Eos®

A Two-Step Approach to Training Earth Scientists in AI

Researchers learned machine learning methods during a boot camp, then applied their new knowledge to real-world research problems during a hackathon.

By **Lexie Goldberger**, **Peishi Jiang**, **Tirthankar "TC" Chakraborty**, **Andrew Geiss,** and **Xingyuan Chen**
29 April 2025

Artificial intelligence (AI) has a multitude of uses in Earth science research, once you know how to use the tools. This image, generated by AI using stock imagery, was inspired by a boot camp for Earth scientists interested in expanding their own AI tool kits. Credit: Nathan Johnson/Pacific Northwest National Laboratory; Inset images (clockwise from lower left): Gorodenkoff/Shutterstock.com, Evgeny_V/Shutterstock.com, Dino Adventure/Shutterstock.com, Nikolay Zaborskikh/Shutterstock.com,

You can't teach an old dog new tricks, but can you teach the current generation of Earth scientists about emerging artificial intelligence and machine learning (AI/ML) methods relevant to their research? From our experience helping run a program intended to do just that at the U.S. Department of Energy's (DOE) Pacific Northwest National Laboratory (PNNL), the answer is yes.

Earth scientists, from those focused on the atmosphere or ocean to those studying the continents or deep subsurface, often work with extremely large—sometimes global—datasets, trying to find patterns among noisy real-world observations. AI/ML **is well suited** for **such tasks**.

> "
>
> *Relatively few Earth scientists have been trained in artificial intelligence and machine learning (AI/ML) methods, meaning unfulfilled opportunities exist to learn from the growing volumes of Earth science data available.*

AI/ML approaches have recently been used, for example, to replace slow, numerical representations of rainfall in a global general circulation model [*Gettelman et al.*, 2021]. Similarly, AI/ML image detection techniques have been used with weather radar datasets to better predict short-term rainfall [*Ji and Xu*, 2024]. Yet relatively few domain scientists in the field have been trained in these methods, meaning unfulfilled opportunities exist to learn from the growing volumes of Earth science data available.

Several hundred data scientists work at PNNL, and for more than a decade, the lab has developed AI/ML approaches to address critical challenges in scientific discovery, energy resilience, and national security. Recent advancements in computational techniques and methodologies have sparked renewed interest in **applying AI/ML across various disciplines**. However, connecting the expertise of PNNL's data scientists to Earth science research at the lab—encompassing atmospheric, hydrological, and environmental sciences—has been a challenge.

Beginning in 2022, researchers at PNNL implemented a two-step approach—a boot camp followed by a hackathon—to prepare their colleagues to incorporate AI/ML into their research effectively. Eighty percent of those who participated in both events are now using ML techniques in their research, and the experience has boosted collaboration between the lab's data scientists and Earth scientists. The program has also led to innovative new projects, and its initial success suggests it may be a useful model for other organizations.

Boot Camp

Prior to PNNL initiating the program, many of the lab's Earth scientists expressed interest in learning more about AI/ML and exploring its applicability for addressing a wide variety of science questions.

Atmospheric science in particular **offers ideal ground** for teaching and applying ML methods because these methods are conducive to tackling many common tasks in the field. For example, they can help fill patchy datasets, such as in time series of satellite imagery [*Appel*, 2024]; correct biases in gridded data (e.g., overestimations of solar radiation reaching Earth in reanalysis products) [*Chakraborty and Lee*, 2021]; merge measurements of atmospheric properties into numerical models [*Krasnopolsky*, 2023]; and iteratively improve models [*Irrgang et al.*, 2021]. Furthermore, the field is ripe with the sort of very large, high-quality datasets that are necessary for applying modern ML methods.

The staff's interest and the clear relevance of AI/ML for their work motivated development of an initial 10-week boot camp, held in fall 2022, with weekly hybrid (online and in-person) sessions attended by 30–50 people. We enlisted 10 in-house data scientists to design **lessons, hands-on tutorials, and activities** covering a range of AI/ML methods and tools.

The first four sessions introduced participants to the basics of ML, with each session building upon the previous one and focusing on more state-of-the-art approaches. The remaining sessions covered popular **deep learning techniques** such as convolutional neural networks (CNNs), generative adversarial networks, transformers, and recurrent neural networks. They also covered topics such as how to use the ML libraries **Keras** and **PyTorch**, which offer the tools to run these models and other useful resources.

> " *As a result of the boot camp approach, participants gained understanding and appreciation of data curation for AI/ML and the full gamut of AI/ML methods they could use in their research.*

To connect the lessons to the participants' research interests, each one featured an Earth science–relevant activity, such as using maps of monthly sea surface temperature anomaly data from NOAA satellites with unsupervised learning algorithms to detect the phases of the El Niño–Southern Oscillation (i.e., El Niño and La Niña). The instructors developed and guided participants through virtual notebook environments that included fundamental information (with references) about the topic of the activity and heavily commented model code that could be run interactively. Time was also allotted for participants to better familiarize themselves with the models by running them in parallel on their own research computing environments.

As a result of the boot camp approach, participants gained understanding and appreciation of data curation for AI/ML and the full gamut of AI/ML methods they could use in their research. One remarked that they were impressed by the diversity of applications for ML and said, "I can tell if I continue to work on this skill, it will open a lot of doors and funding opportunities in the future for me." Another commented, "By the end, I felt my programming skills had improved as well."

Together with colleagues, one scientist at the lab who took part in the training applied knowledge and code directly from the boot camp material in research exploring stochasticity in aerosol-cloud interactions using field campaign data [*Li et al.*, 2024].

The instructors also reported that participating in the boot camp was worthwhile for several reasons. Each of their lessons and student demonstrations were reviewed by the other instructors, which fostered connections among peers knowledgeable in ML. According to one instructor, teaching their fellow staff also "helped provide context of how valuable my expertise is here at the lab."

> "
> *Additional hands-on opportunities were necessary to bridge the gap between learning ML and putting it into practice. So we organized a second learning opportunity—this time a hackathon.*

In addition, creating and presenting the weekly lesson plans to an audience with limited knowledge about AI/ML offered opportunities for instructors to improve their teaching skills. Furthermore, the adaptability of the instructional materials to other domain sciences supports the materials' value, longevity, and easy reuse in future trainings and research.

One year after the boot camp, participant responses to a questionnaire indicated that though many had gained literacy in ML, most had not taken the next step to start incorporating ML methods into their research. The results also showed that additional hands-on opportunities were necessary to bridge the gap between learning ML and putting it into practice. So we organized a second learning opportunity—this time a hackathon—focused on pairing ML experts and data scientists with domain scientists who share common research interests.

The Hackathon

Twenty-five domain and data scientists, many of whom had participated in the boot camp, took part in the 6-week hackathon, which began in January 2024. The domain scientists involved work in various areas of Earth science and as part of DOE projects such as the **Atmospheric Radiation Measurement** user facility and the PNNL-led **Addressing Challenges in Energy: Floating Wind in a Changing Climate** (a DOE Energy Earthshot research center), as well as NASA's **Aerosol Cloud Meteorology Interactions over the western Atlantic Experiment** project.

In preparing for the course, we discovered that these scientists often had trouble formulating research questions suited to ML methods and selecting which ML method to use. Prehackathon brainstorming sessions proved critical to success. During the first prehackathon meeting, the organizing committee gathered participants virtually to group the domain scientists by their topics of interest—vegetation-atmosphere interactions, clouds and precipitation, aerosols and aerosol-cloud interactions, hydrology, and wind energy—and to brainstorm potential research questions to address.

Each of the five groups then pitched project ideas to the participating ML experts and data scientists, who selected which team to join. With the teams assembled, each further workshopped a research question within their topic focus area—as well as which ML methods to use—that they could address within the duration of the hackathon. For example, one team chose to use a CNN model to identify open- versus closed-cell atmospheric convection in radar data, which helps explain distributions of clouds and rainfall.

During the hackathon, all the teams met weekly to discuss progress and exchange ideas for continuing work. This assessment method allowed the domain scientists to engage further with experts in the PNNL ML community, who provided feedback and answers to follow-up questions, such as how to prepare data for use in the ML models. Data preparation proved to be the most time-consuming step for the domain scientists because of the challenges of correctly formatting time series and gridded atmospheric datasets (e.g., temperature, relative humidity, and pressure) before they were fed into the models.

At the end of the 6 weeks, four of the five project groups had successfully processed their data and run them through their models to achieve results related to their initial questions. The fifth group, upon reflection, agreed that selecting an overly broad research question hindered progress on their project. Their experience underscored the importance of clearly defining a focused research question—and an appropriate ML approach—with cross-disciplinary consultation among scientists.

Soon after the hackathon concluded, a representative from each team presented their project during a seminar. A postseminar Q&A about the projects with staff who had not participated in the hackathon was positive and engaging, indicating a base level understanding of AI/ML methods within the division that was not present before the boot camp.

Fostering an AI-Literate Workforce

With growing datasets of Earth observations and ongoing computing advancements, AI/ML is an increasingly useful tool to aid in skillfully assessing conditions and processes in the Earth system.

Jingjing Tian presents results from the hackathon at the HydroML Symposium in May 2024. Her project involved training a convolutional neural network (CNN) model to detect open versus closed convection using weather radar data. Credit: Andrea Starr/Pacific Northwest National Laboratory

At PNNL, more than 20% of the research workforce is advancing AI and its applications in science. The initial goal of the recent training activities was to further grow ML expertise and implementation specifically within the lab's Atmospheric, Climate, and Earth Sciences (ACES) division. The lessons and successes of these activities suggest that other organizations similarly seeking to expand their use of AI/ML may benefit from the model of PNNL's approach.

The boot camp created a long-term, structured environment for a large number of staff to better understand the increasingly complex ML landscape, whereas the follow-up hackathon allowed a smaller group of eager staff to be coached in a faster-paced environment to produce deliverables. The different approaches of the boot camp and the hackathon allowed instructors to meet participants at their preferred comfort level and cater to different learning styles.

“

The different approaches of the boot camp and the hackathon allowed instructors to meet participants at their preferred comfort level and cater to different learning styles.

The results demonstrate that although learning new skills in AI/ML takes time, the effort is worthwhile and a collaborative, cross-disciplinary environment accelerates such learning. Staff self-reported that work done during the boot camp and hackathon had resulted in three conference presentations, including at the **HydroML 2024 Symposium**, and two publications (another is still in preparation).

Furthermore, PNNL reported an uptick in proposals from its Earth scientists for various internal funding opportunities focused on leveraging AI/ML methods. More proposals means more competition for funding, which should drive innovation and ultimately lead to stronger projects moving forward.

Another lesson from our experience was that sourcing instructors from within PNNL (i.e., ML experts who are already colleagues of Earth scientists in the ACES division) facilitated future collaborations between data and domain scientists and new research opportunities that wouldn't have been possible previously. One of the participating AI/ML experts noted to us that "after the hackathon, many lab scientists reached out to me for help in implementing ML/AI algorithms into their work," leading to multiple collaborations.

Hackathon participant **Sha Feng**'s comments offer additional, anecdotal evidence of the success of PNNL's program: "Participating in the hackathon has been a transformative experience," Feng said. "By bridging the gap between atmospheric science and data science, we have created a foundation for future projects that leverage the strengths of both fields."

We plan to continue to bridge such gaps at PNNL—and we support other organizations doing the same—to advance applications of AI/ML to address crucial questions about our planet, from the atmosphere to the ocean to the solid Earth.

Acknowledgments

References

Appel, M. (2024), Efficient data-driven gap filling of satellite image time series using deep neural networks with partial convolutions, *Artif. Intell. Earth Syst.*, *3*, e220055, **https://doi.org/10.1175/AIES-D-22-0055.1**.

Chakraborty, T. C., and X. Lee (2021), Using supervised learning to develop BaRAD, a 40-year monthly bias-adjusted global gridded radiation dataset, *Sci. Data*, *8*(1), 238, **https://doi.org/10.1038/s41597-021-01016-4**.

Gettelman, A., et al. (2021), Machine learning the warm rain process, *J. Adv. Model. Earth Syst.*, *13*(2), e2020MS002268, **https://doi.org/10.1029/2020MS002268**.

Irrgang, C., et al. (2021), Towards neural Earth system modelling by integrating artificial intelligence in Earth system science, *Nat. Mach. Intell.*, *3*, 667–674, **https://doi.org/10.1038/s42256-021-00374-3**.

Ji, C., and Y. Xu (2024), trajPredRNN+: A new approach for precipitation nowcasting with weather radar echo images based on deep learning, *Heliyon*, *10*(18), e36134, **https://doi.org/10.1016/j.heliyon.2024.e36134**.

Krasnopolsky, V. (2023), Review: Using machine learning for data assimilation, model physics, and post-processing model outputs, *Off. Note 513*, 32 pp., Natl. Cent. for Environ. Predict., College Park, Md., **https://doi.org/10.25923/71tx-4809**.

Li, X.-Y., et al. (2024), On the prediction of aerosol-cloud interactions within a data-driven framework, *Geophys. Res. Lett.*, *51*, e2024GL110757, **https://doi.org/10.1029/2024GL110757**.

Author Information

Lexie Goldberger, Peishi Jiang, Tirthankar "TC" Chakraborty, Andrew Geiss, and Xingyuan Chen (**xingyuan.chen@pnnl.gov**), Pacific Northwest National Laboratory, Richland, Wash.

What do you think?

6 Responses

👍
Upvote

😝
Funny

😍
Love

😮
Surprised

😣
Angry

😢
Sad

0 Comments

1 **Login** ▼

G

Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS (?)

Name

♡ **Share**

Best **Newest** **Oldest**