPEN ACCESS

EDITED BY
Jorge Luis Chau,
Leibniz Institute of Atmospheric Physics (LG), Germany

REVIEWED BY
Scott England,
Virginia Tech, United States

*CORRESPONDENCE
Beatriz Sánchez-Cano,
bscmdr1@leicester.ac.uk

SPECIALTY SECTION
This article was submitted to Space Physics,
a section of the journal
Frontiers in Astronomy and Space Sciences

RECEIVED 18 November 2022
ACCEPTED 12 December 2022
PUBLISHED 11 January 2023

CITATION
Sánchez-Cano B (2023), Mars' ionosphere: The key for systematic exploration of the red planet.
Front. Astron. Space Sci. 9:1101945.
doi: 10.3389/fspas.2022.1101945

Mars' ionosphere: The key for systematic exploration of the red planet

Beatriz Sánchez-Cano*

School of Physics and Astronomy, University of Leicester, Leicester, United Kingdom

The near three decades of continuous Mars' exploration has opened the door to the understanding of the Martian space environment, which includes the solar wind, magnetosphere, ionosphere and atmosphere, and is a complex structure with simultaneous downward and upward couplings. However, we do not yet understand many of the physical processes that drive matter and energy flow between these couplings and within the various atmospheric reservoirs (including temporal and spatial changes on short time scales). Although each coupling plays an essential role for the system, understanding the fate of the ionosphere, as a natural sink of both internal (i.e., atmospheric cycles) and external (i.e., solar wind) energy inputs, is the key for a successful future systematic exploration of Mars.

KEYWORDS

mars, aeronomy, plasma, ionosphere, future exploration, article types perspective

1 Introduction

The Martian space environment, which includes the solar wind, magnetosphere, ionosphere and atmosphere, is a complex system with simultaneous downward and upward couplings (e.g., Lillis et al., 2021; Sanchez-Cano et al., 2022), all of which induce large dynamics in the entire system. In the absence of a global intrinsic field at Mars (Acuna, 1999), the solar wind interaction with the planet mainly occurs at high altitude with its upper atmosphere, where the interplanetary magnetic field bends about the planet inducing a magnetosphere (e.g., Vaisberg et al., 2018). The ultimate region where the energy of the solar wind is dissipated is the ionosphere, which is the solar photoionized part of the thermosphere (upper atmosphere) at ~100–500 km altitude. As an ionized medium, the ionosphere is sensitive to electrodynamics and magnetic fields. As a reactive medium, the ionosphere strongly interacts with the chemistry of the neutral atmosphere. This is known as the magnetosphere-ionosphere-thermosphere (M-I-T) coupled system. The solar wind is the major energy input to the M-I-T system at any of the terrestrial planets, and therefore, the major source of upper atmospheric variability. Thus, understanding how the M-I-T system evolve along the long-term variation of the solar wind with the solar cycle of activity (typically ~11 years) is an essential factor that determines the background variability of the space environment of each planet, and in particular for Mars (e.g., Sanchez-Cano et al., 2022).

A particular feature of Mars is that the ionosphere-solar wind interaction is more complex over a region of the southern hemisphere where highly non-uniform crustal magnetic fields are located. These fields, which are of the order of a few tends to hundreds nT at ~400 km (e.g., Langlais et al., 2019), can interact directly with the solar wind producing a "hybrid magnetosphere", i.e., with features of both induced and intrinsic magnetospheres, which affect the whole planet because it changes as the crustal magnetic fields rotate with the planet (Fang et al., 2015). This is a unique aspect of Mars in our Solar System and means that parts of the Martian M-I-T system may behave differently and in shorter time-scales than expected. Moreover, these crustal fields play an important role in guiding plasma motion, producing hemispheric asymmetry in the magnetosphere, ionosphere, and ion escape (Vaisberg et al., 2018).

Although each region of the M-I-T coupling plays an essential role for the system, the continued exploration of Mars for almost three decades has shown that the ionosphere is a natural sink of both internal and external energy inputs. For example, as internal energy sources, the ionosphere is seasonally affected by lower atmospheric cycles such as the water or CO_2 cycles during spring (e.g., Sánchez-Cano et al., 2018), as well as by regional and even sometimes global dust storms (e.g., Fang et al., 2015; Montabone, 2015), and gravity waves (e.g., England et al., 2017). As an external energy source, solar wind energetic particles precipitate into the ionosphere and ionize the middle atmosphere (a region typically not ionized) causing absorption of signals and, therefore, major radio propagation issues (e.g., Sánchez-Cano et al., 2015; Lester et al., 2022), as well as a myriad of different types of auroras (Schneider et al., 2015; Schneider et al., 2021), and gives us an estimation of the level of shielding of the surface from harmful radiations (Guo et al., 2017). Moreover, the presence of different magnetic field features within the ionosphere (either from crustal fields or from the solar wind) significantly changes the thermal balance of the ionosphere, which becomes magnetized, and ion distributions critically depend on the balance of both the thermal and magnetic pressures (e.g., Sanchez-Cano et al., 2020). All these processes are entangled between several regions of the system. Understanding their temporal and spatial variability in order to assess the differences of the processes that control the long-term dynamics of an ionosphere, including the role of the electrodynamics induced by the solar wind at Mars along the solar cycle and the motion and dynamics of the neutral atmosphere, which in turn can also create currents in the ionosphere (Riousset et al., 2014; Collinson et al., 2019), is a critical aspect for exploration of the red planet that we must resolve (Lillis et al., 2021; Sanchez-Cano et al., 2022).

The ionosphere of Mars is the layer of the upper atmosphere that plays a critical role in balancing the energy of the Martian system by coupling the neutral atmosphere with space, and can be considered as a tracer for atmospheric dynamics and also, for

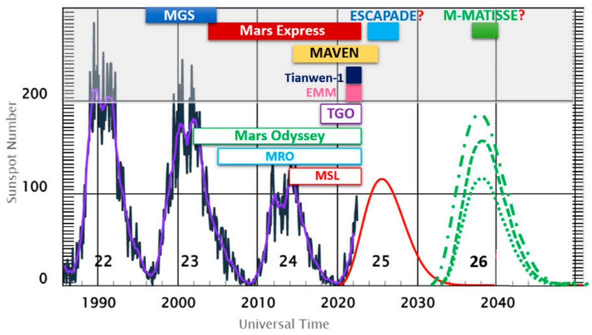

FIGURE 1
Time life of missions at Mars since 1986 with respect to solar cycles 22–26. Missions with ionospheric instrumentation are shown with filled boxes and over the grey band at the top of the Figure. Missions with complementary atmospheric or particle observations are shown with white boxes. ESCAPADE (awaiting for launch window period) and M-MATISSE (under current ESA study) are also shown as potential future missions. The sunspot numbers are plotted in black and with a running average in purple. The red line shows the current prediction for solar cycle 25. The multiple green lines show some potential possibilities for solar cycle 26 with the intention of showing the expected period of maximum solar activity. There has been an active ionospheric mission at Mars since 1996. MGS, Mars Global Surveyor; TGO, ExoMars Trace Gas Orbiter; MSL, Mars Science Laboratory; MRO, Mars Reconnaissance Orbiter; MAVEN, Mars Atmospheric and Volatile EvolutioN; EMM, Emirates Mars Mission; ESCAPADE, The Escape and Plasma Acceleration and Dynamics Explorers; M-MATISSE, Mars Magnetosphere ATmosphere Ionosphere and Space-weather SciencE. This Figure has been modified from the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Centre (https://www.swpc.noaa.gov/products/solar-cycle-progression).

the solar wind-lower atmosphere coupling. The future of Mars exploration, either manned or robotic, will critically depend on a good understanding of ionospheric variability, thus the M-I-T coupling, as it determines the properties of the near-planet environment, and in turn its habitability.

2 Our gained knowledge of Mars' ionosphere thanks to the continuous exploration of Mars

For the last almost 3 decades there has always been an active mission at Mars taking different types of ionospheric observations. This is illustrated in Figure 1, where the lifetime of different missions as well as planned and proposed missions at Mars with respect to the latest solar cycles is plotted. Before the arrival of Mars Global Surveyor (MGS) at Mars, the knowledge of Mars ionosphere was very limited and only basic information was gathered thanks to a few flybys and short-life missions. A critical discovery by MGS was the presence of crustal magnetic fields at Mars (Acuna, 1999) that changed the entire vision that we had of the planet, as well as it showed for the first time the effect of non-migrating waves in the ionosphere, a clear sign of coupling between the lower and upper atmospheres (Bougher, 2001). This, together with its long duration covering the maximum and moderate levels of solar activity in solar cycle 23 opened a new door for planetary plasma physics research. Since then, the continued exploration of Mars by several missions, but in particular, thanks first to Mars Express, and later to the Mars Atmosphere and Volatile EvolutioN (MAVEN) missions, has allowed us to start to appreciate how the ionosphere intertwine with the atmospheric layers and with the solar wind. It has also allowed us to start characterizing the basic properties of the various plasma boundaries and regions that exist in the Martian induced magnetosphere (e.g., the bow shock, magnetic pileup boundary, ionopause (e.g., Hall et al., 2019; Edberg, 2009; Mazelle, 2004; Sanchez-Cano et al., 2020), as well as processes that control the current Martian climate, from the general atmospheric circulation, to the role of photochemistry, clouds, development of dust storms (both local and global), channels of atmospheric escape (e.g., England et al., 2017; Hartwick, 2019; Farahat et al., 2021).

Some remarkable findings are the characterization of the vertical structure of the dayside ionosphere and how it varies with altitude, solar zenith angle, solar flux, Sun-Mars distance, seasons, solar cycle, crustal magnetic fields, dust seasons, and water/CO_2 cycles (Duru et al., 2006; Sanchez-Cano et al., 2013; Sanchez-Cano et al., 2015; Sanchez-Cano et al., 2016; Sanchez-Cano et al., 2018; Mendillo et al., 2013). These missions also found density fluctuations in the topside Martian ionosphere (Gurnett, 2010; Fowler et al., 2017), as well as large ionization layers that last for several days at the bottomside of the ionosphere after cometary dust is deposited, as seen after the flyby of comet Siding-Spring (Gurnett et al., 2015), or even highly sporadic lower magnitude layers found in the bottomside at other times less actives than during this cometary flyby (Peter et al., 2021).

We also have gained significant knowledge of the nightside ionosphere, including terminator transport (e.g., Withers et al., 2012), electron impact ionization down to 90 km (e.g., Lillis et al., 2018), as well as some insights into how the ionosphere is strongly controlled by the inclination of crustal magnetic fields (Němec et al., 2011). Moreover, we now know that the Martian induced tail is highly twisted (DiBraccio et al., 2018), mostly as a result of the complex interaction of the interplanetary magnetic field with the crustal magnetic fields and with the draped fields about the planet (e.g., Fang et al., 2015). The twisted tail may be potentially related to the large levels of electron precipitation observed in the nightside ionosphere, most of which lead to the generation of a myriad of auroras that have been observed at Mars (Schneider et al., 2015, 2021), such as: 1) "discrete auroras", which occur typically over regions of strong vertical crustal magnetic fields (Bertaux, 2005); 2) "diffuse aurora", which are caused by global precipitation of solar energetic particles on the nightside of Mars (e.g., Gerard et al., 2017); 3) "proton aurora", which occurs on the dayside upper atmosphere and it is caused by solar wind proton precipitation (Deighan et al., 2018); 4) and the recently discovered "sinuous aurora" by the Emirates Mars Mission (EMM), which is characterized by elongated serpentine structures of thousands of kilometers in the nightside northern hemisphere, far from intense crustal fields (Lillis et al., 2022), and which origin is still unknown.

An important step forward in our knowledge of Mars as an integrate system has been the synergistic opportunities to use observations from different missions to investigate the variability of the ionosphere, and of the M-I-T coupling. This cooperation, mainly performed from researcher-based efforts, has opened a new door for global investigations of Mars, for the first time at other planet than earth.

3 What do we do not know about Mars' ionosphere?

Despite the great progress in our understanding of the M-I-T coupling at Mars, there are major open questions that still need an answer, especially on the eve of Mars' systematic robotic and human exploration. For example, a few highlights are, the fundamental nature of the plasma boundaries, as formed by systems of currents coupling the solar wind, magnetosphere, and ionosphere, is not well understood. Currents are a natural connection between different regions of a planet (Ramstad et al., 2020), and a quantitative description of their role, on both global and local scales, together with crustal fields, ionosphere, magnetosphere and particle and energy exchanges between regions is still missing (e.g., Sanchez-Cano et al., 2022).

Martian atmospheric losses to space are largely the result of thermal escape of neutral hydrogen and photochemical escape of neutral oxygen. This, together with sputtering, ion outflow, and pickup ion escape, are thought to have controlled the loss of liquid water on Mars over time (Jakosky, 2015), a critical aspect for understanding Martian habitability. However, the main atmospheric loss channels at Mars today cannot be observed because the rates of the escaping neutral hydrogen and oxygen atoms cannot be directly measured with current technology due to their low density and energy (several eV), although significant efforts have been done *via* different routes, such as on water vapour and hydrogen abundance observations, to get a better understanding of this escape mechanism (Holmes et al., 2021).

As mentioned before, aurora formation within the ionosphere on the nightside of Mars thanks to electron precipitation far from crustal fields is still a mystery. Energy deposition from particle precipitation can drive ionospheric structure through ionization. At earth, auroras are explained by direct cusp entry of the solar energetic particles, and by magnetospheric tail reconnection. However at Mars, this has not been confirmed, and could be related to both magnetotail field topology and poorly understood tail and cusp processes related to the localized crustal magnetic field regions.

Moving down into the lower altitudes, the bottom-side ionosphere is scarcely sampled; with practically no observations of the ionosphere below ~70 km (only indirect observations and modelling efforts are available). Thus, a significant gap in our knowledge is present as it is believed this region to play a major role in the controls the amount of energy from the solar wind that is deposited into the atmosphere and eventually into the surface (e.g., Sanchez-Cano et al., 2021; Nakamura et al., 2022). This is particularly important when considering the radio propagation issues that low ionization creates, and therefore, being highly relevant to ground operations and future Mars exploration.

This is only a small sample of the Mars' ionosphere unknowns, which may be essential knowledge for the future of Mars science and exploration, but also, to understand other worlds thought comparative planetology. Having good knowledge of Mars, which is relatively close to earth, easy accessible and has is near-unmagnetised in contrast with earth, is fundamental to extrapolate and compare knowledge to other worlds which may not be that accessible and limited observations can be taken.

4 Discussion: The ionosphere as key to unravel the way forward of Mars' exploration

There are still many aspects that we do not yet understand from Mars environment, including many of the physical processes that drive matter and energy flow between and within the various atmospheric reservoirs (including temporal and spatial changes on short time scales, see Section 3), and although single-spacecraft missions provide a wealth of observations, synergetic and simultaneous multi-point measurements of the system are still missing. Simultaneous multi-point and coordinated measurements are required to determine how energy flows through the induced magnetosphere and into the ionosphere-atmosphere, causing important dynamics and energization (Sanchez-Cano et al., 2021).

From the near 60 years of space exploration at earth, we can draw the need for multi-point missions that have revolutionized our understanding of the terrestrial solar wind-magnetosphere-ionosphere coupling particularly during the last 20 years. At Mars, "ad-hoc" multi-spacecraft studies between existing individual assets have been undertaken and have demonstrated the huge potential that a coordinated mission has, but the instrument suites of existing assets are not designed for multi-point observations and opportunities for such analyses are rare. Thus, dedicated missions with multiple spacecraft having fully coordinated and simultaneous observations at different parts of the Martian system is critical to unravel the key mechanisms that strongly couple its surface with the M-I-T system and the solar wind, which requires at least two well-separated spacecraft with one in the solar wind and the other inside the system, and this is not available at the moment.

Following earth example, several multi-spacecraft missions have been proposed to fulfil this requirement, which will bring us to the next-generation of exploration at Mars where for the first time we will be able to disentangle spatial from temporal variability, and capture variations on short spatial/temporal scales that cannot be resolved from a single spacecraft. The first one of this missions is ESCAPADE (Lillis, 2020), which is a small-class NASA twin-spacecraft Mars orbiter mission which launch is still to be decided, that will provide a global picture of how solar wind energy flows through Mars' unique hybrid magnetosphere to drive ion and sputtering escape. With views to future robotic and manned exploration of Mars, the Mars Magnetosphere Atmosphere Ionosphere and Space-weather Science (M-MATISSE) mission is currently being evaluated by ESA for the next Medium-size mission of opportunity to be launched not earlier than 2037. M-MATISSE is a two-spacecraft orbiting Mars to investigate the dynamic response of the M-I-T coupling to space weather activity (Sanchez-Cano et al., 2022). Moreover, the most ambitious of the missions concepts is the Mars Orbiters for Surface, Atmosphere, and Ionosphere Connections (MOSAIC) (Lillis et al., 2021), which is a Planetary Mission Concept Study mission from the NASA Science Decadal Survey, composed of 10 spacecraft to cover and investigate the system as a whole, with the most detailed ever coordination, covering all regions of the M-I-T coupling, including the surface and subsurface of Mars, and with a permanent monitoring of the Sun and solar wind.

All these missions have in common that the ionosphere is key in their observations. This is because Mars ionosphere is the "sponge" or "porous layer" in between the lower-middle atmosphere and space that facilitates the connection between different regions, where ultimately energy is dissipated, where the strongest dynamics occur, it is the reservoir for atmosphere escape, and it is where the larger part of the radiation filtering for the surface occurs. Moreover, the state of the ionosphere strongly controls communications with the surface and instrument operations in HF and UHF frequencies, as well as ionospheric irregularities can potentially produce scintillation in signal propagation. Therefore, the future of Mars exploration is strongly linked to the fate of the ionosphere, being the key for its future systematic exploration.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

BS-C has written and made the Figures of this paper.

Acknowledgments

Useful discussions concerning the main ideas of this article with Prof Mark Lester, Dr Francois Leblanc and Dr Olivier Witasse are acknowledged. The Mars Upper Atmosphere Network (MUAN) led by Dr Hermann Opgenoorth is greatly acknowledged for all the discussions and activities to bring the best science and facilitate collaborations from teams from all current missions at Mars in order to investigate the upper atmosphere of its interaction with the solar wind. BS-C acknowledges support through STFC Ernest Rutherford Fellowship ST/V004115/1.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acuna (1999). Magnetic field and plasma observations at Mars: Initial results of the Mars global Surveyor mission. *Science* 279 (5357), 1676–1680. doi:10.1126/science.279.5357.167

Bertaux (2005). Discovery of an aurora on Mars. *Nature* 435, 790–794. doi:10.1038/nature03603

Bougher (2001). Mars global Surveyor radio science electron density profiles ß neutral atmosphere implications. *Geophys. Res. Lett.* 28 (16), 3091–3094. doi:10.1029/2001GL012884

Collinson, G., McFadden, J., Mitchell, D., Grebowsky, J., Benna, M., Espley, J., et al. (2019). Traveling ionospheric disturbances at Mars. *Geophys. Res. Lett.* 46, 4554–4563. doi:10.1029/2019GL082412

Deighan, J., Jain, S. K., Chaffin, M. S., Fang, X., Halekas, J. S., Clarke, J. T., et al. (2018). Discovery of a proton aurora at Mars. *Nat. Astron.* 2 (10), 802–807. doi:10.1038/s41550-018-0538-5

DiBraccio, G. A., Luhmann, J. G., Curry, S. M., Espley, J. R., Xu, S., Mitchell, D. L., et al. (2018). The twisted configuration of the martian magnetotail: MAVEN observations. *Geophys. Res. Lett.* 45, 4559–4568. doi:10.1029/2018GL077251

Duru, F., Gurnett, D. A., Averkamp, T. F., Kirchner, D. L., Huff, R. L., Persoon, A. M., et al. (2006). Magnetically controlled structures in the ionosphere of Mars. *J. Geophys. Res.* 111, A12204. doi:10.1029/2006JA011975

Edberg, 2009, Simultaneous measurements of martian plasma boundaries by Rosetta and Mars express, *Planet. Space Sci.*, Volume 57, Issues 8–9. doi:10.1016/j.pss.2008.10.016

England, S. L., Liu, G., Yiğit, E., Mahaffy, P. R., Elrod, M., Benna, M., et al. (2017). MAVEN NGIMS observations of atmospheric gravity waves in the Martian thermosphere. *J. Geophys. Res. Space Phys.* 122, 2310–2335. doi:10.1002/2016JA023475

Fang, X., Ma, Y., Brain, D., Dong, Y., and Lillis, R. (2015). Control of Mars global atmospheric loss by the continuous rotation of the crustal magnetic field: A time-dependent MHD study. *J. Geophys. Res. Space Phys.* 120 (10926– 10), 944. doi:10.1002/2015JA021605

Fang, X., Ma, Y., Lee, Y., Bougher, S., Liu, G., Benna, M., et al. (2020). Mars dust storm effects in the ionosphere and magnetosphere and implications for atmospheric carbon loss. *J. Geophys. Res. Space Phys.* 125, e2019JA026838. doi:10.1029/2019JA026838

Farahat, A., Mayyasi, M., Withers, P., Dayeh, M. A., and Abuelgasim, A. (2021). Effects of the June 2018 global dust storm on the atmospheric composition of the Martian upper atmosphere as observed by MAVEN. *J. Geophys. Res. Planets* 126, e2021JE006868. doi:10.1029/2021JE006868

Fowler, C. M., Andersson, L., Shaver, S. R., Thayer, J. P., Huba, J. P., Lillis, R. J., et al. (2017). MAVEN observations of ionospheric irregularities at Mars. *Geophys. Res. Lett.* 44 (10), 854. doi:10.1002/2017GL075189

Gerard, J. C., Soret, L., Shematovich, V. I., Bisikalo, D. V., and Bougher, S. W. (2017). The Mars diffuse aurora: A model of ultraviolet and visible emissions. *Icarus* 288, 284–294. doi:10.1016/j.icarus.2017.01.037

Guo, J., Slaba, T. C., Zeitlin, C., Wimmer-Schweingruber, R. F., Badavi, F. F., Bohm, E., et al. (2017). Dependence of the Martian radiation environment on atmospheric depth: Modeling and measurement. *J. Geophys. Res. Planets* 122, 329–341. doi:10.1002/2016JE005206

Gurnett, D. A., Morgan, D. D., Persoon, A. M., Granroth, L. J., Kopf, A. J., Plaut, J. J., et al. (2015). An ionized layer in the upper atmosphere of Mars caused by dust impacts from comet Siding Spring. *Geophys. Res. Lett.* 42, 4745–4751. doi:10.1002/2015GL063726

Gurnett (2010). Large density fluctuations in the martian ionosphere as observed by the Mars Express radar sounder. *Icarus* 206, 1. doi:10.1016/j.icarus.2009.02.019

Hall, B. E. S., Sánchez-Cano, B., Wild, J. A., Lester, M., and Holmstrom, M. (2019). The Martian bow shock over solar cycle 23–24 as observed by the Mars Express mission. *J. Geophys. Res. Space Phys.* 124, 4761–4772. doi:10.1029/2018JA026404

Hartwick (2019). High-altitude water ice cloud formation on Mars controlled by interplanetary dust particles. *Nat. Geosci. Vol.* 12, 516–521. doi:10.1038/s41561-019-0379-6

Holmes, J. A., Lewis, S. R., Patel, M. R., Chaffin, M. S., Cangi, E. M., Deighan, J., et al. (2021). Enhanced water loss from the martian atmosphere during a regional-scale dust storm and implications for long-term water loss. *Earth Planet. Sci. Lett.* 571, 117109. ISSN 0012-821X. doi:10.1016/j.epsl.2021.117109

Jakosky (2015). MAVEN Explores the martian upper atmosphere. *SCIENCE* 350 (6261), 643. doi:10.1126/science.aad3443

Langlais, B., Thébault, E., Houliez, A., Purucker, M. E., and Lillis, R. J. (2019). A new model of the crustal magnetic field of Mars using MGS and MAVEN. *J. Geophys. Res. Planets* 124, 1542–1569. doi:10.1029/2018JE005854

Lester, M., Sanchez-Cano, B., Potts, D., Lillis, R., Cartacci, M., Bernardini, F., et al. (2022). The impact of energetic particles on the Martian ionosphere during a full solar cycle of radar observations: Radar blackouts. *J. Geophys. Res. Space Phys.* 127, e2021JA029535. doi:10.1029/2021JA029535

Lillis (2020). Escapade: Coordinated Multi-Point Observations Of Ion And Sputtered Escape From Mars, *51st Lunar and planetary science Conference*. https://www.hou.usra.edu/meetings/lpsc2020/pdf/2470.pdf.

Lillis, R. J., Deighan, J., Brain, D., Fillingim, M., Jain, S., Chaffin, M., et al. (2022). First synoptic images of FUV discrete aurora and discovery of sinuous aurora at Mars by EMM EMUS. *Geophys. Res. Lett.* 49, e2022GL099820. doi:10.1029/2022GL099820

Lillis, R. J., Mitchell, D. L., Steckiewicz, M., Brain, D., Xu, S., Weber, T., et al. (2018). Ionizing electrons on the Martian nightside: Structure and variability. *J. Geophys. Res. Space Phys.* 123, 4349–4363. doi:10.1029/2017JA025151

Lillis, Robert J., Mitchell, David, Montabone, Luca, Heavens, Nicholas, Harrison, Tanya, Stuurman, Cassie, et al. (2021). Mosaic: A Satellite Constellation to enable Groundbreaking Mars climate system science and Prepare for human exploration. *Planet. Sci. J.* 10, 211. doi:10.3847/psj/ac0538

Mazelle (2004). Bow shock and Upstream Phenomena at Mars. *Space Sci. Rev.* 111, 115–181. doi:10.1023/B:SPAC.0000032717.98679.d0

Mendillo, M., Marusiak, A. G., Withers, P., Morgan, D., and Gurnett, D. (2013). A new semiempirical model of the peak electron density of the Martian ionosphere. *Geophys. Res. Lett.* 40, 5361–5365. doi:10.1002/2013GL057631

Montabone (2015). Eight-year climatology of dust optical depth on Mars. *Icarus* 251, 65–95. doi:10.1016/j.icarus.2014.12.034

Nakamura, Y., Terada, N., Leblanc, F., Rahmati, A., Nakagawa, H., Sakai, S., et al. (2022). Modeling of diffuse auroral emission at Mars: Contribution of MeV protons. *J. Geophys. Res. Space Phys.* 127 (1), e2021JA029914. doi:10.1029/2021JA029914

Němec, F., Morgan, D. D., Gurnett, D. A., Duru, F., and Truhlík, V. (2011). Dayside ionosphere of Mars: Empirical model based on data from the MARSIS instrument. *J. Geophys. Res.* 116, E07003. doi:10.1029/2010JE003789

Peter, K., Martin, P., GregorioMolina-Cuberos, J., González-Galindo, F., Olivier, W., Tellmann, S., et al. (2021). The lower dayside ionosphere of Mars from 14 years of Mars radio science observations. *Icarus* 359, 114213. ISSN 0019-1035. doi:10.1016/j.icarus.2020.114213

Ramstad (2020). The global current systems of the Martian induced magnetosphere. *Nat. Astron.* 4, 979–985. doi:10.1038/s41550-020-1099-y

Riousset, J. A., Paty, C. S., Lillis, R. J., Fillingim, M. O., England, S. L., Withers, P. G., et al. (2014). Electrodynamics of the Martian dynamo region near magnetic cusps and loops. *Geophys. Res. Lett.* 41. doi:10.1002/2013GL059130

Sánchez-Cano, Beatriz, Lester, Mark, Andrews, David J., Hermann, Opgenoorth, Lillis, Robert, Leblanc, François, et al. (2021). Mars' plasma system. Scientific potential of coordinated multipoint missions: "The next generation". *Exp. Astron.* doi:10.1007/s10686-021-09790-0

Sánchez-Cano, Beatriz, Lester, Mark, Olivier, Witasse, Blelly, Pierre-Louis, Indurain, Mikel, Cartacci, Marco, et al. (2018). Seasonal, and solar cycle variations of the martian Total electron content (TEC): Is the TEC a good tracer for atmospheric cycles? *J. Geophys. Res. Planets* 10, 1746–1759. doi:10.1029/2018je005626

Sanchez-Cano, B., Lester, M., Leblanc, F., Andrews, D., and Opgenoorth, H. (2022). The M-MATISSE mission: Mars magnetosphere ATmosphere ionosphere and surface SciencE. *44th COSPAR Sci. Assem. Held 16-24 July* 44, 421.

Sánchez-Cano, B., Lester, M., Witasse, O., Milan, S. E., Hall, B. E. S., Blelly, P.-L., et al. (2015). Evidence of scale height variations in the Martian ionosphere over the solar cycle. *J. Geophys. Res. Space Phys.* 120, 913–925. doi:10.1002/2015JA021949

Sánchez-Cano, B., Lester, M., Witasse, O., Milan, S. E., Hall, B. E. S., Cartacci, M., et al. (2016). Solar cycle variations in the ionosphere of Mars as seen by multiple Mars Express data sets. *J. Geophys. Res. Space Phys.* 121, 2547–2568. doi:10.1002/2015JA022281

Sánchez-Cano, B., Narvaez, C., Lester, M., Mendillo, M., Mayyasi, M., Holmstrom, M., et al. (2020). Mars' ionopause: A matter of pressures. *J. Geophys. Res. Space Phys.* 125, e2020JA028145. doi:10.1029/2020JA028145

Sanchez-CanoRadicella, S., Herraiz, M., Witasse, O., and Rodriguez-Caderot, G. (2013). NeMars: An empirical model of the martian dayside ionosphere based on Mars Express MARSIS data. *Icarus* 225 (1), 236–247. doi:10.1016/j.icarus.2013.03.021

Sánchez–Cano, B., Blelly, P.-L., Lester, M., Witasse, O., Cartacci, M., Orosei, R., et al. (2019). Origin of the extended Mars radar blackout of September 2017. *J. Geophys. Res. Space Phys.* 124, 4556–4568. doi:10.1029/2018JA026403

Schneider, N. M., Deighan, J. I., Jain, S. K., Stiepen, A., Stewart, A. I. F., Larson, D., et al. (2015). Discovery of diffuse aurora on Mars. *Science* 350 (6261), aad0313. doi:10.1126/science.aad0313

Schneider, N. M., Milby, Z., Jain, S. K., Gérard, J. C., Soret, L., Brain, D. A., et al. (2021). Discrete aurora on Mars: Insights into their distribution and activity from MAVEN/IUVS observations. *J. Geophys. Res. Space Phys.* 126 (10), e2021JA029428. doi:10.1029/2021JA029428

Vaisberg, O. L., Ermakov, V. N., Shuvalov, S. D., Zelenyi, L. M., Halekas, J., DiBraccio, G. A., et al. (2018). The structure of Martian magnetosphere at the dayside terminator region as observed on MAVEN spacecraft. *J. Geophys. Res. Space Phys.* 123, 2679–2695. doi:10.1002/2018JA025202

Withers, P., Fillingim, M. O., Lillis, R. J., Häusler, B., Hinson, D. P., Tyler, G. L., et al. (2012). Observations of the nightside ionosphere of Mars by the Mars express radio science Experiment (Mars). *J. Geophys. Res.* 117, A12307. doi:10.1029/2012JA018185

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 18 January 2023
DOI 10.3389/fspas.2023.1071809

Check for updates

OPEN ACCESS

EDITED BY
Joseph E. Borovsky,
Space Science Institute, United States

REVIEWED BY
Juan Alejandro Valdivia,
University of Chile, Chile

*CORRESPONDENCE
Nithin Sivadas,
sivadas@cua.edu

SPECIALTY SECTION
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

RECEIVED 16 October 2022
ACCEPTED 02 January 2023
PUBLISHED 18 January 2023

CITATION
Sivadas N (2023), Does the way we do science foster discovery?.
Front. Astron. Space Sci. 10:1071809.
doi: 10.3389/fspas.2023.1071809

Does the way we do science foster discovery?

Nithin Sivadas[1,2]*

[1]Space Weather Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, United States,
[2]Department of Physics, The Catholic University of America, Washington DC, WA, United States

Freedom to explore the unknown is key to scientific discovery. Maximizing modern individualistic measures of scientific productivity like the number of citations and publications may impede the progress of science as a whole.

KEYWORDS

discovery, scientific progress, citation, random walk, diffusion model

1 Scientific discovery

Perhaps it is reasonable to categorize scientific endeavors into two essential elements: discovery and assimilation. Discoveries entail breakthroughs, e.g., finding novel truths about our Universe, constructing new theories that explain phenomena, finding new materials, identifying new pathogens, and uncovering new cures for old diseases. While assimilation involves putting to use this newly found knowledge: e.g., building a global positioning system, advancing a theory to new regimes, or making a medical cure cheaper and more effective. With this framework in mind, we can ask whether the requirements for making discoveries are the same as for assimilating new knowledge.

A perspective I present in this paper is that the requirements for making discoveries are qualitatively different from the assimilation phase of scientific progress. Continuous conceptual breakthroughs of the kind we saw a century ago, like quantum mechanics, relativity, genetic inheritance, psychoanalysis, or statistical mechanics, are not common in the present day (Graeber, 2012). Scientific discoveries appear to happen in bursts with even a certain quasi-periodicity (Jaynes, 1967), and many are accidental (Humberstone, 1943; Gambardella et al., 2005), unlike the directed efforts of implementing existing knowledge in the assimilation stage. Hence, scientific discoveries may benefit from many scientists wandering and exploring freely, driven by curiosity, in the vast n-dimensional space of knowledge. By following this line of thought, we notice two surprising possibilities: 1) The central aspects of the current practice of modern science, like 'publish or perish' or 'citation maximization', might be holding back discoveries and breakthroughs, 2) What appears to be "productive" for scientists, based on our current individualistic metrics of productivity, may be unproductive for scientific discovery as a whole.

2 A thought experiment: Discovery through random walks

I explain the logical basis for these possibilities by constructing a simple thought experiment that captures an aspect of scientific discoveries: its accidental nature. In our toy model, scientists (point particles) carry out random walks to explore an abstract knowledge space subject to non-random pressures associated with the need to maximize citations. The thought experiment's only purpose is to provide an analogy that helps us understand how the pressure to maximize specific individual metrics of agents (in our case, scientists) need not necessarily maximize the goals of the system as a whole.

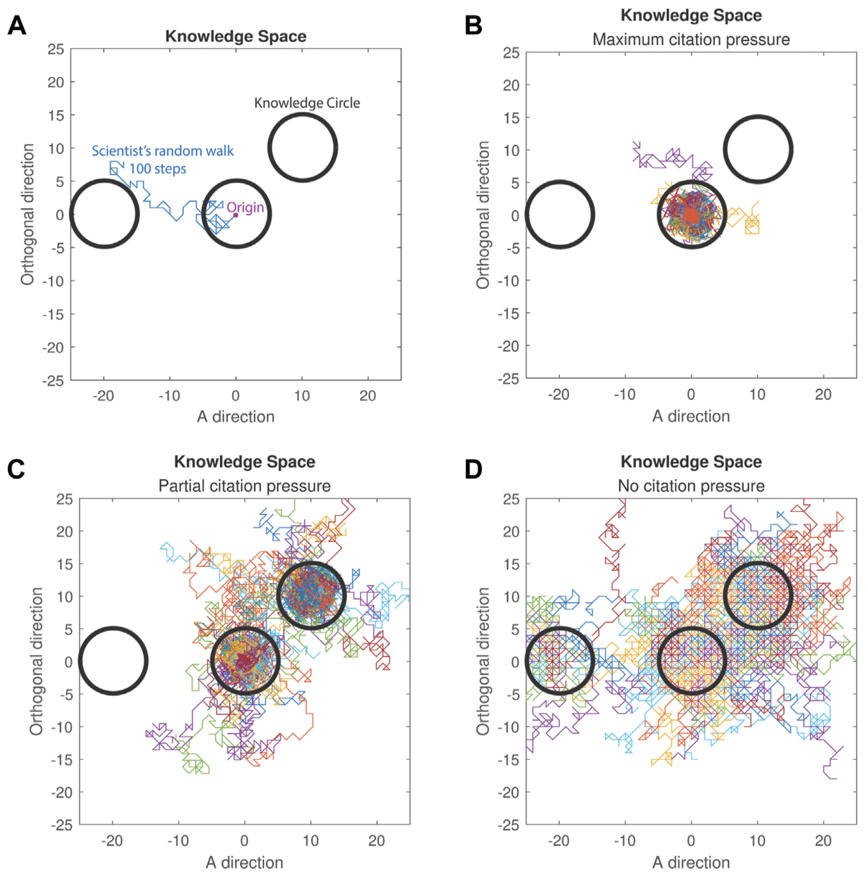

FIGURE 1
Sample runs of the thought experiment showing scientists discover more when citation pressure reduces. **(A)** Shows the abstract knowledge space, with knowledge circles where important breakthroughs reside, and a scientist's random path within the space for a duration of 100 steps. **(B)** Simulation of 100 scientists paths with maximum citation pressure, forcing most scientists to keep exploring within their starting knowledge circle. **(C)** Simulation with partial citation pressure, with scientists discovering a new knowledge circle within the same simulation time frame. **(D)** When there is no citation pressure, scientists discover all knowledge circles within the simulation space during the same simulation time.

Consider an abstract space of knowledge that remains unexplored in the beginning. Such spaces are perhaps n-dimensional, with many degrees of freedom for scientists to traverse along some continuous paths. For simplicity, let us assume scientists are restricted to traveling only along a 2D plane in this abstract space.

Figure 1A depicts such a space of knowledge, with some beneficial knowledge considered by scientists as important and ground-breaking and others considered not-so-useful and not to be of much immediate or apparent value. Regions within black circles represent discoveries of significant magnitude and the scientific field or sub-field that develops around them. These knowledge circles are separated by distance, representing a barrier of knowledge considered "not-so-useful" to be gained before reaching one. In other words, one needs to traverse through knowledge that might not be considered breakthroughs to go from knowing thermodynamics to discovering quantum mechanics. In fact, this construction implies that knowledge considered "not-so-useful" might indeed be actually useful and essential to the making of a discovery. Note that in the knowledge space, knowledge can be uncovered that is not true, and what is considered true might change within the scientific community over time. The model does not discriminate between truth and false knowledge.

In this toy model, all scientists with similar education and basic knowledge start at the center of the coordinate system. The horizontal axis of the plot is simply an arbitrary direction in the knowledge space, with the vertical direction being orthogonal to it. After a scientist appears at the center, they traverse this space through a random walk, with a step in the horizontal or vertical direction that can be either [+1,0,−1] with equal probability. In a random walk, the previous step influences the next step. Though it need not be what happens in reality, in this thought experiment, this is the default motion of a scientist in the knowledge space. It does not imply that the thinking of a scientist is random; instead, it represents scientists' trajectories that are affected by their particular conscious decisions and factors unknown to us.

If a scientist encounters a new knowledge circle, that event is a "discovery". The time taken for the first such event since the start of the simulation is the "time to first discovery". A scientist takes 100 steps in the simulation. After that, it is equally likely that a new scientist will appear at the center of any discovered knowledge circle.

Now, we add some directionality to scientists, perhaps analogous to the real world. For each step a scientist takes, we assume they find something and publish it. If inside a knowledge circle, this publication can gather citations probabilistically from every publication or step by future scientists within the circle. In the toy model, there is a 50–50 chance that a future scientist's publication might cite a scientist's current publication or step.

To model the pressure to maximize one's performance metric in terms of publications and citations, each scientist feels a constant pull towards the center of the knowledge circle they currently exist within. Therefore, scientists in a circle carry out a random walk with a bias toward the center of the circle. I specify the degree of this pull using a parameter called citation pressure, which can vary from no citation pressure to maximum citation pressure. This pressure is analogous to the incentives scientists have towards conducting research that can garner publications and citations from other scientists. Hence it becomes vital to carry out research close to or considered important by other scientists. There are similar pressures scientists feel to propose scientific projects that are close to what other scientists work on and think is essential so that peer reviewers assigned by funding agencies to review your proposals are likely to consider the proposal important.

Assumptions and limitations of the thought experiment: A primary assumption of the thought experiment is that citation pressure forces scientists to remain close to existing fields or breakthroughs. This is described in the toy model as a bias in the random walk towards the center of the knowledge circle. The intuition for this assumption stems from my personal experience. As one attempts to pursue research topics farther away from established areas of research where most scientists spend their time, there are far fewer scientists who may cite your work. But if one does work related to and supportive of most scientists' projects, your paper is more likely to receive citations by them. Pressures to be awarded grants or even to form successful collaborations also appear to incentivize one to remain close to the established research areas, as most of your collaborators and peer reviewers are also likely to work within the established areas of the field. Our toy model allows us to explore what might happen to scientific discovery when these pressures exist in the manner described here or do not exist. A major limitation in the scope of the toy model is that it is not a good model of science, and it is not intended to be. Instead, it is an abstract thought experiment meant to demonstrate certain specific mechanisms of the workings of science. Significant work is needed before one can upgrade this toy model into a realistic model of scientific progress, which is outside the scope of this *perspective* paper (See **Supplementary Material** for details). However, there are studies in the field of the Science of Science (SciSci) that provide results that support the mechanism demonstrated by the toy model Fortunato et al. (2018).

3 Result: Citation pressure delays discovery

We run the above toy model for different citation pressures, with 100 scientists taking 100 steps each. **Figure 1B** shows what happens when citation pressure is set to maximum, i.e., the highest pull towards the center of the knowledge circle. Almost all scientists remain within the starting knowledge circle. In **Figure 1C**, more scientists wander out of the first knowledge circle as citation pressure is lower. One scientist discovers a new knowledge circle at some point, and new scientists appear within the newly discovered knowledge circle. **Figure 1D** shows what happens when there is no citation pressure: scientists will most likely discover all the knowledge circles within the simulation space. These sample runs suggest that discovery is least likely under maximum citation pressure. These figures demonstrate the intuition that if citation pressure makes scientists spend more time close to established research areas, the likelihood of making breakthroughs might decrease. Studies in the field of SciSci appears to agree with this insight as well. A comprehensive review of the field by Fortunato et al. (2018) states "…measurements indicate that scholars are risk-averse, preferring to study topics related to their current expertise, which constrains the potential of future discoveries."

To further investigate this qualitative notion, we delve deeper by running the toy model several times to gather statistics of three relevant and interesting parameters:

1. Time to the first discovery
2. Average citations per scientist
3. Share of time "wasted" by all scientists

The time to first discovery is the simulation time it takes for any scientist to first chance upon a new knowledge circle. The smaller this value is, the more productive science is in making discoveries as a whole. The average citations per scientist is the total citations gathered by all scientists divided by the total number of scientists in that model run. This metric measures how well individual scientists are doing on average, based on the belief that citations are a good measure of a scientist's productivity. The share of time "wasted" by all scientists is the percent of simulation time where scientists' work does not gather any citation. It is the time spent by scientists outside the knowledge circles. The complement of "wasted" time is "productive" time, which is the share of time scientists spend within knowledge circles. The term productive and wasted are placed in quotations to emphasize the subjectivity of this term, i.e., they are perceived to be productive or wasted time, but they need not be in reality. Not many scientists will deny that in the modern academic world, the number of papers and citations is *considered* a proxy for the "productivity" of scientists, even though it does not necessarily mean more progress toward scientific discoveries.

We run the simulation 30 times, each with different citation pressures to generate the statistics displayed in **Figure 2**. **Figure 2A** shows that the time to first discovery increases with citation pressure. In fact, it increases exponentially when citation pressure is greater than 1. Each dot represents data from a simulation run. For high citation pressures, no discovery is made during the entire simulation time, like in **Figure 1B**. Therefore, there are many data points with a time to first discovery greater than or equal to the simulation time of $100 \times 100 = 10^4$ units.

Figure 2B shows that average citations per scientist increase linearly with citation pressure. This confirms that the proxy for citation pressure: the pull towards the center of the knowledge circle, is consistent with an increase in the average number of citations for scientists. **Figure 2C** conveys that the share of time "wasted" decreases with increasing citation pressure, which is consistent with the modern definitions of scientific productivity. As the notion of "productive" time also increases with citation pressure.

By combining **Figures 2A–C**, we can conclude that when most scientists in this toy model appear productive, it takes longer for the society of scientists to discover an entirely new knowledge circle. A more direct description of this result is shown in **Figures 2D, E**. They show that the time to first discovery is indeed delayed as average citations per scientist are higher in the simulation. Similarly, as average citations per scientist are higher, the total share of time scientists are individually productive based on their publications gathering citations is also higher. By combining the results of **Figures 2D, E**, we can conclude that when a large share of scientists' time appears to be "productive", there is a large delay in the time to first discovery. Or

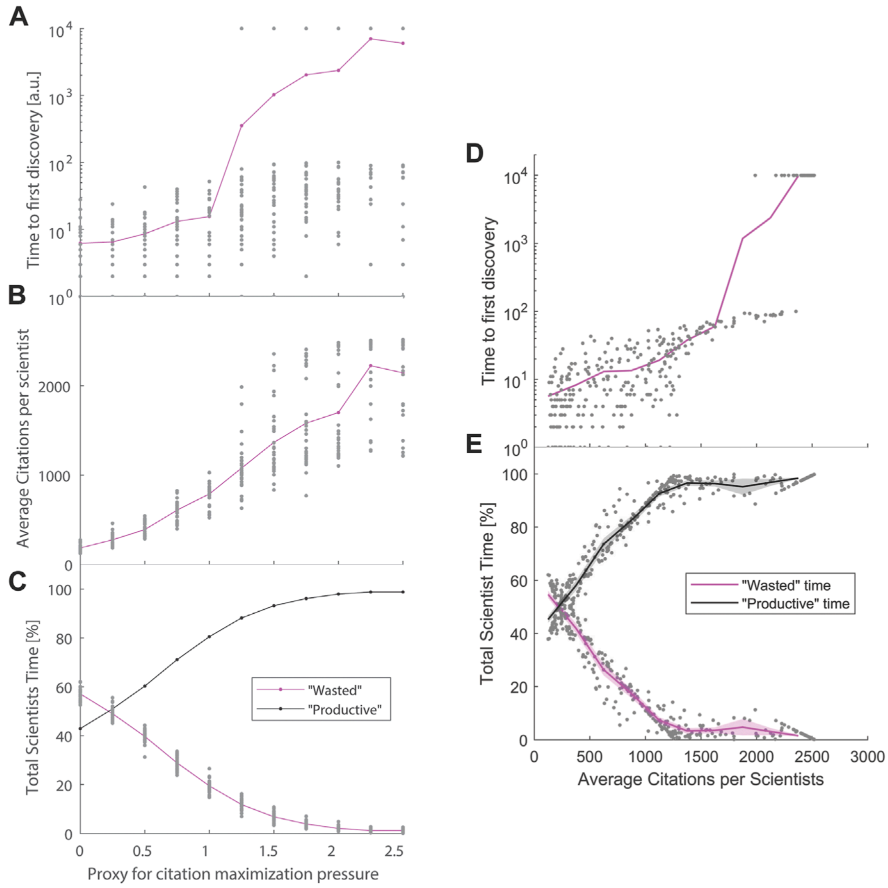

FIGURE 2
Statistics from an ensemble of model runs for different citation pressures. **(A)** The time to the first discovery increases with increasing citation pressure. **(B)** Average citation per scientist tracks the citation pressure parameter in the toy model. **(C)** Fraction of total "productive" time or share of total time spent gathering citations increases with increasing citation pressure. **(D)** Time to first discovery is longer with increasing average citations per scientist. **(E)** Fraction of total "productive" time increases with average citations per scientist.

when science is most productive, i.e., when breakthroughs are being made rapidly, a large share of scientists' time might appear "wasted" as they are wandering outside knowledge circles.

4 The wanderer and the climber

All scientists in the toy model, and perhaps in the modern academic world, spend time wandering and climbing. As wanderers, scientists venture outside knowledge circles, but as climbers, scientists spend time within knowledge circles publishing papers that gather citations. Consider the example of Isaac Newton, who made revolutionary scientific discoveries. Though it is common to ascribe his genius to certain individual characteristics and less to chance, we must consider the possibility that his genius might be attributed to his interest and commitment to wandering (Grassmann, 2022). In fact, a larger fraction of senior physicists are likely to think that "you need to have an innate gift or talent" to "succeed in physics" than students or post-docs (See **Supplementary Material** for details of a survey by Leslie et al. (2015)). Newton spent considerable effort exploring *alchemy* (Schettino, 2017), and the *biblical apocalypse* (Snobelen, 2016). Are we to consider these endeavors to be outside his genius? Or perhaps, was Newton doing what he was good at doing - wandering? Newton most certainly also spent time climbing, aided by his successful discoveries. I choose the word 'climb' here to illuminate other pressures that mimic the citation pressure that biases scientists to be within knowledge circles.

4.1 The ladder

In modern societies, social power hierarchies are pervasive. In one's workplace, for example, one might find "higher ups" who receive more pay, who has more say in decision-making, and perhaps even decides whether you keep your job or not. Rules are in place that grants more power to some people over others. In academia, similar power inequalities exist without question. The review by Fortunato et al. (2018) also mentions the phenomena: "Science often behaves like an economic system with a one-dimensional "currency" of citation counts. This creates a hierarchical system, in which the "rich-get-richer" dynamics suppress the spread of new ideas, particularly those from junior scientists and those who do not fit within the paradigms supported by specific fields." Academic ranks and titles like Dr. or Prof. announce such power differentials. The higher up you are in this power ladder, the more pay you are likely to receive and the more influence you have over the direction of science and even the direction of scientists below the ladder. In such systems, there is a clear incentive to climb the ladder. So how does a scientist climb the ladder fairly? Via widely accepted mechanisms of competition.

4.2 The climbing

The capitalist world widely considers competition necessary because we are told it is for the greater good, i.e., good of the society and the system at large. The promise is that the competition will select *better* scientists. And the better scientists are rewarded with a promotion up the ladder: from a graduate student to a postdoc, to assistant professor, to professor, to Director, and so on. What are the mechanisms of competition? This is, of course, complex, but the primary metric considered during competition within academia is the number of published papers and the number of citations gathered. Other metrics include successful grants, prestigious awards, your network of collaborators, institutional affiliations, your reputation, and the reputation of your mentors. I'd argue that the pressure to maximize the above parameters is roughly aligned with the direction of citation pressure in the toy model, i.e. towards the center of the knowledge circle. For example, one is more likely to develop a network of reputed collaborators or be affiliated to a reputed institution, if one is within established areas of research that garner lot of interest from the scientific community.

Whatever one might think of the utility of competition within science, we might all agree that competition is good, at least in selecting people who are good at the competition. And as I discussed previously, the best way to climb the ladder is to win competitions, so competitions incentivize scientists to spend their time climbing. If climbing is incentivized, then wandering is disincentivized. Wanderers may quickly find themselves at a disadvantage depending on the degree of competitiveness in the system.

If the insight from the toy model is true, then it would imply that it is better for science and scientific discovery for more scientists to spend their time wandering than climbing. But if competition forces and incentivizes scientists to climb rather than wander, we are compelled to ask the question: is competition good for science, or is it just good at producing competitors?

One interesting thing to note here is that no matter the citation pressure, only a maximum of two scientists will discover new knowledge circles in the toy model. The overwhelming majority of wanderers will not chance upon a revolutionary discovery. However, with more wandering scientists, the rate of discovery increases for science as a whole.

5 Conclusion

Science is a collective enterprise. If not, it ought to be. We will all be better off for it. Competition and the pressures it places on scientists beyond a threshold may be detrimental to scientific progress. The more pressure there is in our academic system to maximize citations, the more we might dampen the pace of scientific discovery and breakthroughs. What appears "productive" for individual scientists may not necessarily be productive for science.

To foster discovery, we ought to move towards a more egalitarian social system where free exploration is not discouraged by the rules of interactions that make up the system. The reader might expect me to suggest ways to achieve this, e.g., that we should consider better ways to quantify (or even not quantify) the degree of novelty of a given research or set up a grant solicitation that randomly distributes awards to high-risk proposals. But I refrain from making such recommendations, as they may not lead to what we intend. Before one examines how we can modify the rules to foster scientific discovery and free scientists to explore the unknown, we should investigate the systems-level effects of the rules we set for conducting science in greater detail.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

Author contributions

NS conceptualized, and wrote this manuscript.

Funding

NASA Cooperative Agreement 80NSSC21M0180G: Partnership for Heliophysics and Space Environment Research.

Acknowledgments

This paper was inspired by an online lecture given by Professor David Graeber. I acknowledge useful discussions with David Sibeck, Varsha Subramanyan, Lauri Holappa, Pranay Patil, Dogacan Su Ozturk, Gonzalo Cucho-Padin, Poorya Hosseini, Dolon Bhattacharyya, Maria-Theresia Walach, Gang Kai Poh, Katrina Bossert, and Nitin Prasad.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fspas.2023.1071809/full#supplementary-material

References

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., et al. (2018). Science of science. *Science* 359, eaao0185. doi:10.1126/science.aao0185

Gambardella, B. V. A., Giuri, P., and Mariani, M. (2005). *The value of european patents: Evidence from a survey of european inventors*.

Graeber, D. (2012). Of flying cars and the declining rate of profit. *Baffler* 19, 66–84. doi:10.1162/bflr_a_00003

Grassmann, S. (2022). The scientific world is far too obsessed with "genius". *Jacobin*.

Humberstone, T. L. (1943). Scientific discoveries by accident. *Nature* 151, 215–216. doi:10.1038/151215A0

Jaynes, E. T. (1967). *Foundations of probability theory and statistical mechanics*, 77–101. doi:10.1007/978-3-642-86102-4_6

Leslie, S. J., Cimpian, A., Meyer, M., and Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347, 262–265. doi:10.1126/science.1261375

Schettino, V. (2017). Isaac Newton and alchemy. *Substantia* 1, 69–76. doi:10.13128/SUBSTANTIA-12

Snobelen, S. D. (2016). A time and times and the dividing of time": Isaac Newton, the apocalypse, and 2060 a.d. *Can. J. Hist.* 38 (3), 537–552. doi:10.3138/CJH.38.3.537

frontiers | Frontiers in Astronomy and Space Sciences

TYPE Perspective
PUBLISHED 12 April 2023
DOI 10.3389/fspas.2023.1020921

Check for updates

OPEN ACCESS

EDITED BY
Joseph E. Borovsky,
Space Science Institute, United States

REVIEWED BY
Merav Opher,
Boston University, United States

*CORRESPONDENCE
Eberhard Möbius,
eberhard.moebius@unh.edu

SPECIALTY SECTION
This article was submitted to Space Physics, a section of the journal Frontiers in Astronomy and Space Sciences

RECEIVED 16 August 2022
ACCEPTED 23 March 2023
PUBLISHED 12 April 2023

CITATION
Möbius E (2023), How catching the interstellar wind in the inner solar system led the way on a road to interdisciplinary research between heliophysics and astrophysics.
Front. Astron. Space Sci. 10:1020921.
doi: 10.3389/fspas.2023.1020921

How catching the interstellar wind in the inner solar system led the way on a road to interdisciplinary research between heliophysics and astrophysics

Eberhard Möbius*

Space Science Center, University of New Hampshire, Durham, NH, United States

Combined *in situ* observations of the interstellar wind through the solar system and of its pickup ions (PUIs), implanted after ionization in the solar wind, explain, in comparison with interstellar absorption lines of nearby stars, that the Sun is in an interaction region of the two nearest interstellar clouds. This new finding disrupts the long-held understanding that we are inside the local interstellar cloud (LIC). We discuss how space physics evolved toward such interdisciplinary studies between heliophysics and astrophysics. In 1984, the discovery of interstellar He^+ PUIs exposed the very local interstellar medium to *in situ* diagnostics at 1AU. These PUIs provide the interstellar gas composition and form a stepping stone for the acceleration of ions, especially into anomalous cosmic rays. Using the Sun as a gravitational spectrograph, direct imaging of the neutral interstellar wind, first for He and then for H, O, and Ne, provides the interstellar gas velocity vector and temperature at the heliopause. Combining the interstellar gas flow vectors, those of secondary neutral He and O, and the interstellar magnetic field direction deduced from the interstellar H deflection and termination shock anisotropy seen by the Voyagers provides synergistically the heliosphere's shape, its interaction with the interstellar medium, and constrains our radiation environment. This ISMF organizes the bright Ribbon seen in all-sky images of energetic neutral atoms with the potential to provide its precision determination. The elemental and isotopic composition from PUI and neutral gas observations constrains the galactic evolution and Big Bang cosmology, opening additional interdisciplinary opportunities.

KEYWORDS
pickup ions, interstellar gas flow, interstellar magnetic field, interstellar gas composition, energetic neutral atoms, heliosphere boundary

1 Introduction

For a long time, astronomers have located the Sun inside the local interstellar cloud (LIC) (Bertin et al., 1993; Lallement et al., 1995; Redfield and Linsky, 2008; Frisch et al., 2009), albeit close to its edge. However, a recent study places the Sun in a mixing region between the LIC and the G-Cloud (Swaczyna et al., 2022a). It compares *in situ* measurements of the interstellar neutral (ISN) gas flow vector through the solar system, based on precision analysis of ISN imaging (Wood et al., 2015; Swaczyna et al., 2022b) and pickup ion (PUI) observations (Taut et al., 2018; Bower et al., 2019), with those obtained from stellar absorption lines for nearby interstellar clouds.

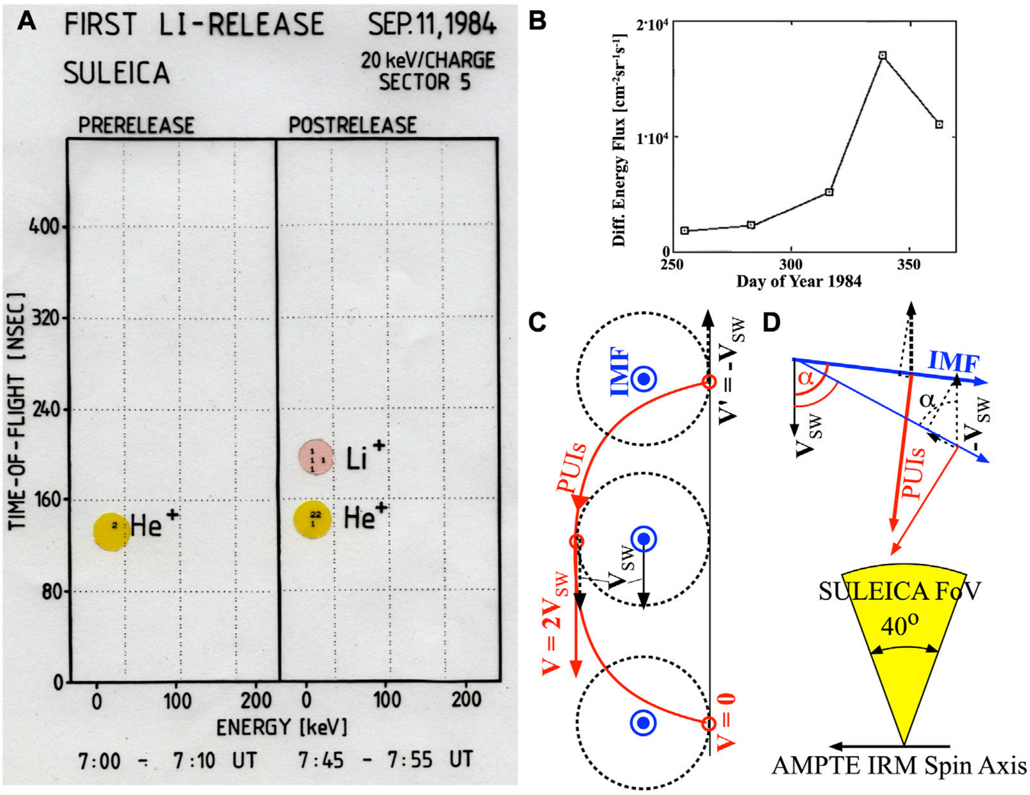

FIGURE 1
(A) Historic TOF versus energy representation of individual He^+ and Li^+ PUIs detected before and after the first AMPTE Li release. **(B)** He^+ PUI flux obtained with AMPTE SULEICA as a function of day of the year in 1984. The flux reaches a maximum in early December when the Earth is downwind of the Sun relative to the ISN gas flow through the solar system [adapted from (Möbius et al., 1985b)], reproduced with permission from Springer Nature. AMPTE IRM was in the SW only from September through January, hence the clipped coverage of PUIs. **(C)** PUI motion in the plane perpendicular to the IMF (the blue circular symbol), gyrating about the IMF in the SW frame (the dashed black line), and moving on a cycloid in the spacecraft frame (the solid red line). After ionization from neutral gas at rest, a PUI (the open red circle) starts with $V = 0$ in the rest frame, while it is injected into the SW with $V' = -V_{SW}$ in the SW frame. Therefore, the PUI gyrates about the IMF with the speed V_{SW} (the dashed black circle). In the rest frame, this gyration is seen as a cycloid trajectory, which starts with $V = 0$, reaches a maximum speed of $2V_{SW}$ or cut-off at the top, where the PUI speed in the SW frame adds to the SW speed, and completes one turn again with $V = 0$. This motion is similar to that of a valve of a bicycle wheel, as seen by an observer at rest. It distinguishes PUIs from the motion and energy distribution of SW ions. Scattering of PUIs in their pitch angle α (see 1D) at fluctuations in the IMF bring PUIs to the cut-off for all IMF angles. Continuous PUI production from the Sun to the observer and adiabatic cooling in the expanding SW fills the velocity space between $2V_{sw}$ and the SW itself. An introduction to PUIs may be found in a recent text book (Hsieh and Möbius, 2022) and review (Zirnstein et al., 2022) and presents a full overview on PUIs in the heliosphere. **(D)** PUI motion in the SW-IMF plane, along with the AMPTE SULEICA FoV. If the IMF is at an angle $\alpha < 90°$ relative to the SW, PUIs are injected into the SW at pitch angle α, resulting in a velocity component parallel to the IMF, which leads to an overall transport of PUIs in the rest frame at angle α relative to the IMF. Shown are sample directions of the PUI motion for IMF orientations that, after initial pickup, lead the ions into the FoV (solid line, $\alpha > 70°$) or miss the FoV (dashed line, $\alpha < 70°$).

PUI observations on New Horizons that the H density in our neighborhood is twice as high (Swaczyna et al., 2020) than that derived from absorption lines in the LIC and G-Cloud (Redfield and Linsky, 2008; Linsky et al., 2022) support this finding. To my knowledge, this is the first time that space physics-based *in situ* observations have diagnosed the interaction of two interstellar clouds, placing the solar system exactly where the action is. Thus, the time has come to conduct genuine interdisciplinary studies between space and astrophysics, and we will map the road that led us here.

2 Discovery of interstellar pickup ions

The journey started on 11 September 1984 at the German Spacecraft Operations Center in Oberpfaffenhofen. After waiting in suspense, the solar wind (SW) conditions appeared to be on target for the first ion cloud release from the Active Magnetospheric Particle Tracer Explorers (AMPTE) Ion Release Module (IRM) (Haerendel et al., 1985), with the interplanetary magnetic field (IMF) almost perpendicular to the SW. After a quick readiness check, the IRM ejected two Li and CuO canisters. They drifted for 10 mins before the mixture ignited and generated an expanding cloud, which silenced the SW for a few seconds like within a comet coma (Häusler et al., 1986). The nearly perpendicular IMF was supposed to pick up the freshly generated Li^+ ions and propel them toward the Earth's magnetosphere. Capable of identifying these ions, the time-of-flight (TOF) spectrograph AMPTE SULEICA (Möbius et al., 1985a) indeed found only seven Li^+ ions before the IMF turned any ions out of its field of view (FOV) (Möbius et al., 1986). However, this disappointing result came with a stunning surprise.

He^+ ions showed up with the Li^+ ions repeatedly. The IMF must have picked them up in the SW, freshly ionized from a gas almost at rest like Li, as shown in Figure 1A, hence called PUIs.

Their fluxes exceeded those for He from the Earth's exosphere at IRM's distance by three orders of magnitude. Memories arose from an astronomy seminar as a student at Bochum about background radiation. My assignment focused on Lyman-α background, which, to my disappointment, originated in our backyard, ISN gas illuminated by the Sun (Bertaux and Blamont, 1971). Contrary to earlier astronomical wisdom, the solar system is not (Fahr, 1968) within a Strömgren sphere (Strömgren, 1939). The ISN gas blows through the solar system as an interstellar wind, too fast for ionization before reaching 1 AU. Interstellar He forms a unique pattern in the inner heliosphere, focused downwind of the Sun (Weller and Meier, 1974; Fahr et al., 1976). I thought, "Wouldn't it not be cool if the He^+ ions were of interstellar origin?" Indeed, they exhibited the predicted behavior with a substantial flux enhancement in early December when the Earth is downwind of the flow (Möbius et al., 1985b) (Figure 1B). The literature from a student seminar proved invaluable for identifying a fundamental discovery, PUIs from the interstellar wind, and so did the humble explorer AMPTE (Krimigis et al., 1982), initiated by Tom Krimigis and Gerhard Haerendel.

The He^+ ions were visible continuously up to a cut-off at $2V_{sw}$ or $4E_{sw}$, as previously predicted (Vasyliunas and Siscoe, 1976) (Figure 1C), contrary to locally injected Li^+ PUIs that entered SULEICA only during favorable IMF orientations (Figure 1D). Continuous He^+ PUI injection into the SW affords its effective scattering in pitch angle α at IMF fluctuations and adiabatic cooling in the expanding SW (Isenberg, 1986; Isenberg, 1987; Möbius et al., 1988), filling the observed spectra up to $4E_{sw}$, as explained in a recent text book (Hsieh and Möbius, 2022) and PUI review (Zirnstein et al., 2022).

3 Diagnostic opportunities with pickup ions and their challenges

This first *in situ* diagnostic method for ISN gas expanded the horizon of space plasma physics into the Sun's galactic neighborhood. It allowed the sampling of interstellar He, H (Gloeckler et al., 1993), N, O, and Ne (Gloeckler and Geiss, 1998), the calculation of their abundance ratios (Gloeckler and Geiss, 2001), and an estimate of the ISN flow velocity and temperature (Möbius et al., 1995).

However, PUIs presented formidable challenges in determining the dynamic ISN parameters precisely. Slower than anticipated pitch-angle scattering made the PUI distributions asymmetric in the SW frame, softened the otherwise sharp cut-off at $2V_{sw}$, and lowered the most easily accessible part of the PUI fluxes above the SW energy (Gloeckler et al., 1995; Möbius et al., 1998). PUI distributions and fluxes varied substantially in response to SW structures, such as stream interaction regions (SIRs) and coronal mass ejections (CMEs) (Möbius et al., 2010). Non-radial PUI transport in the SW affects the shape and location of the focusing cone (Möbius et al., 1996; Chalov and Fahr, 2006; Quinn et al., 2016). Effective acceleration into a suprathermal tail smoothens the PUI cut-off further (Gloeckler et al., 2000; Möbius et al., 2019).

However, one person's trash may be another's treasure. When interstellar PUIs came up as a tentative explanation for the He^+ by AMPTE IRM, Dieter Hovestadt exclaimed "SULEICA has detected the seed particles for the anomalous cosmic rays (ACR)" (Garcia-Munoz et al., 1973; Hovestadt et al., 1973; Klecker, 1995; Jokipii, 1998). ACRs are substantially overabundant in O and Ne over galactic cosmic rays. Dieter pointed to a model that implicated the ISN gas (Fisk et al., 1974) and opened another essential connection for PUIs. They form a particle distribution that enables preferential injection into acceleration to higher energies. The enormous injection efficiency compared to underlying bulk plasma was visible in SIRs (Gloeckler et al., 1994; Morris et al., 2001; Möbius et al., 2002) and CMEs (Kucharek et al., 2003), with remarkable He^+ overabundances over SW He^{2+} in the respective energetic particles. Identifying He^+ PUIs as the source solved a previous mystery: He^+/He^{2+} ratios in energetic interplanetary particles that substantially exceeded the SW ratio (Hovestadt et al., 1984).

Another aspect of the PUI distribution took a surprising turn much later. During a seminar on the PUI discovery, Martin Lee mused whether Alfvén waves that lag behind the SW could reduce the He^+ PUI cut-off energy by ≈ 10–15% from the nominal $4E_{sw}$ as the observations seemed to suggest. However, this conclusion would have been a stretch with SULEICA's 10% energy width, data points spaced by 20% in energy, and integration over a 40 × 45° FOV. Only 15 years later, when analyzing PUI data from CELIAS (Hovestadt et al., 1995) on the SOlar and Heliospheric Observatory (SOHO), centered around June, or upwind of the Sun relative to the ISN flow, the original question entered daylight again but with a twist. The cut-off was ≈15% above the nominal value. SOHO observing upwind and AMPTE IRM downwind suggested an explanation. Because in the SW frame, PUI injection occurs with the vector sum of its local ISN flow and SW velocities, and the cut-off was at a noticeably higher PUI speed on the upwind side than the downwind side (Möbius et al., 1999).

The advent of PLASTIC (Galvin et al., 2008) on the Solar Terrestrial RElations Observatory (STEREO), with superior energy and angle resolution in the PUI regime, enabled a more detailed study of the PUI evolution after their initial injection (Drews et al., 2015). It also turned this earlier discovery into a precision tool to obtain at least the ISN flow direction with much higher fidelity. When increasing ionization losses of the ISN flow from the upwind to the downwind side could not produce the He^+ PUI crescent (Sokół et al., 2016), as proposed earlier (Drews et al., 2012), flux modulation due to the shifting PUI cut-off within the fixed energy window in the analysis became the focal point. This explanation suggested using symmetry in the PUI cut-off shift in the flow axis to determine the ISN flow longitude precisely (Möbius et al., 2015a; Taut et al., 2018; Bower et al., 2019). This measurement has now become a linchpin in obtaining the complete set of dynamic ISN gas parameters in the very local interstellar medium (VLISM) just outside the solar system from local observations the inner heliosphere, complementary to the direct ISN flow observations at 1 AU (Möbius et al., 2009a) with the Interstellar Boundary Explorer (IBEX) (McComas et al., 2009a). These neutral gas measurements became possible after the pioneering observations of the He ISN flow with Ulysses GAS (Witte et al., 1996).

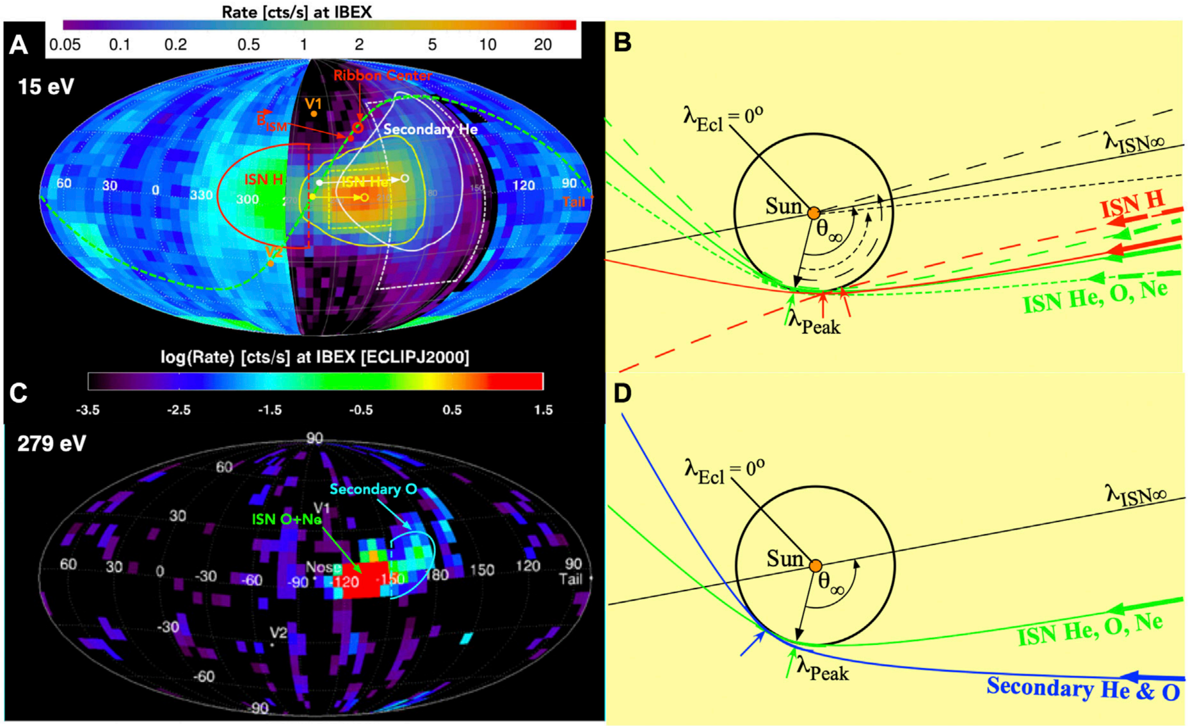

FIGURE 2
Left: Combination of IBEX ENA maps in Mollweide projection that illustrate the interstellar medium information collected at 1 AU. Right: Schematic view of the ISN and ENA trajectories from their source to the observer at 1 AU. **(A)** ISN He and H flow based on H count rates at 15 eV, along with secondary He neutrals [adapted from (Park et al., 2016; Swaczyna et al., 2018)], reproduced with permission from AAS. **(B)** Schematic representation of ISN He, O, and Ne (green), as well as H (red) trajectories in the plane that contains $\vec{V}_{ISN\infty}$, the Sun, and the in-ecliptic location λ_{Peak} where the bulk ISN flow is observed at its perihelion. The ISN flow arrives from $\lambda_{ISN\infty}$ at the heliopause, whose value is connected through the angle swept out by the arriving atoms from infinity to perihelion or true anomaly θ_{∞} to λ_{Peak} and $V_{ISN\infty}$ in Eq. 1; Eq. 1 describes unbound Keplerian trajectories. It describes $V_{ISN\infty}$ as a function of $\lambda_{ISN\infty}$ (samples shown as dashed green lines) for the observed value of λ_{Peak}. An extended range of observer locations for the angular distribution of the ISN flow around the peak constrains $\lambda_{ISN\infty}$ and $V_{ISN\infty}$ separately. For ISN H, λ_{Peak} shifts to larger ecliptic longitude due to the partial compensation of gravitation by solar radiation pressure, increasingly during high solar activity (dashed red) (Rahmanifard et al., 2019). **(C)** ISN O and Ne flow based on O count rates at 279 eV, along with secondary O neutrals [adapted from (Park et al., 2015)], reproduced with permission from AAS. **(D)** Same representation as above for the ISN flow (green) and secondary neutral distributions (blue), whose arrival directions at the heliopause and λ_{Peak} at 1AU are shifted to a smaller longitude. In the maps, the ISN He, O, and Ne flow arrives from the same direction in the sky, deflected westward from the arrival direction outside the heliopause (shown as a yellow dot in panel **A)** due to the Sun's gravitation. In contrast, the ISN H flow is deflected eastward due to radiation pressure. The arrival directions of the ISN flow (yellow dot) and secondary He (white dot), the Ribbon center (open red circle), and the derived $\vec{B}_{ISM}$ direction line up along the $\vec{B}_{ISM}-\vec{V}_{ISN}$ plane (the dashed green line).

4 Catching the neutral interstellar wind directly

The discovery of the interstellar PUIs triggered the invitation by Hans Fahr in 1986 to a series of workshops focused on the interaction between the heliosphere and the surrounding interstellar medium that involved German, Polish, and Soviet groups. The workshops revolved around observational and modeling efforts to understand our galactic neighborhood. During one meeting, Helmut Rosenbauer jokingly regretted that the PUI observations would steal the thunder of Ulysses GAS (Witte et al., 1992), whose launch was still in the future. However, this tiny sensor that measured the ISN He distribution *via* sputtering Li^+ ions off a LiF surface obtained images of the He ISN flow during the transit of Ulysses to Jupiter and during the three fast latitude scans when Ulysses scoped out the 3D structure of the SW and energetic particles. Following the He atoms along their hyperbolic trajectories in the Sun's gravitational field (Fahr, 1974; Wu and Judge, 1979; Lee et al., 2012; Lee et al., 2015) with a tailored fitting technique (Banaszkiewicz et al., 1996) translates the neutral He images into the velocity distribution function outside the heliosphere. The GAS observations of the He ISN flow enabled the most detailed and accurate determination of the ISN flow vector and temperature (Witte, 2004). These values were validated and placed into context with PUI and solar ultraviolet backscattering observations of ISN He within a scientific team at the International Space Science Institute (ISSI). The team effort consolidated the ISN gas parameters (Möbius et al., 2004) and reemphasized the complementary nature of the three *in situ* observation techniques, each affected by different systematic uncertainties.

This collaborative work on the physical state of the interstellar medium dovetailed into the proposal of a potential explorer mission to study the VLISM and its interaction with the heliosphere, the Interstellar Pathfinder, which, in two attempts, almost made it into a

Phase B study, but only "almost". The studies proposed sounded like the perfect interdisciplinary endeavor that could engage two scientific communities, heliophysics and astrophysics, and thus garner multiple support. Instead, the advice was to root the proposal firmly in heliophysics; otherwise, it may fall between all chairs. In a third attempt, the concentration on two neutral atom cameras with the capability to image the boundary of the heliosphere in energetic neutral atoms (ENA) and to simultaneously capture the interstellar wind of He, O, and possibly H under the constraints of a Small Explorer kicked the proposal above the threshold and led to the successful IBEX mission (McComas et al., 2009a).

The combination of mechanical collimation, surface conversion of neutral atoms into negative ions, electrostatic energy analysis, post-acceleration, and a triple time-of-flight measurement in IBEX-Lo (Fuselier et al., 2009) enabled the observation of the He, H, and O ISN flow (Möbius et al., 2009b) and even Ne (Bochsler et al., 2012) (Figures 2A,C). Thus, IBEX went substantially beyond the GAS capabilities, increasing the signal-to-background ratio by orders of magnitude for He and expanding to other species. However, with its observations limited severely in ecliptic longitude to less than 2 months in early spring when the IBEX FOV points to the oncoming flow, approximately parallel to the Earth's orbit, the observations and analysis are subject to degeneracy in the ISN parameter space. It is obvious for an idealized ISN trajectory that passes IBEX precisely perpendicular to the IBEX-Sun line, i.e., reaching its perihelion at the point of observation (Möbius et al., 2012) (Figures 2B,D). Trajectories that start at infinity over a wide range of ISN speeds $V_{ISN\infty}$ and inflow longitudes $\lambda_{ISN\infty}$ fulfill this condition when coupled with the hyperbolic trajectory equation:

$$cos\,(\lambda_{ISN\infty} + 180^o - \lambda_{Peak}) = cos\,\theta_\infty = \frac{-1}{1 + \frac{r_E V^2_{ISN\,\infty}}{GM_S}}. \quad (1)$$

θ_∞ is the angle swept out by the position vector of the atom from infinity to perihelion or the true anomaly of the trajectory; λ_{Peak} is the ecliptic longitude of the observer when seeing the peak ISN flow; M_s is the Sun's mass; G is the gravitational constant; and $r_E = 1$ AU (Lee et al., 2012). Because the inflow latitude $\beta_{ISN\infty}$ and temperature $T_{ISN\infty}$ connect these quantities dynamically, the analysis of IBEX observations leads to a four-dimensional tube in the parameter space (McComas et al., 2012). Observations of the full He distribution over a range in ecliptic longitude constrain the tube in length (Schwadron et al., 2015a; Möbius et al., 2015b; Swaczyna et al., 2015; Swaczyna et al., 2018), and the PUI analysis previously discussed provides a complementary value for λ_{ISN} (Möbius et al., 2015a; Taut et al., 2018; Bower et al., 2019).

5 Synergism between PUI, ISN flow, and ENA observations for the heliosphere and beyond

When IBEX-Lo caught the interstellar wind, IBEX-Hi (Funsten et al., 2009) and Lo (Fuselier et al., 2009) combined took the first all-sky images of the heliospheric boundary in the light of ENAs, thus expanding the emerging field of neutral-atom astronomy to its horizon (Hsieh and Möbius, 2022). At low ENA energies, the IBEX maps reveal secondary He (Kubiak et al., 2014; Kubiak et al., 2016) (Figure 2A) and O (Park et al., 2015; Park et al., 2016) (Figure 2C) interstellar neutral flows, which originate in the region outside the heliopause from the charge exchange of interstellar ions with the ISN gas flow. The secondary neutral flow appears slower and hotter than the pristine ISN flow, hence also referred to as the "warm breeze" (Kubiak et al., 2014). This is because the secondary neutrals mimic the distribution of the interstellar plasma, which slows down and heats in response to the presence of the heliosphere. As the secondary neutral signal is at a level of a few percent of the pristine ISN flow for He and O, its prior analysis is necessary before the velocity distribution of secondary neutrals can be extracted from the observations. Both secondary populations appear substantially deflected relative to the pristine ISN flow in the same direction as ISN H (Lallement et al., 2005) in a plane, dubbed the H-deflection plane, which, based on global heliospheric simulations, contains the interstellar flow velocity $\vec{V}_{ISN}$ and the interstellar magnetic field (ISMF) $\vec{B}_{ISM}$ (Izmodenov et al., 2005), and thus may be termed the $\vec{B}_{ISM} - \vec{V}_{ISN}$ plane. It should be noted that the H ISN flow direction represents a combination of primary H ISN and secondary H and thus shows a deflection (Izmodenov et al., 2005; Lallement et al., 2005) between the primary He (Schwadron et al., 2015a) and O (Schwadron et al., 2016) ISN and the secondary He and O directions (Kubiak et al., 2016; Park et al., 2019). The $\vec{B}_{ISM}$ orientation deduced from the arrangement of the multiple flow directions in the sky was consistent with the heliospheric asymmetry derived from the Voyager 1 and 2 termination shock traversals (Opher et al., 2006; Stone et al., 2008).

Interestingly, the first IBEX ENA sky maps revealed a bright unanticipated Ribbon (McComas et al., 2009b) that traces out a circle in the sky, which conforms with $\vec{B}_{ISM} \cdot \vec{r} = 0$ (Schwadron et al., 2009). $\vec{r}$ indicates the look direction, and $\vec{B}_{ISM}$ is consistent with the orientation found in global heliospheric models constrained by the aforementioned observations (Izmodenov et al., 2005; Lallement et al., 2005; Pogorelov et al., 2009a; Opher et al., 2009). However, the physical processes that conspire to form the Ribbon are less clear and still under debate at this writing. More than a dozen models, involving different source locations and mechanisms, emerged as summarized in an early review (McComas et al., 2014). The frontrunner, which explains most of the observed Ribbon features, appears to be a model that starts with neutral SW entering the VLISM. Next, charge exchange with interstellar H^+ generates PUIs, and those injected into the ISMF at ≈90° are temporarily stored in a ring distribution. This PUI population produces Ribbon ENAs in another charge exchange with ISN H (Heerikhuisen et al., 2010; McComas et al., 2014). A significant challenge for this model is the susceptibility of a PUI ring concentrated at ≈90° to $\vec{B}_{ISM}$ to instabilities (Gary et al., 1986), which would render the intermediate storage time too short for an effective Ribbon production (Florinski et al., 2010). Several different approaches promise to mitigate this challenge, but their coverage goes beyond the scope of this paper, and the interested reader is referred to Hsieh and Möbius (2022). Simulations based on this Ribbon model constrain the orientation of $\vec{B}_{ISM}$ (Zirnstein et al., 2016) consistent with a field topology around the heliosphere that describes correctly (Schwadron et al., 2015b) the observed TeV cosmic ray anisotropies (Abdo et al., 2008; Abdo et al., 2009; Abbasi et al., 2011). It also agrees with field directions obtained from starlight polarization in our extended neighborhood (Frisch et al., 2022). Remarkably, the Ribbon model also constrains the magnetic field strength (Zirnstein et al., 2016) to values that agree with an earlier determination based on the Voyager TS crossings (Gloeckler et al., 1997).

However, the magnetic field direction measured by Voyager 1 and 2 outside the heliopause, starting in line with the heliospheric field, and the Ribbon-derived direction are still far from each other. After first seemingly approaching the IBEX direction (Schwadron et al., 2015c), it turned back again in response to heliospheric disturbances (Schwadron et al., 2018). Also, the ENA belt, discovered by Cassini INCA at higher energies (Krimigis et al., 2009), appears oriented differently in the sky than the Ribbon, and the heliosheath thickness and heliosphere shapes differ as derived from the ENAs in different energy regimes (Dialynas et al., 2017; Schwadron and Bzowski, 2018). Whether and how the remote and *in situ* observations of $\vec{B}_{ISM}$ will converge, or the Ribbon and Belt are one or separate phenomena, remain open questions. Also, the debates on the heliospheric topology (Pogorelov et al., 2015; Opher et al., 2020) and the precise extrapolation to the undisturbed $\vec{B}_{ISM}$ (Zirnstein et al., 2016; Izmodenov and Alexashov, 2020) continue.

Yet, combining the O, He, and H ISN flow arrival directions at the heliopause with the He and O secondary neutrals and the ISMF direction arranges them along an arc that represents the plane, which organizes the deflection of the interstellar plasma flow around the heliosphere (Pogorelov et al., 2009b; Izmodenov et al., 2009). Thus, together with $\vec{B}_{ISM}$ and $\vec{V}_{ISN}$, the secondary neutral flow observations effectively constrain the shape of the heliopause and the plasma flow around it in global heliosphere models, while the TS and heliopause crossings of the two Voyager spacecraft (Stone et al., 2008; Krimigis et al., 2013; Stone et al., 2013; Burlaga and Ness, 2014) provide a linchpin on the absolute size of the heliosphere.

6 Foray into the VLISM, a truly interdisciplinary endeavor between space plasma and astrophysics

With the fidelity of the ISN flow and secondary neutral observations and analysis achieved to date, catching the interstellar wind in the inner heliosphere now connects to several aspects of astrophysics in our galactic neighborhood. For example, a detailed analysis of the secondary He flow provides the He^+ density, and thus, the ratio of ionized and neutral He and H in the VLISM (Bzowski et al., 2019), which, in turn, constrains the radiation environment in the Sun's neighborhood (Slavin and Frisch, 2008). Improvements in the precision of the locally obtained interstellar flow parameters (Swaczyna et al., 2022b) and densities (Swaczyna et al., 2020) now strongly support an earlier suspicion, i.e., that the Sun is neither inside the LIC nor the G-cloud proper (Redfield and Linsky, 2008). Most likely, the solar system traverses an interaction region between these two adjacent interstellar clouds in our immediate galactic neighborhood (Swaczyna et al., 2022a). A recently discovered anisotropy in the ISN He distribution (Wood et al., 2019) may even be a sign of incomplete mixing of the two cloud populations and thus provide insights into the kinetics of the interstellar cloud interaction. Furthermore, obtaining abundances of the ISN species that make it inside the heliosphere through PUI (Gloeckler and Geiss, 2001) and ISN sampling, in particular, for O and Ne (Park et al., 2014), provides a window on the processing of matter in the Milky Way over time (Prantzos et al., 1998). With the interstellar $^3He/^4He$ ratio from PUIs (Gloeckler and Geiss, 1996) and the D/H ratio from ISN observations (Rodriguez-Moreno et al., 2013), we even touch upon cosmology and Big Bang nucleosynthesis (Schramm et al., 1998). In summary, combining space physics-based *in situ* diagnostics and astronomy-based spectroscopy finally enables genuine interdisciplinary research opportunities.

By placing a powerful suite of sensors for ENAs, PUIs, energetic particles, interstellar dust, and Lyman-α radiation at the Lagrangian point L1, while monitoring the interplanetary environment, the Interstellar Mapping and Acceleration Probe (IMAP) (McComas et al., 2018) will bring our understanding of the heliosphere and its place in the interstellar medium to the next level. A future dream of an Interstellar Probe that will venture into the VLISM proper and unravel thus far inaccessible ion populations and related interaction processes has recently moved closer to a realization after a detailed scientific and technical feasibility study (McNutt et al., 2021), (Brandt et al., 2023).

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found at: http://ibex.swri.edu/researchers/publicdata.shtml.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The author is very grateful to the many individuals at various institutions that made the many spacecraft missions successful whose observations contributed substantially to the results compiled in this paper. He also gratefully acknowledges the many fruitful collaborations with wonderful colleagues that triggered new insights in many ways. Support for the work toward this manuscript by the IBEX and IMAP projects and by NASA Grant 80NSSC18K1212 is thankfully acknowledged.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abbasi, R., Abdou, Y., Abu-Zayyad, T., Adams, J., Aguilar, J. A., Ahlers, M., et al. (2011). Observation of anisotropy in the arrival directions of galactic cosmic rays at multiple angular scales with IceCube. *Astrophys. J.* 740, 16.

Abdo, A. A., Allen, B., Aune, T., Berley, D., Blaufuss, E., Casanova, S., et al. (2008). Discovery of localized regions of excess 10-TeV cosmic rays. *Phys. Rev. Lett.* 101, 221101. doi:10.1103/physrevlett.101.221101

Abdo, A. A., Allen, B., Aune, T., Berley, D., Casanova, S., Chen, C., et al. (2009). The large-scale cosmic-ray anisotropy as observed with Milagro. *Astrophys. J.* 698, 2121–2130. doi:10.1088/0004-637x/698/2/2121

Banaszkiewicz, M., Witte, M., and Rosenbauer, H. (1996). Determination of interstellar helium parameters from theULYSSES-NEUTRALGAS experiment: Method of data analysis. *Astron. Astrophys. Suppl.* 120, 587–602. doi:10.1051/aas:1996311

Bertaux, J. L., and Blamont, J. E. (1971). Evidence for a source of an extraterrestrial hydrogen lyman-alpha emission. *Astron. Astrophys.* 11, 200–217.

Bertin, P., Lallement, R., Ferlet, R., and Vidal-Madjar, A. (1993). Detection of the local interstellar cloud from high-resolution spectroscopy of nearby stars: Inferences on the heliospheric interface. *J. Geophys. Res.* 98, 15193–15198. doi:10.1029/93ja01179

Bochsler, P., Petersen, L., Möbius, E., Schwadron, N. A., Wurz, P., Scheer, J. A., et al. (2012). Estimation of the neon/oxygen abundance ratio at the heliospheric termination shock and in the local interstellar medium from IBEX observations. *Astrophys. J. Suppl.* 198, 13. doi:10.1088/0067-0049/198/2/13

Bower, J., Moebius, E., Berger, L., Farrugia, C., Keilbach, D., Lee, M. A., et al. (2019). Effect of rapid changes of solar wind conditions on the pickup ion velocity distribution. *J. Geophys. Res.* 124, 6418–6437. doi:10.1029/2019ja026781

Brandt, P. C., Provornikova, E., Bale, S. D., Cocoros, A., DeMajistre, R., Dialynas, K., et al. (2023). Future exploration of the outer heliosphere and very local interstellar medium by interstellar Probe. *Space Sci. Rev.* 219, 18. doi:10.1007/s11214-022-00943-x

Burlaga, L. F., and Ness, N. F. (2014). Voyager 1 observations of the interstellar magnetic field and the transition from the heliosheath. *Astrophys. J.* 784, 146. doi:10.1088/0004-637x/784/2/146

Bzowski, M., Czechowski, A., Frisch, P. C., Fuselier, S. A., Galli, A., Grygorczuk, J., et al. (2019). Interstellar neutral helium in the heliosphere from IBEX observations. VI. The He^+ density and the ionization state in the Very Local Interstellar Matter. *Astrophys. J.* 882, 60. doi:10.3847/1538-4357/ab3462

Chalov, S. V., and Fahr, H-J. (2006). Pickup interstellar helium ions in the region of the solar gravitational cone. *Astron. Lett.* 32, 487–494. doi:10.1134/s1063773706070061

Dialynas, K., Krimigis, S. M., Mitchell, D. G., Decker, R. B., and Roelof, E. C. (2017). The bubble-like shape of the heliosphere observed by Voyager and Cassini. *Nat. Astron.* 1, 0115. doi:10.1038/s41550-017-0115

Drews, C., Berger, L., Taut, A., Peleikis, T., and Wimmer-Schweingruber, R. F. (2015). 2D He^+pickup ion velocity distribution functions: STEREO PLASTIC observations. *Astron. Astrophys.* 575, A97. doi:10.1051/0004-6361/201425271

Drews, C., Berger, L., Wimmer-Schweingruber, R. F., Bochsler, P., Galvin, A. B., Klecker, B., et al. (2012). Inflow direction of interstellar neutrals deduced from pickup ion measurements at 1 AU. *J. Geophys. Res.* 117, A09106. doi:10.1029/2012ja017746

Fahr, H-J. (1968). Charge-transfer interactions between solar wind protons and neutral particles in the vicinity of the Sun. *Nature* 219, 473–474. doi:10.1038/219473a0

Fahr, H-J., Lay, G., and Blum, P. W. (1976). Planetary cones of focused interstellar gases. *Astron. Astrophys.* 52, 363–372.

Fahr, H-J. (1974). The extraterrestrial UV-background and the nearby interstellar medium. *Space Sci. Rev.* 15, 483–540. doi:10.1007/bf00178217

Fisk, L. A., Kozlovsky, B., and Ramaty, R. (1974). An interpretation of the observed oxygen and nitrogen enhancements in low-energy cosmic rays. *Astrophys. J.* 190, L35. doi:10.1086/181498

Florinski, V., Zank, G. P., Heerikhuisen, J., Hu, Q., and Khazanov, I. (2010). Stability of a pickup ion ring-beam population in the outer heliosheath: Implications for the IBEX Ribbon. *Astrophys. J.* 719, 1097–1103. doi:10.1088/0004-637x/719/2/1097

Frisch, P. C., Bzowski, M., Grün, E., Izmodenov, V., Kruger, H., Linsky, J. L., et al. (2009). The galactic environment of the Sun: Interstellar material inside and outside of the heliosphere. *Space Sci. Rev.* 146, 235–273. doi:10.1007/s11214-009-9502-0

Frisch, P. C., Piirola, V., Berdyugin, A. B., Heiles, C., Cole, A., Hill, K., et al. (2022). Whence the interstellar magnetic field shaping the heliosphere? *Astrophys. J. Suppl.* 259, 48. doi:10.3847/1538-4365/ac5750

Funsten, H. O., Guthrie, A. A., Harper, R. W., Dunn, G., Ellis, S., Everett, D., et al. (2009). The interstellar boundary explorer high energy (IBEX-Hi) neutral atom imager. *Space Sci. Rev.* 146, 75–103. doi:10.1007/s11214-009-9504-y

Fuselier, S. A., Ghielmetti, A. G., Hertzberg, E., Clark, G., Crew, G. B., Dunn, G., et al. (2009). The IBEX-lo sensor. *Space Sci. Rev.* 146, 117–147. doi:10.1007/s11214-009-9495-8

Galvin, A. B., Kistler, L., Popecki, M. A., Farrugia, C. J., Simunac, K. D. C., Ellis, L., et al. (2008). The plasma and suprathermal ion composition (plastic) investigation on the STEREO observatories. *Space Sci. Rev.* 136, 437–486. doi:10.1007/s11214-007-9296-x

Garcia-Munoz, M., Mason, G. M., and Simpson, J. A. (1973). A new test for solar modulation theory: The 1972 may-july low-energy galactic cosmic-ray proton and helium spectra. *Astrophys. J.* 182, L81–L84. doi:10.1086/181224

Gary, S. P., Hinata, S., Madland, C. D., and Winske, D. (1986). The development of shell-like distributions from newborn cometary ions. *Geophys. Res. Lett.* 13, 1364–1367. doi:10.1029/gl013i013p01364

Gloeckler, G., Fisk, L. A., and Geiss, J. (1997). Anomalously small magnetic field in the local interstellar cloud. *Nature* 386, 374–377. doi:10.1038/386374a0

Gloeckler, G., Fisk, L. A., Zurbuchen, T. H., and Schwadron, N. A. (2000). "Sources, injection and acceleration of heliospheric ion populations," in ACE 2000 Symposium (Melville: AIP), 221–228.

Gloeckler, G., and Geiss, J. (1996). Abundance of ^{3}He in the local interstellar cloud. *Nature* 381, 210–212. doi:10.1038/381210a0

Gloeckler, G., Galvin, A. B., Geiss, J., Balsiger, H., Balsiger, H., von Steiger, R., et al. (1993). Detection of interstellar pick-up hydrogen in the solar system. *Science* 261, 70–73. doi:10.1126/science.261.5117.70

Gloeckler, G., and Geiss, J. (2001). Heliospheric and interstellar phenomena deduced from pickup ion observations. *Space Sci. Rev.* 97, 169–181. doi:10.1023/a:1011867320416

Gloeckler, G., and Geiss, J. (1998). Interstellar and inner source pickup ions observed with SWICS on ULYSSES. *Space Sci. Rev.* 86, 127–159. doi:10.1023/a:1005019628054

Gloeckler, G., Geiss, J., Roelof, E. C., Fisk, L. A., Ipavich, F. M., Ogilvie, K. W., et al. (1994). Acceleration of interstellar pickup ions in the disturbed solar wind observed on Ulysses. *J. Geophys. Res.* 99, 17637–17644. doi:10.1029/94ja01509

Gloeckler, G., Schwadron, N. A., Fisk, L. A., and Geiss, J. (1995). Weak pitch angle scattering of few MV rigidity ions from measurements of anisotropies in the distribution function of interstellar pickup H^+. *Geophys. Res. Lett.* 22, 2665–2668. doi:10.1029/95gl02480

Haerendel, G., Bauer, O. H., Ertl, M., Foppl, H., Kaiser, K. h., Lieb, W., et al. (1985). The Li/Ba release experiments of the Ion release Module. *IEEE Trans. Geosci. Remote Sens.* GE-23, 253–258. doi:10.1109/tgrs.1985.289523

Häusler, B., Woolliscroft, L. J., Anderson, R. R., Gurnett, D. A., Holzworth, R. H., Koons, H. C., et al. (1986). Plasma waves observed by the IRM and UKS spacecraft during the AMPTE solar wind lithium releases: Overview. *J. Geophys. Res.* 91, 1283–1300. doi:10.1029/ja091ia02p01283

Heerikhuisen, J., Pogorelov, N. V., Zank, G. P., Crew, G. B., Frisch, P. C., Funsten, H. O., et al. (2010). Pick-up ions in the outer heliosheath: A possible mechanism for the interstellar boundary EXplorer Ribbon. *Astrophys. J.* 708, L126–L130. doi:10.1088/2041-8205/708/2/l126

Hovestadt, D., Gloeckler, G., Klecker, B., Scholer, M., and Ipavich, F. M. (1984). Survey of He^+/He^{2+} abundance ratios in energetic particle events. *Astrophys. J.* 282, L39–L42. doi:10.1086/184300

Hovestadt, D., Hilchenbach, M., Bürgi, A., Klecker, B., Laeverenz, P., Scholer, M., et al. (1995). CELIAS - charge, element and isotope analysis system for SOHO. *Sol. Phys.* 162, 441–481. doi:10.1007/bf00733436

Hovestadt, D., Vollmer, O., Gloeckler, G., and Fan, C. Y. (1973). Differential energy spectra of low-energy (<8.5 MeV per nucleon) heavy cosmic rays during solar quiet times. *Phys. Rev. Lett.* 31, 650–653. doi:10.1103/physrevlett.31.650

Hsieh, K. C., and Möbius, E. (2022). *Neutral-atom astronomy – plasma diagnostics from the aurora to the interstellar medium.* Hackensack, NJ: World Scientific Publishing Co.

Isenberg, P. A. (1987). Evolution of interstellar pickup ions in the solar wind. *J. Geophys. Res.* 92, 1067–1074. doi:10.1029/ja092ia02p01067

Isenberg, P. A. (1986). Interaction of the solar wind with interstellar neutral hydrogen: Three-fluid model. *J. Geophys. Res.* 91, 9965–9972. doi:10.1029/ja091ia09p09965

Izmodenov, V. V., and Alexashov, D. B. (2020). Magnitude and direction of the local interstellar magnetic field inferred from Voyager 1 and 2 interstellar data and global heliospheric model. *Astron. Astrophys.* 633, L12. doi:10.1051/0004-6361/201937058

Izmodenov, V. V., Alexashov, D. B., and Myasnikov, A. V. (2005). Direction of the interstellar H atom inflow in the heliosphere: Role of the interstellar magnetic field. *Astron. Astrophys.* 437, L35–L38. doi:10.1051/0004-6361:200500132

Izmodenov, V. V., Malama, Y. G., Ruderman, M. S., Chalov, S. V., Alexashov, D. B., Katushkina, O. A., et al. (2009). Kinetic-gasdynamic modeling of the heliospheric interface. *Space Sci. Rev.* 146, 329–351. doi:10.1007/s11214-009-9528-3

Jokipii, J. R. (1998). Insights into cosmic-ray acceleration from the study of anomalous cosmic rays. *Space Sci. Rev.* 86, 161–178.

Klecker, B. (1995). The anomalous component of cosmic rays in the 3-D heliosphere. *Space Sci. Rev.* 72, 419–430. doi:10.1007/bf00768815

Krimigis, S . M., Haerendel, G., McEntire, R . W., Paschmann, G ., and Bryant, D. A. (1982). The active magnetospheric particle tracer Explorers (AMPTE) program. *Eos Trans. AGU* 63, 843. doi:10.1029/eo063i045p00843

Krimigis, S. M., Decker, R. B., Roelof, E. C., Hill, M. E., Armstrong, T. P., Gloeckler, G., et al. (2013). Search for the exit: Voyager 1 at heliosphere's border with the galaxy. *Science* 341, 144–147. doi:10.1126/science.1235721

Krimigis, S. M., Mitchell, D. G., Roelof, E. C., Hsieh, K. C., and McComas, D. J. (2009). Imaging the interaction of the heliosphere with the interstellar medium from Saturn with Cassini. *Science* 326, 971–973. doi:10.1126/science.1181079

Kubiak, M. A., Bzowski, M., Sokòl, J. B., Swaczyna, P., Grzedzielski, S., Alexashov, D. B., et al. (2014). Warm breeze from the starboard bow: A new population of neutral helium in the heliosphere. *Astrophys. J. Suppl.* 213, 29. doi:10.1088/0067-0049/213/2/29

Kubiak, M. A., Swaczyna, P., Bzowski, M., Sokol, J. M., Fuselier, S. A., Galli, A., et al. (2016). Interstellar neutral helium in the heliosphere from IBEX observations. IV. Flow vector, Mach number, and abundance of the Warm Breeze. *Astrophys. J. Suppl.* 223, 25. doi:10.3847/0067-0049/223/2/25

Kucharek, H., Möbius, E., and Li, W. (2003). On the source and acceleration of energetic He+: A long-term observation with ACE/SEPICA. *J. Geophys. Res.* 108, 8040. doi:10.1029/2003ja009938

Lallement, R., Ferlet, R., Lagrange, A. M., Lemoine, M., and Vidal-Madjar, A. (1995). Local cloud structure from HST-GHRS. *Astron. Astrophys.* 304, 461–474.

Lallement, R., Quemerais, E., Bertaux, J. L., Ferron, S., Koutroumpa, D., and Pellinen, R. (2005). Deflection of the interstellar neutral hydrogen flow across the heliospheric interface. *Science* 307, 1447–1449. doi:10.1126/science.1107953

Lee, M. A., Kucharek, H., Möbius, E., Wu, X., Bzowski, M., and McComas, D. J. (2012). AN analytical MODEL of interstellar gas in the heliosphere tailored to *interstellar boundary explorer* observations. *Astrophys. J. Suppl.* 198, 10. doi:10.1088/0067-0049/198/2/10

Lee, M. A., Möbius, E., and Leonard, T. (2015). The analytical structure of the primary interstellar helium distribution function in the heliosphere. *Astrophys. J. Suppl.* 220, 23. doi:10.1088/0067-0049/220/2/23

Linsky, J. L., Redfield, S., Ryder, D., and Chasan-Taber, A. (2022). Inhomogeneity within local interstellar clouds. *Astron. J.* 164, 106. doi:10.3847/1538-3881/ac816b

McComas, D. J., Alexashov, D., Bzowski, M., Fahr, H., Heerikhuisen, J., Izmodenov, V., et al. (2012). The heliosphere's interstellar interaction: No bow shock. *Science* 336, 1291–1293. doi:10.1126/science.1221054

McComas, D. J., Allegrini, F., Bochsler, B., Bzowski, M., Christian, E. R., Crew, G. B., et al. (2009). Global observations of the interstellar interaction from the interstellar boundary explorer (IBEX). *Science* 326, 959–962. doi:10.1126/science.1180906

McComas, D. J., Allegrini, F., Bochsler, P., Bzowski, M., Collier, M., Fahr, H., et al. (2009). IBEX - interstellar boundary explorer. *Space Sci. Rev.* 146, 11–33. doi:10.1007/s11214-009-9499-4

McComas, D. J., Christian, E. R., Schwadron, N. A., Fox, N., Westlake, J., Allegrini, F., et al. (2018). Interstellar mapping and acceleration Probe (IMAP): A new NASA mission. *Space Sci. Rev.* 214, 116. doi:10.1007/s11214-018-0550-1

McComas, D. J., Lewis, W. S., and Schwadron, N. A. (2014). IBEX's Enigmatic Ribbon in the sky and its many possible sources. *Rev. Geophys.* 52, 118–155. doi:10.1002/2013rg000438

McNutt, R. L., Paul, M. V., and Brandt, P. C. (2021). *Interstellar Probe – humanity's journey to interstellar space*. NASA Solar and Space Physics Mission Concept Study Report. Laurel, MD: Johns Hopkins Applied Physics Laboratory.

Möbius, E., Bochsler, P., Bzowski, M., Crew, G. B., Funsten, H. O., Fuselier, S. A., et al. (2009). Direct observations of interstellar H, He, and O by the interstellar boundary explorer. *Science* 326, 969–971. doi:10.1126/science.1180971

Möbius, E., Bochsler, P., Bzowski, M., Heirtzler, D., Kubiak, M. A., Kucharek, H., et al. (2012). Interstellar gas flow parameters derived from IBEX-Lo observations in 2009 and 2010: Analytical analysis. *Astrophys. J. Suppl.* 198, 11. doi:10.1088/0067-0049/198/2/11

Möbius, E., Bower, J., Aly, A., Berger, L., Farrugia, C., Galvin, A. B., et al. (2019). Observation of suprathermal tails of He+ pickup ions across solar wind compression regions with STEREO PLASTIC. *J. Phys. Conf. Ser.* 1332, 012011. doi:10.1088/1742-6596/1332/1/012011

Möbius, E., Bzowski, B., Chalov, S., Fahr, H. J., Gloeckler, G., Izmodenov, V., et al. (2004). Synopsis of the interstellar He parameters from combined neutral gas, pickup ion and UV scattering observations and related consequences. *Astron. Astrophys.* 426, 897–907. doi:10.1051/0004-6361:20035834

Möbius, E., Bzowski, M., Frisch, P. C., Fuselier, S. A., Heirtzler, D., Kubiak, M. A., et al. (2015). Interstellar flow and temperature determination with IBEX: Robustness and sensitivity to systematic effects. *Astrophys. J. Suppl.* 220, 24. doi:10.1088/0067-0049/220/2/24

Möbius, E., Gloeckler, G., Hovestadt, D., Ipavich, F. M., Klecker, B., Scholer, M., et al. (1985). The time-of-flight spectrometer SULEICA for ions of the energy range 5-270 keV/charge on AMPTE IRM. *IEEE Trans. Geoscience Remote Sens.* 23, 274–279. doi:10.1109/tgrs.1985.289527

Möbius, E., Hovestadt, D., Klecker, B., Scholer, M., Gloeckler, G., and Ipavich, F. M. (1985). Direct observation of He+ pick-up ions of interstellar origin in the solar wind. *Nature* 318, 426–429. doi:10.1038/318426a0

Möbius, E., Hovestadt, D., Klecker, B., Scholer, M., Gloeckler, G., Ipavich, F. M., et al. (1986). Observation of lithium pickup ions in the 5 to 20 keV energy range following the AMPTE solar wind releases. *J. Geophys. Res.* 91, 1325–1332. doi:10.1029/ja091ia02p01325

Möbius, E., Klecker, B., Bochsler, P., Gloeckler, G., Kucharek, K., Simunac, K., et al. (2010). "He pickup ions in the inner heliosphere-diagnostics of the local interstellar gas and of interplanetary conditions," in Pickup ions throughout the heliosphere and beyond: Proc. of the 9th Ann. Int. Astrophys. Conf., AIP Conf. Proc., 37–43.

Möbius, E., Klecker, B., Hovestadt, D., and Scholer, M. (1988). Interaction of interstellar pick-up ions with the solar wind. *Astrophys. Space Sci.* 144, 487–505. doi:10.1007/bf00793200

Möbius, E., Kucharek, H., Clark, G., O'Neill, M., Petersen, L., Bzowski, M., et al. (2009). Diagnosing the neutral interstellar gas flow at 1 AU with IBEX-lo. *Space Sci. Rev.* 146, 149–172. doi:10.1007/s11214-009-9498-5

Möbius, E., Lee, M. A., and Drews, C. (2015). Interstellar flow longitude from the symmetry of the pickup ion cut-off at 1 AU. *Astrophys. J.* 815, 20. doi:10.1088/0004-637x/815/1/20

Möbius, E., Litvinenko, Y., Grünwaldt, H., Aellig, M. R., Bogdanov, A., Ipavich, F. M., et al. (1999). Direct evidence of the interstellar gas flow velocity in the pickup ion cut-off as observed with SOHO CELIAS CTOF. *Geophys. Res. Lett.* 26, 3181–3184. doi:10.1029/1999gl003644

Möbius, E., Morris, D., Popecki, M. A., Klecker, B., Kistler, L. M., and Galvin, A. B. (2002). Charge states of energetic (≈0.5 MeV/n) ions in corotating interaction regions at 1 AU and implications on source populations. *Geophys. Res. Lett.* 29, 1016. doi:10.1029/2001gl013410

Möbius, E., Rucinski, D., Hovestadt, D., and Klecker, B. (1995). The helium parameters of the very local interstellar medium as derived from the distribution of He+ pickup ions in the solar wind. *Astron. Astrophys.* 304, 505–519.

Möbius, E., Rucinski, D., Isenberg, P. A., and Lee, M. A. (1996). Determination of interstellar pickup ion distributions in the solar wind with SOHO and CLUSTER. *Ann. Geophys* 14, 492–496. doi:10.1007/s00585-996-0492-x

Möbius, E., Rucinski, D., Lee, M. A., and Isenberg, P. A. (1998). Decreases in the anti-sunward flux of interstellar pickup He+ associated with radial interplanetary magnetic field. *J. Geophys. Res.* 103, 257–265. doi:10.1029/97ja02771

Morris, D., Möbius, E., Lee, M. A., Popecki, M. A., Klecker, B., Kistler, L. M., et al. (2001). Implications for source populations of energetic ions in Co-rotating interaction regions from ionic charge states. In: Solar and galactic composition. *AIP Conf. Proc.* 598, 201.

Opher, M., Loeb, A., Drake, J., and Toth, G. (2020). A small and round heliosphere suggested by magnetohydrodynamic modelling of pick-up ions. *Nat. Astron.* 4, 675–683. doi:10.1038/s41550-020-1036-0

Opher, M., Richardson, J. D., Toth, G., and Gombosi, T. I. (2009). Confronting observations and modeling: The role of the interstellar magnetic field in voyager 1 and 2 asymmetries. *Space Sci. Rev.* 143, 43–55. doi:10.1007/s11214-008-9453-x

Opher, M., Stone, E. C., and Liewer, P. C. (2006). The effects of a local interstellar magnetic field on voyager 1 and 2 observations. *Astrophys. J.* 640, L71–L74. doi:10.1086/503251

Park, J., Kucharek, H., Möbius, E., Galli, A., Kubiak, M. A., Bzowski, M., et al. (2016). IBEX observations of secondary interstellar helium and oxygen distributions. *Astrophys. J.* 833, 130. doi:10.3847/1538-4357/833/2/130

Park, J., Kucharek, H., Möbius, E., Galli, A., Livadiotis, G., Fuselier, S. A., et al. (2015). Statistical analysis of the heavy neutral atoms measured by IBEX. *Astrophys. J. Suppl.* 220, 34. doi:10.1088/0067-0049/220/2/34

Park, J., Kucharek, H., Möbius, E., Leonard, T., Bzowski, M., Sokol, J. M., et al. (2014). The Ne to O abundance ratio of the interstellar medium from IBEX-Lo observations. *Astrophys. J.* 795, 97. doi:10.1088/0004-637x/795/1/97

Park, J., Kucharek, H., Paschalidis, J., Szabo, A., Heirtzler, D., Mobius, E., et al. (2019). The characterization of secondary interstellar neutral oxygen beyond the heliopause: A detailed analysis of the *IBEX*-lo oxygen observations. *Astrophys. J.* 880, 4. doi:10.3847/1538-4357/ab264a

Pogorelov, N. V., Borikov, S. N., Heerikhuisen, J., and Zhang, M. (2015). *Heliotail. Astrophys. J. Lett.* 812, L6. doi:10.1088/2041-8205/812/1/l6

Pogorelov, N. V., Borovikov, S. N., Zank, G. P., and Ogino, T. (2009). Three-dimensional features of the outer heliosphere due to coupling between the interstellar and interplanetary magnetic fields. III. The effects of solar rotation and activity cycle. *Astrophys. J.* 696, 1478–1490. doi:10.1088/0004-637x/696/2/1478

Pogorelov, N. V., Heerikhuisen, J., Zank, G. P., and Borikov, S. N. (2009). Influence of the interstellar magnetic field and neutrals on the shape of the outer heliosphere. *Space Sci. Rev.* 143, 31–42. doi:10.1007/s11214-008-9429-x

Prantzos, N. (1998). "The chemical evolution of the Milky Way disk," in *Primordial nuclei and their galactic evolution*. Editors N. Prantzos, M. Tosi, and R. von Steiger (Bern, Switzerland: Space Science Series of ISSI), 225.

Quinn, P., Schwadron, N. A., Möbius, E., Gorbi, M., and McComas, D. J. (2016). Transport of helium pickup ions within the focusing cone: Reconciling STEREO observations with IBEX. *Astrophys. J.* 824, 142. doi:10.3847/0004-637x/824/2/142

Rahmanifard, F., Möbius, E., Schwadron, N. A., Galli, A., Richards, N., Kucharek, H., et al. (2019). Radiation pressure from interstellar hydrogen observed by *IBEX* through solar cycle 24. *Astrophys. J.* 887, 217. doi:10.3847/1538-4357/ab58ce

Redfield, S., and Linsky, J. (2008). The structure of the local interstellar medium. IV. Dynamics, morphology, physical properties, and implications of cloud-cloud interactions. *Astrophys. J.* 673, 283–314. doi:10.1086/524002

Rodriguez-Moreno, D. F., Wurz, P., Saul, L., Bzowski, M., Kubiak, M. A., Sokol, J. M., et al. (2013). Evidence of direct detection of interstellar deuterium in the local interstellar medium by IBEX. *Astron. & Astrophys.* 557, A125. doi:10.1051/0004-6361/201321420

Schramm, D. N. (1998). "Big Bang nucleosynthesis and the density of baryons in the universe," in *Primordial nuclei and their galactic evolution*. Editors N. Prantzos, M. Tosi, and R. von Steiger (Bern, Switzerland: Space Science Series of ISSI), 3.

Schwadron, N. A., Bzowski, M., Crew, G. B., Gruntman, M., Fahr, H., Fichtner, H., et al. (2009). Comparison of interstellar boundary explorer observations with 3D global heliospheric models. *Science* 326, 966–968. doi:10.1126/science.1180986

Schwadron, N. A., Bzowski, M., McComas, D. J., and Moebius, E. (2018). The local interstellar magnetic field observed by voyager 1 and IBEX. *J. Phys. Conf. Ser.* 1100, 012021. doi:10.1088/1742-6596/1100/1/012021

Schwadron, N. A., and Bzowski, M. (2018). The heliosphere is not round. *Astrophys. J.* 862, 11. doi:10.3847/1538-4357/aacbcf

Schwadron, N. A., Frisch, P. C., Adams, F. R., Christian, E. R., Desiati, P., Funsten, H. O., et al. (2015). A consistent scenario for the IBEX Ribbon, anisotropies in TeV cosmic rays, and the local interstellar medium. *ASTRA Proc.* 2, 9–16. doi:10.5194/ap-2-9-2015

Schwadron, N. A., Möbius, E., Lee, M. A., Bochsler, P., Bzowski, M., Fuselier, S. A., et al. (2016). Determination of interstellar O parameters using the first two years of data from theinterstellar *boundary explorer*. *Astrophys. J.* 828, 81. doi:10.3847/0004-637x/828/2/81

Schwadron, N. A., Möbius, E., Leonard, T., Fuselier, S. A., McComas, D. J., Heirtzler, D., et al. (2015). Determination of interstellar He parameters using 5 years of data from the Interstellar Boundary Explorer–Beyond closed form approximations. *Astrophys. J. Suppl.* 220, 25. doi:10.1088/0067-0049/220/2/25

Schwadron, N. A., Richardson, J., Burlaga, L., McComas, D. J., and Moebius, E. (2015). Triangulation of the interstellar magnetic field using voyager 1, SOHO/SWAN and IBEX. *Astrophys. J. Lett.* 813, L20. doi:10.1088/2041-8205/813/1/l20

Slavin, J. D., and Frisch, P. C. (2008). The boundary conditions of the heliosphere: Photoionization models constrained by interstellar and *in situ* data. *Astron. Astrophys.* 491, 53–68. doi:10.1051/0004-6361:20078101

Sokól, J. M., Bzowski, M., Kubiak, M. A., and Möbius, E. (2016). Solar cycle variation of interstellar neutral He, Ne, O density and pickup ions along the Earth's orbit. *MNRAS* 458, 3691–3704. doi:10.1093/mnras/stw515

Stone, E. C., Cummings, A. C., McDonald, F. B., Heikkila, B. C., Lal, N., and Webber, W. R. (2008). An asymmetric solar wind termination shock. *Nature* 454, 71–74. doi:10.1038/nature07022

Stone, E. C., Cummings, A. C., McDonald, F. B., Heikkila, B. C., Lal, N., and Webber, W. R. (2013). Voyager 1 observes low-energy galactic cosmic rays in a region depleted of heliospheric ions. *Science* 341, 150–153. doi:10.1126/science.1236408

Strömgren, B. (1939). The physical state of interstellar hydrogen. *Astrophys. J.* 89, 526–547. doi:10.1086/144074

Swaczyna, P., Bzowski, M., Kubiak, M. A., Sokol, J. M., Fuselier, S. A., Galli, A., et al. (2018). Interstellar neutral helium in the heliosphere from *IBEX* observations. V. Observations in *IBEX*-lo ESA steps 1, 2, and 3. *Astrophys. J.* 854, 119. doi:10.3847/1538-4357/aaabbf

Swaczyna, P., Bzowski, M., Kubiak, M. A., Sokol, J. M., Fuselier, S. A., Heirtzler, D., et al. (2015). Interstellar neutral helium in the heliosphere from *IBEX* observations. I. Uncertainties and backgrounds in the data and parameter determination method. *Astrophys. J. Suppl.* 220, 26. doi:10.1088/0067-0049/220/2/26

Swaczyna, P., Kubiak, M. A., Bzowski, M., Bower, J., Fuselier, S. A., Galli, A., et al. (2022). Very local interstellar medium revealed by a complete solar cycle of interstellar neutral helium observations with IBEX. *Astrophys. J. Suppl.* 259, 42. doi:10.3847/1538-4365/ac4bde

Swaczyna, P., McComas, D. J., Zirnstein, E. J., Sokol, J. M., Elliott, H. A., Bzowski, M., et al. (2020). Density of neutral hydrogen in the Sun's interstellar neighborhood. *Astrophys. J.* 903, 48. doi:10.3847/1538-4357/abb80a

Swaczyna, P., Schwadron, N. A., Möbius, E., Bzowski, M., Frisch, P. C., Linsky, J. L., et al. (2022). Mixing interstellar clouds surrounding the Sun. *Astrophys. J. Lett.* 937, L32. doi:10.3847/2041-8213/ac9120

Taut, A., Berger, L., Drews, C., Morris, E., Heidrich-Meisner, V., Keilbach, D., et al. (2018). Challenges in the determination of the interstellar flow longitude from the pickup ion cutoff. *Astron. Astrophys.* 611, A61. doi:10.1051/0004-6361/201731796

Vasyliunas, V. M., and Siscoe, G. L. (1976). On the flux and the energy spectrum of interstellar ions in the solar system. *J. Geophys. Res.* 81, 1247–1252. doi:10.1029/ja081i007p01247

Weller, C. S., and Meier, R. R. (1974). Observations of helium in the interplanetary/interstellar wind: The solar-wake effect. *Astrophys. J.* 193, 471–476. doi:10.1086/153182

Witte, M., Banaszkiewicz, M., and Rosenbauer, M. (1996). Recent results on the parameters of the interstellar helium from the ULYSSES/GAS experiment. *Space Sci. Rev.* 78, 289–296. doi:10.1007/bf00170815

Witte, M. (2004). Kinetic parameters of interstellar neutral helium. Review of results obtained during one solar cycle with the Ulysses/GAS-instrument. *Astron. Astrophys.* 426, 835–844. doi:10.1051/0004-6361:20035956

Witte, M., Rosenbauer, M., Keppler, E., Keppler, E., Fahr, H., Hemmerich, P., et al. (1992). The interstellar neutral-gas experiment on Ulysses. *Astron. Astrophys. Suppl. Ser.* 92, 333–348.

Wood, B., Müller, H-R., and Witte, M. (2015). Revisiting Ulysses observations of interstellar helium. *Astrophys. J.* 801, 62. doi:10.1088/0004-637x/801/1/62

Wood, B. E., Müller, H. R., and Möbius, E. (2019). Evidence for asymmetry in the velocity distribution of the interstellar neutral helium flow observed by IBEX and Ulysses. *Astrophys. J.* 881, 55. doi:10.3847/1538-4357/ab2e74

Wu, F. M., and Judge, D. L. (1979). Temperature and flow velocity of the interplanetary gases along solar radii. *Astrophys. J.* 231, 594–605. doi:10.1086/157221

Zirnstein, E. J., Heerikhuisen, J., Funsten, H. O., Livadiotis, G., McComas, D. J., and Pogorelov, N. V. (2016). Local interstellar magnetic field determined from the interstellar boundary explorer Ribbon. *Astrophys. J. Lett.* 818, L18. doi:10.3847/2041-8205/818/1/l18

Zirnstein, E. J., Möbius, E., Zhang, M., Bower, J., Elliott, H. A., McComas, D. J., et al. (2022). *In situ* Observations of Interstellar Pickup Ions from 1 au to the Outer Heliosphere. *Space Sci. Rev.* 218, 28. doi:10.1007/s11214-022-00895-2

Frontiers in
Astronomy and Space Sciences

Explores planetary science and extragalactic astronomy in all wavelengths

Advances the understanding of our universe - from planetary science to extragalactic astronomy, to high-energy and astroparticle physics.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact