

1
2

3

4
5
6

7
8
9
10

Applications of Statistical Methods and Machine Learning in the Space Sciences

Bala Poduval^{1,2}, Karly M. Pitman², and The SOC Team³

¹University of New Hampshire, Durham, NH, USA

²Space Science Institute, Boulder, CO, USA

³Listed in the Appendix

Key Points:

- machine learning
- space science
- statistical methods

Corresponding author: Bala Poduval, bala.poduval@unh.edu

Abstract

Applications and methods of artificial intelligence (AI) are becoming powerful tools for scientific investigations in the space sciences, particularly in the analysis and interpretation of the plethora of spacecraft and ground-based observations of the near-Earth and faraway space. *Applications of Statistical Methods and Machine Learning in the Space Sciences* was a virtual conference held during 17-21 May 2021 to bring together experts in AI and in the various subfields of space sciences to further explore the utility of AI, machine learning, and statistical analysis techniques while sharing the current status of applications in these fields. The conference concluded by emphasizing the scope of AI techniques available to the space science community for addressing outstanding problems with great success as revealed in the number of research works presented.

SSI Virtual Conference

Emerging trends in artificial intelligence (AI), data science, and machine learning are revolutionizing the way scientific problems in the space sciences are placed in context and addressed. AI techniques are indeed shown to be highly promising in modeling and data analysis. The techniques of AI vary over a wide range including, but not limited to, machine learning (ML), neural networks and deep learning, data and text mining, natural language processing, knowledge discovery and decision making, expert system modeling, genetic algorithms, fuzzy logic, and computer vision and pattern recognition.

Applications of Statistical Methods and Machine Learning in the Space Sciences (presentation materials available at <http://spacescience.org/workshops/mlconference2021.php>), a fully virtual conference hosted by Space Science Institute's Center for Data Science and sponsored by the National Science Foundation (NSF), was held during 17-21 May 2021. This event brought together experts in various disciplines of the space sciences (such as solar physics and aeronomy, planetary and exoplanetary sciences, geology, astrobiology and astronomy) and industry to leverage the advancements in statistics, data science, methods of artificial intelligence (AI) and information theory to improve analytic models and their predictive capabilities utilizing the enormous data in these fields.

This multidisciplinary conference provided a vibrant forum for senior scientists and industry professionals, early career researchers and students to present their latest results using a wide variety of techniques and methods in advanced statistics, to enhance their knowledge in the recent trends in AI and to participate in a platform for future collaborations. The conference covered a breadth of topics, such as advanced statistical methods, deep learning and neural networks, time series analysis, Bayesian methods, feature identification and feature extraction, physics-based models combined with machine learning techniques and surrogate models, space weather prediction and other domain topics where AI is applied, model validation and uncertainty quantification, turbulence and nonlinear dynamics in space plasma, physics informed neural networks, information theory and, data reconstruction and data assimilation.

A summary of current efforts on applying ML methods in the field of space sciences in comparison with those efforts in other fields of natural sciences and recommendations for machine learning in planetary science to funding agencies and the planetary community can be found in Azari et al. (2021). Figure 1 of Azari et al. (2021) illustrated that heliophysics and space physics had the highest percentage of published works discussing machine learning in 2020, followed by astrophysics and Earth science. Azari et al. (2021) concluded with recommendations in the next decade for supporting a data-rich future for planetary science.

AI methods have already been applied in addressing various problems in the field solar-terrestrial physics since the 1990s (Newell et al., 1991; Lundstedt, 1992, 1996, 2006; Wintoft & Lundstedt, 1997; Wing et al., 2005). These included classifications of auro-
 ral particle precipitation, predictions of solar wind velocity, geomagnetic disturbances, and K_p . Information theory has proved useful in establishing linear and nonlinear relationships and causalities in the studies of solar and space physics (Wing et al., 2016, 2018). Early attempts to apply ML techniques involve the forecast of geomagnetic indices (e.g. Wu & Lundstedt, 1996; Wu & H. Lundstedt, 1997), of the relativistic electrons at geosynchronous orbits (e.g. Stringer et al., 1996), and of solar eruptions (Fozzard et al., 1988; Camporeale et al., 2019).

The “International Workshop on Artificial Intelligence Applications in Solar-Terrestrial Physics” held in 1993 was one of the first of its kind which focused on “neural network applications of Multi-Layer-Error-Back-Propagation (MLBP) and Self-Organizing Map (SOM) neural nets and traditional expert systems and fuzzy expert systems” (<http://www.lund.irf.se/HeliosHome/proceed.html>). Unlike this and other conferences on machine learning (e.g. Camporeale & SOC-ML-Helio, 2020), the SSI virtual conference had an emphasis on understanding the physics and dynamics of systems while seeking accurate solutions using ML methods (“black box” versus “interpretable” models). Furthermore, this virtual conference highlighted the interdisciplinary nature of machine learning applications in space sciences, the main theme of the conference. The research works presented revealed close collaborations among space science, statistics, computer science and AI, showcasing how these experts can collaborate to soundly improve their models and predictions.

The virtual conference was hosted by Space Science Institute’s Center for Data Science (CDS) and served as an initiative to bring together domain experts in space sciences and highly skilled corporate talents sharing a common interest in data science and machine learning. CDS aims to inspire the scientific community to utilize key insights on emerging technologies, transforming this possibility into reality. Space Science Institute hosted 219 registered participants from more than 25 countries over Zoom for this event. Participants did not volunteer their demographic data but based on 103 of the conference registrants for whom the conference organizers could reasonably determine their backgrounds, 32 were female, 43 were from underrepresented minorities, and 45 were early career (student to within 5 years of Ph.D. date). There were 79 oral and 28 e-poster presentations in addition to interactive sessions demonstrating data processing and machine learning methods. The virtual conference featured 14 keynote speakers, 50% of whom were female and 5 were early career (within 5 years of earning their Ph.D.) scientists. Links to these presentations and the recordings are available at the conference website (<http://spacescience.org/workshops/mlconference2021.php>). Moreover, a collection of the works presented at this virtual conference, along with new contributions from the broader scientific community in the form of original research articles, reviews/mini-reviews, brief reports and commentaries on the present scenario of AI applications in the space sciences, and scope of statistical methods and machine learning (ML) in the various fields of space sciences are being compiled as a topical collection of *Frontiers in Astronomy and Space Sciences* (<https://www.frontiersin.org/research-topics/25408/>).

The highlight of the conference was the lively discussion sessions. The virtual conference designated 45 minutes each day for live discussion sessions to discuss AI and ML trends in specific fields of space science and to encourage cross-disciplinary approaches to problems in different fields. Discussions were distributed among different topics and centered on the applicability of Statistical Methods and ML in Magnetospheric Studies, Astronomy, Heliophysics, Planetary Sciences and Exoplanets, and Aeronomy, Turbulence and Nonlinear Dynamics. Moreover, these sessions highlighted the importance and the impact of a few fundamental aspects in all the space science domains such as interpretability and explainability of ML models, reproducibility, and the need and avail-

ability of AI-ready data. These designated sessions addressed: the challenges of BIG DATA and *small data sets*; how to handle overfitting; uncertainties and gaps in the data sets and how they are incorporated into the models; supervised and unsupervised ML; how to compare models. These discussions defined and emphasized the necessity of “AI-ready” data in all the disciplines of space sciences and the participants shared information on the various data sets currently available and what are the steps to be taken to create better and more concise AI-ready data. We believe that these discussion sessions were particularly helpful for the students, early career researchers and early ML practitioners who constituted a substantial fraction of the conference attendees because these sessions covered links and access to a number of educational, software and data resources. The discussions revealed the interdisciplinary nature of ML applications in space sciences and how this virtual conference presented itself as a platform for connecting the various components of this fast emerging, dynamical trend of AI applications.

One of the most significant outcomes of this virtual conference is the multi-authored white paper on AI-readiness (Poduval et al., 2022) of the numerous space science data for AI/ML applications that is currently in preparation for submission to The National Academies of Science, Engineering, and Medicine’s Decadal Survey for Solar and Space Physics (Heliophysics) 2024-2033 (<https://www.nationalacademies.org/our-work/decadal-survey-for-solar-and-space-physics-heliophysics-2024-2033>).

We conclude this brief report by emphasizing the emerging trends in ML applications in the space sciences and how the scientific community is embracing these novel, powerful techniques for exploiting the enormous volume of spacecraft and ground-based data while exploring how to explain and interpret the ML methods more scientifically.

Funding

The material presented here is based upon work supported by the National Science Foundation under Award No. AGS - 2114219. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments

The authors wish to thank the National Science Foundation’s Aeronomy program for supporting this virtual event and James Harold and Evaldas Vidugiris for their technical assistance during the virtual conference.

References

- Azari, A. R., Biersteker, J. B., Dewey, R. M., Doran, G., Forsberg, E. J., Harris, C. D. K., . . . co signatories, . (2021). Integrating machine learning for planetary science: Perspectives for the next decade. *Submitted to the NRC Planetary and Astrobiology Decadal Survey*. (<https://baas.aas.org/pub/2021n4i128/release/1?readingCollection=7272e5bb>)
- Camporeale, E., Chu, X., Agapitov, O. V., & Bortnik, J. (2019). On the generation of probabilistic forecasts from deterministic models. *Space Weather*, **17**, 1166–1207.
- Camporeale, E., & SOC-ML-Helio. (2020). ML-helio: An emerging community at the intersection between heliophysics and machine learning. *J. Geophys. Res.*, **125**. (doi:10.1029/2019JA027502)
- Fozzard, R., Bradshaw, G., & Ceci, L. (1988). *Advances in neural information processing systems*, 248–255.
- Lundstedt, H. (1992). Neural networks and predictions of solar-terrestrial effects. *Planet. Space Sci.*, **40**(4), 457–464.

- Lundstedt, H. (1996). Solar origin of geomagnetic storms and predictions. *JATP*, **58**(7), 821–830.
- Lundstedt, H. (2006). Solar activity modelled and forecasted: A new approach. *Adv. Space Res.*, **38**, 862–867.
- Newell, P. T., Wing, S., Meng, C. I., & Sigillito, V. (1991). The auroral oval position, structure and intensity of precipitation from 1984 onwards: An automated on-line data base. *J. Geophys. Res.*, **96**, 5877–5882.
- Poduval, R. M., B., Walker, R., Shneider, C., Himes, M., Wintoft, P., Kapali, S., ... Guidoni, D. (2022). Space science data and ai-readiness by 2050. *Abstract Submitted to the Heliophysics Decadal Survey White Paper Concepts*. (abstract:<https://www.lpi.usra.edu/decadal.whitepaper.proposals/heliophysics/>)
- Stringer, G., Heuten, I., Salazar, C., & Stokes, B. (1996). Radiation belts: Models and standards. *J. Geophys. Res.*, **97**(7-8), 291–295.
- Wing, S., Johnson, J., & Vourlidas, A. (2018). Information theoretic approach to discovering causalities in the solar cycle. *Ap. J.*, **854**, 85. (doi:10.3847/1538-4357/aaa8e7)
- Wing, S., Johnson, J. R., Camporeale, E., & Reeves, G. D. (2016). Information theoretical approach to discovering solar wind drivers of the outer radiation belt. *J. Geophys. Res.*, **121**, 9378–9399. (doi:10.1002/2016JA022711)
- Wing, S., Johnson, J. R., Jen, J., Meng, C.-I., Sibeck, D. G., Bechtold, K., ... Takahashi, K. (2005). Kp forecast models. *J. Geophys. Res.*, **110**, A04203. (doi:10.1029/2004JA010500)
- Wintoft, P., & Lundstedt, H. (1997). Prediction of daily average solar wind velocity from solar magnetic field observations using hybrid intelligent systems. *Phys. Chem. Earth*, **22**(7-8), 612–622.
- Wu, J.-G., & H. Lundstedt, H. (1997). Solar origin of geomagnetic storms and predictions. *J. Geophys. Res.*, **102**(A7), 14,255.
- Wu, J.-G., & Lundstedt, H. (1996). *GRL*, **23**(4), 319–322.

Appendix A: Authors and affiliations in alphabetical order

- Michael Balikhin, University of Sheffield, UK
- Joe Borovsky, Space Science Institute, Boulder, CO, USA
- Raffaella D’Amicis, Institute for Space Astrophysics and Planetology, Rome
- Maria Dainotti¹, National Astronomical Observatory of Japan, Japan
- Manolis Georgoulis, Academy of Athens, Greece
- Jay Johnson, Andrews University, MI, USA
- Karly Pitman, Space Science Institute, Boulder, CO, USA
- Bala Poduval¹, University of New Hampshire, Durham, NH, USA
- Ralph Shuping, Space Science Institute, Boulder, CO, USA
- Olga Verkhoglyadova, NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
- Simon Wing¹, Johns Hopkins University, Laurel, MD, USA
- Peter Wintoft, Swedish Institute of Space Physics, Sweden

¹ Affiliate of Space Science Institute, Boulder, CO