JSWSC

Available online at:
www.swsc-journal.org

Topical Issue - Space Weather research in the Digital Age and across the full data lifecycle

PREFACE

OPEN ∂ ACCESS

Space Weather research in the Digital Age and across the full data lifecycle: Introduction to the Topical Issue

Ryan M. McGranaghan[1,*] , Enrico Camporeale[2,3] , Manolis Georgoulis[4] ,
and Anastasios Anastasiadis[5]

[1] Atmosphere and Space Technology Research Associates (ASTRA), Louisville, 80027 CO, USA
[2] CIRES, University of Colorado, Boulder, 80309 CO, USA
[3] NOAA Space Weather Prediction Center, Boulder, 80305 CO, USA
[4] Academy of Athens, 106 79 Athens, Greece
[5] National Observatory of Athens, 15236 Penteli, Greece

Abstract – The onset and rapid advance of the Digital Age have brought challenges and opportunities for scientific research characterized by a continuously evolving data landscape reflected in the four V's of big data: volume, variety, veracity, and velocity. The big data landscape supersedes traditional means of storage, processing, management, and exploration, and requires adaptation and innovation across the full data lifecycle (i.e., collection, storage and processing, analytics, and representation). The Topical Issue, "Space Weather research in the Digital Age and across the full data lifecycle", collects research from across the full data lifecycle (collection, management, analysis, and communication; collectively "Data Science") and offers a tractable compendium that illustrates the latest computational and data science trends, tools, and advances for Space Weather research. We introduce the paradigm shift in Space Weather and the articles in the Topical Issue. We create a network view of the research that highlights the contribution to the change of paradigm and reveals the trends that will guide it hereafter.

Keywords: Data science / machine learning / forecasting / space weather / network / transdisciplinary

1 Introduction: A paradigm shift and data science network

Paradigm shifts arise when the dominant paradigm under which normal science operates is rendered incompatible with new phenomena, facilitating the adoption of a new theory or paradigm (Kuhn, 1962).

Under the pretext of a new paradigm emerging in the study of space physics (McGranaghan et al., 2017), a group of Space Weather researchers, solar physicists, atmospheric scientists, and several data scientists, who are less often present at Space Weather meetings, gathered in a large lecture hall during the 15th European Space Weather Week. The organizing theme was to unveil the current challenges facing Space Weather forecasting.

The onset and rapid advance of the Digital Age have brought challenges and opportunities for scientific research characterized by a continuously evolving data landscape reflected in the four V's of big data: volume, variety, veracity, and velocity. The big data landscape supersedes traditional

means of storage, processing, management, and exploration and requires adaptation and innovation across the full data lifecycle (i.e., collection, storage and processing, analytics, and representation). Indeed, our science and community face a paradigm shift in which data science becomes increasingly inextricable from scientific discovery. Despite the dramatic change, Space Weather research has not fully embraced available approaches and techniques to facilitate discovery and knowledge from data, or data science. Too often, data science is used synonymously with machine learning or artificial intelligence in general, but this special issue defines it more capaciously to include the full data lifecycle:

> Data Science = Scalable architectural approaches, techniques, software, and algorithms which alter the paradigm by which data are collected, managed and analyzed [and communicated]. *D. Crichton, NASA Jet Propulsion Laboratory, private correspondence.*

Space Weather research in the Digital Age and across the full data lifecycle is one of the outcomes of the presentations and discussion in that lecture hall, including connections
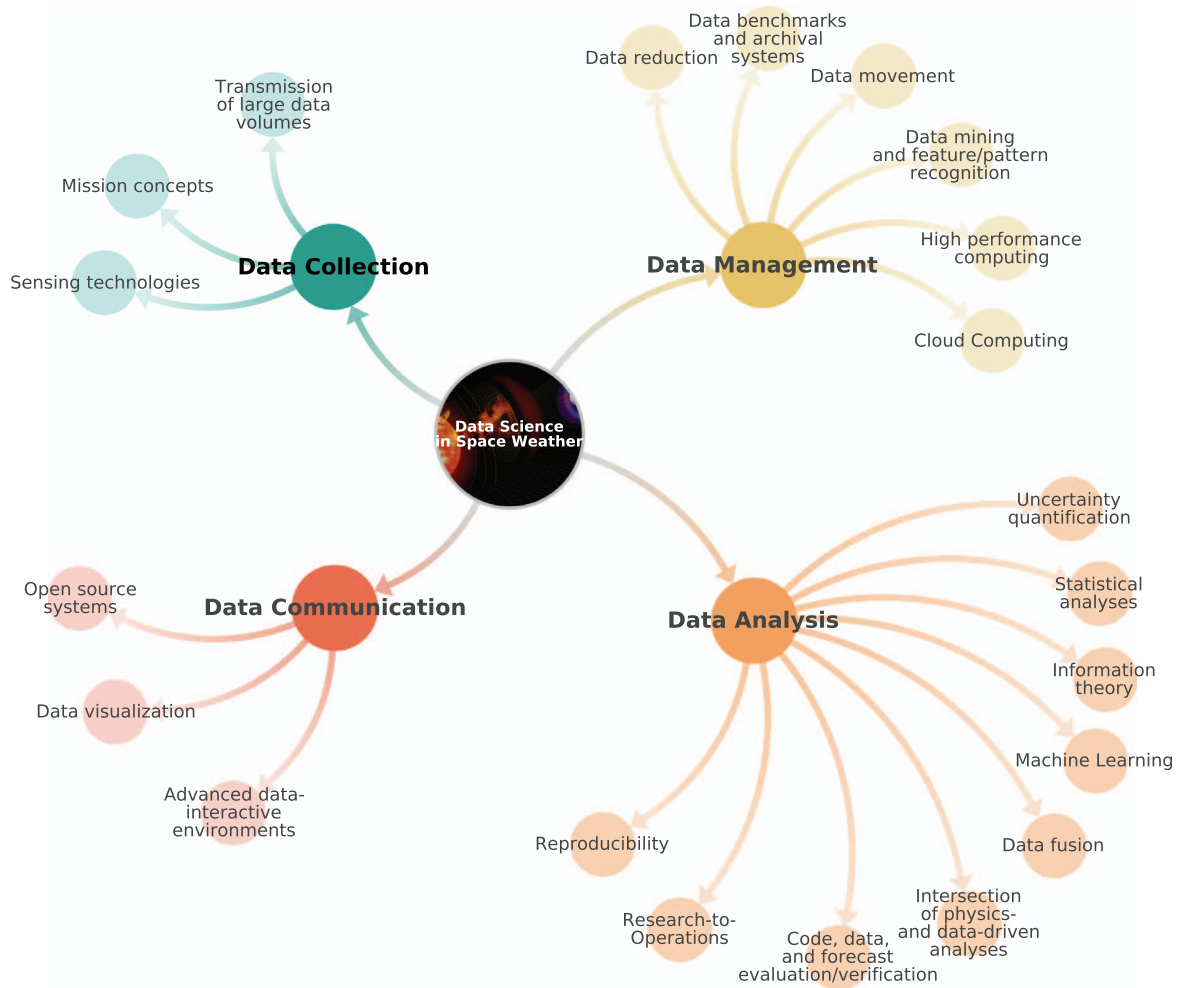
*Corresponding author: ryan.mcgranaghan@colorado.edu

Fig. 1. The "network" of data science topics solicited by the Topical Issue and a useful taxonomy for future research. Image created using kumu.io.

reaching beyond it. The topical issue offers a tractable compendium that illustrates the latest computational and data science trends, tools, and advances for Space Weather research.

2 The Topical Issue

The Topical Issue serves as a resource from which researchers and practitioners across the different disciplines involved in Space Weather data science can learn and ideate. Topics solicited covered the full spectrum of the data lifecycle: collection, management, analysis, and communication. These headings provide a useful guide to the breadth of activity that data science encompasses. Each can be further subdivided to form a taxonomy of the role of data science in Space Weather understanding and/or forecasting. The editors of the special issue synthesized community conversations, activities, and needs in designing this collection, the result of which is shown as a "network" in Figure 1.

The papers in this collection cover some, but not all, of these topics. However, this network can be a useful guide to the future development of Space Weather data science. Below we outline the contributions in the collection and their relation to this network. First, a pair of papers push the boundaries of geomagnetic index prediction. Chakraborty & Morley (2020) have created a model that generates a probabilistic 3-hour-ahead prediction of the planetary disturbance K_p index with uncertainty bounds for both storms ($K_p \leq 5-$) and non-storm cases. Their two-layer long short term memory (LSTM) machine learning model shows both that probabilistic forecasts can be produced, and that including operationally available information about solar irradiance enhances the ability of predictive models to capture the onset of geomagnetic storms. Park et al. (2021) combined machine learning and empirical approaches to produce an operational model for the disturbance storm time (*Dst*) index, demonstrating the value of integrating machine learning into other forms of Space Weather modeling. An important aspect of improving machine learning models is to identify the most important data to give to the model or the process of featurization.

Deshmukh et al. (2020), on the other hand, use machine learning with topological mathematics to improve the input parameters for solar flare prediction algorithms. They compute the topology and geometry of structures in 2D solar photospheric magnetograms, successfully extend the relevant feature

set to include characteristics of the magnetic field based purely on the geometry and topology of 2D magnetogram images. Their extended feature set improves the prediction accuracy of a neural-net-based flare-prediction method.

Results from machine learning application papers in the collection highlight three important areas of future development: (1) treating Space Weather prediction as a probabilistic, rather than the deterministic, problem (Camporeale et al., 2019); (2) allowing machine learning to augment, rather than replace, traditional Space Weather prediction methods; and (3) exploring roles for machine learning outside of predictive algorithms to aid Space Weather (e.g., by determining and ranking the most useful input features).

Machine learning is intended to model input-output systems with high nonlinearity, albeit with the challenge of interpreting the results. An area of study that can unravel causal relationships is information theory, which involves treating the physical and parameter space impacted by Space Weather as a complex adaptive system. Information theory provides methods to determine the information flow among various parameters, causalities, to untangle the drivers, and provide observational constraints that can help guide the development of theories and physics-based models (Wing & Johnson, 2019). Using a nonlinear multi-scale dynamical systems approach, Alberti et al. (2020) uncover a distinct behavior of Space Weather manifestations in Earth's magnetosphere on short (minutes) and long (days) timescales. Their results suggest that different physical processes are typical for both regimes from a more chaotic (with "dynamic anomalies") and less predictable behavior in the fast regime (short time scales) and less chaotic behavior in the slow component (long time scales). Space weather is an excellent natural testbed for the trend of exploring machine learning and information-theoretic approaches in tandem, seeking a convergence that may be critical in creating explainable artificial intelligence (XAI, Gunning et al., 2019).

The Topical Issue also contains frontier scientific research obtained in synergy with data science methods. Georgoulis et al. (2021) summarize the Flare Likelihood and Region Eruption Forecasting (FLARECAST) Project. They give an account of progress and challenges in solar flare prediction within a diverse consortium that has made openly available all comprehensive data, codes, and infrastructure spawned in the project. Owens et al. (2020) use 40+ years of observation-driven solar wind model results to accurately predict near-Earth solar wind conditions, by identifying the representativity error associated with measuring the solar wind at different heliographic latitudes. Their findings make possible solar wind data assimilative schemes that will be important to provide the driver information for Space Weather. Tang et al. (2020) apply three machine learning methods to predict a physical quantity that drives much of Space Weather activity: ionospheric total electron content (TEC) in the ionosphere. They discover that memory in Space Weather observational data is key to the prediction of TEC and that long short-term memory (LSTM) neural networks can, in certain applications, produce more accurate predictions of TEC in Space Weather storms. Cesaroni et al. (2020) also apply machine learning to the prediction of TEC. They construct a 24-hour global prediction model based on nonlinear autoregressive neural networks with external input (NARX). They rigorously evaluate the efficacy of this model for TEC prediction, which is a prerequisite for all Space Weather

studies involving machine learning, and identify a pathway from research to operations for these models. Their model is currently implemented on the Ionosphere Prediction Service (IPS) (Vadakke Veettil et al., 2019), a prototype platform to support different classes of GNSS users. Data science offers a means for the Space Weather and Earth science communities to collaborate. Rogers et al. (2020) look at ground-level manifestations of Space Weather, presenting a multi-parameter global statistical model of extreme horizontal geomagnetic field fluctuations (dBH/dt). Their statistical model is useful for power grid utilities seeking to assess the risk of geomagnetically induced currents in the grid and a demonstration of the value of direct partnership between Space Weather researchers and power grid utilities.

Finally, the Topical Issue is a harbinger of the foreseen transition toward open science (National Academies of Sciences, Engineering and Medicine, 2018). Bhatt et al. (2020), provide a glimpse into open Space Weather science by demonstrating the Reproducible Software Environment (Resen). Resen is an open-source tool enabling computationally reproducible scientific results for the geospace science community. They reveal how the goals of open science (namely, to remove barriers in accessing, processing, analyzing, and visualizing data) can be accomplished for the Space Weather community by combining two mainstream open-source software tools, Docker (Merkel, 2014) and JupyterHub (Kluyver et al., 2016). Resen facilitates computationally reproducible research results and effective collaboration among researchers. It serves as a powerful example of the open science future for Space Weather.

3 The future

The contributions to this topical issue are not exhaustive nor a complete image of the fusion between data science and Space Weather. Beyond the individual use cases presented herein, many more will emerge in the coming years.

Two trends are discernible in the Topical Issue's collection, indicating the apparent paradigm shift in Space Weather science:

First, the challenge for data science to treat Space Weather as a knowledge and wisdom field, rather than simply as a data field. The problem in Space Weather is not one of information, but one of access. This is because everything is siloed: datasets, disciplines, people, projects, institutions. For instance, thermospheric density data are separate from solar imagery. The community of data contributors is fragmented, leaving many observations unknown and often inaccessible to researchers, and research analyses with significant data gaps even where observations may exist. Yet each new bit of information is further testament to the interconnectedness of the solar-terrestrial system. From the cultural side, teams consist of individuals involved in tens of associated activities: proposals, committees, papers, presentations, conferences, organizations; yet these myriad projects are not linked or converged. Disconnected teams fail to realize their potential or even to be aware of each other's activities; their progress is inaccessible and unusable. The distinctions between the silos are artificial. Lack of awareness and usability make reuse and collective progress impossible. Data science can create a more cohesive community, better capable of linking and understanding the resources (people, capabilities, assets,

contents, data, models) available and bridging to related disciplines in a valuable, minimally burdensome, and fully searchable way, allowing the seamless exchange of information and knowledge. This linking, moving from data to knowledge and wisdom, is vital to a more flourishing community.

Second, this brighter future relies on embracing the burgeoning field of Open Science. Open Science objectifies increased rigor, accountability, and reproducibility for research and is based on the principles of inclusion, fairness, equity, and sharing National Academies of Sciences, Engineering and Medicine (2018). Open Science spans all disciplines and can offer the set of tools and know-how for the modern research environment, in the understanding of an equally rigorous and transparent comparative juxtaposition of models relying on the same data and performance thereof.

References

Alberti T, Lekscha Jaqueline, Consolini G, De Michelis P, Donner RV. 2020. Disentangling nonlinear geomagnetic variability during magnetic storms and quiescence by timescale dependent recurrence properties. *J Space Weather Space Clim* **10**: 25. https://doi.org/10.1051/swsc/2020026.

Bhatt A, Valentic T, Reimer A, Lamarche L, Reyes P, Cosgrove R. 2020. Reproducible Software Environment: a tool enabling computational reproducibility in geospace sciences and facilitating collaboration. *J Space Weather Space Clim* **10**: 12. https://doi.org/10.1051/swsc/2020011.

Camporeale E, Chu X, Agapitov OV, Bortnik J. 2019. On the generation of probabilistic forecasts from deterministic models. *Space Weather* **17(3)**: 455–475. https://doi.org/10.1029/2018SW002026.

Cesaroni C., Spogli L., Aragon-Angel A., Fiocca M., Dear V., De Franceschi G., Romano V. 2020. Neural network based model for global Total Electron Content forecasting. *J Space Weather Space Clim* **10**: 11. https://doi.org/10.1051/swsc/2020013.

Chakraborty S, Morley SK. 2020. Probabilistic prediction of geomagnetic storms and the Kp index. *J Space Weather Space Clim* **10**: 36. https://doi.org/10.1051/swsc/2020037.

Deshmukh V, Berger TE, Bradley E, Meiss JD. 2020. Leveraging the mathematics of shape for solar magnetic eruption prediction. *J Space Weather Space Clim* **10**: 13. https://doi.org/10.1051/swsc/2020014.

Georgoulis MK, Bloomfield DS, Piana M, Massone AM, Soldati M, et al. 2021. The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era. *J Space Weather Space Clim* **11**: 39. https://doi.org/10.1051/swsc/2021023.

Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. 2019. XAI-Explainable artificial intelligence. *Sci Robot* **4(37)**: eaay7120. https://doi.org/10.1126/scirobotics.aay7120.

Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, et al. 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: *Positioning and power in academic publishing: players, agents and agendas.* Loizides F, Schmidt B (Eds.) IOS Press. pp. 87–90.

Kuhn T. 1962. The structure of scientific revolutions. *IEUS* **2(2)**: 52–77 and 111–136.

McGranaghan RM, Bhatt A, Matsuo T, Mannucci AJ, Semeter JL, Datta-Barua S. 2017. Ushering in a new Frontier in geospace through data science. *J Geophys Res Space Phys* **122(12)**: 12586–12590. https://doi.org/10.1002/2017JA024835.

Merkel D. 2014. Docker: lightweight linux containers for consistent development and deployment. *Linux J* **2014(239)**: 2. https://doi.org/10.5555/2600239.2600241.

National Academies of Sciences, Engineering and Medicine. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research.* The National Academies Press, Washington, DC. ISBN 978-0-309-47624-9. https://doi.org/10.17226/25116, URL https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century.

Owens MJ, Lang M, Riley P, Lockwood M, Lawless AS. 2020. Quantifying the latitudinal representivity of in situ solar wind observations. *J Space Weather Space Clim* **10**: 8. https://doi.org/10.1051/swsc/2020009.

Park W, Lee J, Kim K-C, Lee J, Park K, et al. 2021. Operational Dst index prediction model based on combination of artificial neural network and empirical model. *J Space Weather Space Clim* **11**: 38. https://doi.org/10.1051/swsc/2021021.

Rogers NC, Wild JA, Eastoe EF, Gjerloev JW, Thomson AWP. 2020. A global climatological model of extreme geomagnetic field fluctuations. *J Space Weather Space Clim* **10**: 5. https://doi.org/10.1051/swsc/2020008.

Tang R, Zeng F, Chen Z, Wang J-S, Huang C-M, Wu Z. 2020. The comparison of predicting storm-time ionospheric TEC by three methods: ARIMA, LSTM, and Seq2Seq. *Atmosphere* **11(4)**: https://doi.org/10.3390/atmos11040316.

Vadakke Veettil S, Cesaroni C, Aquino M, De Franceschi G, Berrili F, et al. 2019. The ionosphere prediction service prototype for GNSS users. *J Space Weather Space Clim* **9**: A41. https://doi.org/10.1051/swsc/2019038.

Wing S, Johnson JR. 2019. Applications of information theory in solar and space physics. *Entropy* **21(2)**: 40. https://doi.org/10.3390/e21020140.