



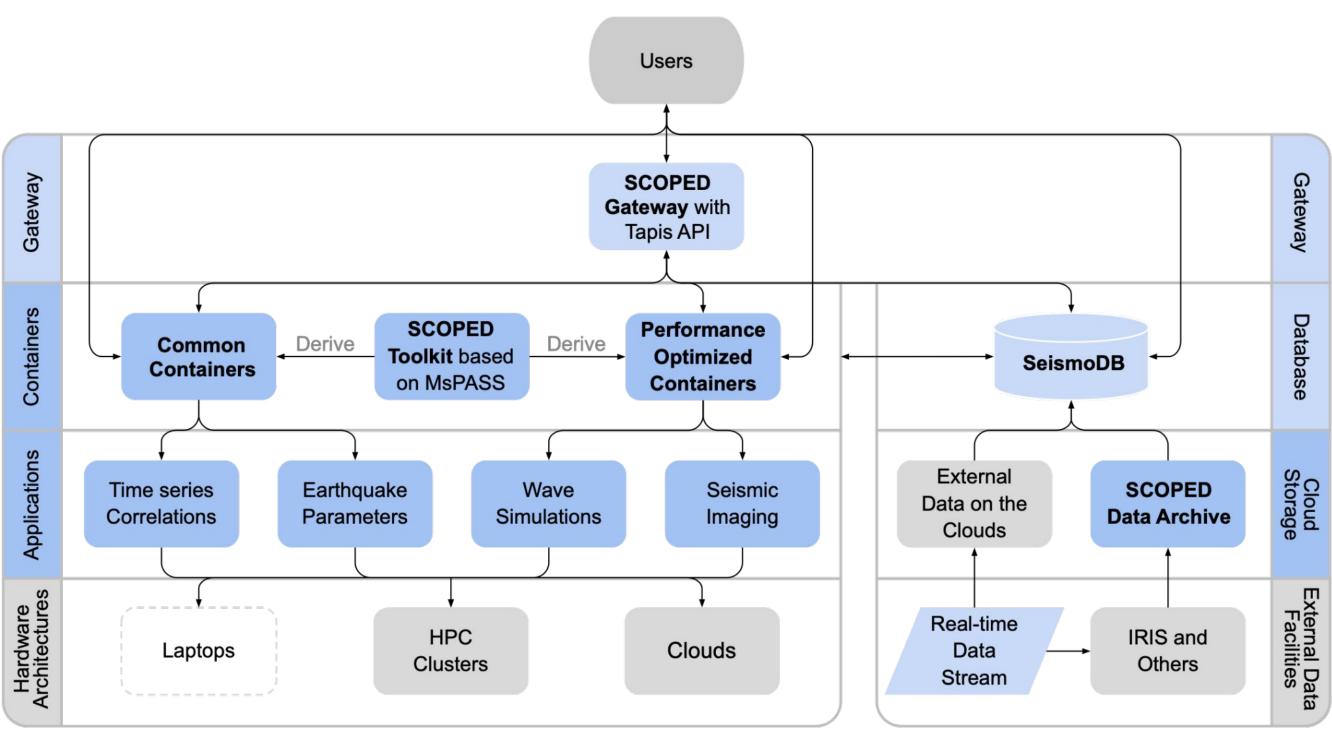


PI TEAM:

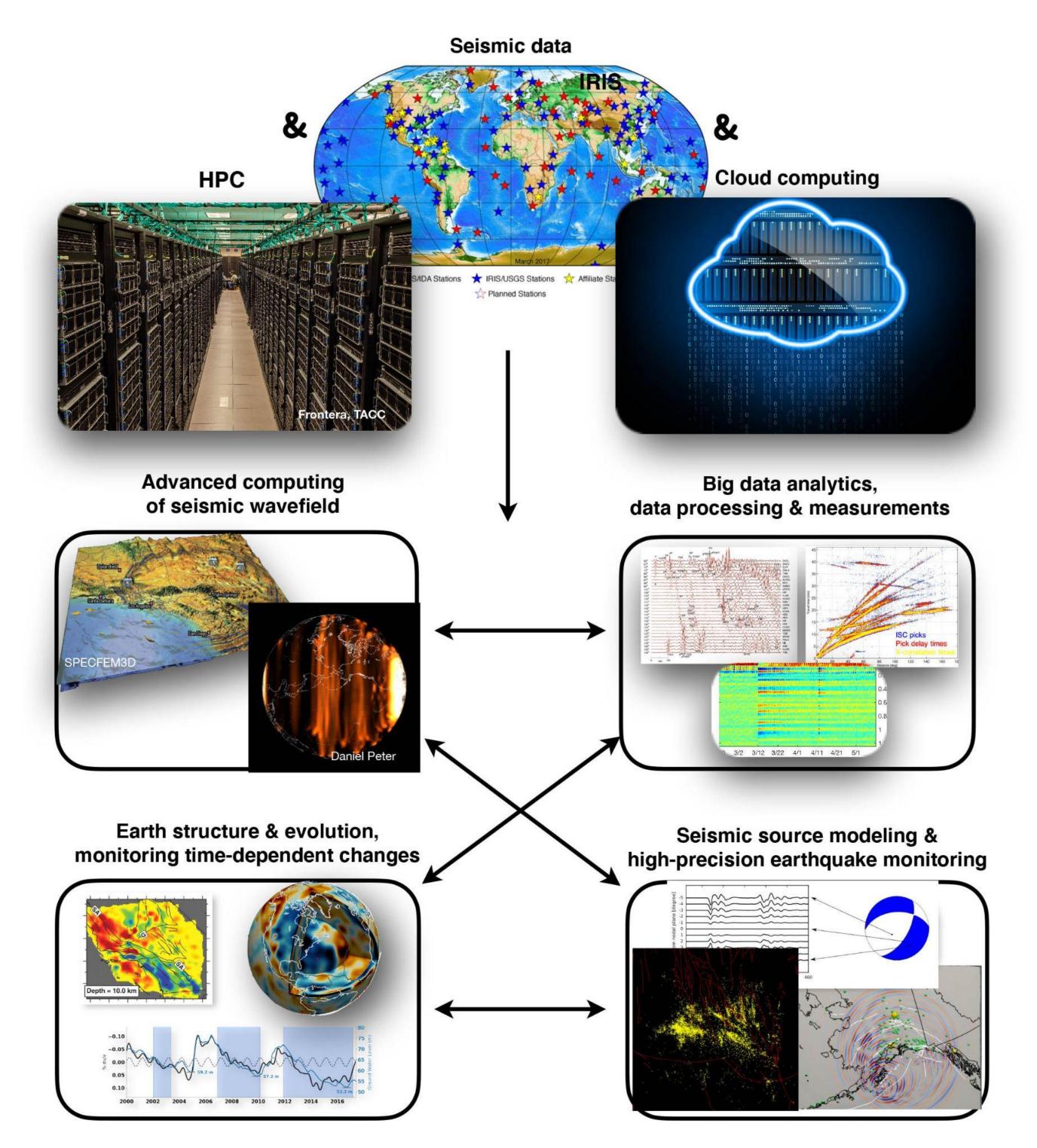
Ian Wang (iwang@tacc.utexas.edu): high-performance seismology, filesystem, database Ebru Bozdag (bozdag@mines.edu): full waveform modeling and tomography Marine Denolle (mdenolle@uw.edu): ambient noise seismology and cloud computing Felix Waldhauser (felixw@ldeo.columbia.edu): earthquake source precisions and catalogs Carl Tape (ctape@alaska.edu) [PI]: full waveform modeling, earthquake sources and uncertainties

Overview

Seismology is a data and model-intensive field. Data-driven computations benefit from Cloud Computing because observational seismology relies on horizontal scalability. Model-driven computation benefits from High-Performance Computing because simulations necessitate large memory and a high number of workers for efficient parallelization. Our project SCOPED (Seismic COmputational Platform for Empowering Discovery) aims to bridge both observational and theoretical fields by building a CyberInfrastructure (software, data) as a service to the seismic community. The challenges ahead are the benchmarking of open-source codes in cloud and HPC environments, the containerization of these codes adapted to various computing architectures, and the choice of data formats adapted to both Cloud and HPC computing, among others.



A schematic diagram of the proposed SCOPED Cyberinfrastructure that combines computation (left) and data (right). The SCOPED deliverables will be either fully functional (dark blue) or prototyped (light blue). The external components that SCOPED will interact with (gray) are either computing facilities, data archives, and users.



Representation of the unified and interconnected SCOPED Cyberinfrastructure described in the science components and cyberinsfrastructure. At the top, the global map of seismic stations represents one of several featured sources of data sets. Data are hosted and processed on HPC and on Cloud clusters, which overly pillars of wavefield simulations (left column) and data processing and analysis (right column). Cross-disciplinary research will facilitate the investigations of outstanding questions about Earth and fault dynamics

SCOPED: Seismic COmputational Platform for Empowering Discovery [Year 2]

NSF Cyberinfrastructure for Sustained Scientific Innovation (CSSI) Frameworks Project

[TAAC/UT: OAC-CSSI 2103494] [MINES: OAC-CSSI 2103621] [UW: OAC-CSSI 2103701] [Lamont/Columbia: OAC-CSSI 2103741] [UAF: OAC-CSSI 2104052]

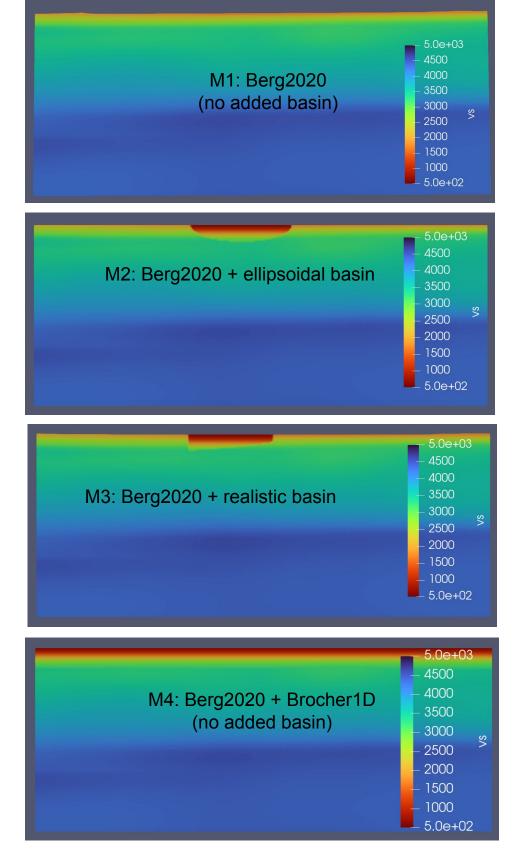


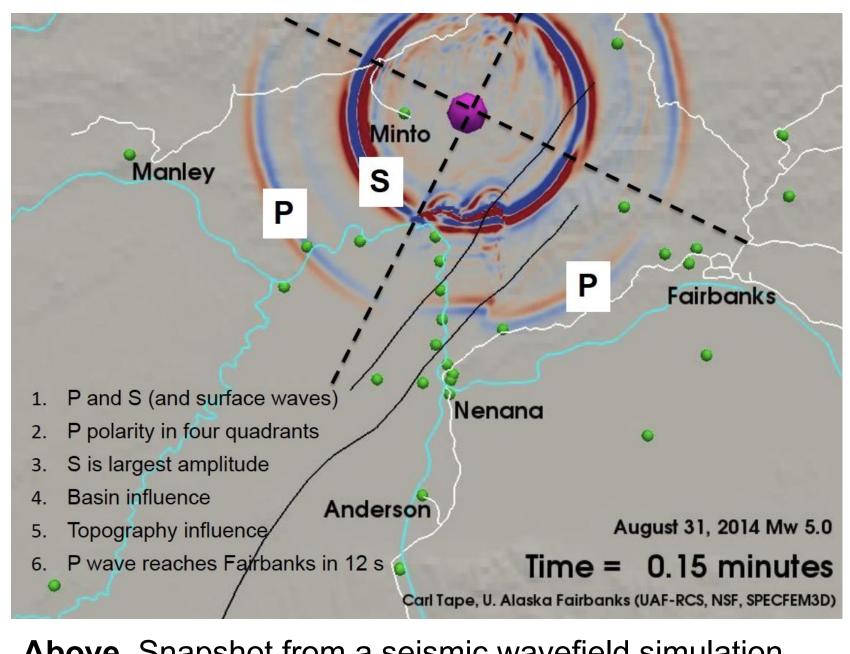
Advancing science with SCOPED codes

SCOPED focuses on four seismological applications enabled by several open-source software packages. We seek to improve these packages, while also containerizing them for easier use by others.

Seismic wavefield modeling with spectral-element method simulations

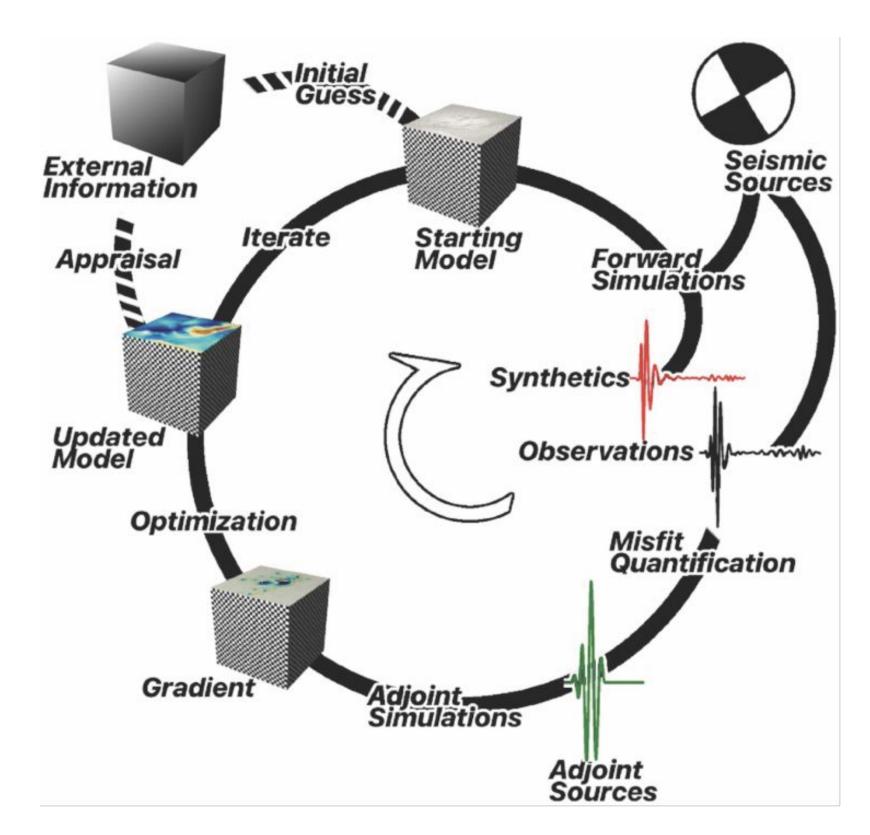
Wavefield simulations within 3D Earth models provide synthetic seismograms that can be compared with recorded seismograms, either to better understand earthquake sources or to improve the subsurface characterization of Earth's structure.





Above. Snapshot from a seismic wavefield simulation, showing an S wavefront entering into a deep sedimentary basin. Left. Cross sections of four 3D models used by Yuan Tian to investigate the influence of basin structures on the seismic wavefield.

Seismic imaging using recorded seismograms and simulated seismograms



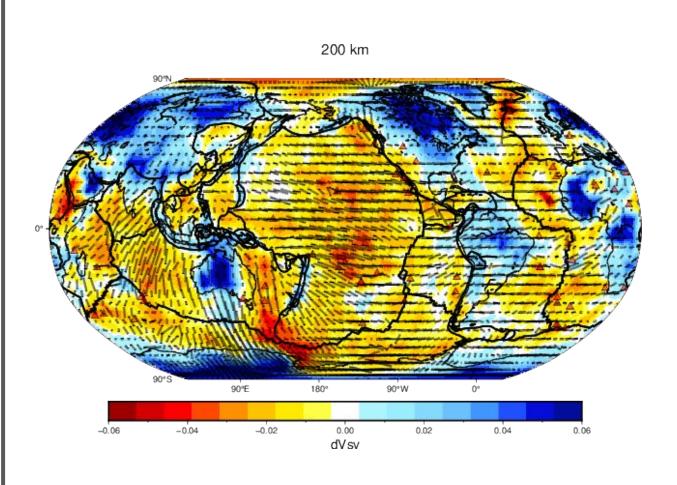
Seismic imaging of the whole Earth

Iterative workflow for seismic imaging using 3D wavefield simulations and adjoint methods (Chow et al., 2020). This workflow is automated with the help of Pyatoa software.

The seismic imaging workflow represents an optimization problem whereby the misfit between recorded and simulated seismograms is minimized while iterating toward a more accurate representation of the subsurface Earth properties (Vp, Vs, density).

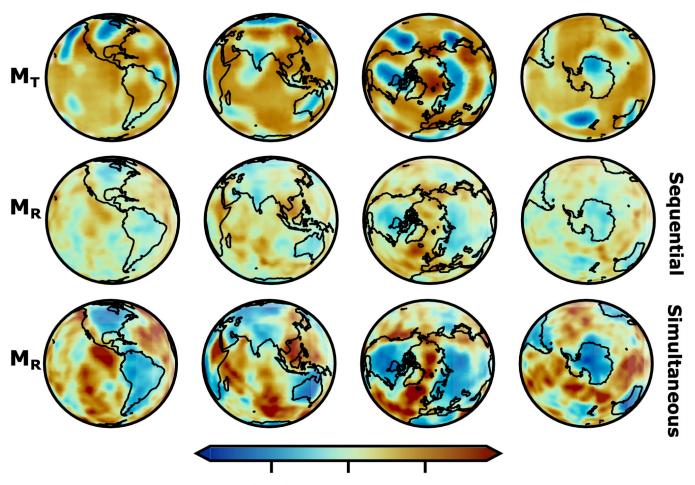
Seismic imaging of the entire Earth is computationally challenging, as it requires thousands of simulations within a large domain, as well as massive storage of temporary volumetric fields. Furthermore, the physics of Earth's interior must be addressed properly in the inverse problem as well as global data coverage problem.

We participated in *Texascale days on TACC's Frontera system* and successfully performed two iterations scaling up our simulations on ~8000 nodes with our current resolution. The main takeaways are: 1) GPU computing is necessary to further increase number of earthquakes or the resolution of simulations. 2) **ADIOS library** and compression help reduce the **IO** challenges.



Azimuthally anisotropic global adjoint model GLAD-AZI-M50 (Bozdag, Örsvuran et al. in prep.)

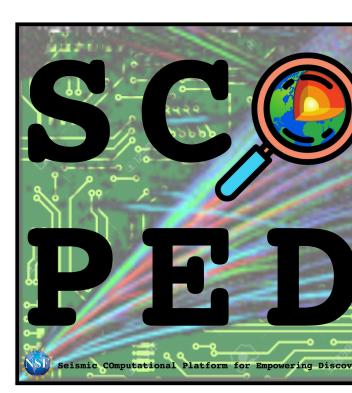
constructed on TACC's Frontera system.



-25.0 0.0 25.0 δQ_{μ}^{-1} [%]

Towards anelastic global FWI. We performed 3D global tests on TACC's Frontera system to explore the parameter trade-offs between wavespeeds and attenuation (Carmona et al. in revision for Geophys. J. Int.).

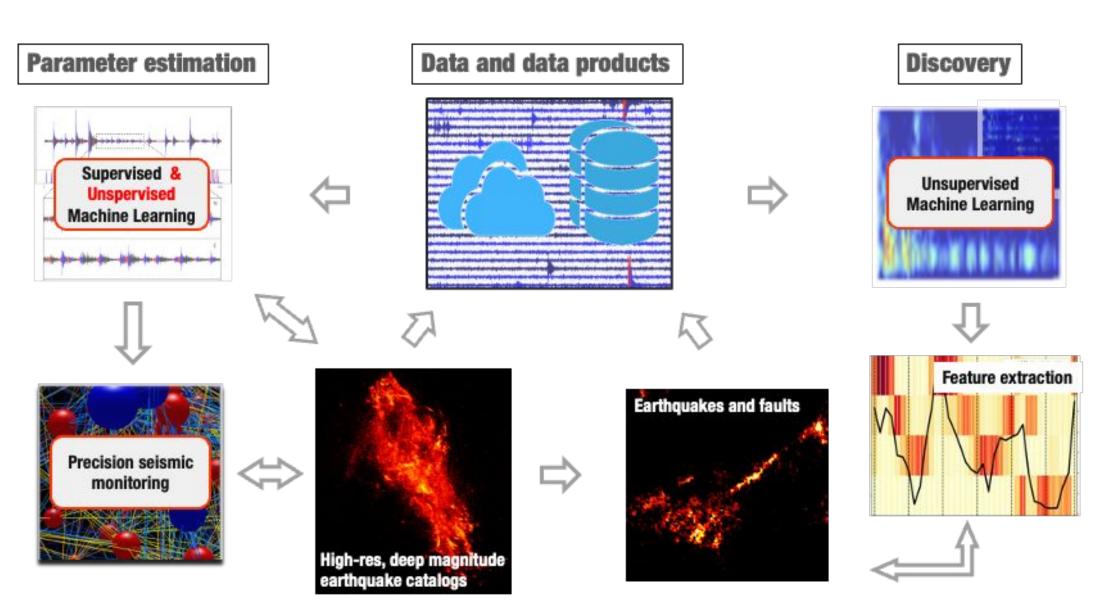
project website: https://seisscoped.org



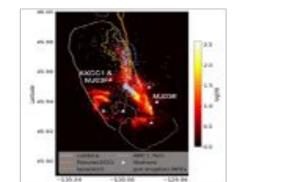
Ambient noise seismology We delivered a cloud-optimized solution for for Distributed Acoustic Sensing storage The surface of the Earth is in a constant state of and processing that uses TileDB or Zarr. (Ni low-amplitude ground motion that can be used for et al, in press in SRL) monitoring and for seismic imaging. HPC - Filesystems for large-scale ambient noise cross correlations tested seismic stations in California - target 100TB of data on Frontera (pers. comm. Kurama Okubo, Frontera). Use large hierarchica formats (H5) for data output, maximize local SSD or RAM for intermediate data products. Deploy over 500+ cores. High throughput on scratch filesystems. Parallelization in time (e.g., days of data) using distributed memory and station and station pairs using shared memory on nodes. We use SLURM to manage jobs on cluster. **Cloud - Object Storage:** When raw data is stored on object storage (e.g., S3). Optimal when using millions of small objects (e.g., mseed/npy day files < 1MB) due to the writing and parallelization needs. We use AWS Batch to manage jobs on instances. Use S3-optimized instances for improved throupughts. SCEDC Earthscope SCOPED is supporting the refactoring of Noisepy, an SCOPED र्योक ambient noise seismology python package. The parallelization now uses distributed and shared py [tile]DB Lambda - serverless - Streaming memory from S3 memory. The back-end data formats are flexible and user-defined, we tested on H5, Numpy-tar NPZ, and Per day Zarr cloud optimized data formats such as TileDB and Zarr Data sorted by time chunk Joisepy will be delivered to the community in fall 2023 with extensive testing and contributing guidelines. It cross correlate() has extensive continuous integration and improved Pmap: processing/fft/xcorr code coverage, while preserving the original work of the package from Jiang and Denolle (2020) Zarr [tile]DB Zarr We will demonstrate this on the S3-hosted SCED Data sorted by stack() Pmap: pair data set. We have demonstrated such workflow on station pair Per channel-pair AWS for Distributed Acoustic Sensing.

Precision seismology: large-scale, high-resolution catalog production and analysis

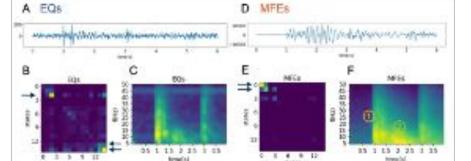
Data driven workflows to enable discovery



Discovery of precursory mixed-frequency earthquakes (MFE) prior and leading up to 2015 eruption of Axial volcano



Supervised ML detectior cross-correlation timing, high-res Blue dots show earthquakes. mixed-frequency earthquake: with unsupervised ML. Wang et al., 2023

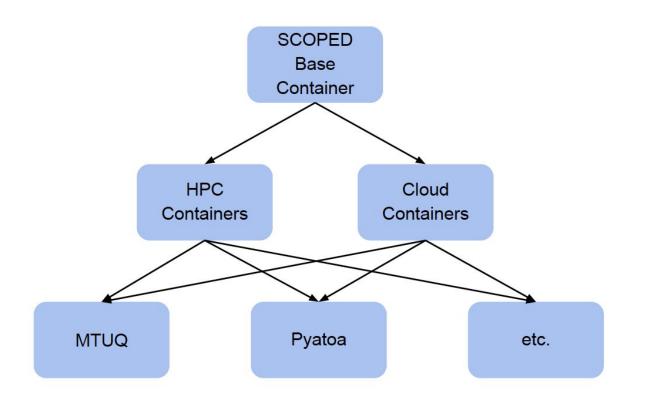


spectrograms (C and F) using 100 events in each cluster.

Pre-eruption evolution of MFEs in ~1 week, ~2 days, and 8 hours time scale. Dashed red curve shows binned seismic moment of the MFEs leading up to the (dashed black line). Arrows in B) point to northward and southward migration of MFEs prior and leading up to the eruption. Wang et al. 2023

SCOPED Container Registry

We now have adjtomo, MTUQ, pysep, specufex, and MsPASS added to the SCOPED Container Registry built with the SCOPED Base Container. We also added an external software, SeisSol, to our container collection using the base container.



SCOPED Gateway

The SCOPED Gateway prototype has been created through the utilization of Tapis. Currently, we have integrated MTUQ and MsPASS as HPC applications added into the gateway. Ultimately, this will enable all SCOPED users to access containerized applications and execute them on either Cloud or HPC systems.

SCOPED												
								Incomposition and a second second				
			System List					+ Create a New System				
Disettoard C	Dashboard for		B Stampada area				••• = = <		deserves and some hard to solve the set	rantera, esecularata (MJD10003), Secilara mijoba	ATTN: A	0 Ó
	② You are not logged in			s.edu) System Details					unrekens Fobu sisinings			
			Portera conc				SCOPED					jadetevi
	Please log in to use SCOPED						(Darbeard					@ +
	Proceed to login (0)		B spd-tpc-jacksont (stampood-face steam)		IN .		E System	Files / Frontena.exec / scratch1 / 0905	a06301145		Cary Tables Decred	GP + Upload Faller
	Processo to eight			tenant: scoped			E1 794	C Surpade ever	D Nave		Stee	Last Medified
				id;			EE ANPI	Promota even	D D 2000407207203050	MT_ampRude_ration	4118	6712002313.32
				Stampede.ex	ec		E .444	E] apt-tpt-jadaam	D 🛅 200040720-200000	MT_ansi	4.018	1011010103 10.02
• • • · · ·	> 0 🔒 seisscoped org/tapis-u//#/		⊕ ₫ + ©	description:					D 👌 2000-0720-203300	WT_Insufficial proj	42.1 iB	£7112/0023 13-31
· · · ·	Interception of Meetingson (A)		0 0 7 0	Stampede2 H	IPC system for SCOPED					FMT_Builtoop	4.0.48	62112/2023 15:32
SCOPED			jacksont@scoped *	systemType:					D 🛅 2000407201200304	MT_nagear	4.038	821122022312322
				LINUX					D 🗁 2000-0720-003300	Mit_sealer	4334	E21(10033-13-32
Deshboard	Dashboard for SCOPED			owner:					D D 20004072072003070	MT_orgins.jeon	180-3 lighes	60112/0023 13:30
Oyatama	Dashooard for SCOPED			jacksont host:					D 🖸 2000407207203000	MT_soutoes'	4.018	60112/0129 12:32
	low more than the second se	(Internet and Internet)		stampede2.ta	soc utexas edu					MT_stations (son	0.718	0710/0020 10:02
Files	Systems	E Files			5 n	i setsceped.org/aps-		4	⊕	0	4.0.18	61112/0025 13:52
Apps	0 million	Files and lable as 0 and a								and a state	4.618	60112/0429 13:32
Jobs	3 systems	Files available on 3 system		SCOPED					jacksort@scop	od * andar	4118	62110/2022 13:32
	View TAPIS systems	Access files available on TAPIS s	ystems	() Desitioned	Apps					are a reduction'	4.2 18	E0112/0020 13:02
				E Systems	white					allow y	4218	E010/20129 13:32
	Go to Systems	Go to Files	0	E Res	SE smuq vila					supervised.	536.8 kB	87%3/8823 13:31 87%3/8823 13:22
				EE Assa	88 564PASS v1.0	Job Launcher					6101 bytes	6712002515.22
	SB Applications	🖾 Jobs		El John	88 Bhread-juckaset v0.10			Summary			125.0 bytes	8713/0023 12:32
						Job Submission		Job Neme	edt		414	61712/2022 13:22
	3 apps	100 jobs View status and details for previously launched SCOPED jobs				Successfully submitted job c7 b18a-33e2do498bc9-007	772deds-61s9-4e25-	MiPASS-1.0 MiPASS Containerized; currently support	ta running a MaPASS		2118	671129125 19:22
	View SCOPED applications and launch jobs			-		0100-0000000000000	·	Jupyter Notebook on Fromera Application: MoPADS v1.0				
						This is a preview of the json job		Execution Options	edit			
	Go to Applications					may copy it for future reference	20 SUDHISSION Data. YOU 28.	Frontera asses tacisticornal		C interest open of the	NAME AND AND AND ADDRESS OF A	
		Go to Jobs	٢			Simplified	D Copy	Command Profix: system detault	1004E			(LANSING MALE)
								File Inputs	0.444	ite .		
								File Input Arrays		E MARTINE JACTOR		
								Arguments	11		805-803940-138564-007	
								App: /usribbin/tini is ig /usribbin/blantin Container:		ED REPORTS COLUMN	601-060846116804-007	
								Environment Variables		E MARK I CARDINE STOP		
								TERM : xterm-256oslor		E sena is ones more		
								Scheduler Options Scheduler Profile: none selected		El administrationed and an and a second and		
								Scheduler Profile: nono selected Scheduler Args:			nomeniki monorek nominy a ShiPASS Janyike Bahatanak mi	Pomera
								Archiving		E main month MAG		
								Archive System ID: default Archive System Directory: default		E manification Sample American		
								Archive On App Error: false Includes: 0		E manufit man 205-04079302140	1042	
								Decludes: 0		E) manager in program in province		
								Job Submission		E manual free lands		

Lamont-Doherty Earth Observatory Columbia University | Earth Institute



COLORADOSCHOOLOFMINES.



Training users of SCOPED codes

Jupyterhul

SCOPED contributed toward two training workshops: **Specfem** on October 5–7, 2022 (Bryant Chow, Carl Tape), and High-Performance Seismology Cybertraining on May 9–12 (Marine Denolle). These free workshops were attended by hundreds and included downloadable containerized software for users to gain research-level experience.

	•••			Zoom Meeting	8/4-8///				
	🥏 💿 Recording 💵 🗖				III View		Participants (186)		
2022 SPECF	EM virtual	workshop fo	or users (Oc	ctober 5–7,	2022)	Q Search			
						HC H	ost Carl Tape (Host, me)	o 🔏 🗅	
containerized	S MA Read	organize	4			🚭 на	ost - Amanda McPherson (Co-	-host) 🔏 🗔	
software			Gaurav Tomar			HA Ho	ost Aakash Gupta (Co-host)	<u>¥</u> 🕞	
running on	Host Ebru Bozdag	Host Carl Tape	🔏 Gaurav Tomar	Host Aakash Gupta	∦ Host - Yuan Tian	Вно	ost Bryant Chow (Co-host, Gue	est) 🔏 🗅	
user's laptop	Harris and		· _			Je T	roen Tromp (Guest)	₽ C4	
host team of 10		do			7 51	LV Lu	iis Vazquez (Guest)	₽ □:	
nost team of to	Jeroen Tromp speak	deve oper daniel peter	Cristian	发 Joses Omojola	🔏 Host - Amanda McPherson	AC Aa	aron C Hirsch (Guest)	¥ 🕫	
breakout rooms		fead		0		AK Aa	aron Kearns (Guest)	¥ 74	
for 1-1 help	1/8	host				AK Ab	oolfazl Khan Mohammadi (Gue	st) 🔏	
	🔏 jorge	Host Bryant Chow	Host Andrea Riano	Host Julien Thurin	Host Ridvan Örsvuran		lam Ringler (Guest)	¥ 🕫 1	
poll questions	9.5	The question				Ali	i Imran Sandhu (Guest)	¥ 🕫	
to gauge pace	and S an	-	R	zoya	Shubham Tiwari	Invi	ite Mute All	More	
	Host Félix Rodríguez Cardozo	Luis Vazquez	🔏 Li Ren	🔏 zoya	🔏 Shubham Tiwari		ie Mule All	More	
post-workshop						×	Chat		
surveys for	VS	Toufik	Ali Imran Sandhu	Jonas Kintner	Kuan-Fu Feng		to Everyone ease raise your hand if you	8:44 AM I have a	
assessment	🗶 vs	🗶 Toufik	🌋 Ali Imran Sandhu	📕 Jonas Kintner	🌋 Kuan-Fu Feng		estion for Jeroen.		
							lick Reactions from the zoc en Raise Hand.)	om bar,	
	186						2 Who can see your messages? Recording On		
	🤌 n 🖬 n 🕎 🚉 186 n 🗭 n 🖬 n 🖬 🔟 🔲 🚭 🕄 \cdots 📻						vone 💌	Ů ☺ …	
	Unmute Stop Video	- and the second state of		Pause/Stop Recording Reactions	Apps More		ssage here		
					total size is 175561 speaduo	16 1 00			

Zoom screenshot from a free virtual workshop on Specfem3D, held October 5-7, 2022, and hosted by UAF. The screenshot shows a range of participants, including the organizer (Carl Tape), the guest speaker (Jeroen Tromp), the principal developer (Daniel Peter), the lead host (Bryant Chow), a participant asking a question (Luis Vasquez), and the host team: Ebru Bozdag, Aakash Gupta, Yuan Tian, Amanda McPherson, Andrea Riano, Julien Thurin, Ridvan Orsvun, and Felix Cardozo. At this starting time of the workshop, there were 186 participants.

Summary of attendance

There were over 200 interested participants, 85 online participants the first day, then a steady 30-50 participants throughout the week. We had 175 Slack members and over 2.6K slack messages during the live debugging sessions. The project resulted in a Youtube Channel and are writing a JupyerBook with the materials



rn about the application he

MsPASS: Massive Parallel Analysis System for

• Core implementation in Python and C++

• Scalable parallel processing framework using

• All components are containerized that can be

pulled by Docker or Singularity for distribution

• Data provenance support at global and object

Data management using MongoDB

Seismologists:

levels

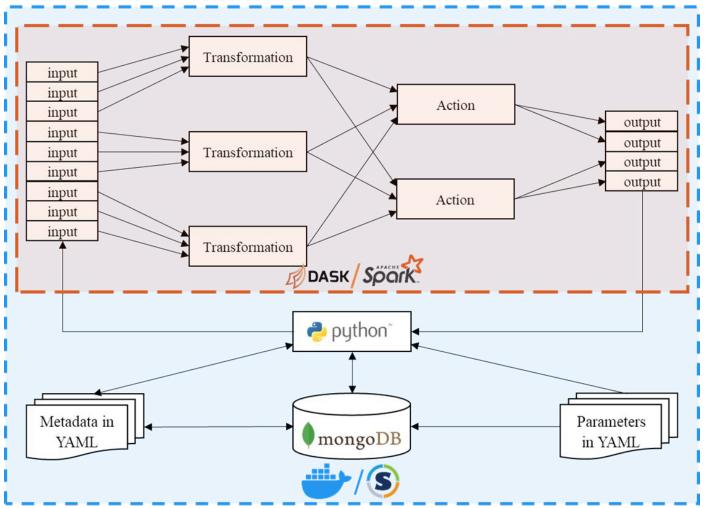
Spark or Dask



Establishing the SCOPED platform

We have established a base software container for SCOPED that contains basic Python packages, as well as ObsPy, pysep, and MsPASS (see below). The container can run on the gateway, offering both batch and interactive modes through Jupyter Lab.

Input l_____/ l_____/ l_____/ l_____/



References and Links

SeisNoise: https://github.com/tclements/SeisNoise.j NoisePy: https://github.com/mdenolle/NoisePy Ambient Noise on the cloud: https://github.com/Denolle-Lab/C4-Project.jl MSPASS: <u>https://www.mspass.org/</u> Mtuq: https://uafgeotools.github.io/Mtug/index.html Specfem2d: https://github.com/geodynamics/Specfem2d Specfem3d: https://github.com/geodynamics/Specfem3d Specfem3d Globe: https://github.com/geodynamics/Specfem3d Globe Pysep: https://pysep.readthedocs.io/en/latest/ Pyatoa: https://pyatoa.readthedocs.io/en/latest/: Seisflows: https://seisflows.readthedocs.io/en/latest SPECUFEX: https://github.com/Specufex/specufex

BPMF: https://github.com/ebeauce/Seismic BPMF

HYPODD: https://www.ldeo.columbia.edu/~felixw/HYPODD/

PHASENET: https://github.com/wayneweigiang/PhaseNet

QUAKEFLOW: https://github.com/wayneweigiang/QuakeFlow

SCOPED container registry: https://github.com/orgs/SeisSCOPED/packages SCOPED base container:

https://github.com/SeisSCOPED/container-base The SCOPED project is supported by the National

SCOPED website: <u>https://seisscoped.org/</u>

Science Foundation via the Cyberinfrastructure for Sustained Scientific Innovation program of the Office of Advanced Cyberinfrastructure; award numbers to each PI institution are listed at the top.