

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/336463250>

Thoughtfully Using Artificial Intelligence in Earth Science

Article in *Eos Transactions American Geophysical Union* · October 2019

DOI: 10.1029/2019EO135235

CITATIONS

4

READS

64

3 authors, including:



Imme Ebert-Uphoff

Colorado State University

87 PUBLICATIONS 2,311 CITATIONS

[SEE PROFILE](#)



Savini Samarasinghe

Colorado School of Mines

9 PUBLICATIONS 71 CITATIONS

[SEE PROFILE](#)

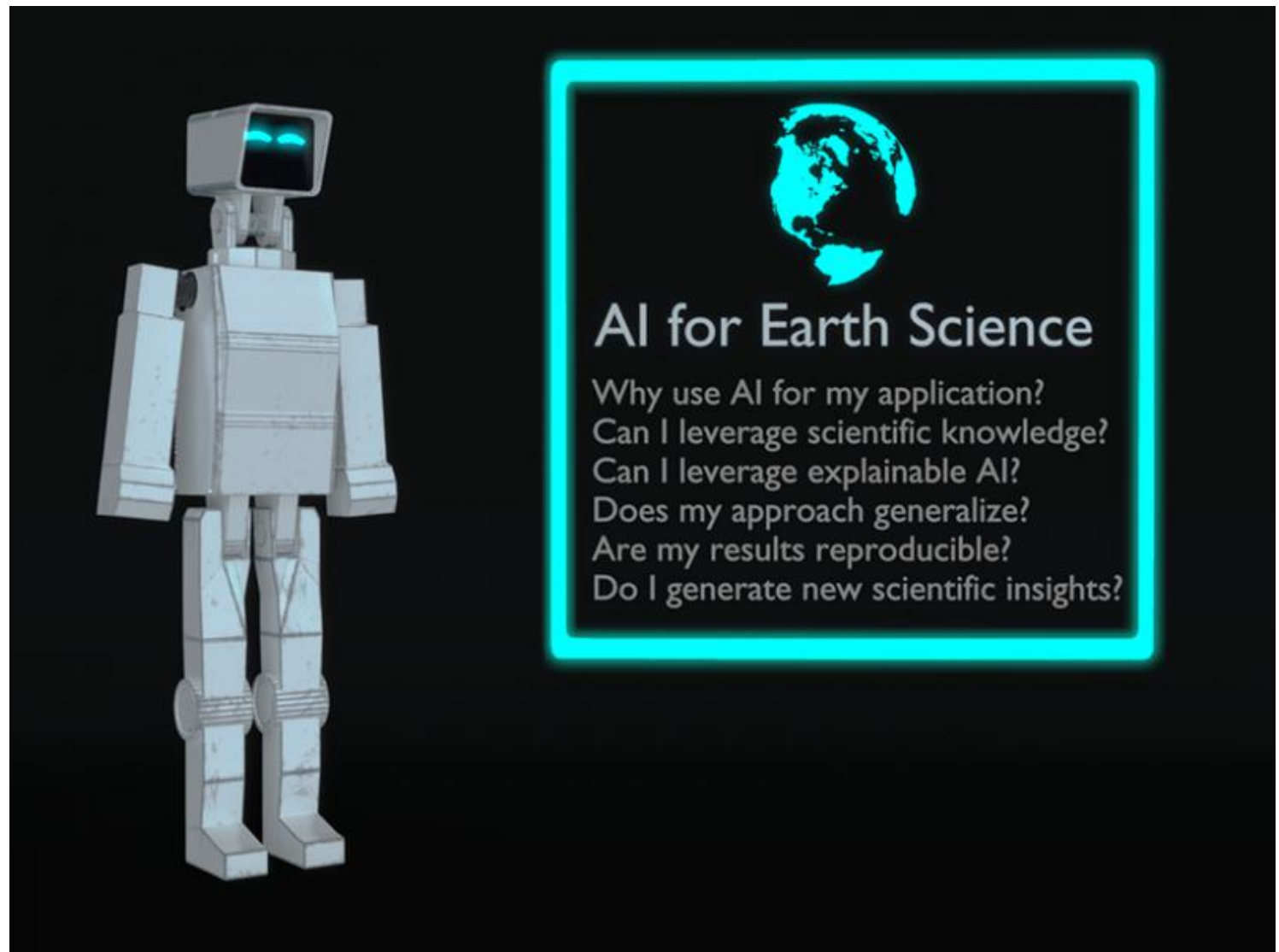
Some of the authors of this publication are also working on these related projects:



Machine learning & Causal Discovery for the Geosciences [View project](#)

Thoughtfully Using Artificial Intelligence in Earth Science

Deriving scientific insights from artificial intelligence methods requires adhering to best practices and moving beyond off-the-shelf approaches.



Before choosing an appropriate artificial intelligence approach for an Earth science application, key questions must be considered. Artificial intelligence, integrated with existing scientific knowledge and best practices for application, can provide scientifically sound analyses and predictions for use in the Earth sciences. Credit: Niels Uphoff © 2019

By Imme Ebert-Uphoff, Savini M. Samarasinghe, and Elizabeth A. Barnes © 11 October 2019

Artificial intelligence (AI) methods have emerged as useful tools in many Earth science domains (e.g., climate models (<https://eos.org/research-spotlights/next-generation-climate-models-could-learn-improve-on-the-fly>), weather prediction (<https://eos.org/opinions/artificial-intelligence-may-be-key-to-better-weather-forecasts>), hydrology (<https://eos.org/editors-vox/deep-learning-a-next-generation-big-data-approach-for-hydrology>), space weather (<https://eos.org/meeting-reports/space-weather-in-the-machine-learning-era>), and solid Earth (<https://eos.org/meeting-reports/machine-learning-in-geoscience-riding-a-wave-of-progress>)). AI methods are being used for tasks of prediction, anomaly detection, event classification, and onboard decision-making on satellites (<https://eos.org/articles/new-eyes-on>

wildfires), and they could potentially provide high-speed alternatives for representing subgrid processes in climate models [*Rasp et al.* (<https://www.pnas.org/content/115/39/9684>), 2018; *Brenowitz and Bretherton* (<https://doi.org/10.1029/2019MS001711>), 2019].

Although the use of AI methods has spiked dramatically in recent years, we caution that their use in Earth science should be approached with vigilance and accompanied by the development of best practices for their use. Without best practices, inappropriate use of these methods might lead to “bad science,” which could create a general backlash in the Earth science community against the use of AI methods. Such a backlash would be unfortunate because AI has much to offer Earth scientists, helping them sift through and gain new knowledge from ever-increasing amounts of data. Thus, it is time for the Earth science community to develop thoughtful approaches for the use of AI.

Easy Access to Powerful New Methods

Setting up and running experiments using AI methods used to require sophisticated computer science knowledge. This is no longer the case.

Setting up and running experiments with AI methods used to require sophisticated computer science knowledge. This is no longer the case. The recent success of AI in other domains has spurred the development of free and highly efficient software packages that are extremely easy to learn and use. Even complex artificial neural networks can be set up in a few lines of code, and countless tutorials and examples are available to guide the novice user. Furthermore, as algorithms become more efficient and computational power gets cheaper and more available on the cloud, access to high-performance computing is no longer a limiting factor. All of these developments bring powerful AI methods to Earth scientists’ fingertips.

Earth scientists have a long tradition of using methods based on physics (e.g., dynamical models) and sophisticated statistics (e.g., empirical orthogonal function analysis and spectral analysis). They have thus accepted statistical methods, which are a type of data-driven method, as useful tools. However, the sudden rise of AI methods—another type of data-driven method—in Earth science, coupled with a terminology and culture unfamiliar to Earth scientists, may make AI methods seem more foreign than they actually are. AI simply provides an extended set of new data-driven methods, many of which are derived from statistical principles. For example, one basic type of artificial neural network (deep learning) is essentially a linked series of linear regression models interspersed with scalar nonlinear transformations.

We address here the question of how best to leverage both physics-based and data-driven methods simultaneously by outlining several proposed steps for researchers. For brevity we use only the term “AI methods” below, although most of our discussion applies equally to all data-driven methods.

Step 1: Ask Guiding Questions

We suggest that Earth scientists ask themselves the following questions before choosing a specific AI approach:

Why exactly do I want to use AI for my application? Is this application for prediction, understanding, or both? The answer is important for choosing an AI method that satisfies the desired trade-off between transparency and performance.

How can scientific knowledge be integrated into the AI method? There are many ways to combine expert scientific knowledge of underlying physical processes (e.g., physics and chemistry) and AI methods; every effort should be made to merge these approaches, as discussed more below in step 2.

Which tools from explainable AI are available? The emerging field of explainable AI (XAI) provides many new tools for the visualization and interpretation of AI methods [*Samek et al.* (<https://link.springer.com/book/10.1007%2F978-3-030-28954-6>), 2019]. *McGovern et al.* (<https://doi.org/10.1175/BAMS-D-18-0195.1>) [2019], for example, show the enormous potential of these tools for weather-related applications. These tools have the potential to transform the use of AI methods in Earth science by increasing transparency and thus building trust in their reasoning.

Does my approach generalize to address all conditions in which it will be used? AI methods rely on “training data” to learn the characteristics of a system. Special attention must be paid to testing and ensuring that the resulting AI model works under changing conditions, including regime shifts. Generalization can be greatly enhanced by fusing scientific knowledge and can be tested by methods ranging from cross validation to the AI technique of generating adversarial examples.

Is my approach reproducible? Am I following the findable, accessible, interoperable, and reusable (FAIR (<https://eos.org/agu-news/enabling-fair-data-across-the-earth-and-space-sciences>)) data principles? Is my method easily accessible to the community?

What am I hoping to learn in terms of science? Because of the complexity and abstract nature of many AI techniques, the answer to this question is crucial to set up the problem and method so that there is something to learn from the AI approach and to be able to interpret the results. The availability of explainable AI tools discussed above creates many novel opportunities to gain new scientific insights into Earth science processes (E. A. Barnes et al., Viewing forced climate patterns through an AI lens, submitted to *Geophysical Research Letters*, 2019).

We encourage researchers to thoroughly reflect on these questions to select the best AI method for their application. Furthermore, to promote substantial advances in scientific research, editors of Earth science journals may need to create guidelines for the review of AI-focused manuscripts to ensure that findings are explained clearly and placed in the context of existing Earth science. Likewise, editors might encourage comparison to standard or simpler approaches to discern the scientific advances that AI offers.

Step 2: Explore Fusing Scientific Knowledge into AI

The field of theory-guided data science investigates ways in which AI and scientific knowledge can be combined into hybrid algorithms that incorporate the best of both worlds.

The field of theory-guided data science investigates ways in which AI and scientific knowledge can be combined into hybrid algorithms that incorporate the best of both worlds [*Karpatne et al.* (<https://ieeexplore.ieee.org/document/7959606>), 2017]. Earth science applications have several properties that make it imperative to integrate scientific knowledge as much as possible, such as

- the desire of Earth scientists to gain scientific insights rather than just “get numbers” from an algorithm
- the availability of extensive existing knowledge
- the high complexity of the Earth system
- the limited sample size and lack of reliable labels in many Earth science applications

Integrating scientific knowledge into AI approaches greatly improves transparency because the more scientific knowledge is used, the easier it is to follow the reasoning of the algorithm. Generalization, robustness, and performance are also improved because the scientific knowledge can fill many gaps left by small sample sizes, as explained below.

How can we integrate scientific knowledge? One excellent practice uses a two-step approach. For a given task, first identify all subtasks that can easily and efficiently be addressed by physics-driven methods, and apply those. Although this guideline may appear obvious, it is often overlooked, likely because applying an off-the-shelf AI algorithm to the entire task appears to require less work than carefully analyzing and addressing several subtasks.

For remaining subtasks, consider using AI algorithms while still leveraging scientific knowledge. Most AI methods have some kind of optimization procedure at their core; that is, they search for an optimal model over a large parameter space. Without using scientific knowledge, the search space is often large, and many solutions may exist, only some of which might be physically meaningful. By leveraging known physical relationships (e.g., by including them as constraints in the optimization problem), the optimization is guided toward only physically meaningful solutions (T. Beucler et al. (<https://arxiv.org/abs/1909.00912v2>), Enforcing analytic constraints in neural-networks emulating physical systems, arXiv:1909.00912). This approach offers additional benefits: Convergence tends to be faster because of the smaller search space, and the resulting method tends to generalize better because of the use of established physical relationships. Thus, although this approach requires more work at the onset, it tends to result in much better overall solutions.

Step 3: Foster Interdisciplinary Collaboration and Education

Innovative approaches, such as fusing scientific knowledge and AI methods, require deep knowledge integration across disciplinary boundaries [Pennington (<https://link.springer.com/article/10.1007/s13412-015-0354-5>), 2015]. This integration is best achieved by close collaboration between Earth scientists and AI researchers (see [guidelines for such collaborations](https://eos.org/opinions/three-steps-to-successful-collaboration-with-data-scientists) (<https://eos.org/opinions/three-steps-to-successful-collaboration-with-data-scientists>)). The [Association for Interdisciplinary Studies](https://oakland.edu/ais/) (<https://oakland.edu/ais/>) organizes conferences and provides general strategies for effective interdisciplinary collaboration.

There is a tremendous need to develop guidelines and best practices to prepare the future Earth science workforce for innovative, interdisciplinary research bridging Earth science and AI.

It might not always be possible to find suitable collaborators, so one option is to join learning communities, such as the National Science Foundation–sponsored EarthCube Research Coordination Network [IS-GEO: Intelligent Systems Research to Support Geosciences](https://is-geo.org/) (<https://is-geo.org/>). The increasing number of sessions (e.g., coordinated by [AGU's Earth and Space Science Informatics](https://essi.agu.org/) (<https://essi.agu.org/>) section), workshops (e.g., [Climate Informatics](http://climateinformatics.org/) (<http://climateinformatics.org/>)), and conferences (e.g., the American Meteorological Society's [Conference on Artificial Intelligence for Environmental Science](https://annual.ametsoc.org/index.cfm/2020/programs/conferences-and-symposia/19th-conference-on-artificial-intelligence-for-environmental-science/) (<https://annual.ametsoc.org/index.cfm/2020/programs/conferences-and-symposia/19th-conference-on-artificial-intelligence-for-environmental-science/>)) dedicated to AI research in Earth science is encouraging, yet there is still a large need for additional events that engage Earth science and AI researchers *simultaneously* and build bridges between these communities.

Furthermore, there is a tremendous need to develop guidelines and best practices to educate the future Earth science workforce to be well prepared for innovative, interdisciplinary research bridging Earth science and AI. Numerous institutions are starting to incorporate data science and AI courses into their curricula (e.g., Cornell's [Institute for Computational Sustainability](https://computational-sustainability.cis.cornell.edu/) (<https://computational-sustainability.cis.cornell.edu/>) and National Science Foundation Research Traineeship programs at the [University of Chicago](https://epic.uchicago.edu/education/programs-concentrations/data-science-energy-and-environmental-research) (<https://epic.uchicago.edu/education/programs-concentrations/data-science-energy-and-environmental-research>), [University of California, Berkeley](http://ds421.berkeley.edu/) (<http://ds421.berkeley.edu/>), and [Northwestern University](https://www.ideas.ciera.northwestern.edu/ids/) (<https://www.ideas.ciera.northwestern.edu/ids/>)). The community can support these efforts by collecting and disseminating educational resources and by developing guidelines on which topics are most beneficial to integrate into Earth science education [Pennington *et al.* (<https://doi.org/10.1007/s11625-019-00735-3>), 2019].

The availability of AI methods provides many new and exciting research avenues for Earth scientists, but it also requires the community to reflect on when and how these methods should be used because using such methods without careful consideration can lead to bad science. The two most promising safeguards against such bad science are the integration of scientific knowledge into AI methods and the use of visualization tools to maximize their transparency.

Acknowledgments

Support was provided by National Science Foundation grants AGS-1749261 (E.A.B.) and AGS-1445978 (I.E.).

References

- Brenowitz, N. D., and C. S. Bretherton (2019), Spatially extended tests of a neural network parametrization trained by coarse-graining, *J. Adv. Model. Earth Syst. (JAMES)*, **11**, 2,728–2,744, <https://doi.org/10.1029/2019MS001711> (<https://doi.org/10.1029/2019MS001711>).
- Karpatne, A., et al. (2017), Theory-guided data science: A new paradigm for scientific discovery from data, *IEEE Trans. Knowl. Data Eng.*, **29**(10), 2,318–2,331, <https://doi.org/10.1109/TKDE.2017.2720168> (<https://doi.org/10.1109/TKDE.2017.2720168>).
- McGovern, A., et al. (2019), Making the black box more transparent: Understanding the physical implications of machine learning, *Bull. Am. Meteorol.*, <https://doi.org/10.1175/BAMS-D-18-0195.1> (<https://doi.org/10.1175/BAMS-D-18-0195.1>), in press.

Pennington, D. (2015), A conceptual model for knowledge integration in interdisciplinary teams: Orchestrating individual learning and group processes, *J. Environ. Stud. Sci.*, 6(2), 300–312, <https://doi.org/10.1007/s13412-015-0354-5> (<https://doi.org/10.1007/s13412-015-0354-5>).

Pennington, D., et al. (2019), Bridging sustainability science, Earth science, and data science through interdisciplinary education, *Sustainability Sci.*, <https://doi.org/10.1007/s11625-019-00735-3> (<https://doi.org/10.1007/s11625-019-00735-3>), in press.

Rasp, S., et al. (2018), Deep learning to represent subgrid processes in climate models, *Proc. Natl. Acad. Sci. U. S. A.*, 115(39), 9,684–9,689, <https://doi.org/10.1073/pnas.1810286115> (<https://doi.org/10.1073/pnas.1810286115>).

Samek, W., et al. (2019), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, Springer, Cham, Switzerland, <https://doi.org/10.1007/978-3-030-28954-6> (<https://doi.org/10.1007/978-3-030-28954-6>).

Author Information

Imme Ebert-Uphoff (iebert@colostate.edu (<mailto:iebert@colostate.edu>)), Cooperative Institute for Research in the Atmosphere and Department of Electrical and Computer Engineering, Colorado State University, Fort Collins; Savini M. Samarasinghe, Department of Electrical and Computer Engineering, Colorado State University, Fort Collins; and Elizabeth A. Barnes, Department of Atmospheric Science, Colorado State University, Fort Collins

Citation: Ebert-Uphoff, I., S. M. Samarasinghe, and E. A. Barnes (2019), Thoughtfully using artificial intelligence in Earth science, *Eos*, 100, <https://doi.org/10.1029/2019EO135235>. Published on 11 October 2019.

Text © 2019. The authors. [CC BY-NC-ND 3.0](#)

Except where otherwise noted, images are subject to copyright. Any reuse without express permission from the copyright owner is prohibited.

This article does not represent the opinion of AGU, *Eos*, or any of its affiliates. It is solely the opinion of the author.
