

# Training the Next Generation of Seismologists: Delivering Research-Grade Software Education for Cloud and HPC Computing Through Diverse Training Modalities

Marine A. Denolle<sup>\*</sup>  *et al.*

## Abstract

With the rise of data volume and computing power, seismological research requires more advanced skills in data processing, numerical methods, and parallel computing. We present the experience of conducting training workshops in various forms of delivery to support the adoption of large-scale high-performance computing (HPC) and cloud computing, advancing seismological research. The seismological foci were on earthquake source parameter estimation in catalogs, forward and adjoint wavefield simulations in 2D and 3D at local, regional, and global scales, earthquake dynamics, ambient noise seismology, and machine learning. This contribution describes the series of workshops delivered as part of research projects, the learning outcomes for participants, and lessons learned by the instructors. Our curriculum was grounded on open and reproducible science, large-scale scientific computing and data mining, and computing infrastructure (access and usage) for HPC and the cloud. We also describe the types of teaching materials that have proven beneficial to the instruction and the sustainability of the program. We propose guidelines to deliver future workshops on these topics.

**Cite this article as** Denolle, M. A., C. Tape, E. Bozdağ, Y. Wang, F. Waldhauser, A.-A. Gabriel, J. Braunmiller, B. Chow, L. Ding, K.-F. Feng, *et al.* (2025). Training the Next Generation of Seismologists: Delivering Research-Grade Software Education for Cloud and HPC Computing Through Diverse Training Modalities, *Seismol. Res. Lett.* **XX**, 1–15, doi: [10.1785/0220240413](https://doi.org/10.1785/0220240413).

## Introduction

Seismological research is advancing rapidly with the rise of computational power and big data, similar to other branches of geosciences (Morra *et al.*, 2021). Seismological research encompasses a vast range of scientific inquiries and methodological practices. Driven by often sparse but fundamental observations of earthquake phenomena at all spatial and temporal scales, seismological research has historically relied mostly on first-principle theories that supported observations. Higher education in earthquake sciences builds on this rich legacy. Most undergraduate and graduate curricula are centered around foundational textbooks, such as “Introduction to Seismology” by Shearer (2019), “Introduction to Seismology, Earthquakes, and Earth Structure” by Stein and Wysession (2009), or advanced seismological theory, such as “Quantitative Seismology” by Aki and Richards (2002). These theoretical foundations for seismological research are typically taught in class lecture settings.

Numerical methods and the rise of high-performance computing (HPC) have fueled the development of computational seismology, notably to solve the wave equation in complex

media (e.g., Olsen and Archuleta, 1996; Bao *et al.*, 1998; Graves, 1998; Komatitsch and Vilotte, 1998; Komatitsch *et al.*, 2002) and coupled to complex source models for purposes of physics-based ground-motion simulations (e.g., Graves *et al.*, 2011) and for seismic imaging (e.g., Liu and Gu, 2012; Tromp, 2020). As examples, the open-source SPECFEM package (e.g., Komatitsch and Tromp (2002a); Komatitsch *et al.* (2004), see [Data and Resources](#)) has supported a new era of passive-source (earthquake, ambient noise) full-waveform inversion (FWI) (e.g., Tape *et al.*, 2009; Peter *et al.*, 2011; Bozdağ *et al.*, 2016; Chow *et al.*, 2020) and the open-source SeisSol software (see [Data and Resources](#)) enables realistic simulations of 3D earthquake rupture dynamics (e.g., Käser *et al.*, 2010; Pelties *et al.*, 2012, 2014; Krenz *et al.*, 2021; Gabriel *et al.*, 2023; Uphoff *et al.*, 2024).

Full author list and affiliations appear at the end of this article.

<sup>\*</sup>Corresponding author: [mdenolle@uw.edu](mailto:mdenolle@uw.edu)

© Seismological Society of America

Big data seismology is also vastly expanding because continuous seismic data are recorded by more and more permanent stations worldwide, tens of thousands at the time of writing, and many more in temporary deployments. New methods emerged to include array processing (e.g., [Rost and Thomas, 2009](#)), ambient field (noise) seismology (e.g., [Nakata et al., 2019](#)), and machine learning (e.g., [Kong et al., 2019](#); [Mousavi and Beroza, 2022](#)). Discoveries of new tectonic and environmental phenomena invigorate the collection of large seismic datasets, leading to an exponential growth in data volumes and bringing our community to an era of petabyte-scale archives ([Arrowsmith et al., 2022](#)). Novel computing infrastructures such as cloud computing are particularly well suited for big data seismological research ([MacCarthy et al., 2020](#); [Krauss et al., 2023](#); [Ni et al., 2023](#)).

The broad adoption of open-source software in seismology based on Python (e.g., [Beyreuther et al., 2010](#)) or Julia (e.g., [Jones et al., 2020](#)), as well as version control hosted on GitHub, Bitbucket, and GitLab is transforming research practice and standards ([Barker et al., 2022](#); [Chue Hong et al., 2022](#)). Scientific journals require publicly hosted repositories or software availability. The Jupyter project ([Pérez and Granger, 2007](#); [Pimentel et al., 2019](#)) encompasses a suite of interactive computing tools: JupyterLab, a modern, integrated development environment that unifies notebooks, code editors, and more; Jupyter Notebook, the classic, document-focused interface for interactive computing; and JupyterHub, a server that enables multiuser access to these environments.

Educational approaches responding to the rise of computational and big data seismology have mostly leveraged advanced theoretical seismology and well-established numerical methods at the graduate student level. Computational Infrastructure for Geodynamics (CIG) has established best practices for both software development and training workshops ([Computational Infrastructure for Geodynamics \[CIG\], 2016a,b](#)). “Computational Seismology” by Heiner Igel ([Igel, 2017](#)) and the associated Coursera course on “Computers, Waves, Simulations: A Practical Introduction to Numerical Methods using Python” (see [Data and Resources](#)) has effectively equipped STEM graduate students with the skills needed to solve the wave equation with a syllabus that blends numerical methods with seismological research problems. The textbook provides Jupyter Notebooks, is entirely open source in Python, and can be run for simple problems from the associated Binder hub ([Krischer et al., 2018](#)). Despite this, we see a growing gap between higher education curricula and research practice. Open science and novel cyberinfrastructure (CI) present opportunities to train students and researchers in current research practices.

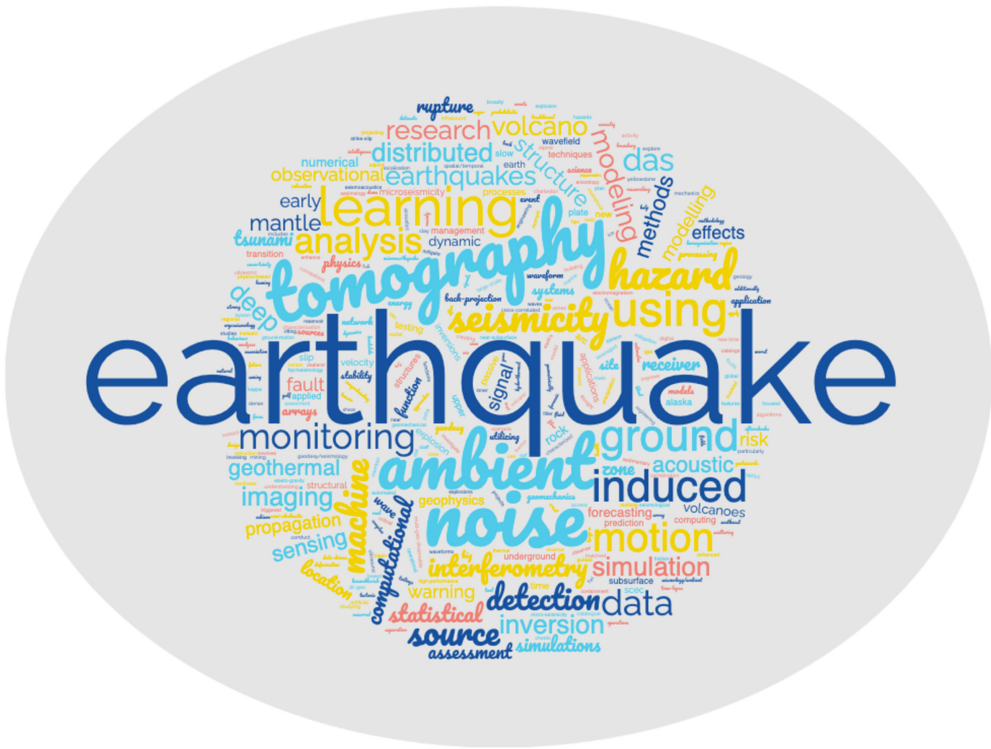
The COVID-19 pandemic has transformed education: students and teachers had to transition from in-person to remote, online learning. Several efforts have contributed to improving remote access to seismology education, such as the ROSES

program ([Brudzinski et al., 2021](#)). These contributions have democratized education through pedagogical approaches that analyze small datasets, utilize approximate solutions, or perform modest simulations using single nodes and Jupyter Notebooks. However, a gap remains in the adoption of advanced computing platforms, such as HPC infrastructure and cloud computing. This article presents recent developments by the project Seismological COmputational Platform to Empower Discovery (SCOPED, [Tape et al., 2022](#); [Wang et al., 2023](#); [Denolle et al., 2024](#)) and collaborations with other projects (e.g., the Statewide California Earthquake Center (SCEC), EarthScope, and the European projects Geo-Inquire, DT-GEO and ChEESE-2P, [Folch et al., 2023](#)) to help close that gap for students and researchers with multimodal educational efforts.

The goal of the SCOPED project is to develop a CI that enables hybrid model–data research in seismology by utilizing cloud and HPC infrastructures, open-source software, and containerization. Software containerization is a lightweight virtualization of software and its dependencies into a portable, isolated environment, ensuring consistency across different computing environments. Research enabled by SCOPED includes (1) machine learning-enhanced earthquake source characterization and catalog building, (2) FWI for source mechanisms, (3) FWI for Earth imaging across scales, and (4) time-lapse imaging of the subsurface. The SCOPED community expressed their research interest, which we illustrate with Figure 1. This article discusses the workshops held as part of the SCOPED project (Table 1), and in particular, by its use of containers. The main goal was for workshop participants to learn about research software and how to access and use HPC resources, clusters from HPC centers, and resources from the cloud.

## A Broad Survey of the Seismology Community

As part of SCOPED, we ran multiple surveys to gauge the community interest in using widely employed seismological software in advanced computing environments such as HPC and the cloud before the workshops (see Table 1). The workshops were delivered in virtual, hybrid, and in-person formats with a target audience of researchers with advanced graduate-level skills in seismology. Surveys and workshops were announced in domain-specific mailing lists, such as the EarthScope Consortium and SCEC, and social media platforms (X, formerly Twitter and LinkedIn). The surveys were tailored to each workshop, and the data collected mainly focused on familiarity with the required technical skills. Overall, we received 976 responses, although some may come from the same individuals. Our workshops had a total of 574 participants, with over 130 joining in person. Some of the surveys presented here have a broad community reach, whereas our postevent surveys were limited to workshop participants. We found that the timing and frequency of surveys had an



**Figure 1.** Word cloud illustrating the participant-reported research areas of the 2023 CyberTraining workshop. The color version of this figure is available only in the electronic edition.

effect on the response rate. Postevent surveys were successful only if participants completed them during the event. Because of differences in response rates for pre- and postevent surveys,

**TABLE 1**  
**Dates and Attendance Modes of the Workshops**

Name	Date (yyyy/mm)	Attendance Mode	Range of Participant Attendance
MTUQ	2022/04	Virtual	77
SPECFEM	2022/10	Virtual	50–183
Users			
SPECFEM	2022/10	Hybrid	~30
Developers			
HPS	2023/04	Virtual	30–80
CyberTraining			
SSA	2024/04	In-person	80
SCOPED	2024/05	Hybrid	100
MsPASS	2024/06	Virtual	54

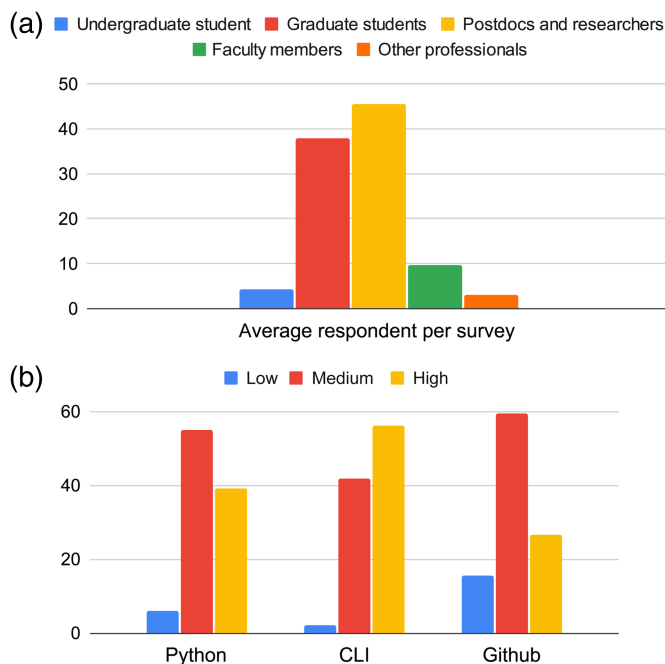
HPS, high-performance seismology cybertraining; MTUQ, moment tensor estimates and uncertainty quantification from broadband seismic data; MsPASS, massive parallel analysis system for seismology; SPECFEM, spectral finite-element method; SSA, Seismological Society of America.

our analysis combines common questions and categories from both types. The response rate was above 96% for requests during the workshop, whether the meeting was in person or virtual, whereas it was 13% in the case of the 2024 SSA workshop.

Survey questions were designed to minimize the impostor syndrome as suggested by Huppenkothen *et al.* (2018). For instance, we asked participants about their familiarity with shell scripting in various forms: “How familiar are you with computing programming from a command line (i.e., within a terminal window)?” with the response fields of “No experience, Some Experience, Extensive Experience.” We also asked about their familiarity with version control with questions such as “All of my active

research projects over the past year are on GitHub with many check-ins.” Another example to assess their proficiency in Python was “I use Python in my life” with the multiple-choice answer “several hours a week and mostly in the classroom,” “several hours a day in my research,” “all and every day!,” and “Never-ever.” We also gathered preliminary knowledge about the technical skill levels of the survey respondents. We emphasized that our surveys were a “self-assessment,” which likely provided a biased response.

Our surveys canvassed career levels of interested workshop participants, which is illustrated in Figure 2a. The surveys included multiple-choice questions with various career levels and sometimes received multiple answers. For instance, participants responded to both “graduate student” and “research scientist” or added an additional category of “PhD candidate.” Although some surveys distinguished between “postdoctoral researcher” and “research scientist,” we have grouped these two categories because they both represent researchers with advanced technical skills and dedicated project-based research. Out of the 976 survey responses, the demographic of the surveyed community exhibited a great majority of graduate students and researchers, with a significant participation (15%) of faculty members. Undergraduate students have a distinctly lower participation level, likely due to our choice of communication channels and the required technical skills advertised in the announcements.



**Figure 2.** Proportion in percent (%) of participants as a summary over seven workshop surveys for (a) career levels of survey respondents and (b) self-assessment of proficiency in computing tools. CLI, command line interface. The color version of this figure is available only in the electronic edition.

We use three skill-based proficiency levels as a relative metric for the seismological community's readiness for working with our large-scale software in Figure 2b. Overall, most participants felt comfortable with command-line interface (CLI) tools. 90% of the participants reported having a sufficient level of familiarity with running Jupyter Notebooks and using Docker images. 40% of the participants declared being sufficiently expert in Python for their research, though a majority declared having a medium level of familiarity. Interestingly, version control using Git ranks last in our assessment because most participants report having a medium level of comfort, and 16% declare having no experience with GitHub.

## Novel CI CI on HPC

The primary computing environment benefiting large-scale seismology today is HPC, which is enabled by clusters of thousands of tightly connected nodes managed by large computing centers, such as the resources used in our workshops, the Texas Advanced Computing Center (TACC) and the San Diego Supercomputing Center. Clusters are designed for parallelized workflows. Many seismological applications, such as FWI, seismic imaging, and earthquake cycle simulations, follow a single-program multiple data model, in which the same code runs across multiple nodes, each processing a different subset of the data. These HPC workflows require optimized software with

efficient parallel scaling, as well as proficiency in job scheduling, memory management, and storage architecture. HPC centers often provide training to help researchers develop and optimize these computational techniques.

Over the course of the workshops, we have trained participants in various aspects of HPC. The lectures covered training on the fundamentals of HPC, including how to write allocation proposals for HPC resources (e.g., what elements to include in a National Science Foundation (NSF) access proposal or how to obtain Amazon Web Services (AWS) education or Cloudbank resources), as well as the parallelization of workflows leveraging shared or distributed memory architectures. We also trained a few selected groups of participants, ~80 total, to access and run forward and adjoint simulations to compute 3D synthetic seismograms and data sensitivity kernels with SPECFEM3D\_GLOBE (Komatitsch and Tromp, 2002a,b) for FWI and dynamic rupture simulations with SeiSol (Käser *et al.*, 2010) on the Frontera system (Stanzione *et al.*, 2020) at the TACC.

## CI on cloud computing

Cloud computing is a new paradigm for computing, where users rent hardware from commercial computing centers such as AWS or Microsoft Azure, which provide on-demand and à la carte hardware choices. Computing is done on “virtual machines” (VM), an abstraction of hardware that contains up to a few hundred central processing unit (CPU) cores, up to a few graphic processing units (GPUs), and a tunable amount of memory. Maximum-size instances can have up to ~200 cores, 10 GPUs, and 1 TB of memory and are primarily designed for big-data processing, such as training complex machine learning models. VMs have a preloaded operating system on which users install dependencies from scratch, from Docker images, or from previously saved virtual images.

Cloud computing is still in its infancy in seismology, and user access remains a challenge (Krauss *et al.*, 2023). We trained participants in cloud computing concepts, including its design for interacting with storage and performing large-scale deployments. We presented diverse strategies for using cloud resources to workshop participants. They accessed Google Colab (see Data and Resources) provided by Google Cloud Platform, which is a pre-configured Python-based Jupyter Notebook, and learned how to customize it by manually installing additional dependencies. Accessibility is a significant benefit of the Colab approach, as VM specifications can be easily modified through the Google Colab web interface. The free version of Google Colab is limited in size (e.g., a few CPUs, 12 GB of RAM, and 50 GBs of storage).

The SCOPED project chose AWS as the cloud provider due to the availability of large seismic datasets already hosted on AWS simple storage service (S3) (northern and southern California data centers as Northern California Earthquake Data Center, NCEDC, and Southern California Earthquake Data Center, SCEDC, NCEDC, 2014; Yu *et al.*, 2021). The workshop covered (1) various ways to access AWS cloud resources,



Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
HC22222	i-01cc78bcf846cf45d	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
claudio	i-0cd3001ef7305051b	Running	t2.xlarge	2/2 checks passed	View alarms	us-west-2b
	i-030d5fad8a4766c4c	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
aaaaaaa	i-01b5f77657b555f2	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
nthapa_event	i-0551f45b26de0780f	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
Stevens_Event_Classific...	i-0f563d76c6c5bad96	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
KIM	i-0d9ede39d65e913b3	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
jyotis	i-0a39d5fb3b60913f5	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
ahutko_Fri_PM	i-0771f324131c1b310	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
utpal-ml	i-09ebfd2c2848dac1c	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
kyungmin	i-075e7aa4e6d94cdd3	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
akash_kharita_scoped	i-05a7331a4a3649315	Running	c5.12xlarge	2/2 checks passed	View alarms	us-west-2b
rodrigo-test-friday-afte...	i-01876e0adf800fbfc	Running	t2.2xlarge	2/2 checks passed	View alarms	us-west-2b

(2) how to launch an AWS computing resource on the elastic-computing (EC2) referred to as an instance from scratch via the web console, (3) how to install basic research software into their instances, and (4) how to run research-grade problems on the Cloud. We used the typical AWS web console to deploy computing resources during one of our workshops and illustrated it in Figure 3. We taught basic concepts and practiced popular tools for software environments and versioning, such as git, Docker, and conda. We note that significant effort was required to simplify and prepare instructions for streamlined access and use of AWS instances. In particular, it is not trivial to open and access a Jupyter Lab, and we curated the training materials to achieve this in our Jupyter Book (HPS; high-performance seismology, SeisSCOPED, 2024).

In addition, we taught various ways to conduct research workflows on the cloud: cloud-native workflows that incorporate cloud services as part of the design (e.g., NoisePy, Jiang and Denolle, 2020, HPS), and, alternatively, workflows that are lifted-and-shifted migrated to the cloud, for example, the Lamont–Doherty Earth Observatory earthquake catalog production workflow (see Data and Resources) that includes algorithms for event detection and phase arrival-time measurements (QuakeFlow, Zhu *et al.*, 2023), discrimination (SpecUFEx, Holtzman *et al.*, 2018; Sawi *et al.*, 2022), and relocation (HypoDD, Waldhauser and Ellsworth, 2000).

## Open-source and containerized software

The SCOPED platform gathers open-source software that tackles big data and large-scale software research. Currently, SCOPED includes full waveform modeling and inversion, machine learning-aided earthquake catalog building and

**Figure 3.** Web browser screenshot showing Amazon Web Services (AWS) instances created by workshop participants during the 2024 SCOPED workshop, illustrating multiple instances running simultaneously on the same allocation. Instance participants chose names. The color version of this figure is available only in the electronic edition.

source characterization, ambient field seismology, and earthquake dynamic rupture simulations.

The underpinning strategy for deploying our software is containerization, which enhances portability and exploits negligible computing overhead (Wang *et al.*, 2019) once successfully containerized. Containers are isolated images of software and its dependencies that can be deployed on various operating systems and hardware (Docker, Singularity). Containers promote long-term sustainability and reproducibility of the computing analysis. To grow our user and developer community, SCOPED’s flagship software is containerized with tutorials provided in the form of Jupyter Notebooks. We developed a SeisSCOPED container registry in which the container base holds minimum dependencies. Additional dependencies can be added to the container base; for instance, an HPC-specific container loads modules for the Message-Processing Interface, and a cloud-specific container includes cloud-provider CLI packages. One significant advantage of the containerized software approach is long-term stability; workshop users and future students alike can leverage the same workshop container and its pinned software dependencies, training materials, and test data. Another powerful use of notebooks is integrating shell scripting within notebook cells using system commands. As an example, one can deploy

parallelized Python scripts on Azure Pool (Krauss *et al.*, 2023), AWS batch resources in a single notebook, or run parallelized SPECfEM simulations through Jupyter Lab (HPS).

Although HPC favors the use of scripting, compiled executables, and minimal container sizes, training materials benefit from attaching small test data and notebooks for documentation and visualization of results. Containers may add IPython and ipykernel dependencies to support Jupyter Notebooks and small test data to a given container. Especially for cloud computing, opening Jupyter Notebooks from remote servers can pose a challenge in group settings, especially on cloud instances. Throughout workshops, our team came up with the following command line to easily allow access to Jupyter Notebooks from a container by fixing the token and IP address:

```
1 sudo docker run -p 80:8888 -rm -it ghcr.io/
  seisscoped/noisepy:centos7_jupyterlab\
2 nohup jupyter lab --no-browser --ip=0.0.0.0 --
  allow-root --
  IdentityProvider.token=scoped &
```

Here, `IdentityProvider.token=scoped` gives a specific token (this avoids users tracking it in the long logs printed on the terminal), `allow-root` grants root access for users inside the container volume, `ip=0.0.0.0` tells the server to listen on all available network interfaces, and `nohup jupyter lab --no-browser &` starts Jupyter Lab in the background without attempting to open a browser because launching a browser is not possible on a remote virtual machine that lacks a graphical user interface and is protected by a private IP address. This small code snippet was designed to accelerate research rather than being hung up on infrastructure.

## Open education

Open education, a set of practices and principles aimed at making learning opportunities more accessible and equitable for everyone, is a promising future direction for higher education as research becomes increasingly specialized and training materials require extensive, globally distributed expertise. Jupyter Book is an appropriate platform for collaborative research education, as many instructors can contribute, and students receive up-to-date materials. Such an example is shown in Figure 4. The challenge remains in curating training materials, as many come from complex research literature and free, non-peer-reviewed online materials.

We are compiling a dynamic textbook titled “High-Performance Seismology” (HPS) that the workshop instructors have contributed to HPS (SeisSCOPED, 2024).

## SCOPED-Related Events

### Virtual events

We have conducted several virtual events, which offer great potential for democratizing access to advanced computing

globally. To maximize participation, we structured events into short sessions (~45 min) with adequate breaks and scheduled them at times that accommodate participants across various time zones. Pre-event surveys of user locations helped select optimal event times, ensuring broad participation. In addition, we recorded the training events and made them asynchronously available on our SCOPED Youtube channel to address time zone conflicts. Over the channel’s lifetime and until 16 February 2025, it has seen 2500 views with 250 hr of course content.

In April 2022, we organized a two-day workshop on moment tensor estimation using the open-source MTUQ software (Thurin *et al.*, 2023). The first day featured a 2 hr session introducing key concepts and tutorials. The second day consisted of a 4 hr session that demonstrated how to calculate a library of Green’s functions for a specified 1D layered model using a frequency–wavenumber code (Zhu and Rivera, 2002) and obtain a seismic moment tensor solution. Attendance was strong, with 78 on day 1 and 68 on day 2, indicating sustained interest in the more detailed content. In preparation, software containers for four systems (Windows/PC, Linux, Mac OS Intel, and Mac OS Apple Silicon processors) were developed and tested, resulting in high success rates for participants running the examples.

Building on the success of the previous workshop, we held a three-day SPECfEM users’ workshop in October 2022. Each of the three daily, 4 hr sessions had a specific focus: the forward wavefield (day 1), sensitivity kernels (day 2), and seismic imaging (day 3). Each session included short (20 min) science lectures, 45 min tutorials that participants could run locally using predownloaded software containers, and wrap-up discussion sessions. Participation ranged from 187 attendees in the day 1 opening seminar to 63 in the day 3 discussion (Fig. 5).

This was the first SPECfEM workshop featuring seismic imaging, providing a natural progression from synthetic seismograms (day 1) to sensitivity kernels (day 2) to iterative tomographic inversion using SeisFlows and Pyatoa (day 3) (Modrak *et al.*, 2018; Chow *et al.*, 2020). Crafting a pedagogical but research-grade notebook took a dedicated effort, given the differences in objectives (performance and robustness versus clarity and interactivity). We took some steps to exemplify but downsize large-scale processes to retain the same outcome. In the big-data analysis, this meant choosing the duration of the experiment (e.g., 1 day of data) or the spatial extent (e.g., number of stations) to be feasible on 2–4 GB worth of RAM, but using the same software tools for a 100 times bigger scale.

In 2023, we held a four-day virtual training workshop in collaboration with the SCEC and several European and NSF-funded projects. Each day focused on a specific theme, starting with an opening day of lectures on open science, reproducibility, software best practices, and an introduction to HPC and Cloud Computing. Subsequent days were divided into sub-disciplines and platforms. The workshop attracted over 200 interested participants, with 80 joining on Zoom at the workshop’s start, though attendance varied due to time zone



High Performance  
Seismology

Search this book...

#### ABOUT

SCOPED

Team

#### SCOPED Events

SCOPED Workshop (2024)

SSA Workshop Data Mining and  
Cloud 101 (2024)

SCEC HPS Workshop (2023)

Event Code of Conduct

#### PRELIMINARY RESOURCES

Preliminary Coding Work

Preliminary Seismology

#### CLOUD COMPUTING

Introduction

AWS 101

AWS S3

#### AMBIENT SEISMIC FIELD

Introduction

NoisePy tutorial: SCEDC

NoisePy tutorial: Visualization

NoisePy tutorial: Monitoring

NoisePy tutorial: AWS Batch

NoisePy tutorial: Coiled

#### QUAKE CATALOG BUILDING

Quake Catalog Building

Lamont ML Catalog

SpecUFEx Tutorial: Amatrice, Italy  
October 2016

Surface Event Detection

#### ML SEISMOLOGY (GEOSMART)

Machine Learning for Seismology



Content

Schedule

## SSA Workshop Data Mining and Cloud 101 (2024)

This workshop will introduce participants to cloud computing, from concept and best practices to practice, for two main approaches of data mining in seismology: correlation seismology and machine learning. Participants will learn how to port their Python scripts from their laptops to the cloud, analyze their intermediate data products, and download the final data product. Participants will learn ambient noise seismology software noise and run it on cloud-hosted data sets of broadband seismometers and distributed acoustic sensing data. Participants will learn machine learning in seismology (earthquake catalog building and data discovery of various geohazards). The workshop curriculum is supported by the NSF project SCOPED.



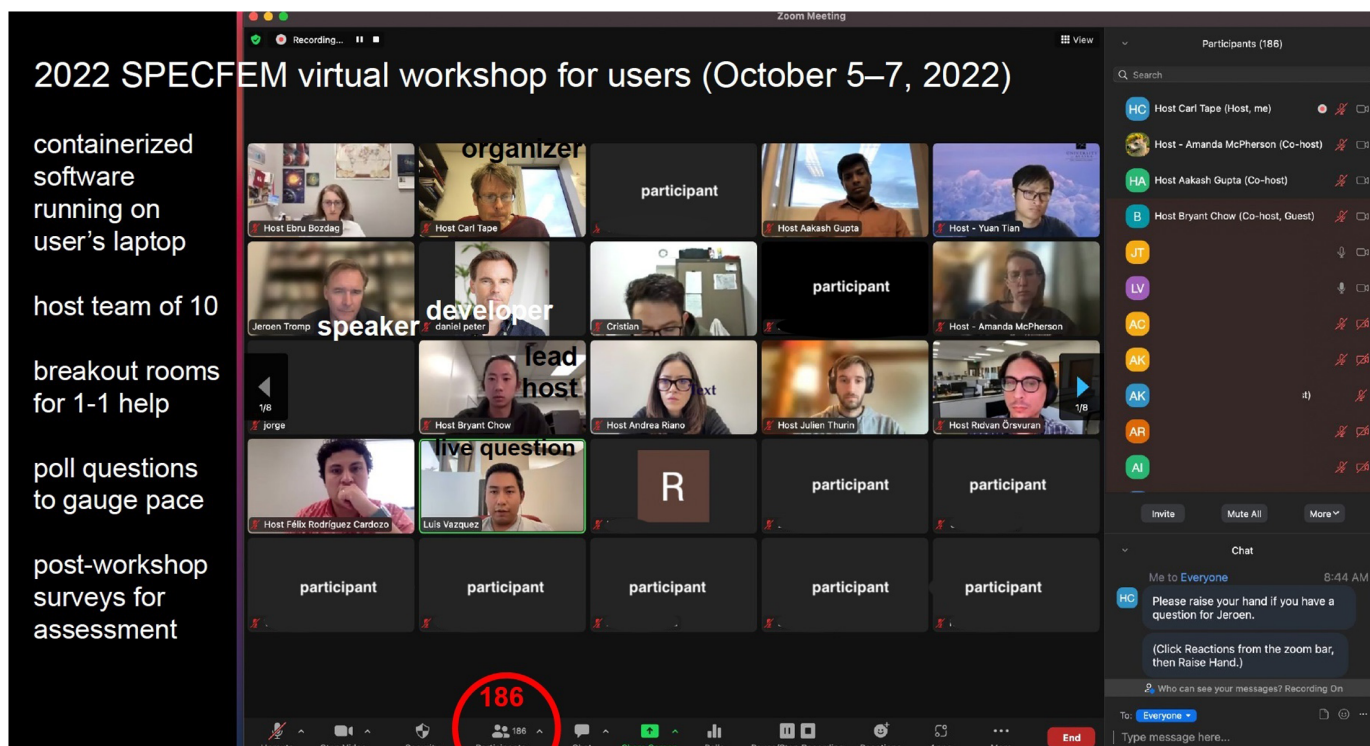
## Schedule

Time	Topics	Instructors	Link to notebook or slides
9:00-9:30m	Welcome	Marine Denolle and Felix Waldhauser	
10:00-11:15am	Cloud 101	Yiyu Ni, Zoe Krauss, Marine Denolle	<a href="https://github.com/SeisSCOPED/seis_cloud">https://github.com/SeisSCOPED/seis_cloud</a> , book
11:15-12:30	Ambient Noise	Yiyu Ni, Kuan-Fu Feng, Marine Denolle	<a href="https://github.com/SeisSCOPED/noisepy">https://github.com/SeisSCOPED/noisepy</a> , book

challenges. Because the workshop was designed for tool adoptions and relatively fast-paced, participants were exposed to diverse topics in seismology, including earthquake simulations focusing on dynamic rupture (SeisSol, Käser *et al.*, 2010; Uphoff *et al.*, 2024) and wave propagation (SPECFEM, Komatitsch *et al.*, 2002), machine learning phase picking (ELEP, Yuan *et al.*, 2023), earthquake probabilistic forecasting

**Figure 4.** A page of the HPS Jupyter Book for the SSA 2024 workshop, which embeds a Google slide presentation for the introduction presentation, the schedule of the specific workshop, and links to relevant book pages. The color version of this figure is available only in the electronic edition.





(pyCSEP, Savran *et al.*, 2022), and user access to Community Earth Models maintained by SCEC (e.g., Plesch *et al.*, 2007; Small *et al.*, 2017).

The 2024 MsPASS training short course was hosted in collaboration with EarthScope during the week of 8 July 2024, as part of their 2024 Technical Short Course series. The event featured two hours of lectures and hands-on sessions over three days. Participants could attend the course in real time or access recordings on YouTube afterward. Daily homework assignments were given, and an optional final project was given. The application-based enrollment process received 99 valid applications, from which 53 participants were accepted to attend. The cohort was notably diverse, with 38% self-identifying as underrepresented in the geoscience community and 26% identifying as female. This short course was the first event on EarthScope's GeoLab platform, a new experimental cloud-based Jupyter Lab platform hosted by EarthScope. MsPASS was the first application to run parallel processing workflows on the GeoLab platform (Wang *et al.*, 2022). Participants were exposed to topics such as using MsPASS to process waveform data in the cloud, managing datasets with a document database, and executing data processing workflows in parallel. Six participants completed all the homework assignments, four had partial submissions, and one completed the optional project.

### In-person events

We held a one-day workshop at the SSA meeting in Anchorage in 2024 that was in-person only, a fast-paced event with an introduction to cloud computing and research workflow. We successfully had 80 participants launch their own cloud

**Figure 5.** Annotated zoom screenshot from the SPECfEM virtual workshop for users (5 October 2022). At this stage in the workshop, there were 186 participants (red circle). The annotations and windows show the hosts, the speaker, the lead software developer, the lead host/instructor, the organizer, and a participant asking a question. The color version of this figure is available only in the electronic edition.

instances on AWS, where they detected earthquakes in cloud-hosted SCEDC data and output data products to a shared MongoDB database. Participants also ran machine learning workflows for earthquake catalog building, including supervised and unsupervised learning approaches for event-type classification. The SSA participants enrolled using a first-come, first-served approach, and communication with the participants was not as well established as in the other SCOPED events. This slowed down the initial steps of the setup, and advanced participants had to follow the pace of beginners.

Our most recent SCOPED workshop was a five-day hybrid meeting at the University of Washington in Seattle in May 2024 (see Data and Resources). About 50 participants, including instructors, attended the workshop in person (example of room layout in Fig. 6), along with a varying number of online participants (on average about 50 per day). The training program was led by the research groups of the SCOPED PIs and Dr. Alice-Agnes Gabriel from the University of California San Diego, supported by several NSF- and European-funded projects. Day 1 covered subjects and practica with an introduction to HPC and cloud computing and best practices for





**Figure 6.** In-person component of the 2024 SCOPED workshop. Participants engaged in live exercises. The OWL camera and directional microphone (bottom center), together with Zoom (speaker's laptop at lower left), enabled hybrid participation. Participants' posters can be seen on the walls. The color version of this figure is available only in the electronic edition.

developing and maintaining open-source software. Day 2 was dedicated to 2D and 3D wave simulations with SPECFEM packages and a tutorial on the introduction to FWI by SeisFlows and moment tensor inversions with the MTUQ software. Day 3 focused on 3D dynamic rupture and finite-source earthquake simulations with SeisSol and 3D wave simulations and computation of 3D adjoint data sensitivity kernels (Tromp *et al.*, 2005; Bozdağ *et al.*, 2011) for FWI on a one-chunk mesh with SPECFEM3D\_GLOBE (Komatitsch and Tromp, 2002a,b). Day 4 focused on high-precision earthquake catalogs (Wang *et al.*, 2024), where they combined machine learning algorithms (QuakeFlow, Zhu *et al.*, 2023) with large-scale cross-correlation and double-difference methods (HypoDD, Waldhauser and Ellsworth, 2000; Waldhauser and Schaff, 2008) and demonstrated the use of unsupervised machine learning (SpecUFEx, Holtzman *et al.*, 2018; Sawi *et al.*, 2022). The participants also attended a session on MsPASS (Wang *et al.*, 2022) to learn how to manage large datasets on HPC and the Cloud. Day 5 addressed ambient noise seismology on the cloud (NoisePy, Jiang and Denolle, 2020), and the trainees were given tutorials on machine learning workflows for seismology on the cloud. All the workshop tutorials were prepared in Jupyter Notebooks hosted on GitHub, containerized versions of the open-source SCOPED software were used, and lectures were recorded and uploaded to the SCOPED YouTube channel. All 3D simulations on day 3 were performed by the trainees on the Frontera system, and the observational seismology tutorials were on AWS.

### Learning objectives and outcomes

Our main learning objectives were for participants to (1) be able to explain the fundamental principles of HPC and cloud computing in seismology, (2) apply appropriate computing

resources (e.g., HPC clusters, AWS instances) to execute research workflows, (3) compare different computational strategies for seismological research (e.g., traditional local computing versus cloud/HPC-based approaches), and (4) evaluate their efficiency in handling large-scale seismological data. Surveys following the SSA and HPS CyberTraining workshops enabled us to evaluate some of these learning outcomes. In the SSA survey, we evaluated the learning outcomes of each module, which were about cloud computing and research-grade applications in ambient noise seismology and machine learning in earthquake catalog building. Eleven survey respondents out of eighty participants noted improved cloud computing skills and overall self-reported positive learning outcomes with the workshop. Future improvements in workshop materials and delivery mechanisms will enhance the impact of the training.

In the HPS CyberTraining survey, 23 participants responded and expressed positive learning outcomes, with 70% ranking their satisfaction 5/5 and 62.5% indicating that the workshop was a valuable use of their time (rank 5/5). Positive learning outcomes were on Docker and reproducible and open science, frontier seismological topics, and HPC and cloud computing. Several participants expressed verbally or via the survey that the pace was fast and that instructors should slow down when going through code blocks in notebooks,

along with improving participant-led exercises in the notebooks with empty cells.

## A Guide to Advanced Computing Workshops

The development of teaching materials requires dedicated effort from both faculty and participants. In-person workshops need a participant-to-assistant ratio of about 15 to 1 for effective debugging. Recruiting participants with similar technical skill levels ensures consistent progress, or additional instruction time can be provided for beginners.

### Surveys

Evaluation surveys can be helpful in quantifying learning outcomes, and crafting them with consideration can benefit professional educators and evaluators. Employing a more standardized approach to the surveys may improve their usefulness. For example, metrics such as “None, Little, Moderate, Quite a bit, Complete” for levels 1 through 5 are similar to the Likert 6-point, “strongly disagree, disagree, somewhat disagree, somewhat agree, agree, strongly agree.” More standard metrics, such as the Likert 6-point, will be incorporated in future surveys to provide a more nuanced measure of fields (Huppenkothen *et al.*, 2018). Allowing only one response per question is essential because multiple answers can hinder postevent quantitative analysis.

### Content

Each of the SCOPED workshops had various designs and, overall, was packed with tutorials. We developed or used several forms of pedagogy for training workshops, which individual workshops may have combined:

- Scientific lectures, especially those that motivate the use of advanced computing resources;
- lectures on CI, research ethics, and software best practices;
- core package tutorials to train participants in using a specific software in its generic form;
- research-grade workflow tutorials with assisted walk-throughs; and
- group or participant-led activity (hackathon-style).

Taken together, these pedagogies may form a module, a self-contained unit that includes (1) a brief lecture (20–30 min) introducing key concepts and their relevance to scientific applications, (2) a hands-on tutorial (60–90 min) in which participants apply the concepts through structured exercises, such as running computations on cloud platforms, setting up HPC environments, or analyzing seismic datasets, and (3) a guided exploration and Q&A (30–60 min) to allow participants to troubleshoot their workflows and gain deeper insights. Modules form half-day activities. In-person workshops may combine two modules per day, with 3 hr sessions striking a good balance between depth and independent exploration

without too much cognitive fatigue; virtual workshops may spread these modules over multiple days and time zones to help attract foreign participants. This strategy worked particularly well for the SPECfEM and MTUQ workshops.

### Before the event

**Instructor coordination.** Instructor coordination is critical for workshop effectiveness, which was not consistently undertaken in our workshops. We recommend *preworkshop planning meetings* to define learning objectives, align instructional materials, and anticipate challenges, *mock run-throughs* with non-participating students or colleagues to identify unclear instructions and potential bottlenecks in execution, *role assignment among instructors*, enduring that each focuses on specific tasks such as concept explanation, hands-on support, or software troubleshooting, and *diverse instructor background* to represent different expertise area (e.g., computational scientists in HPC/Cloud, seismologists form scientific applications, and software developers) to allow for a more comprehensive learning experience.

**Materials and platforms.** We used Jupyter Books or Google Docs as shareable, open platforms to organize the workshop schedule and share training materials, with a clear first page with the schedule that links to the sources (e.g., YouTube recording, Zoom links, GitHub repositories, etc.). To date, our teaching materials are still available, and YouTube videos are still watched. Communication platforms like Slack or Teams, or other forms of group communication and direct messaging, allow for rapid, practical communication among instructors and between participants and organizers. Preworkshop materials (e.g., prerequisite tutorials, recorded lectures, and videos from previous workshops, and software installation guides) help participants familiarize themselves with foundational concepts beforehand. Instructors may predownload data, for example, pre-processed data and static visualization, to ensure the workshop runs smoothly even with unforeseen technical issues such as loss of network connectivity. To accommodate different learning styles, workshops may provide slides and annotated code for visual learners, interactive coding exercises for experiential learners, and may incorporate discussion-based problem-solving for verbal learners.

**Accounts.** We found that the workshop ran more smoothly when participants’ accounts on the computing resources were set up days in advance. We provide guidance to automatically create user accounts for AWS on the HPS book. It is important to remind participants that workshop computing resources are *temporary*. Educational allocations at HPC centers are typically provided when supercomputer center research scientists are involved in the workshop. For our workshop, users chose a simple username, for example, the participants’ email address or its prefix (e.g., <yourID>email.edu), as well as a

single generic password to avoid manual and complicated intervention. Cloud accounts can be created at any time and managed during the workshop. For instance, a cloud manager can reassign policies, roles, and temporary passwords during the event if needed. Through surveys, emails, and possibly “Day 0” virtual help sessions, instructors may find it useful to ensure that computing setups (accounts, software containers) are working in advance for all virtual participants.

Participants were made aware that these accounts were temporary and provided with guidelines on how to access these platforms in the future.

### First day

This is the day to onboard participants, ensure that the accounts to HPC and the cloud are set up and accessible, install ancillary software, and download and test workshop containers to ensure they perform as expected on the participants’ platforms. These tasks can also be done prior to the workshop to free up actual workshop time. Hybrid workshops can be challenging to deliver. They require multiple cameras and microphones for large rooms, attention to remote attendance, and interaction with remote participants. For hybrid events, organizers may find it beneficial to ensure that there are sufficient staff/instructors online who can help manage remote participants. Engagement can be improved with frequent polling.

### During the event

The feedback on workshops has been positive, especially for focused, single-tool, and single-platform sessions. For virtual workshops, a helpful strategy is to ask, “Are you ready to move on?” with the options “Yes, Almost, and No.” This helps pace the session and provides instant feedback on participant experience, showing engagement levels and areas needing assistance.

Some of the tutorials, especially those for the core software, included additional cells in the Jupyter Notebooks so that participants could test various parameters independently, which was implemented in SPECfEM, MTUQ, and SeisSol. Other full-stack, research-grade tutorials (e.g., machine learning-aided earthquake catalog building or ambient noise seismology) included advanced workflows tailored to specific use cases, making it challenging to strike a balance between teaching fundamental concepts and realistic scenarios.

Teaching cloud computing and HPC strategies for research-grade analysis can be challenging for participants if the content is not relevant to their work. In several tutorials, we demonstrated how to adapt a homegrown software stack based on a specific platform to provision a cloud instance. Such an approach allows researchers to upload and deploy their own software stacks on the cloud, ensuring flexibility and independence without imposing a specific platform or software style. Although this requires coordination among instructors, it

enables each researcher to bring up their preferred tools, reflecting the natural workflow of scientific research.

### Postevent

Surveying the participants is a good way to measure learning outcomes. We found that most participants will not fill out postevent surveys unless asked **at the time of the events**, both for virtual and in-person meetings. To improve the evaluation, the exit survey may benefit from having similar questions to the incoming survey. Leaving an empty box at the end allows participants to speak freely about what worked and what needs improvement.

Assessments need to be more quantitative, with more structured responses than were provided in many of the surveys we ran. Some respondents provided multiple answers to the same question, posing additional challenges in the analysis of the survey during postprocessing. Further automation of the survey, such as more rigorous Python-based postprocessing, will improve the reproducibility of the survey analysis.

### Conclusions

The diversity of workshops is essential to reach multiple pedagogical goals. Large attendance in virtual meetings allows for a global reach and democratization of training and access to computing resources. The size of these virtual meetings was not optimal for spontaneous communication and career network—although future workshops could take this into account. We found that at that scale (200+ participants), it was easier to have participants run containers and software locally, whereas, for smaller, virtual meetings, it is possible to provision remote participants with temporary cloud accounts.

In-person meetings are well suited for career development, building collaborations, and providing participants with more advanced computing resources, which may be limited to certain countries. These in-person meetings can run longer than virtual events, with the caveat that organizers may consider pacing the delivery of the materials more slowly than they anticipate and even include participant-led hackathons for better learning outcomes and stronger cohort building.

Advanced computing with projects such as services for Jupyter Hubs (e.g., Infrastructure-as-a-Service such as 2i2c that support centralized servers running Jupyter Lab or Notebooks with multiple-user access), or Python projects that manage distributed cloud resources such as Coiled, and the up-and-coming science gateways (e.g., McLennan and Kennell, 2010; Marru *et al.*, 2011; Stubbs *et al.*, 2021) promote ease of access to resources, potentially benefiting the user community. Nevertheless, training the community in the concepts of cloud computing and HPC for new *developers* remains important so that they can continue innovating solutions for large-scale computing for seismological research and that their expertise lasts beyond the lifetime of specific Iaas.



Our efforts in conducting these workshops reflect a positive outlook for seismological research in the twenty-first century. Because big seismic data become more widely accessible, seismologists at all career levels desire to pursue training in HPC and Cloud computing. We highlight the benefits of our workshop model by uniting CI and research professionals skilled in HPC and Cloud computing. They leverage large-scale computing to solve seismological problems. Through these workshops and their associated teaching materials, we can disseminate that collective knowledge in an open, sustainable, and reproducible manner, thereby accelerating the pace of seismological discovery.

## Data and Resources

The survey data were collected through Google Forms responses. Because of the lack of anonymity in the responses, the authors decided not to share the original data. All SCOPED educational materials are open-source (e.g., <https://seisscoped.org/HPS-book/intro.html>, last accessed May 2025). Video recordings of our workshops are available on SCOPED YouTube channel. The open-source SPECfem package is available at <https://specfem.org/> (last accessed May 2025). The open-source SeisSol software is available at <https://seissol.org/> (last accessed May 2025). The Coursera course is available on “Computers, Waves, Simulations: A Practical Introduction to Numerical Methods using Python”. The Google Colab can be accessed at <https://colab.research.google.com/> (last accessed May 2025) provided by Google Cloud Platform, which is a preconfigured Python-based Jupyter Notebook. The most recent SCOPED workshop was a five-day hybrid meeting at the University of Washington in Seattle in May 2024 (<https://seisscoped.org/workshop-2024/>, last accessed October 2024).

## Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

## Acknowledgments

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The authors are grateful to Associate Editor Alan Kafka and reviewers Brad Aagaard and Clara Yoon for their constructive feedback. The authors also thank Theresa Sawi at the U.S. Geological Survey (USGS) for her contribution to the SCOPED project, the training workshops, and editing the article. This work is supported by the Seismic Computational Platform for Empowering Discovery (SCOPED) project under the National Science Foundation (Award Numbers OAC-2103701 (UW), OAC-2104052(UAF), OAC-2103621 (CSM), OAC-2103741 (CU), and OAC-2103494 (UT)). The events were also supported by the eScience Institute, an SCEC Grant 22162. The MTUQ workshop was partly sponsored by the Geophysical Detection for Nuclear Proliferation (GDNF) University Affiliated Research Center (UARC) Task Order 7, funded by the Air Force Research Laboratory under Contract HQ0034-20-F-0284. The HPS CyberTraining and the SCOPED training were additionally supported by the Southern California Earthquake Center and USGS (SCEC Project 22162), by the National Science Foundation

(MTMOD, Grant Number EAR-2121568, CSA-LCCF, Grant Number OAC-2139536, QUAKEWORX, Grant Number OAC-2311208), the European Union’s Horizon 2020 research and innovation programme (TEAR ERC Starting Grant Number 852992) and Horizon Europe (ChEESE-2P, Grant Number 101093038; DT-GEO, Grant Number 101058129; and Geo-INQUIRE, Grant Number 101058518). Waveform data, metadata, or data products for this study were accessed through the Northern California Earthquake Data Center (NCEDC), doi: [10.7932/NCEDC](https://doi.org/10.7932/NCEDC). The authors thank the Texas Advanced Computing Center (TACC) at the University of Texas at Austin for providing computational resources on “Frontera” system (Stanzione *et al.*, 2020). The authors thank Zhengtang Yang, Weiming Yang, and Jinxin Ma, they contributed to the development of MsPASS and maintained the SCOPED containers. The authors thank Dave A. May for a guest lecture on “Open and Reproducible Science” during the HPS CyberTraining. The authors thank all SeisSol developers, specifically Thomas Ulrich, Iris Christadler, Mathilde Marchandon, Jeena Yun, Nicolas Hayek, and Jonatan Glehman, for their support in preparing and during the HPS CyberTraining and the SCOPED training. The group also thanks the SCEC and specifically Pablo Iturrieta, Jose Bayona, Phil Maechling, Fabio Silva, Mei-Hui Su, and Scott Callaghan for giving a full day of tutorials in the 2023 HPS CyberTraining workshop, tutorials that are also integrated into the HPS book. The 2024 MsPASS training short course was also sponsored by the EarthScope Consortium. The authors thank Sarah Wilson, Robert Weekly, Chad Trabant, Melissa Weber, Gillian Haberli, and Tammy Bravo, as well as other staff members of the EarthScope Consortium who provided technical and pedagogical support throughout the event.

## References

- Aki, K., and P. G. Richards (2002). *Quantitative Seismology*, 2nd ed., University Science Books, 704 pp.
- Arrowsmith, S. J., D. T. Trugman, J. MacCarthy, K. J. Bergen, D. Lumley, and M. B. Magnani (2022). Big data seismology, *Rev. Geophys.* **60**, no. 2, e2021RG000769, doi: [10.1029/2021RG000769](https://doi.org/10.1029/2021RG000769).
- Bao, H., J. Bielak, O. Ghattas, L. F. Kallivokas, D. R. O’Hallaron, J. R. Shewchuk, and J. Xu (1998). Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers, *Comp. Meth. Appl. Mech. Eng.* **152**, nos. 1/2, 85–102.
- Barker, M., N. P. Chue Hong, D. S. Katz, A.-L. Lamprecht, C. Martinez-Ortiz, F. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P. A. Martinez, *et al.* (2022). Introducing the fair principles for research software, *Sci. Data* **9**, no. 1, 622.
- Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81**, no. 3, 530–533.
- Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire (2016). Global adjoint tomography: First-generation model, *Geophys. J. Int.* **207**, 1739–1766.
- Bozdağ, E., J. Trampert, and J. Tromp (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.* **185**, no. 2, 845–870.
- Brudzinski, M., M. Hubenthal, S. Fasola, and E. Schnorr (2021). Learning in a crisis: Online skill building workshop addresses immediate pandemic needs and offers possibilities for future trainings, *Bull. Seismol. Soc. Am.* **92**, no. 5, 3215–3230.

- Chow, B., Y. Kaneko, C. Tape, R. Modrak, and J. Townend (2020). An automated workflow for adjoint tomography—Waveform misfits and synthetic inversions for the North Island, New Zealand, *Geophys. J. Int.* **223**, 1461–1480.
- Chue Hong, N. P., D. S. Katz, M. Barker, A.-L. Lamprecht, C. Martinez, F. E. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P. A. Martinez, *et al.* (2022). Fair principles for research software (fair4rs principles), *Zenodo* doi: [10.15497/RDA00068](https://doi.org/10.15497/RDA00068).
- Computational Infrastructure for Geodynamics (CIG) (2016a). Software development best practices for the CIG community, available at [https://github.com/geodynamics/best\\_practices/blob/master/SoftwareDevelopmentBestPractices.md](https://github.com/geodynamics/best_practices/blob/master/SoftwareDevelopmentBestPractices.md) (last accessed July 2023).
- CIG (2016b). Software training best practices for the CIG community, available at [https://github.com/geodynamics/best\\_practices/blob/master/TrainingBestPractices.md](https://github.com/geodynamics/best_practices/blob/master/TrainingBestPractices.md) (last accessed July 2023).
- Denolle, M., F. Waldhauser, C. Tape, E. Bozdog, and I. Wang (2024). Scoped update: A cloud and HPC software platform for computational seismology, *AGU Fall Meeting 2023, Seismology/Advances in Theoretical and Computational Seismology I Poster*, San Francisco, California, 11–15 December 2023, Poster No. 0288, id. S11C-0288.
- Folch, A., C. Abril, M. Afanasiev, G. Amati, M. Bader, R. M. Badia, H. B. Bayraktar, S. Barsotti, R. Basili, F. Bernardi, *et al.* (2023). The EU center of excellence for exascale in solid earth (cheese): Implementation, results, and roadmap for the second phase, *Future Gener. Comput. Syst.* **146**, 47–61.
- Gabriel, A.-A., T. Ulrich, M. Marchandon, J. Biemiller, and J. Rekoske (2023). 3d dynamic rupture modeling of the 6 February 2023, Kahramanmaraş, Türkiye  $M_w$  7.8 and 7.7 earthquake doublet using early observations, *Seism. Res. Lett.* **94**, no. 4, 342–356.
- Graves, R. W. (1998). Three-dimensional finite-difference modeling of the San Andreas Fault: Source parameterization and ground-motion levels, *Bull. Seismol. Soc. Am.* **88**, no. 4, 881–897.
- Graves, R., T. H. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner, *et al.* (2011). Cybershake: A physics-based seismic hazard model for southern California, *Pure Appl. Geophys.* **168**, 367–381.
- Holtzman, B., A. Paté, J. Paisley, F. Waldhauser, and D. Repetto (2018). Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field, *Sci. Adv.* **4**, no. 5, eaao2929, doi: [10.1126/sciadv.aao2929](https://doi.org/10.1126/sciadv.aao2929).
- Huppenkothen, D., A. Arendt, D. W. Hogg, K. Ram, J. T. VanderPlas, and A. Rokem (2018). Hack weeks as a model for data science education and collaboration, *Proc. Natl. Acad. Sci.* **115**, no. 36, 8872–8877.
- Igel, H. (2017). *Computational Seismology: A Practical Introduction*, Oxford University Press, Oxford, United Kingdom.
- Jiang, C., and M. A. Denolle (2020). NoisePy: A new high-performance python tool for ambient-noise seismology, *Seismol. Res. Lett.* **91**, no. 3, 1853–1866.
- Jones, J. P., K. Okubo, T. Clements, and M. A. Denolle (2020). Seisio: A fast, efficient geophysical data architecture for the Julia language, *Seismol. Res. Lett.* **91**, no. 4, 2368–2377.
- Käser, M., C. Castro, V. Hermann, and C. Pelties (2010). Seissol—A software for seismic wave propagation simulations, *High Performance Computing in Science and Engineering, Garching/Munich 2009: Transactions of the Fourth Joint HLRB and KONWIHR Review and Results Workshop*, Leibniz Supercomputing Centre, Garching/Munich, Germany, 8–9 December 2009, 281–292, Springer.
- Komatitsch, D., and J. Tromp (2002a). Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.* **149**, 390–412.
- Komatitsch, D., and J. Tromp (2002b). Spectral-element simulations of global seismic wave propagation—II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophys. J. Int.* **150**, 308–318.
- Komatitsch, D., and J.-P. Vilotte (1998). The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. Seismol. Soc. Am.* **88**, no. 2, 368–392.
- Komatitsch, D., Q. Liu, J. Tromp, P. Süß, C. Stidham, and J. H. Shaw (2004). Simulations of ground motion in the Los Angeles basin based upon the spectral-element method, *Bull. Seismol. Soc. Am.* **94**, no. 1, 187–206.
- Komatitsch, D., J. Ritsema, and J. Tromp (2002). The spectral-element method, Beowulf computing, and global seismology, *Science* **298**, 1737–1742.
- Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft (2019). Machine learning in seismology: Turning data into insights, *Seismol. Res. Lett.* **90**, no. 1, 3–14.
- Krauss, Z., Y. Ni, S. Henderson, and M. Denolle (2023). Seismology in the cloud: Guidance for the individual researcher, *Seismica* **2**, no. 2, doi: [10.26443/seismica.v2i2.979](https://doi.org/10.26443/seismica.v2i2.979).
- Krenz, L., C. Uphoff, T. Ulrich, A.-A. Gabriel, L. S. Abrahams, E. M. Dunham, and M. Bader (2021). 3d acoustic-elastic coupling with gravity: The dynamics of the 2018 Palu, Sulawesi earthquake and tsunami, *Proc. Int. Conf. for High Performance Computing, Networking, Storage and Analysis, SC '21*, Association for Computing Machinery, New York, New York.
- Krischer, L., Y. A. Aiman, T. Batholomäus, S. Donner, M. van Driel, K. Duru, K. Garina, K. Gesele, T. Gunawan, S. Hable, *et al.* (2018). *seismo-live: An educational online library of Jupyter notebooks for seismology*, *Seismol. Res. Lett.* **89**, no. 6, 2413–2419.
- Liu, Q., and Y. J. Gu (2012). Seismic imaging: From classical to adjoint tomography, *Tectonophysics* **566/567**, 31–66.
- MacCarthy, J., O. Marcillo, and C. Trabant (2020). Seismology in the cloud: A new streaming workflow, *Seismol. Res. Lett.* **91**, no. 3, 1804–1812.
- Marru, S., L. Gunathilake, C. Herath, P. Tangchaisin, M. Pierce, C. Mattmann, R. Singh, T. Gunarathne, E. Chinthaka, R. Gardler, *et al.* (2011). Apache Airavata: A framework for distributed applications and computational workflows, *Proc. of the 2011 ACM Workshop on Gateway Computing Environments, GCE'11*, Association for Computing Machinery, New York, New York, 21–28.
- McLennan, M., and R. Kennell (2010). Hubzero: A platform for dissemination and collaboration in computational science and engineering, *Comput. Sci. Eng.* **12**, no. 2, 48–53.
- Modrak, R. T., D. Borisov, M. Lefebvre, and J. Tromp (2018). SeisFlows—Flexible waveform inversion software, *Comput. Geosci.* **115**, 88–95.
- Morra, G., E. Bozdog, M. Knepley, L. Räss, and V. Vesselinov (2021). A tectonic shift in analytics and computing is coming, *Eos* **102**, doi: [10.1029/2021EO159258](https://doi.org/10.1029/2021EO159258).
- Mousavi, S. M., and G. C. Beroza (2022). Deep-learning seismology, *Science* **377**, no. 6607, eabm4470, doi: [10.1126/science.abm4470](https://doi.org/10.1126/science.abm4470).

- Nakata, N., L. Gualtieri, and A. Fichtner (2019). *Seismic Ambient Noise*, Cambridge University Press, Cambridge, United Kingdom.
- Ni, Y., M. A. Denolle, R. Fatland, N. Alterman, B. P. Lipovsky, and F. Knuth (2023). An object storage for distributed acoustic sensing, *Seismol. Res. Lett.* **95**, no. 1, 499–511.
- Northern California Earthquake Data Center (NCEDC) (2014). *Northern California Earthquake Data Center* [Dataset], doi: [10.7932/NCEDC](https://doi.org/10.7932/NCEDC).
- Olsen, K. B., and R. J. Archuleta (1996). Three-dimensional simulation of earthquakes on the Los Angeles fault system, *Bull. Seismol. Soc. Am.* **86**, no. 3, 575–596.
- Pelties, C., J. de la Puente, J.-P. Ampuero, G. B. Brietzke, and M. Käser (2012). Three-dimensional dynamic rupture simulation with a high-order discontinuous Galerkin method on unstructured tetrahedral meshes, *J. Geophys. Res.* **117**, no. B2, doi: [10.1029/2011JB008857](https://doi.org/10.1029/2011JB008857).
- Pelties, C., A.-A. Gabriel, and J.-P. Ampuero (2014). Verification of an ADER-DG method for complex dynamic rupture problems, *Geosci. Model Develop.* **7**, no. 3, 847–866.
- Pérez, F., and B. E. Granger (2007). Ipython: A system for interactive scientific computing, *Comput. Sci. Eng.* **9**, no. 3, 21–29.
- Peter, D., D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Locher, F. Magnoni, Q. Liu, C. Blitz, et al. (2011). Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.* **186**, 721–739.
- Pimentel, J. F., L. Murta, V. Braganholo, and J. Freire (2019). A large-scale study about quality and reproducibility of Jupyter notebooks, *2019 IEEE/ACM 16th Int. Conf. on Mining Software Repositories (MSR)*, 507–517.
- Plesch, A., J. H. Shaw, C. Benson, W. A. Bryant, S. Carena, M. Cooke, J. Dolan, G. Fuis, E. Gath, L. Grant, et al. (2007). Community Fault Model (CFM) for southern California, *Bull. Seismol. Soc. Am.* **97**, no. 6, 1793–1802.
- Rost, S., and C. Thomas (2009). Improving seismic resolution through array processing techniques, *Surv. Geophys.* **30**, 271–299.
- Savran, W. H., J. A. Bayona, P. Iturrieta, K. M. Asim, H. Bao, K. Bayliss, M. Herrmann, D. Schorlemmer, P. J. Maechling, and M. J. Werner (2022). pyCSEP: A python toolkit for earthquake forecast developers, *Seismol. Res. Lett.* **93**, no. 5, 2858–2870.
- Sawi, T., B. Holtzman, F. Walter, and J. Paisley (2022). An unsupervised machine-learning approach to understanding seismicity at an alpine glacier, *J. Geophys. Res.* **127**, no. 12, e2022JF006909, doi: [10.1029/2022JF006909](https://doi.org/10.1029/2022JF006909).
- SeisSCOPED (2024). High-performance seismology, SeisSCOPED, Dynamic textbook, available at <https://seisscoped.org/HPS-book> (last accessed May 2025).
- Shearer, P. M. (2019). *Introduction to Seismology*, Cambridge University Press, Cambridge, United Kingdom.
- Small, P., D. Gill, P. J. Maechling, R. Taborda, S. Callaghan, T. H. Jordan, K. B. Olsen, G. P. Ely, and C. Goulet (2017). The SCEC Unified Community Velocity Model software framework, *Seismol. Res. Lett.* **88**, no. 6, 1539–1552.
- Stanzione, D., J. West, R. T. Evans, T. Minyard, O. Ghattas, and D. K. Panda (2020). Frontera: The evolution of leadership computing at the national science foundation, *Practice and Experience in Advanced Research Computing 2020: Catch the Wave*, PEARC '20, New York, New York, Association for Computing Machinery, 106–111.
- Stein, S., and M. Wysession (2009). *An Introduction to Seismology, Earthquakes, and Earth Structure*, John Wiley & Sons, Hoboken, New Jersey.
- Stubbs, J., R. Cardone, M. Packard, A. Jamthe, S. Padhy, S. Terry, J. Looney, J. Meiring, S. Black, M. Dahan, et al. (2021). Tapis: An api platform for reproducible, distributed computational research, in *Advances in Information and Communication*, K. Arai (Editor), Springer International Publishing, Cham, Switzerland, 878–900.
- Tape, C., E. Bozdog, M. Denolle, F. Waldhauser, and I. Wang (2022). SCOPED: Seismic COmputational Platform for Empowering Discovery [Year 1], *Zenodo* doi: [10.5281/zenodo.6862979](https://doi.org/10.5281/zenodo.6862979).
- Tape, C., Q. Liu, A. Maggi, and J. Tromp (2009). Adjoint tomography of the southern California crust, *Science* **325**, 988–992.
- Thurin, J., J. Braunmiller, F. R. R. Cardozo, L. Ding, Q. Liu, A. McPherson, R. Modrak, and C. Tape (2023). MTUQ: A high-performance Python package for moment tensor estimation and uncertainty quantification, *Abstract at 2023 SSA Annual Meeting*, San Juan, Puerto Rico, 17–20 April.
- Tromp, J. (2020). Seismic wavefield imaging of earth's interior across scales, *Nat. Rev. Earth Environ.* **1**, 40–53.
- Tromp, J., C. Tape, and Q. Liu (2005). Seismic tomography, adjoint methods, time reversal, and Banana-Doughnut kernels, *Geophys. J. Int.* **160**, 195–216.
- Uphoff, C., L. Krenz, T. Ulrich, S. Wolf, A. Knoll, S. David, D. Li, R. Dorozhinskii, A. Heinecke, S. Wollherr, et al. (2024). SeisSol (v1.3.1), *Zenodo* doi: [10.5281/zenodo.15117916](https://doi.org/10.5281/zenodo.15117916).
- Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the Northern Hayward fault, California, *Bull. Seismol. Soc. Am.* **90**, no. 6, 1353–1368.
- Waldhauser, F., and D. P. Schaff (2008). Large-scale relocation of two decades of northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.* **113**, no. B8, B08311, doi: [10.1029/2007JB005479](https://doi.org/10.1029/2007JB005479).
- Wang, I., E. Bozdog, denolle, and F. Waldhauser (2023). SCOPED: seismic computational platform for empowering discovery [year 2], doi: [10.6084/m9.figshare.24212847.v1](https://doi.org/10.6084/m9.figshare.24212847.v1).
- Wang, K., F. Waldhauser, D. Schaff, M. Tolstoy, W. S. D. Wilcock, and Y. J. Tan (2024). Real-time detection of volcanic unrest and eruption at axial seamount using machine learning, *Seismol. Res. Lett.* **95**, no. 5, 2651–2662.
- Wang, Y., R. T. Evans, and L. Huang (2019). Performant container support for HPC applications, *Practice and Experience in Advanced Research Computing 2019: Rise of the Machines (Learning)*, PEARC '19, Association for Computing Machinery, New York, New York.
- Wang, Y., G. L. Pavlis, W. Yang, and J. Ma (2022). MsPASS: A data management and processing framework for seismology, *Seismol. Res. Lett.* **93**, no. 1, 426–434.
- Yu, E., A. Bhaskaran, S. Chen, Z. E. Ross, E. Hauksson, and R. W. Clayton (2021). Southern California earthquake data now available in the AWS cloud, *Seismol. Res. Lett.* **92**, no. 5, 3238–3247.
- Yuan, C., Y. Ni, Y. Lin, and M. Denolle (2023). Better together: Ensemble learning for earthquake detection and phase picking, *IEEE Trans. Geosci. Remote Sens.* **61**, 1–17, doi: [10.1109/TGRS.2023.3320148](https://doi.org/10.1109/TGRS.2023.3320148).



- Zhu, L., and L. A. Rivera (2002). A note on the dynamic and static displacements from a point source in multilayered media, *Geophys. J. Int.* **148**, 619–627.
- Zhu, W., A. B. Hou, R. Yang, A. Datta, S. M. Mousavi, W. L. Ellsworth, and G. C. Beroza (2023). Quakeflow: A scalable machine-learning-based earthquake monitoring workflow with cloud computing, *Geophys. J. Int.* **232**, no. 1, 684–693.

## Authors and Affiliations

**Marine A. Denolle:** Department of Earth and Space Sciences, University of Washington, Seattle, Washington, U.S.A., <https://orcid.org/0000-0002-1610-2250>; **Carl Tape:** Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, U.S.A., <https://orcid.org/0000-0003-2804-7137>; **Ebru Bozdağ:** Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, U.S.A.; Department of Geophysics, Colorado School of Mines, Golden, Colorado, U.S.A.; **Yinzhi Wang:** Texas Advanced Computing Center, The University of Texas at Austin, Austin, Texas, U.S.A., <https://orcid.org/0000-0001-8505-0223>; **Felix Waldhauser:** Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, U.S.A., <https://orcid.org/0000-0002-1286-9737>; **Alice-Agnes Gabriel:** Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California at San Diego, San Diego, California, U.S.A. Geophysics, Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität (LMU), Munich, Germany, <https://orcid.org/0000-0003-0112-8412>; **Jochen Braunmiller:** School of Geosciences, University of South Florida, Tampa, Florida, U.S.A., <https://orcid.org/0000-0002-8541-6560>; **Bryant Chow:** Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, U.S.A., <https://orcid.org/0000-0002-3901-4755>; **Liang Ding:** Natural Resources Canada, Ottawa, Ontario, Canada, <https://orcid.org/0000-0002-3047-9556>; **Kuan-Fu Feng:** Department of Earth and Space Sciences, University of Washington, Seattle, Washington, U.S.A., <https://orcid.org/0009-0002-0765-4499>; **Ayon Ghosh:** Department of Geophysics, Colorado School of Mines, Golden, Colorado, U.S.A., <https://orcid.org/0000-0001-9627-1849>; **Nathan Groebner:** School of Geosciences, University of South Florida, Tampa, Florida, U.S.A., <https://orcid.org/0000-0002-6535-1739>; **Aakash Gupta:** Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, U.S.A., <https://orcid.org/0000-0002-2776-6431>; **Zoe Krauss:** Department of Earth and Space Sciences, University of Washington, Seattle, Washington, U.S.A., <https://orcid.org/0000-0002-5933-6823>; **Amanda M. McPherson:** Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, U.S.A.; **Masaru Nagaso:** Department of Applied Mathematics

and Statistics, Colorado School of Mines, Golden, Colorado, U.S.A., <https://orcid.org/0000-0002-3566-6174>; **Zihua Niu:** Geophysics, Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität (LMU), Munich, Germany, <https://orcid.org/0000-0003-4213-3322>; **Yiyu Ni:** Department of Earth and Space Sciences, University of Washington, Seattle, Washington, U.S.A., <https://orcid.org/0000-0001-5181-9700>; **Rüdván Örsvuran:** Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, U.S.A., <https://orcid.org/0000-0002-5098-7515>; **Gary Pavlis:** Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, Indiana, U.S.A.; **Felix Rodriguez-Cardozo:** School of Geosciences, University of South Florida, Tampa, Florida, U.S.A.; **Theresa Sawi:** Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, U.S.A. U.S. Geological Survey, NASA Ames Research Park, Mountain View, California, U.S.A., <https://orcid.org/0000-0001-5938-6743>; **David Schaff:** Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, U.S.A.; **Nico Schliwa:** Geophysics, Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität (LMU), Munich, Germany, <https://orcid.org/0009-0004-7825-9458>; **David Schneller:** Chair of Scientific Computing in Computer Science, Department of Computer Science, Technical University of Munich (TUM), Garching, Germany, <https://orcid.org/0000-0002-2145-6133>; **Qibin Shi:** Department of Earth and Space Sciences, University of Washington, Seattle, Washington, U.S.A.; **Julien Thurin:** Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, U.S.A., <https://orcid.org/0000-0001-9982-3768>; **Chenxiao Wang:** Texas Advanced Computing Center, The University of Texas at Austin, Austin, Texas, U.S.A., <https://orcid.org/0009-0008-4031-1782>; **Kaiwen Wang:** Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, U.S.A., <https://orcid.org/0000-0001-7838-6984>; **Jeremy Wing Ching Wong:** Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California at San Diego, San Diego, California, U.S.A., <https://orcid.org/0000-0001-5523-8145>; **Sebastian Wolf:** Chair of Scientific Computing in Computer Science, Department of Computer Science, Technical University of Munich (TUM), Garching, Germany; and **Congcong Yuan:** Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, U.S.A., <https://orcid.org/0000-0002-1877-5539>

Manuscript received 31 October 2024

Published online 5 June 2025