

Real-time Bayesian inference at extreme scale: A digital twin for tsunami early warning applied to the Cascadia subduction zone

Stefan Henneking

Oden Institute, The University of Texas at Austin

Collaboration Team: Sreeram Venkat, Milinda Fernando, Omar Ghattas (UTA),
Veselin Dobrev, John Camier, Tzanio Kolev (LLNL),
Alice Gabriel (UCSD)

MFEM Community Workshop
September 10–11, 2025 — Portland, OR

Collaboration Team



Stefan Henneking (UTA)



Sreeram Venkat (UTA)



Milinda Fernando (UTA)



Omar Ghattas (UTA)



Alice-Agnes Gabriel (UCSD)



Veselin Dobrev (LLNL)



John Camier (LLNL)



Tzanio Kolev (LLNL)

H., Venkat, Dobrev, Camier, Kolev, Fernando, Gabriel, and Ghattas, *Real-time Bayesian inference at extreme scale: A digital twin for tsunami early warning applied to the Cascadia subduction zone, 2025.*



Real-time Bayesian inference at extreme scale: A digital twin for tsunami early warning applied to the Cascadia subduction zone

Stefan Henneking

Oden Institute

The University of Texas at Austin

stefan@oden.utexas.edu

Sreeram Venkat

Oden Institute

The University of Texas at Austin

svenkat@utexas.edu

Veselin Dobrev

Center for Applied Scientific Computing

Lawrence Livermore National Laboratory

dobrevl@llnl.gov

John Camier

Center for Applied Scientific Computing

Lawrence Livermore National Laboratory

camier1@llnl.gov

Tzario Kolev

Center for Applied Scientific Computing

Lawrence Livermore National Laboratory

kolev1@llnl.gov

Milinda Fernando

Oden Institute

The University of Texas at Austin

milinda@oden.utexas.edu

Alice-Agnes Gabriel
Scripps Institution of Oceanography
University of California San Diego
al gabriel@ucsd.edu

Omar Ghattas
Oden Institute, Walker Department of Mechanical Engineering
The University of Texas at Austin
omar@oden.utexas.edu

Abstract—We present a Bayesian inversion-based digital twin that employs acoustic pressure data from seafloor sensors, along with 3D coupled acoustic-gravity wave equations, to infer earthquake-induced spatiotemporal seafloor motion in real time and forecast tsunami propagation toward coastlines for early warning with quantified uncertainties. Our target is the Cascadia zone, with one billion parameters. Computing the posterior mean alone would require 50 years on a 512 GPU machine. Instead, exploiting the shift invariance of the parameter-to-observable map and devising novel parallel algorithms, we induce a fast offline-online decomposition. The offline component requires days to solve a global system of partial differential equations (PDEs) using a large-scale unstructured mesh finite element method (FEM), while this part of the computation to the full El Capitan system (43,520 GPUs) with 92% weak parallel efficiency. Moreover, given real-time data, the online component exactly solves the Bayesian inverse and forecasting problems in 0.2 seconds on a modest GPU system, a ten-billion-fold speedup.

Index Terms—Bayesian inverse problems, data assimilation, uncertainty quantification, digital twins, tsunami early warning, finite elements, real-time GPU supercomputing

I. JUSTIFICATION FOR ACM GORDON BELL PRIZE

Fastest time-to-solution of a PDE-based Bayesian inverse problem with 1 billion parameters in 0.2 seconds, a ten-billion-fold speedup over SoA. Largest-to-date unstructured mesh FE simulation with 55.5 trillion DOF on 43,520 GPUs, with 92% weak and 79% strong parallel efficiencies in scaling over a 128× increase of GPUs on the full-scale *El Capitan* system.

This research was supported by DOD MURI grants FA9550-21-1-0084 and FA9550-21-1-0327, DOE ASCR grant DE-SC0021471, and NSF grants DGE-2137420, EAR-2225286, EAR-2121568, OAC-2119536. Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-PROC-200475). This research used resources from the Texas Advanced Computing Center (TACC), the National Energy Research Scientific Computing Center (NERSC), ALCC ERCCAP0300871, and Livermore Computing.

II. PERFORMANCE ATTRIBUTES

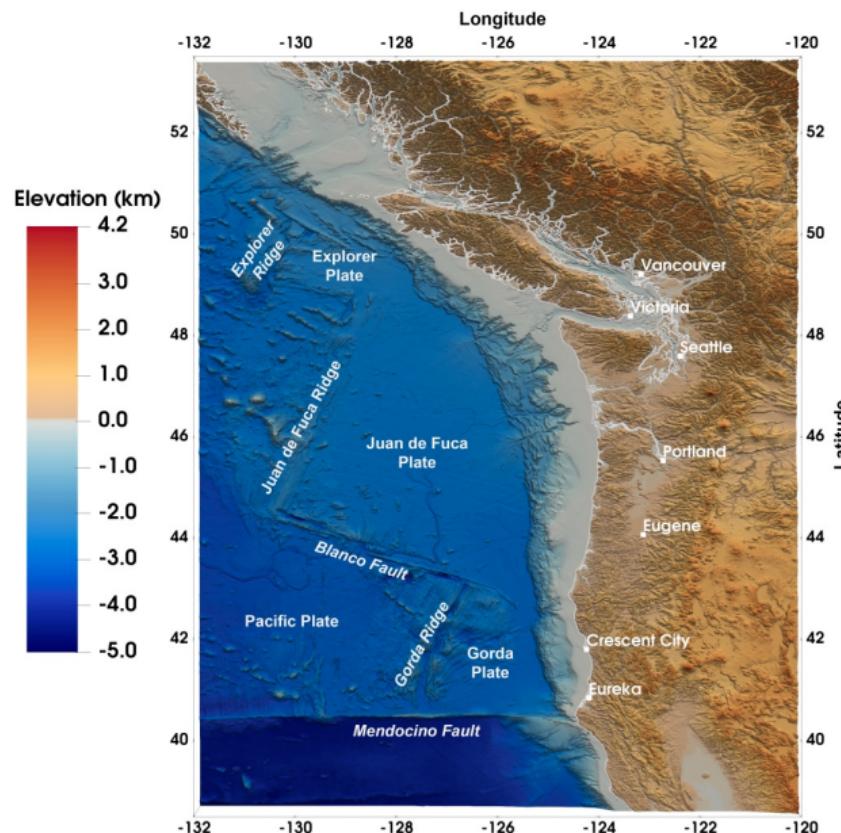
Performance attribute	This submission
Category of achievement	Scalability, time-to-solution, peak performance
Type of method used	Bayesian inversion, real-time computing, FEM
Results reported based on	Whole application including I/O
Precision reported	Double precision
System scale	Results measured on full-scale system
Measurement mechanism	Timers, DOF throughput, FLOP count

III. OVERVIEW OF THE PROBLEM

A. Tsunami forecasting and the Cascadia subduction zone

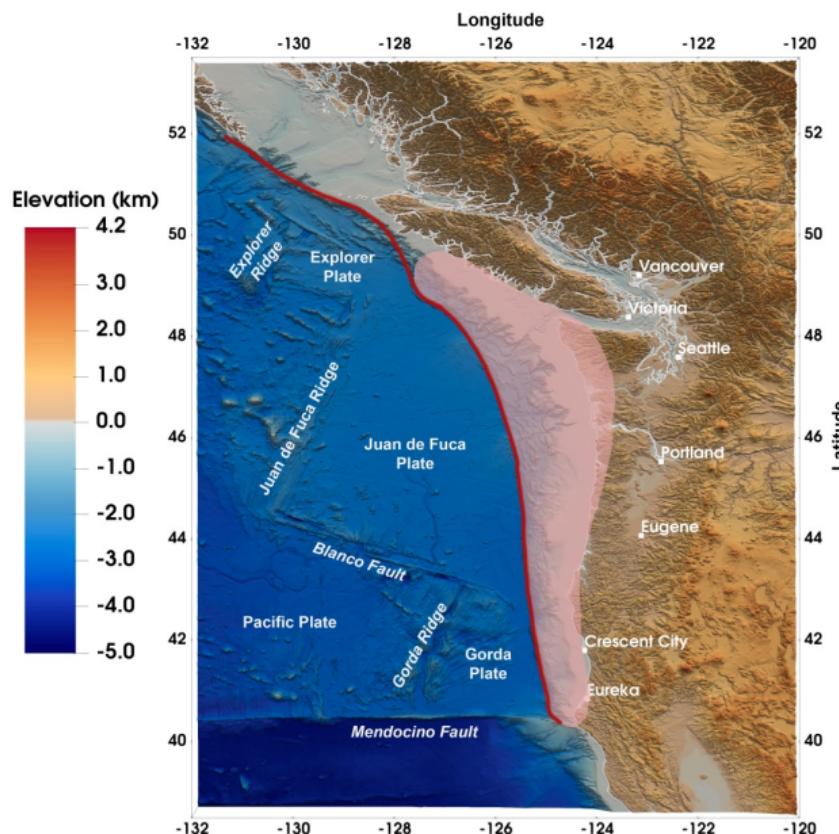
Tsunamis are rare events but can be extremely deadly and cause catastrophic socioeconomic losses. In the past century tsunamis have claimed more than 400,000 lives, and individual tsunamis can result in hundreds of billions of dollars in losses, making them among the costliest natural disasters. State-of-the-art (SoA) tsunami early warning systems [1] rely on rapid characterization of source parameters primarily from seismic data [2], issuing alerts within minutes but based on simplified assumptions about the event. These systems typically infer earthquake moment magnitude and hypocenter location, and then trigger tsunami forecasts using precomputed scenarios or simplified fault models, assuming instantaneous, uniform slip on a predefined fault geometry. This approach is inadequate in the near field [3], where destructive tsunami waves can arrive on-shore in under ten minutes. In addition, current methods fail to capture the complexities of earthquake rupture dynamics [4], including variable slip distributions and slow rupture speeds. This can lead to delayed, missed, or false warnings, especially when early seismic data misrepresent the true tsunami potential, such as during the 2011 Tohoku, Japan tsunami or the 2024 Cape Mendocino, California earthquake,

Cascadia Subduction Zone



GEBCO bathymetry data of the Cascadia Subduction Zone (CSZ) with $\sim 250\text{m}$ resolution.

Cascadia Subduction Zone



The shaded area illustrates the “locked” part of the interface between the North American and subducting plate. The red line marks the seaward edge where the subducting plates begin their descent beneath the North American Plate.

Evidence for 1700 Cascadia Megathrust M9.0 Earthquake



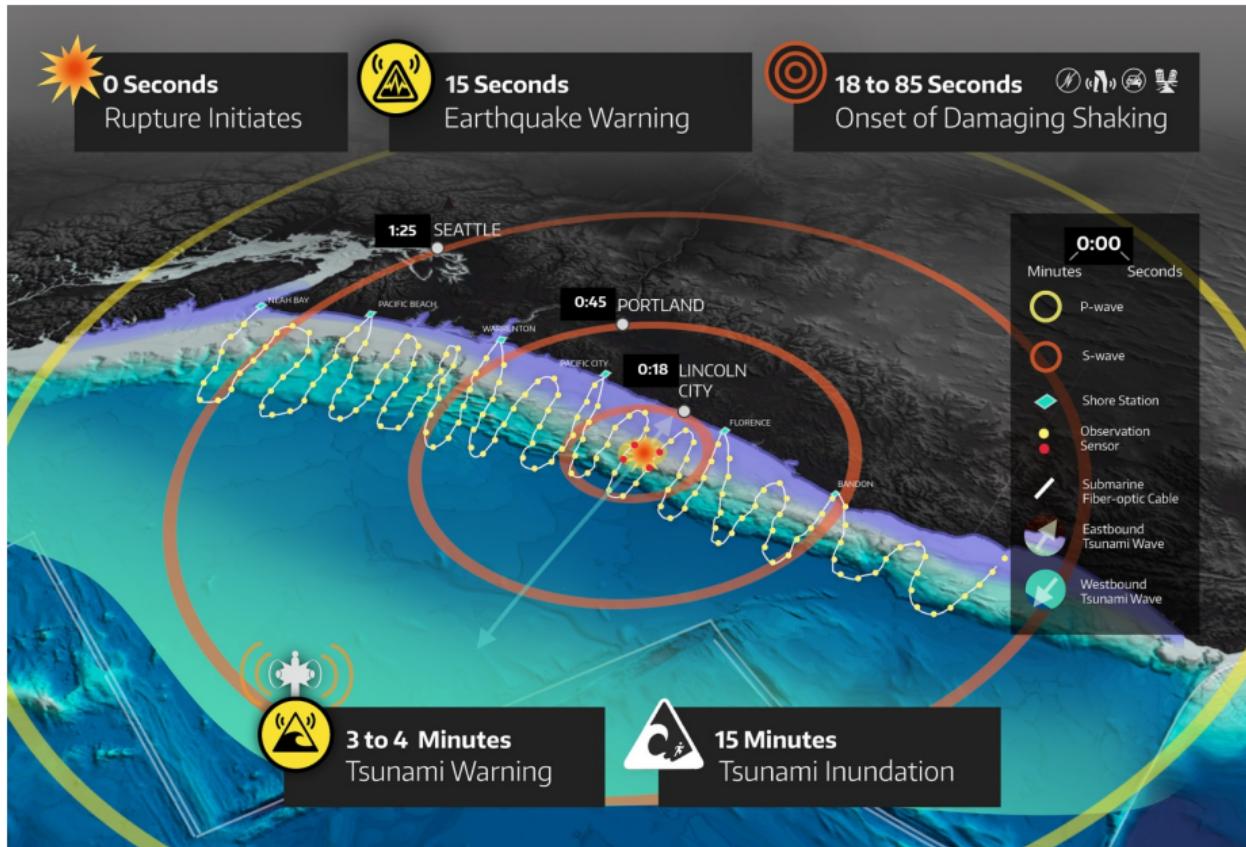
Neskowin Ghost Forest



Alternating layers in pit, Tofino, BC (Clague, 97)

- Cascadia subduction zone has seen major ruptures in 1700 AD, 1310 AD, 810 AD, 400 AD, 170 BC and 600 BC
- Current estimate: 37% probability of M8.2+ event within 50 years; 10–15% probability that entire Cascadia subduction zone will rupture with an M9+ event

Cascadia Subduction Zone: Proposed Sensor Network¹



¹D. Schmidt et al., 2019

Current forecasting models rely on shallow water equations for tsunami propagation

- Efficient computation
- Works well in the far-field (observing hydrostatic pressure changes)
- But **does not make use of near-field pressure transients** from hydroacoustic waves

Cascadia subduction zone digital twin:

- **Predict tsunami propagation** by inversion of near-field pressure data recorded during a megathrust rupture via **inference of seafloor velocity**
- Employ **high fidelity model** of acoustic/gravity wave propagation—solution and inverse operators do not admit low rank approximation (Kolmogorov n -width problem)
- **Quantify uncertainty** in inversion & prediction via **Bayesian inference**
- Solve inverse problem in **real time** (order of seconds) to provide early warning

Advantages: Earlier & more accurate forecasting

Challenges: Billion-parameter inverse problem governed by billions DOF forward model

Outline

- 1 Introduction
- 2 Forward Model
- 3 Inference and Prediction Model
- 4 Application to Cascadia Subduction Zone

- Mixed formulation for pressure and velocity unknowns:²

$$\begin{aligned}\rho \frac{\partial \vec{u}}{\partial t} + \nabla p &= 0 \quad \text{in } \Omega \times (0, T) \\ \frac{1}{K} \frac{\partial p}{\partial t} + \nabla \cdot \vec{u} &= 0 \quad \text{in } \Omega \times (0, T)\end{aligned}$$

- Boundary conditions:

- ▶ ODE-type BC (sea surface):

$$p = \rho g \eta, \quad \partial \eta / \partial t = \vec{u} \cdot \vec{n} \quad \text{on } \Gamma_{\text{surface}} \times (0, T)$$

- ▶ Seafloor velocity (boundary source):

$$-\partial b / \partial t = \vec{u} \cdot \vec{n} \quad \text{on } \Gamma_{\text{bottom}} \times (0, T)$$

- ▶ First-order absorbing BC (outgoing waves):

$$\vec{u} \cdot \vec{n} = p / \rho c \quad \text{on } \Gamma_{\text{absorb}} \times (0, T), \quad c = \sqrt{K / \rho}$$

- Initial conditions (homogeneous)
- Implemented in MFEM with high-order finite elements and RK4 time-stepping

²Lotto and Dunham, 2015

- A weak form of the governing equations in the space-time domain is formally derived by multiplying with test functions $(\vec{\tau}, v)$ and integrating-by-parts the second equation; the ODE boundary condition is imposed weakly.

Find $\vec{u} \in \mathcal{U}$, $p \in \mathcal{V}$,

such that

$$\int_0^T \int_{\Omega} \left[\rho \frac{\partial \vec{u}}{\partial t} \cdot \vec{\tau} + \nabla p \cdot \vec{\tau} \right] d\vec{x} dt = 0, \quad \vec{\tau} \in \mathcal{U},$$

$$\int_0^T \int_{\Omega} \left[\frac{1}{K} \frac{\partial p}{\partial t} v - \vec{u} \cdot \nabla v \right] d\vec{x} dt +$$

$$\int_0^T \int_{\Gamma_{\text{surface}}} \frac{1}{\rho g} \frac{\partial p}{\partial t} v d\vec{x} dt + \int_0^T \int_{\Gamma_{\text{absorb}}} \frac{1}{\rho c} p v d\vec{x} dt = \int_0^T \int_{\Gamma_{\text{bottom}}} \frac{\partial b}{\partial t} v d\vec{x} dt, \quad v \in \mathcal{V}.$$

- Function spaces: $\mathcal{U} = (L^2(\Omega))^d \times L^2(0, T)$, $\mathcal{V} = H^1(\Omega) \times L^2(0, T)$ (+ homogeneous ICs)
- Surface gravity wave height η recovered from surface BCs

- Time-stepping operator:

$$[\mathbf{u}_\delta \ \mathbf{p}_\delta]^T = \mathbf{M}^{-1} \left(-\mathbf{A} [\mathbf{u} \ \mathbf{p}]_i^T + [\mathbf{f} \ \mathbf{g}]_i^T \right),$$

where $[\mathbf{u} \ \mathbf{p}]_i^T$ and $[\mathbf{f} \ \mathbf{g}]_i^T$ are respectively the state and right-hand-side (RHS) vectors at time instance i , and $[\mathbf{u}_\delta \ \mathbf{p}_\delta]$ is the state increment used by RK4.

- The mass matrix \mathbf{M} and stiffness matrix \mathbf{A} are discretizations of the block operators M and A defined by:

$$\left(M \begin{bmatrix} \vec{u} \\ p \end{bmatrix}, \begin{bmatrix} \vec{\tau} \\ v \end{bmatrix} \right) := \begin{bmatrix} (\rho \vec{u}, \vec{\tau}) & 0 \\ 0 & (K^{-1}p, v) + \langle (\rho g)^{-1}p, v \rangle_{\partial\Omega_s} \end{bmatrix}$$

and

$$\left(A \begin{bmatrix} \vec{u} \\ p \end{bmatrix}, \begin{bmatrix} \vec{\tau} \\ v \end{bmatrix} \right) := \begin{bmatrix} 0 & (\nabla p, \vec{\tau}) \\ -(\vec{u}, \nabla v) & \langle (\rho c)^{-1}p, v \rangle_{\partial\Omega_a} \end{bmatrix},$$

where $\vec{u}, \vec{\tau} \in (L^2(\Omega))^d$ and $p, v \in H^1(\Omega)$; (\cdot, \cdot) denotes the (component-wise) $L^2(\Omega)$ inner product, and $\langle \cdot, \cdot \rangle_{\partial\Omega}$ is the $L^2(\partial\Omega)$ inner product over (part of) the boundary $\partial\Omega$.

Convergence of Modal Solutions (2D)

Let k_n denote the wavenumber and ω the angular frequency of the wave. The homogeneous solutions satisfying the state system with source $\partial b(x, t)/\partial t = 0$ and prescribed frequency ω are given by:³

$$\begin{aligned} p(x, y, t) &= \sin(k_n x) \sin(\omega t) \left(\sinh(\kappa_n y) + \frac{g\kappa_n}{\omega^2} \cosh(\kappa_n y) \right), \\ u_x(x, y, t) &= \frac{k_n p_0}{\rho \omega u_0} \cos(k_n x) \cos(\omega t) \left(\sinh(\kappa_n y) + \frac{g\kappa_n}{\omega^2} \cosh(\kappa_n y) \right), \\ u_y(x, y, t) &= \frac{\kappa_n p_0}{\rho \omega u_0} \sin(k_n x) \cos(\omega t) \left(\cosh(\kappa_n y) + \frac{g\kappa_n}{\omega^2} \sinh(\kappa_n y) \right), \\ \eta(x, t) &= \frac{\kappa_n p_0}{\rho \omega^2 l_0} \sin(k_n x) \sin(\omega t). \end{aligned}$$

Each solution (mode) must satisfy the dispersion relation

$$\omega^2 = g\kappa_n \tanh(\kappa_n H), \quad \text{where } k_n = \sqrt{\kappa_n^2 + (\omega/c)^2}.$$

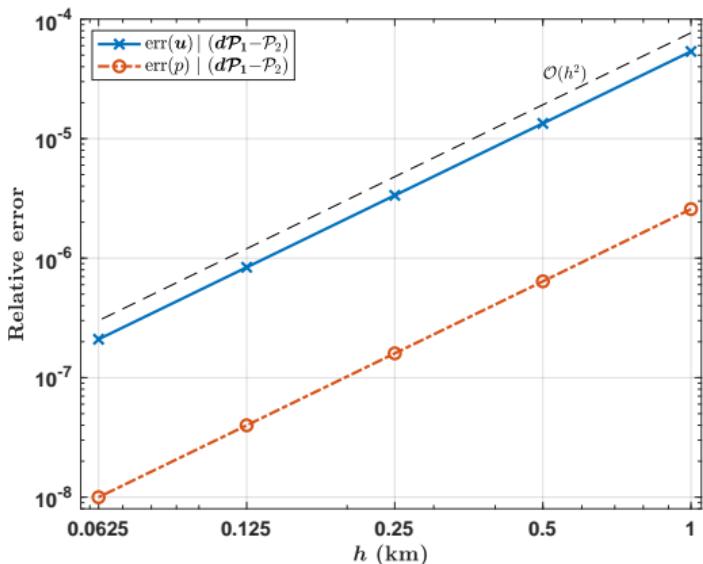
Given frequency ω , there are infinitely many $\kappa_n \in \mathbb{C}$, $n = 0, 1, 2, \dots$, satisfying the dispersion relation:

- Surface-gravity wave mode: $k_0 > 0$ ($\kappa_0 > 0$)
- Propagating acoustic-gravity modes: $k_n > 0$ ($|\kappa_n| < \omega/c$), $n = 1, \dots, K$, [$K = \omega H/(\pi c) + 1/2$]
- Decaying (evanescent) acoustic modes: $k_n^2 < 0$ ($|\kappa_n| > \omega/c$), $n = K + 1, \dots$

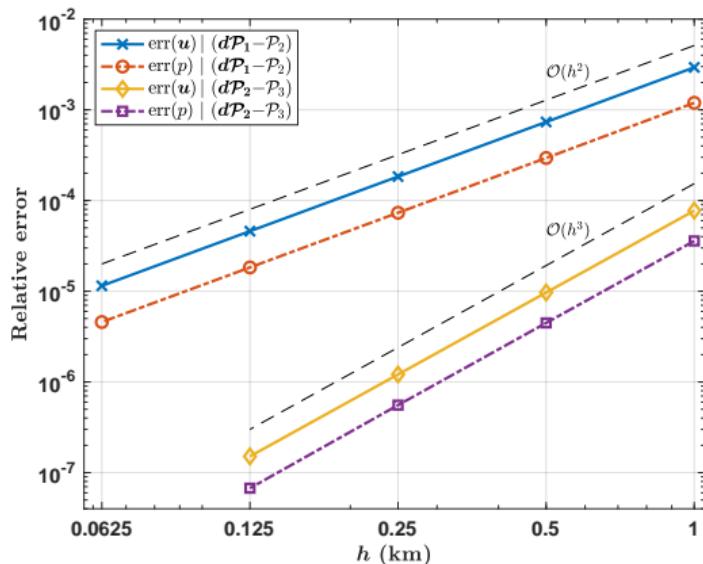
³Lotto and Dunham, 2015

Convergence of Modal Solutions (cont.)

Surface gravity wave mode



First acoustic-gravity wave mode



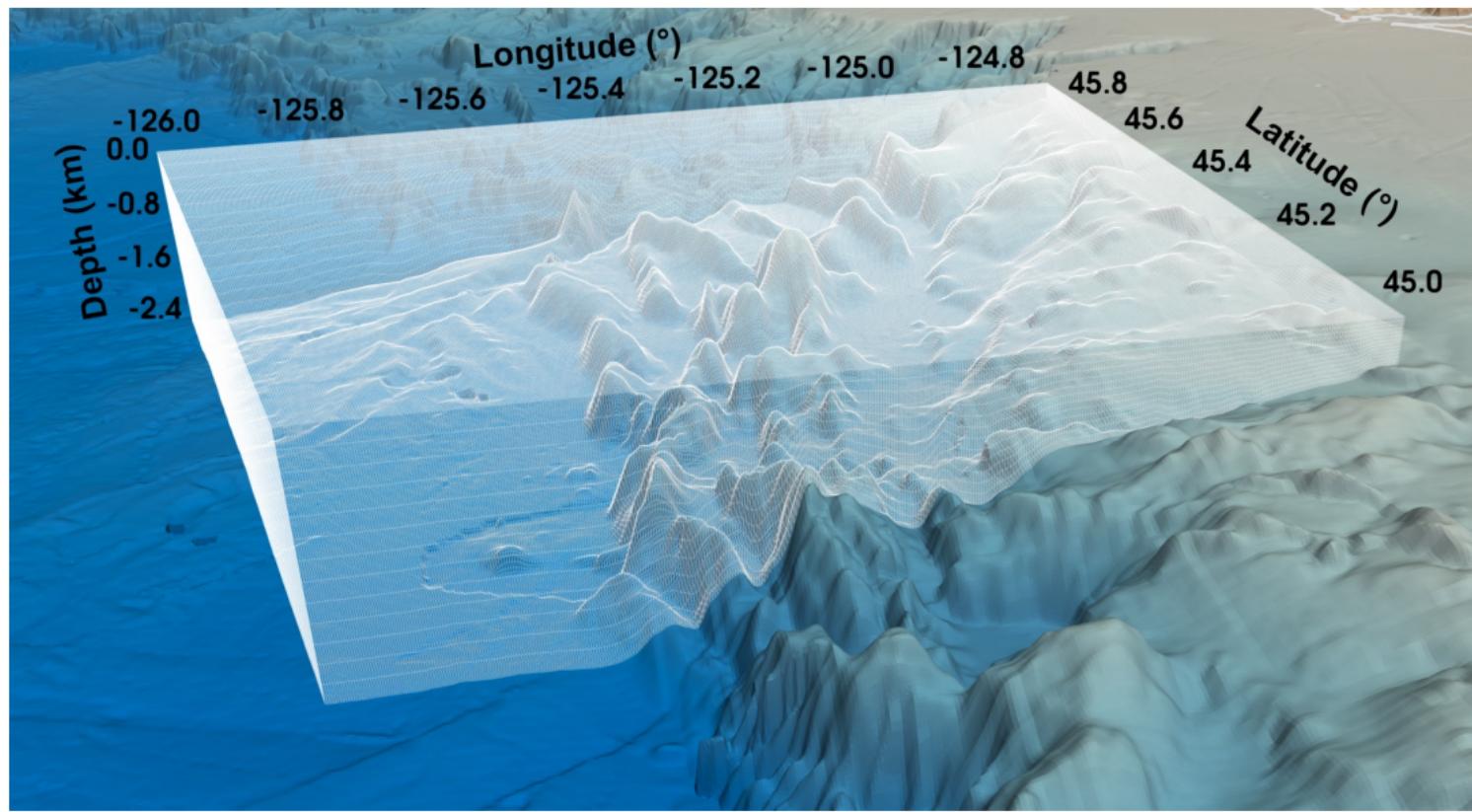
Error of modes in a bounded compressible ocean domain.^{4,5} Optimal convergence rates are attained for $(d\mathcal{P}_{r-1}, \mathcal{P}_r)$ elements for discontinuous velocity \mathbf{u} and continuous pressure p when $\Delta t \sim h^{r/2}$.

$$\text{err}(\mathbf{u}) := \|\mathbf{u}(T) - \mathbf{u}_{h,N}\|, \quad \text{err}(p) := (\|p(T) - p_{h,N}\|^2 + \|\nabla p(T) - \nabla p_{h,N}\|^2)^{1/2}$$

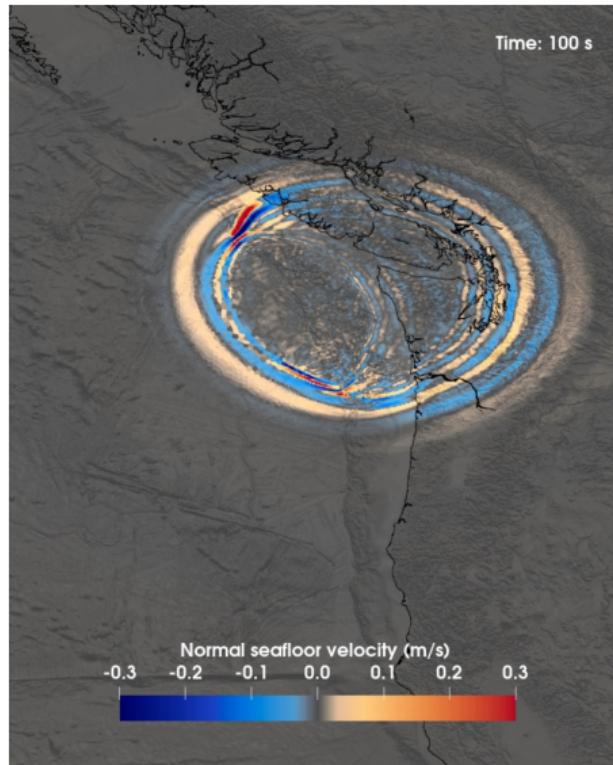
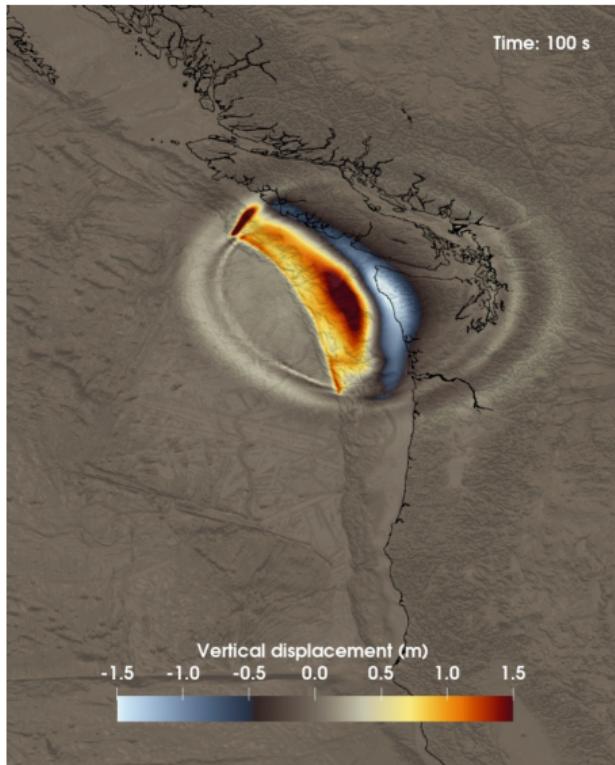
⁴ $\Omega = (0, L) \times (-H, 0)$, $L = 100$ km, $H = 5000$ m. Numerical solutions to the modes are computed by prescribing boundary data $p = \rho g \eta$ on Γ_s and sound-soft BC $p = 0$ on Γ_a for wavenumber $k_n = n\pi/L$, $n = 1$.

⁵using 2nd-order accurate Crank–Nicolson time-stepping.

3D Mesh Must Respect Bathymetry

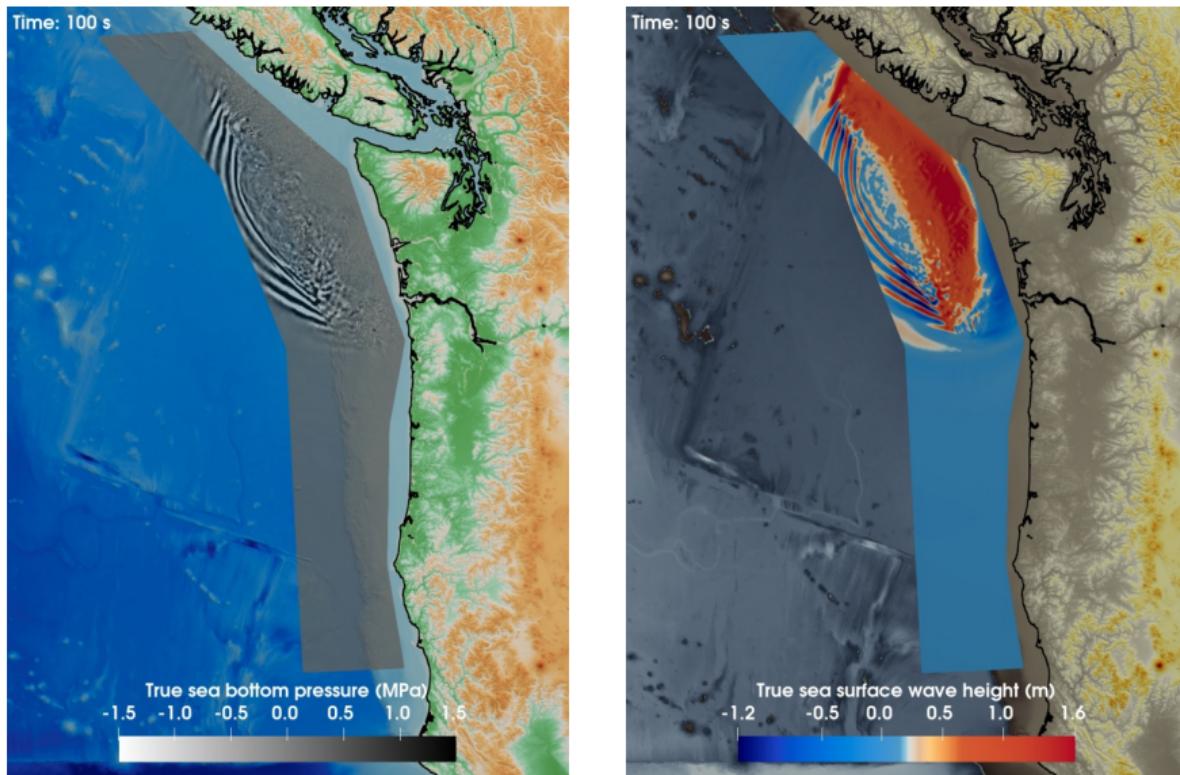


Cascadia Subduction Zone Margin-Wide Rupture Scenario



Rendering of data obtained from 3D Cascadia dynamic rupture simulation (Glehran et al., 2024).

Cascadia Subduction Zone Margin-Wide Rupture Scenario



Rendering of the **sea bottom pressure transients** and the **surface wave heights** obtained from the coupled acoustic-gravity PDE forward model.

Outline

- 1 Introduction
- 2 Forward Model
- 3 Inference and Prediction Model
- 4 Application to Cascadia Subduction Zone

Inverse problem:

Given pressure recordings from sensors on the seafloor,
infer the spatiotemporal seafloor motion in the subduction zone

Forecasting problem:

Given inferred spatiotemporal seafloor motion,
forward predict tsunami wave heights at specified coastal locations

- Parameters m : spatiotemporal seafloor motion
- Data d : spatiotemporal pointwise pressure observations
- Parameter-to-observable map $\mathcal{F} : m \mapsto d$
 - ▶ Solve forward PDE using given parameter field as boundary condition
 - ▶ Extract pressure values at sensor locations
- QoI q : spatiotemporal pointwise surface wave height predictions
- Parameter-to-QoI map $\mathcal{F}_q : m \mapsto q$
- Bayes' rule: $d\mu_{\text{post}} / d\mu_{\text{prior}} \propto \pi_{\text{like}}(d|m)$
- Gaussian prior: $m \sim \mathcal{N}(m_{\text{prior}}, \Gamma_{\text{prior}})$, $\Gamma_{\text{prior}} := (\alpha_1 I - \alpha_2 \Delta_{2D})^{-2}$
- Likelihood: $\pi_{\text{like}}(d|m) \propto \exp\left(-\frac{1}{2} \|\mathcal{F}m - d\|_{\Gamma_{\text{noise}}^{-1}}^2\right)$
- Observations: $d = \mathcal{F}m + \nu$, $\nu \sim \mathcal{N}(0, \Gamma_{\text{noise}})$
- Goal: Characterize the posteriors $\mu_{\text{post}} = \mathcal{N}(m_{\text{map}}, \Gamma_{\text{post}})$ and $\mu_{\text{post}(q)} = \mathcal{N}(q_{\text{map}}, \Gamma_{\text{post}(q)})$
 - ▶ $\Gamma_{\text{post}} := \left(\mathcal{F}^* \Gamma_{\text{noise}}^{-1} \mathcal{F} + \Gamma_{\text{prior}}^{-1}\right)^{-1}$ (*inverse of the Hessian of the negative log-posterior*)
 - ▶ $m_{\text{map}} = \Gamma_{\text{post}} \left(\mathcal{F}^* \Gamma_{\text{noise}}^{-1} d + \Gamma_{\text{prior}}^{-1} m_{\text{prior}}\right)$ ⇒ **Need a solver to rapidly compute Hessian action**

Autonomous dynamical system:

- Evolution does not depend explicitly on independent variable (e.g. time)
- Here: the mapping $m(\vec{x}, t + \tau) \mapsto d(\vec{x}, t + \tau)$ is the same as $m(\vec{x}, t) \mapsto d(\vec{x}, t)$

The discrete p2o and p2q maps \mathbf{F} and \mathbf{F}_q are **shift-invariant** with respect to time-stepping:

$$\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_3 \\ \vdots \\ \mathbf{d}_{N_t} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_{21} & \mathbf{F}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_{31} & \mathbf{F}_{21} & \mathbf{F}_{11} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{F}_{N_t,1} & \mathbf{F}_{N_t-1,1} & \cdots & \mathbf{F}_{21} & \mathbf{F}_{11} \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \\ \vdots \\ \mathbf{m}_{N_t} \end{bmatrix}, \quad \begin{aligned} \mathbf{d}_i &:= \mathbf{d}(\vec{x}, t_i) \in \mathbb{R}^{N_d} \\ \mathbf{m}_i &:= \mathbf{m}(\vec{x}, t_i) \in \mathbb{R}^{N_m} \end{aligned}$$

$\Rightarrow \mathbf{F}$ is **block Toeplitz** ($N_t \times N_t$ blocks, $\mathbf{F}_{ij} \in \mathbb{R}^{N_d \times N_m}$, $N_d \ll N_m$).

- Obtain \mathbf{F} by only N_d (number of sensors) adjoint solves (last block column of \mathbf{F}^*)
- Compactly store \mathbf{F} : $\mathcal{O}(N_m N_d N_t)$
- Efficiently apply \mathbf{F} and \mathbf{F}^* matvecs in Fourier space: $\mathcal{O}(N_m N_d N_t \log N_t)$

- Now we have a fast algorithm for applying \mathbf{F} , \mathbf{F}^* and \mathbf{F}_q , \mathbf{F}_q^* to vectors.
How can we use this to enable real-time inference and prediction?
- Split inversion into **Offline** and **Online** phases:⁶
 - Phase 1 (Offline):** Construct p2o and p2q maps from adjoint PDE solves
N_d + N_q adjoint PDE solves requires a scalable solver and supercomputing!
 - Phase 2 (Offline):** Compute compact representation of Γ_{post}
Apply Woodbury formula to reformulate inverse Hessian in the data space.
 - Phase 3 (Offline):** Compute QoI uncertainties and data-to-QoI map
Exploit autonomous structure of p2o map to enable FFT-based Hessian matvecs.⁷
 - Phase 4 (Online):** Compute parameter and QoI MAP points in real time
- Offline/Online decomposition allows us to **exactly** solve inference and prediction problems in $\mathcal{O}(\text{seconds})$ using the high-fidelity model
 - No surrogates or reduced-order models necessary**

⁶H., Venkat, Ghattas. “*Goal-oriented real-time Bayesian inference for linear autonomous dynamical systems with application to digital twins for tsunami early warning,*” arXiv:2501.14911 (2025)

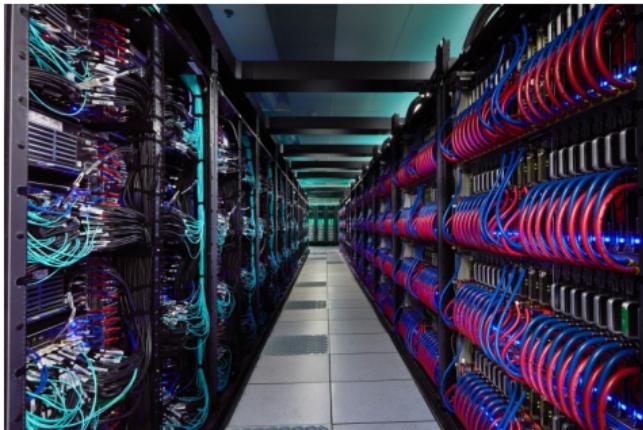
⁷Venkat, Fernando, H., Ghattas. “*Fast and scalable FFT-based GPU-accelerated algorithms for block-triangular Toeplitz matrices with application to linear inverse problems governed by autonomous dynamical systems,*” To Appear in SIAM Journal on Scientific Computing (2025)

El Capitan Supercomputer at Lawrence Livermore National Laboratory



El Capitan:

- **#1 Supercomputer** (June 2025 TOP500)
- 11,136 nodes with **44,544 AMD MI300A**
- System peak: **2.75 ExaFLOP/s (dp)**
- Sustained performance: **1.74 ExaFLOP/s (dp)**
- System memory: **5.70 PB**
- Vendor: **HPE**



Alps and Perlmutter Supercomputers

Alps:

- **#8 Supercomputer** (6/25 TOP500)
- 2,688 nodes with **10,752 NVIDIA GH200**
- System peak: **574.8 PetaFLOP/s** (dp)
- Sustained: **434.9 PetaFLOP/s** (dp)
- Vendor: **HPE**



Alps at the Swiss National Supercomputing Centre (CSCS)

Perlmutter:

- **#25 Supercomputer** (6/25 TOP500)
- 1,536 nodes with **6,144 NVIDIA A100**
- System peak: **113.0 PetaFLOP/s** (dp)
- Sustained: **79.2 PetaFLOP/s** (dp)
- Vendor: **HPE**



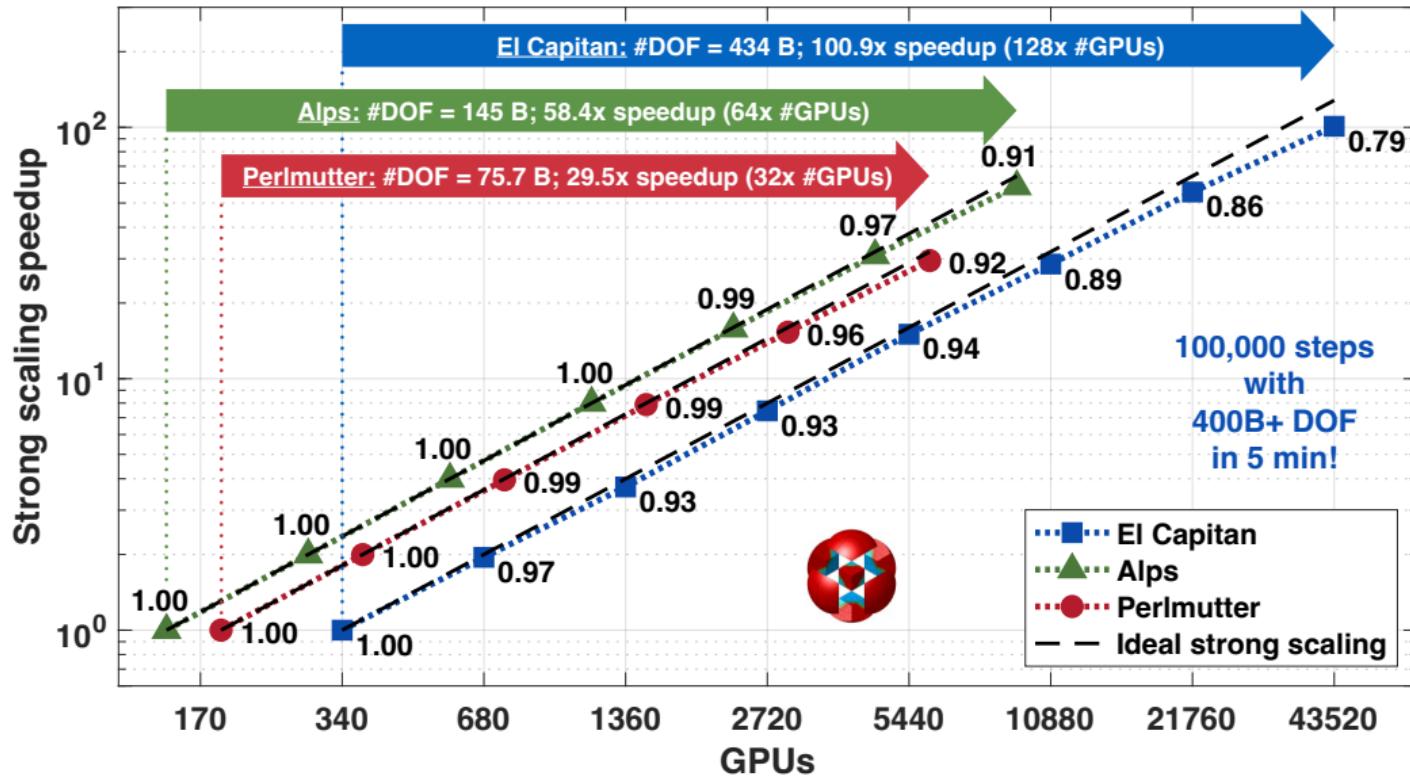
Perlmutter at the National Energy Research Scientific Computing Center (NERSC)

Scalability setup on the *El Capitan*, *Alps* and *Perlmutter* supercomputers

Nodes	GPUs	Processor grid	Weak scaling	Strong scaling
			Mesh elements	Elements/GPU
85	340	$5 \times 17 \times 4$	1,693,450,240	4,980,736
10,880	43,520	$80 \times 136 \times 4$	216,761,630,720	38,912
36	144	$2 \times 18 \times 4$	566,231,040	3,932,160
2,304	9,216	$16 \times 144 \times 4$	36,238,786,560	61,440
47	188	$1 \times 47 \times 4$	295,698,432	1,572,864
1,504	6,016	$8 \times 188 \times 4$	9,462,349,824	49,152

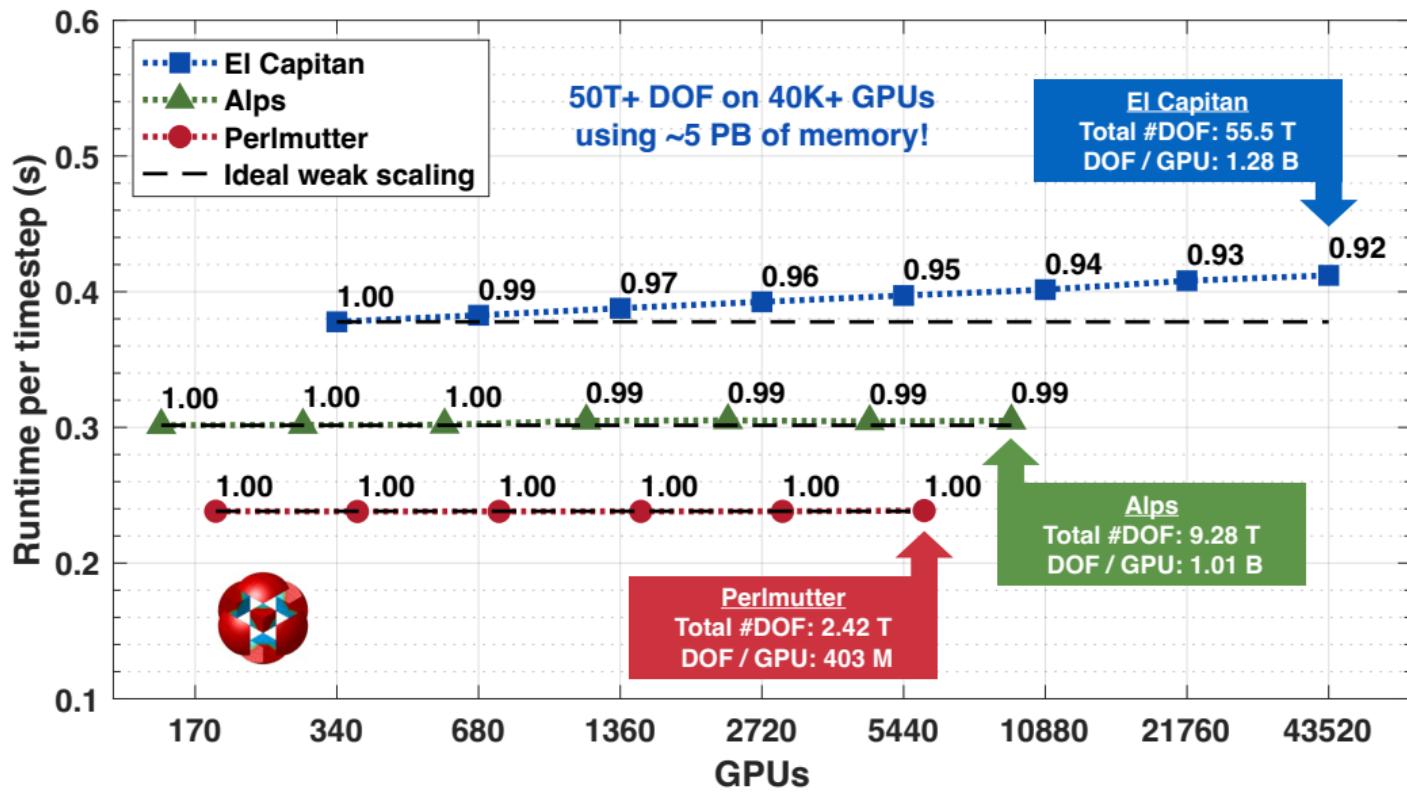
Adjoint PDE Scaling Results (MFEM)

Strong scaling of the acoustic–gravity model on the *El Capitan*, *Alps*, and *Perlmutter* supercomputers



Adjoint PDE Scaling Results (MFEM)

Weak scaling of the acoustic-gravity model on the *El Capitan*, *Alps*, and *Perlmutter* supercomputers



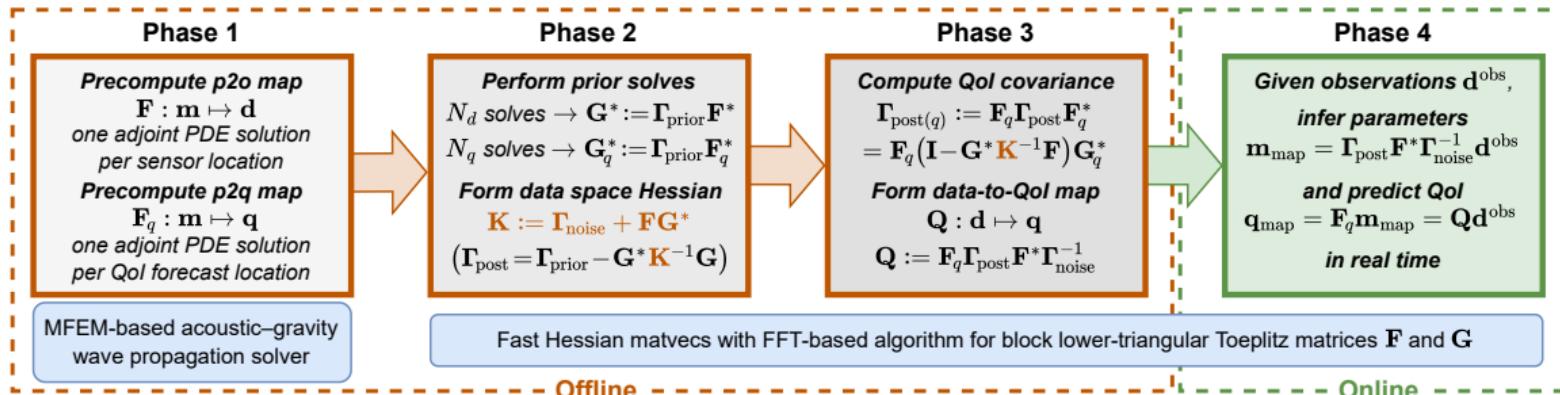
Outline

- 1 Introduction
- 2 Forward Model
- 3 Inference and Prediction Model
- 4 Application to Cascadia Subduction Zone

Inverse problem configuration:

- **Spatial resolution:** 300 m
- **Simulation time:** $T = 420$ s
- **Temporal parameter, data, and Qol dimension:** $N_t = 420$ (1 Hz frequency)
- **Spatial parameter dimension:** $N_m = 2\,416\,530$
- **Number of sensors:** $N_d = 600$
- **Number of Qol spatial locations:** $N_q = 21$
- **Total parameter dimension:** $N_m N_t = 1\,014\,942\,600$
- **Total data dimension:** $N_d N_t = 252\,000$
- **Total Qol dimension:** $N_q N_t = 8\,820$
- **Total state dimension:** 3 744 141 189 spatial DOFs, with 336 000 RK4 timesteps

Real-Time Bayesian Inference Framework: Workflow



The inverse solution is decomposed into **precomputation (offline) phases** that are executed just once, and a **real-time (online) phase** of parameter inference and QoI prediction that is executed when an earthquake occurs and data are acquired.

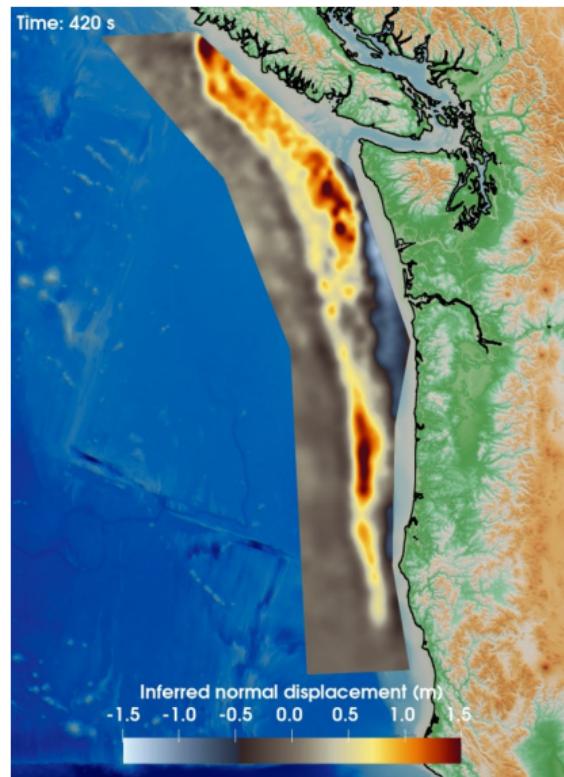
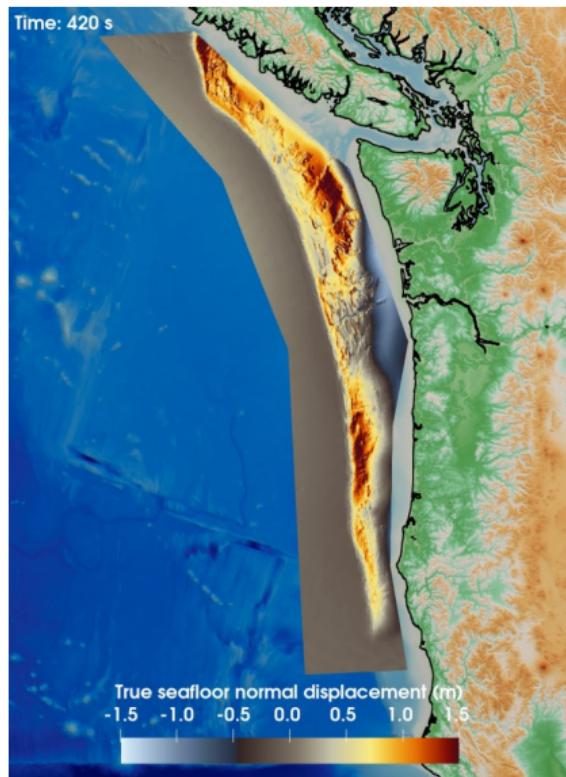
CSZ Margin-Wide Rupture Scenario: Compute Time

Compute time for each phase of the inference and prediction performed on *Perlmutter-GPU* nodes.

Time-to-solution for the **online computation** (Phase 4) is less than 0.2 seconds.

Phase	Task	#GPUs	Compute time
1	form $\mathbf{F} : \mathbf{m} \mapsto \mathbf{d}$	512	$600 \times 52 \text{ m} \sim 520 \text{ h}$
	form $\mathbf{F}_q : \mathbf{m} \mapsto \mathbf{q}$	512	$21 \times 52 \text{ m} \sim 18 \text{ h}$
2	form $\mathbf{G}^* := \boldsymbol{\Gamma}_{\text{prior}} \mathbf{F}^*$	16	$600 \times 4.5 \text{ s} \sim 45 \text{ m}$
	form $\mathbf{G}_q^* := \boldsymbol{\Gamma}_{\text{prior}} \mathbf{F}_q^*$	16	$21 \times 4.5 \text{ s} \sim 1.5 \text{ m}$
	form $\mathbf{K} := \boldsymbol{\Gamma}_{\text{noise}} + \mathbf{F}\mathbf{G}^*$	512	$252k \times 24 \text{ ms} \sim 100 \text{ m}$
	factorize \mathbf{K}	25	22 s
3	compute $\boldsymbol{\Gamma}_{\text{post}(q)}$	512	$8,820 \times 150 \text{ ms} \sim 25 \text{ m}$
	compute $\mathbf{Q} : \mathbf{d} \mapsto \mathbf{q}$	512	$8,820 \times 150 \text{ ms} \sim 25 \text{ m}$
4	infer parameters \mathbf{m}_{map}	512	< 0.2 s
	predict QoI \mathbf{q}_{map}	1	< 1 ms

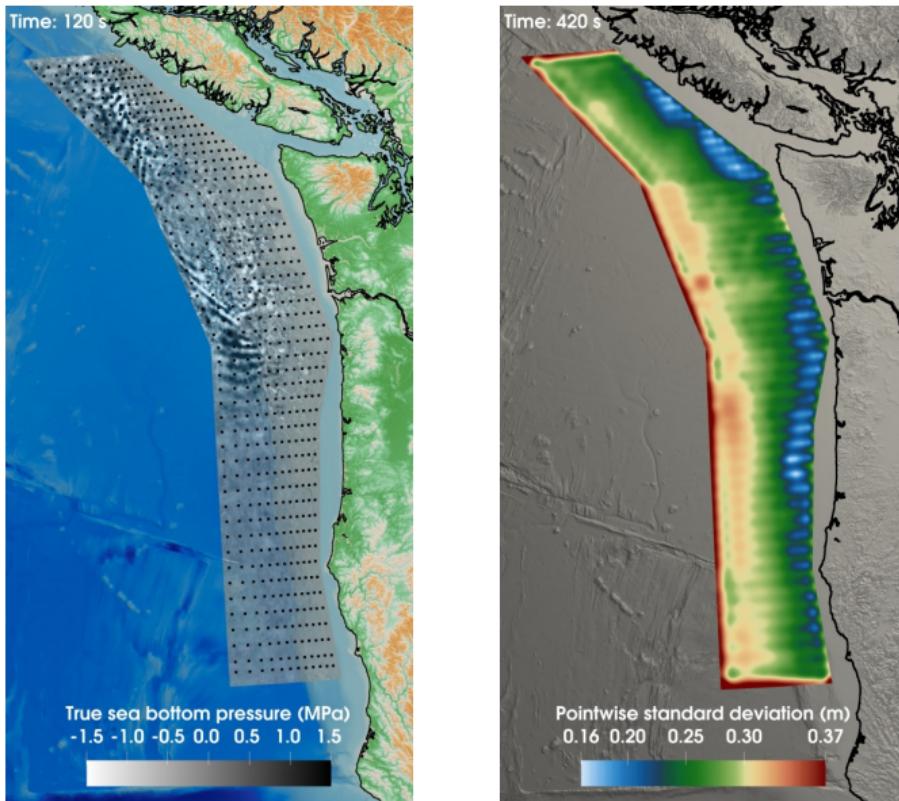
CSZ Margin-Wide Rupture Scenario: Inferred Parameter Field (Real-Time)



Normal seafloor displacement (time-integrated parameter field).

Left: True field; Right: Inferred mean from synthetic observations.

CSZ Margin-Wide Rupture Scenario: Parameter Uncertainty

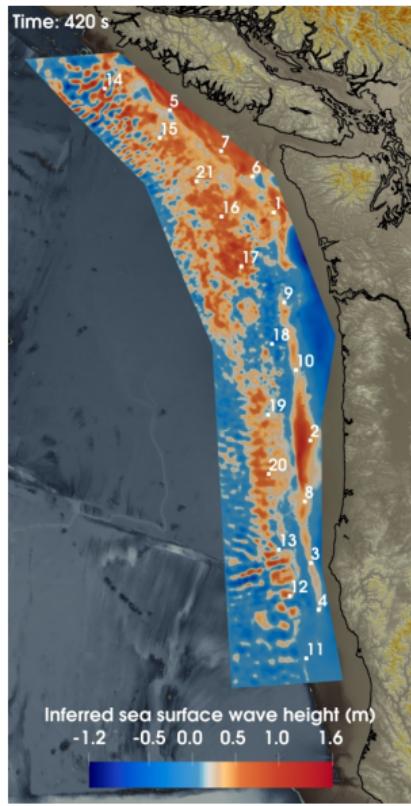


Left: Snapshot of true seafloor acoustic pressure field with 600 hypothesized sensor locations.

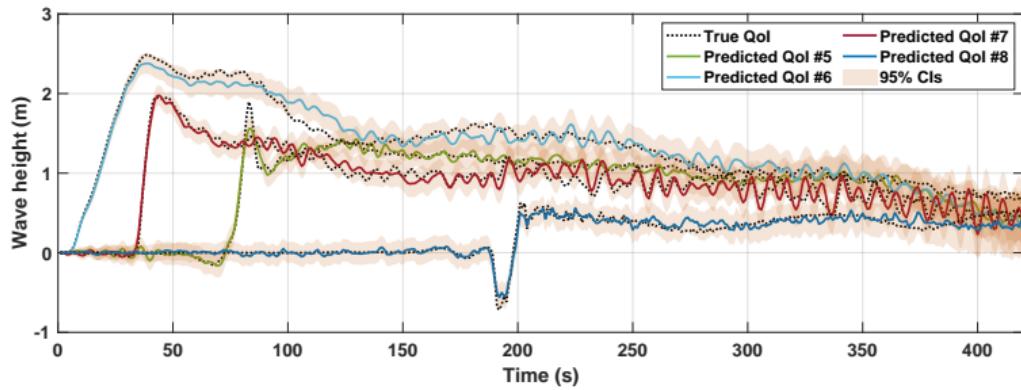
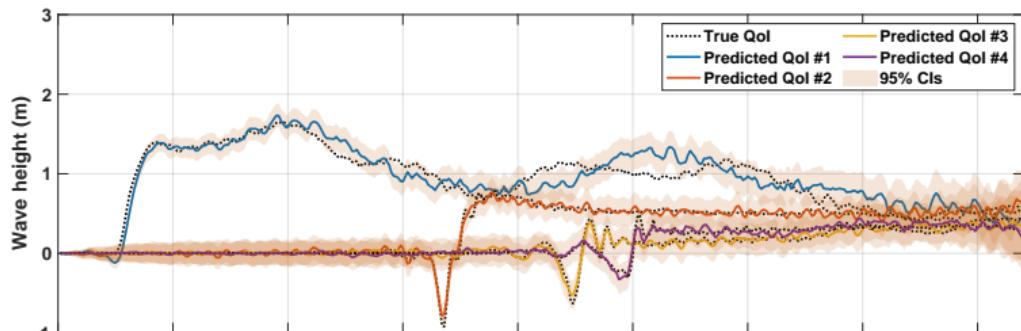
Right: Uncertainties plotted as pointwise standard deviations in meters of seafloor normal displacement.

CSZ Margin-Wide Rupture Scenario: QoI Prediction (Real-Time)

QoI forecasting locations



Real-time QoI predictions with uncertainties illustrated as 95% credible intervals (CIs) inferred from noisy, synthetic data of 600 hypothesized seafloor acoustic pressure sensors for a margin-wide rupture in the CSZ



In this talk:

- Bayesian inverse model to infer tsunamigenic seafloor motion directly from near-field pressure data using an acoustic–gravity model in the deep ocean
- Large-scale real-time inversion algorithms for linear autonomous dynamical systems, exploiting block Toeplitz structure of p2o map \mathbf{F} , enabling fast construction and compact storage of \mathbf{F} and efficient FFT-based \mathbf{F} and \mathbf{F}^* matvecs
- MAP point computed exactly (up to discretization error) in < 0.2 seconds, for inverse problem with $\mathcal{O}(10^9)$ parameters and $\mathcal{O}(10^9)$ states (with $\mathcal{O}(10^5)$ timesteps)
- Wave heights at critical locations and their uncertainties computed in a fraction of a second by exploiting structure of push forward of posterior to QoI

Ongoing & future work:

- Construct data-driven prior for Cascadia subduction zone from numerous rupture scenarios on the fault system
- Solve optimal experimental design problem for placement of seafloor pressure sensors
- Exploit autonomous dynamical system structure in other inverse problems of interest

Literature reference



Thank you for your attention

This research was supported in part by DOE ASCR grants DE-SC0023171 and DE-SC0021239. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 2137420. Supercomputing resources were provided by the Texas Advanced Computing Center (TACC) at The University of Texas at Austin on its *Frontera*, *Stampede3*, *Lonestar6*, and *Vista* systems, by CSCS on its *Alps* system, by NERSC on its *Perlmutter* system, and by Lawrence Livermore National Laboratory on its *Tuolumne* and *El Capitan* systems.