JGR Machine Learning and Computation

EDITORIAL
10.1029/2024JH000184

Key Points:

- This editorial introduces *JGR: Machine Learning & Computation* to the community

Correspondence to:
E. Camporeale,
enrico.camporeale@colorado.edu

Citation:
Camporeale, E., Marino, R., & the Editorial Board (2024). Our vision for JGR: Machine Learning and Computation. *Journal of Geophysical Research: Machine Learning and Computation*, *1*, e2024JH000184. https://doi.org/10.1029/2024JH000184

Received 29 FEB 2024
Accepted 29 FEB 2024

Our Vision for JGR: Machine Learning and Computation

E. Camporeale[1,2], R. Marino[3], and the Editorial Board[4]

[1]Queen Mary University of London, London, UK, [2]Space Weather Technology, Research and Education (SWx TREC), University of Colorado, Boulder, CO, USA, [3]Laboratoire de Mécanique des Fluides et d'Acoustique, UMR 5509, CNRS, École Centrale de Lyon, INSA Lyon, Université Claude Bernard Lyon 1, Ecully, France, [4]See Appendix A

Abstract This editorial introduces the inaugural issue of the *Journal of Geophysical Research: Machine Learning and Computation* to the scientific community, elucidating the motivations and vision behind its establishment. The landscape of computational tools for geoscientists has undergone a rapid transformation in the last decade, akin to a new scientific revolution challenging the traditional scientific method. The paradigm shift emphasizes the integration of data-driven methods and the possibility of predicting and/or reproducing the evolution of natural phenomena with computers as the fourth pillar of scientific discovery, sparking debates on trustworthiness, and ethical implications. The data science revolution is fueled by the convergence of advancements, including the big-data revolution, GPU market expansion, and significant investments in Artificial Intelligence and high performance computing by both institutional and private players. This transformation has given rise to a trans-disciplinary community that has investigated a wide range of questions under the lens of machine learning (ML) approaches and has generally advanced the field of computational methods within the broader geosciences community, the core of the American Geophysical Union (AGU) membership. Responding to an unmet demand in the existing worldwide editorial offer, the *Journal of Geophysical Research: Machine Learning and Computation* aims to serve as an intellectual crucible, fostering collaborations across multiple geophysical disciplines and data scientists. The journal welcomes papers with strong methodological developments that allow for geoscience advancements grounded in specific computational and data-driven methods, leveraging ML as well as innovative computational strategies, and leading to breakthrough discoveries and original scientific outcomes. Authors are encouraged to balance succinctness in introducing methods with a thorough exploration of the novelty of the work proposed and its future applications placing special emphasis on the connection between the data science approach and the scientific outcome, considering a broad readership. Emphasis on result reproducibility aligns with AGU guidance, inviting active participation from the community in shaping geophysical research in the era of machine learning and computation.

1. Context

In this editorial, we are delighted to introduce the inaugural issue of the *Journal of Geophysical Research: Machine Learning and Computation* to the scientific community. The purpose of this discourse is to elaborate on the underlying intentions and the vision that propelled the establishment of this scholarly initiative.

It is manifestly apparent that a conspicuous transformation in the computational tool kit available to geoscientists occurred with remarkable rapidity within the last couple of decades. Since the dawn of computing, the physical sciences have been continuously adopting the latest advances in terms of hardware, software, and method developments. This continuous assimilation has often been accompanied by the establishment of novel journals deemed as apt conduits for the aggregation and dissemination of up-to-date knowledge. However, many in the community believe that what we are currently witnessing is much more than adding a few more tools to our belt, with Artificial Intelligence (AI) and the most advanced techniques for data analysis and simulation having become accessible and scalable, thanks to open science policies. Rather, the current transformation is more akin to a new scientific revolution that is shaking the foundation of the well-established paradigm we refer to as the *scientific method*. Without delving extensively into the captivating history of the original scientific revolution, often referred to as the Copernican Revolution, spanning multiple centuries and featuring the contributions of numerous extraordinary intellects, and thus oversimplifying the complex narrative, it is pertinent to acknowledge that the paradigm of scientific discovery has traditionally hinged upon a well-established cycle encompassing observations, hypotheses, predictions, and experiments.

A pivotal facet in the evolution of hypothesis formulation, dating back to the seminal work of Newton in his *Principia*, is the adoption of a mathematical framework. This choice not only reinforces the notion that mathematics is the natural language of science but also, at times, constrains these hypotheses within the confines of human intelligibility. Furthermore, hypotheses often exhibit a deliberate alignment with overarching theories that have withstood the test of time from an experimental standpoint, becoming rigid elements of the discovery process.

Over time, the complexity of the systems under study has escalated, often surpassing the convenience of "closed-form" mathematical solutions. Consequently, discernible efforts have been made toward the necessity of computational approaches. Notably, this evolution coincided with the development of computers, ushering in a new scientific revolution that gradually unfolded from the 1950s onward. This transformative journey has underscored the indispensability of computational methods to the extent that computing is now widely acknowledged as the third pillar of scientific discovery (theory and laboratory experiments and/or observations being the first two).

Here, we posit that a novel foundational component is being incorporated into the process of scientific discovery: data-driven methods are emerging as the fourth pillar. This proposition has sparked considerable debate and controversy, with one prevalent argument asserting that the scientific community should avoid the inclination toward marketing and the practice of re-branding existing concepts to garner renewed popularity (or, in other words, "Isn't machine learning just fancy statistics?"). Similarly, many argue that experiments, measurements, and observations—collectively referred to as data—have long held a crucial role in scientific inquiry, tracing back at least to Kepler's law of planetary motion (derived from Tycho Brahe's observations) and Galileo's experiments involving objects dropped from the leaning tower of Pisa to disprove the 2000-year-old Aristotelian view. In fact, every model, along with its supporting hypothesis, must undergo testing against observations, a principle famously known as Popper's falsifiability criterion (Popper, 2005). Given this perspective, one may contemplate whether modern methods categorized within the realm of "data science" are indeed introducing any novelty. We believe that the current demonstrations of data-driven methods are in fact novel and that we are at the beginning of a new era of applied computation in the physical sciences.

2. Data-Driven Methods in Science

The current data science revolution is fueled by the convergence of three distinct advancements, rendering our current era unparalleled in human history. The first catalyst is the well-known big-data revolution and subsequent *datafication* of our society (Biltgen & Ryan, 2016), which is typically encapsulated by the concept that, for the last decade or so, approximately 90% of the data in a given domain has been produced in the prior 2 years (i.e., an exponential growth). The second important impulse to the data science revolution is the proliferation of the GPU market and the increased availability of parallel and cloud computing hardware architectures to the broader consumer base, which is almost certainly going to increase with the advent of quantum computing. Lastly, the third pivotal factor involves substantial investments in AI by prominent IT companies, followed by the subsequent creation and release of open-source software in the realms of machine learning and AI.

The unprecedented combination of these three enabling factors has given rise to a novel community comprising a blend of machine learning experts, theorists, and practitioners. This community has, in essence, embarked on addressing a myriad of challenges that can be translated into machine learning inquiries. Likewise, new questions have been addressed with novel methodologies that reflect advances in fundamental sciences, tailored to mine information from data. Undoubtedly, this trans-disciplinary community initially prioritized challenges with potential profitability, such as recommendation systems and computer vision. However, it did not take long for the science community to notice and start adopting the latest advancements in ML and data-driven methods.

The list of recent advancement in the physical sciences that have been propelled by data-driven methods is long, including astrophysics and cosmology, material science, climate and weather prediction, nuclear physics, computational biology, plasma physics, geology, hydrology, and others. More importantly, all areas that form the core of AGU science have been profoundly impacted.

Returning momentarily to the examination of what constitutes a hypothesis in the context of scientific discovery and considering the assertion that the community has traditionally leaned toward the Newtonian perspective of mathematically describable hypotheses, it is evident that a new paradigm is emerging. The paradigm is heretical to some, and it fundamentally asserts the supremacy of data. A prime example of this new paradigm is seen in the prediction of complex nonlinear systems, for example, global weather forecasting, where "black-box" neural network models have demonstrated comparable performance to large-scale physics-based numerical prediction models while executing a few orders of magnitude times faster. The new paradigm is obviously not restricted to geoscience; AlphaFold has shown the ability to predict the structure of proteins with atomic precision at a large scale and within minutes, essentially solving a 50-year grand challenge and showcasing AI's potential to greatly expedite scientific breakthroughs, thereby propelling human advancement (Jumper et al., 2021).

As is typical with any new and possibly dissenting perspective, this emerging paradigm continues to ignite a spirited debate, delving into crucial issues such as the reproducibility and trustworthiness of outcomes derived from ML models. More broadly, the discourse encompasses considerations regarding the ethical implications associated with the widespread adoption of AI (Stall et al., 2023).

Finally, the idea of making the best use of both our centuries-old knowledge and the recent advancements in ML is rapidly making its way into the current discourse. This concept, often referred to as physics-guided or physics-informed machine learning, holds the promise of harnessing the strengths and mitigating the weaknesses inherent in each individual approach (Cuomo et al., 2022; Karniadakis et al., 2021; Karpatne et al., 2022; Shen et al., 2023). Complementary developments of high performance computing applications, which ingest ML, GPUs, quantum, and cloud computing as integral parts of broader solution approaches, hold similar promises.

3. The Need for a New Journal

In light of these considerations, a significant segment of the American Geophysical Union (AGU) community has encountered a void in the prevailing publishing landscape. Particularly, scientists who were pushing the envelope in their field, initially by merely figuring out what problems were potentially suited to be better tackled by ML and novel computational strategies and eventually, by proposing major innovations that built on unprecedented uses of data-driven methods and high performance computing, realized that their work was not an ideal fit for any of the existing scientific journals.

This means, on the one hand, that potentially ground-breaking work has, in many cases, been deemed out of scope in traditional journals or been reviewed and assessed by non-expert peers. On the other hand, when published, these contributions may not receive the optimal exposure required to fertilize their respective domains. Furthermore, the dispersion across various publishers and journals of articles belonging to the same family of works (e.g., application of ML methods in the geosciences) hampers the potential for cross-fertilization across contiguous fields.

The envisioned role of *JGR: Machine Learning and Computation* is thus to fill this vacuum, aspiring to serve as an intellectual crucible, fostering collaborative endeavors that bridge the traditional divide between geophysics and computational science. We aim to publish research that develops and explores innovative data-driven and computational methodologies based on mathematical models, statistical analysis, machine learning, and AI, with the aim of advancing knowledge in the domain of Earth and space sciences. A recurrent question in a prospective author's mind would be whether their work is more appropriate for a specialized journal in their discipline (e.g., one of the other journals in the *Journal of Geophysical Research* family) or for this new journal. There is no straightforward answer, and we expect that in some cases, manuscripts may contain elements to potentially fit well in a few AGU journals depending also on their length, the emphasis given to the methodological approach implemented and its originality, and the scientific fallout it produces. However, we welcome submissions of works that either have a strong component of methodological development or whose advancement in addressing an open question in geosciences is firmly anchored in a specific data-driven method or novel computational domain, possibly stemming from advances in physical and mathematical sciences.

We encourage authors to be succinct in introducing well-known methods or textbook material. One should assume that the readership is well-versed in data science, numerical, and computational techniques. Therefore, manuscripts do not need to review the basic concepts of well-known methodologies (e.g., models such as decision trees, neural networks, deep learning methodologies, or numerical schemes in general) but rather be self-contained in referencing appropriate papers and textbooks, focusing on describing the novelty of the approach and the original scientific outcome of the research presented.

At the same time, we strongly encourage authors to devote some space to explaining the novelty and the prospective future applications of the approach they put forth, knowing that experts in specific machine learning, data-driven, or advanced computational methods form a relatively small niche within a community. Therefore, published papers should strive to captivate and engage a broader readership by ensuring their content remains intriguing and accessible to a wider audience.

Finally, we strongly emphasize the reproducibility of the results and their validation against existing methods and models. Hence, we align with the Guidance for Authors Submitting to AGU Journals (https://data.agu.org/resources/agu-data-software-sharing-guidance), and we expect that all data and software necessary to reproduce any part of a manuscript is made available to editors and reviewers upon submission.

We invite the AGU community to actively participate in this exciting scholarly venture, contributing to the collective repository of knowledge that will undoubtedly shape the trajectory of geophysical research in an era in which machine learning and computational approaches exploiting emerging technologies will increasingly play a prominent role in science.

Appendix A

Thomas Berger[1], Yangkang Chen[2], Patricio Clark Di Leoni[3], Doris Folini[4], Xie Hu[5], Xiaofeng Li[6], Yangxiaoyue Liu[7], Donald D. Lucas[8], Ryan M. McGranaghan[9], Markus Reichstein[10], Chaopeng Shen[11], Steve Tobias[12], Yixin Wen[13]

[1]University of Colorado at Boulder, Space Weather Technology, Research, and Education Center, [2]The University of Texas at Austin, Jackson School of Geosciences, Bureau of Economic Geology, [3]Universidad de San Andrés, [4]Institute for Atmospheric and Climate Science, ETH Zurich, [5]Peking University, College of Urban and Environmental Sciences, [6]Institute of Oceanology, Chinese Academy of Sciences, [7]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, [8]Lawrence Livermore National Laboratory, Livermore, California, [9]NASA Jet Propulsion Laboratory, California Institute of Technology, [10]Max Planck Institute for Biogeochemistry, [11]The Pennsylvania State University, Civil and Environmental Engineering, [12]School of Mathematics, University of Leeds, [13]The University of Florida, Department of Geography, Meteorology and AI Lab

Data Availability Statement

There is no additional data or software to declare.

References

Biltgen, P., & Ryan, S. (2016). *Activity-based intelligence: Principles and applications*. Artech House.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, *92*(3), 88. https://doi.org/10.1007/s10915-022-01939-z

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Karpatne, A., Kannan, R., & Kumar, V. (2022). *Knowledge guided machine learning: Accelerating discovery using scientific knowledge and data*. CRC Press.

Popper, K. (2005). *The logic of scientific discovery*. Routledge.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, *4*(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Stall, S., Cervone, G., Coward, C., Cutcher-Gershenfeld, J., Donaldson, T. J., Erdmann, C., et al. (2023). *Ethical and responsible use of AI/ML in the earth, space, and environmental sciences*. Authorea Preprints. Retrieved from https://essopenarchive.org/doi/pdf/10.22541/essoar.168132856.66485758/v1