The University of Texas at Austin
Oden Institute for Computational Engineering and Sciences

Real-Time High-Fidelity Bayesian Inference and Forecasting of Tsunamis
A Digital Twin for Tsunami Early Warning Applied to the Cascadia Subduction Zone

Sreeram Venkat, Stefan Henneking, and Omar Ghattas · *Oden Institute for Computational Engineering and Sciences – The University of Texas at Austin*

OPTIMUS
BEYOND FORWARD SIMULATION

Lawrence Livermore National Laboratory
UC San Diego

Overview

Approach
Infer earthquake-induced seafloor velocity from pressure observations and forecast tsunami wave heights with quantified uncertainty in real time.

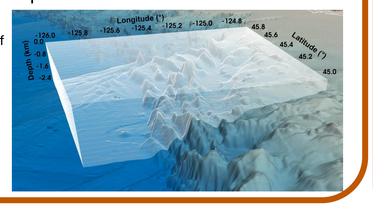
Contributions
- **Real-time Bayesian inversion and prediction** using a high-fidelity acoustic–gravity model without surrogates
- **FFT-based GPU algorithms** enabling fast Hessian matvecs and offline–online decomposition
- **Over 1 billion parameters inferred in 0.2 seconds**[1] (**2025 ACM Gordon Bell Prize**)
- **Scalable Bayesian optimal sensor placement**[7] (**SC26 Best Paper & Best Student Paper Finalist**)

Interpretation & Outlook
- Offline precomputation enables real-time inference; new regions require adapted bathymetry, sensors, priors, and regional precomputation.
- Next: validate with observational data and assess robustness to model and sensor errors.

[1]Henneking et al. Real-time Bayesian inference at extreme scale. *SC25*. **Gordon Bell Prize.**

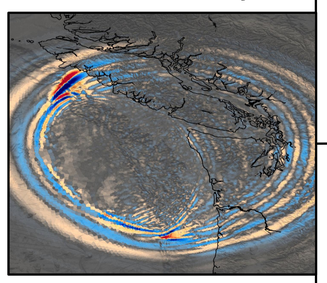
Challenges

- Massive problem scale
  - Forward problem: wave propagation (3D space + time)
  - Parameter: infinite-dimensional field (2D space + time)
- Sparse & noisy data: ill-posed inverse problem
- Real-time: inference and prediction needed in seconds
- Hyperbolic PDE: surrogate modeling is difficult
- Inference and prediction: involves solving a linear system with the **Hessian** (of the negative log posterior)
  - Cannot form dense Hessian matrix ($10^9 \times 10^9$)
  - Traditional matrix-free methods use a pair of (expensive) PDE solves per Hessian action
  - Parameter-to-observable (p2o) map $\mathcal{F}$ and its adjoint $\mathcal{F}^*$ are the main components of the Hessian



Representative block of a 3D multi-block hexahedral mesh of the Cascadia subduction zone, depicting bathymetry-adapted meshing. GEBCO's 15-arc-second gridded bathymetry data is used to generate the mesh. Bathymetry scale is exaggerated for ease of visualization.

Real-Time Inversion and Prediction Framework



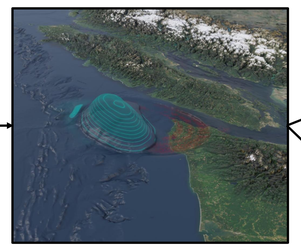
Tsunami Simulation

Parameter-to-observable map $\mathcal{F}$

Forward PDE

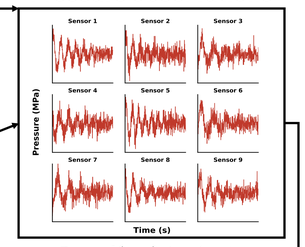
Problem size

N_m	~	10^6
N_d	~	10^2
N_q	~	10^1
N_w	~	10^9
N_t	~	10^3



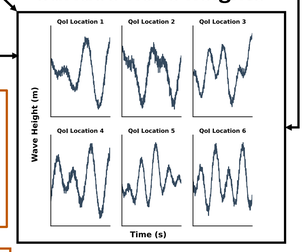
$p(x_i,t)$

$\eta(x_j,t)$



Data: $d(x_i,t), i = 1 \dots N_d$
Discrete Dimension: $N_d \times N_t$

$\mathcal{Q}$



Qoi: $q(x_j,t), j = 1 \dots N_q$
Discrete Dimension: $N_q \times N_t$

Parameter-to-QoI map $\mathcal{F}_q$

Acoustic–gravity PDE model[4]

$$\begin{cases} \rho\partial_t \vec{u} + \nabla p = 0, & \text{Ocean Domain} \\ K^{-1}\partial_t p + \nabla\cdot\vec{u} = 0, & \text{Ocean Domain} \\ p = \rho g\eta,\ \partial_t\eta = \vec{u}\cdot\vec{n}, & \text{Ocean Surface} \\ \vec{u}\cdot\vec{n} = m, & \text{Ocean Floor} \\ \vec{u}\cdot\vec{n} = Z^{-1}p, & \text{Lateral Boundaries} \end{cases}$$

Bayesian inference
- Approach: Treat parameter as a random field
- Goal: Characterize the posterior distribution

Bayes' theorem

$$\pi_{\text{post}}(\mathbf{m}|\mathbf{d}) \propto \pi_{\text{like}}(\mathbf{d}|\mathbf{m})\,\pi_{\text{prior}}(\mathbf{m})$$

Bayesian linear inverse problem

Gaussian prior

$$\pi_{\text{prior}}(\mathbf{m}) \propto \mathcal{N}(\mathbf{m}_{\text{prior}}, \boldsymbol{\Gamma}_{\text{prior}})$$

Additive Gaussian noise model

$$\mathbf{d} = \mathbf{F}\mathbf{m} + \boldsymbol{\nu}, \qquad \boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}})$$

Likelihood

$$\pi_{\text{like}}(\mathbf{d}|\mathbf{m}) \propto \exp\left(-\frac{1}{2}\|\mathbf{F}\mathbf{m} - \mathbf{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}}\right)$$

Gaussian posterior

$$\pi_{\text{post}}(\mathbf{m}|\mathbf{d}) \propto \mathcal{N}(\mathbf{m}_{\text{map}}, \boldsymbol{\Gamma}_{\text{post}})$$

$$\boldsymbol{\Gamma}_{\text{post}} = \left(\mathbf{F}^*\boldsymbol{\Gamma}^{-1}_{\text{noise}}\mathbf{F} + \boldsymbol{\Gamma}^{-1}_{\text{prior}}\right)^{-1}$$

$$\mathbf{m}_{\text{map}} = \boldsymbol{\Gamma}_{\text{post}}\left(\mathbf{F}^*\boldsymbol{\Gamma}^{-1}_{\text{noise}}\mathbf{d} + \boldsymbol{\Gamma}^{-1}_{\text{prior}}\mathbf{m}_{\text{prior}}\right)$$

Goal-oriented inference[6]

Predict quantities of interest (QoI):

$$\pi_{\text{post}(q)} \propto \mathcal{N}\left(\mathbf{q}_{\text{map}}, \boldsymbol{\Gamma}_{\text{post}(q)}\right)$$

$$\boldsymbol{\Gamma}_{\text{post}(q)} = \mathbf{F}_q\boldsymbol{\Gamma}_{\text{post}}\mathbf{F}_q^*$$

$$\mathbf{q}_{\text{map}} = \mathbf{F}_q\mathbf{m}_{\text{map}} = \mathbf{Q}\mathbf{d}$$

(assume $\mathbf{m}_{\text{prior}} = 0$)

[4]Lotto & Dunham. Tsunami generation in a compressible ocean. *Comput. Geosci.*, 2015.

Phase 1 — *Precompute p2o map* $\mathbf{F} : \mathbf{m} \mapsto \mathbf{d}$ *one adjoint PDE solution per sensor location*; *Precompute p2q map* $\mathbf{F}_q : \mathbf{m} \mapsto \mathbf{c}$ *one adjoint PDE solution per QoI forecast location*

MFEM-based acoustic–gravity wave propagation solver

~540 hours on 512 A100 GPUs

Phase 2 — *Perform prior solves* N_d *solves* $\rightarrow \mathbf{G}^* := \boldsymbol{\Gamma}_{\text{prior}}\mathbf{F}^*$; N_q *solves* $\rightarrow \mathbf{G}_q^* := \boldsymbol{\Gamma}_{\text{prior}}\mathbf{F}_q^*$; *Form data space Hessian* $\mathbf{K} := \boldsymbol{\Gamma}_{\text{noise}} + \mathbf{F}\mathbf{G}^*$

FFTMatvec for fast parameter-to-observable and parameter-to-QoI map actions

~2.5 hours on 512 A100 GPUs

Phase 3 — *Compute QoI covariance* $\boldsymbol{\Gamma}_{\text{post}(q)} := \mathbf{F}_q\boldsymbol{\Gamma}_{\text{post}}\mathbf{F}_q^* = \mathbf{F}_q\mathbf{G}_q^* - \mathbf{F}_q\mathbf{G}^*\mathbf{K}^{-1}\mathbf{G}\mathbf{F}_q^*$; *Form data-to-QoI map* $\mathbf{Q} : \mathbf{d} \mapsto \mathbf{q}$, $\mathbf{Q} := \mathbf{F}_q\mathbf{G}^*\mathbf{K}^{-1}$

~1 hour on 512 A100 GPUs

Phase 4 — *Given observations* $\mathbf{d}^{\text{obs}}$, *infer parameters* $\mathbf{m}_{\text{map}} = \mathbf{G}^*\mathbf{K}^{-1}\mathbf{d}^{\text{obs}}$ *and predict QoI* $\mathbf{q}_{\text{map}} = \mathbf{F}_q\mathbf{m}_{\text{map}} = \mathbf{Q}\mathbf{d}^{\text{obs}}$ *in real time*

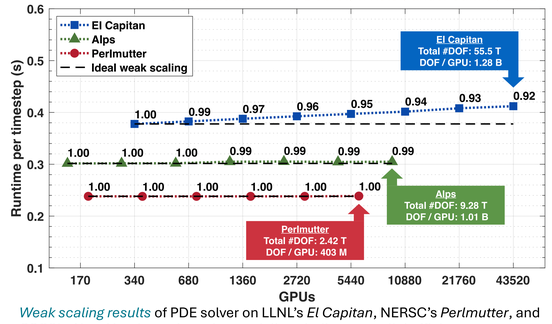
$\mathbf{m}_{\text{map}}$: 0.2s on 512 A100 GPUs
$\mathbf{q}_{\text{map}}$: < 1ms on 1 A100 GPU

Offline — Online

Scalable PDE Solver

- Adjoint PDE solver implemented with high-order finite elements and RK4 time-stepping using MFEM
- Scales to many thousands of processors (CPU and GPU)
- The entire LLNL *El Capitan* system was used to solve a PDE with over **55.5 trillion** unknowns — **largest-to-date unstructured mesh finite element simulation**

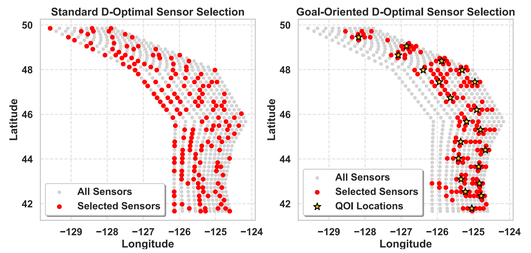
Collaboration with V. Dobrev, T. Kolev, and J. Camier, LLNL



Weak scaling results of PDE solver on LLNL's *El Capitan*, NERSC's *Perlmutter*, and CSCS's *Alps*. Numbers along the graph lines indicate parallel efficiency.

Optimal Sensor Placement

- Standard method: maximize the expected information gain (EIG) in the inferred parameters with respect to sensor locations
- Goal-oriented method: maximize the EIG in the forecasted wave heights
- Use greedy algorithm to subsample initial configuration
  - Objective function computation is parallel and uses block Schur complement updates[7]



Comparison of sensor selection strategies for selecting 175 out of 600 possible sensors. Left: standard algorithm maximizing EIG in inferred parameters. Right: goal-oriented algorithm maximizing EIG in the QoI.

[7]Venkat et al. Sensor placement for tsunami early warning. *SC26*. **Best Paper & Best Student Paper Finalist.**

Fast Hessian Matvecs

- The mapping from parameters to data is **time-invariant**
- The discrete $\mathbf{F}$ matrix is **block lower-triangular Toeplitz:**

$$\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_3 \\ \vdots \\ \mathbf{d}_{N_t} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_{21} & \mathbf{F}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_{31} & \mathbf{F}_{21} & \mathbf{F}_{11} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{F}_{N_t,1} & \mathbf{F}_{N_t-1,1} & \cdots & \mathbf{F}_{21} & \mathbf{F}_{11} \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \\ \vdots \\ \mathbf{m}_{N_t} \end{bmatrix},$$

$$\mathbf{d}_i = \mathbf{d}(\vec{x}, t_i) \in \mathbb{R}^{N_d},\ \mathbf{m}_i = \mathbf{m}(\vec{x}, t_i) \in \mathbb{R}^{N_m}$$

- $\mathbf{F}$ has blocks of size $N_d \times N_m$ and requires **only N_d adjoint PDE solves (one per sensor)** to precompute
- Matrices are **block-diagonal in Fourier space**
- Parameter-to-QoI map $\mathbf{F}_q$ has a similar structure

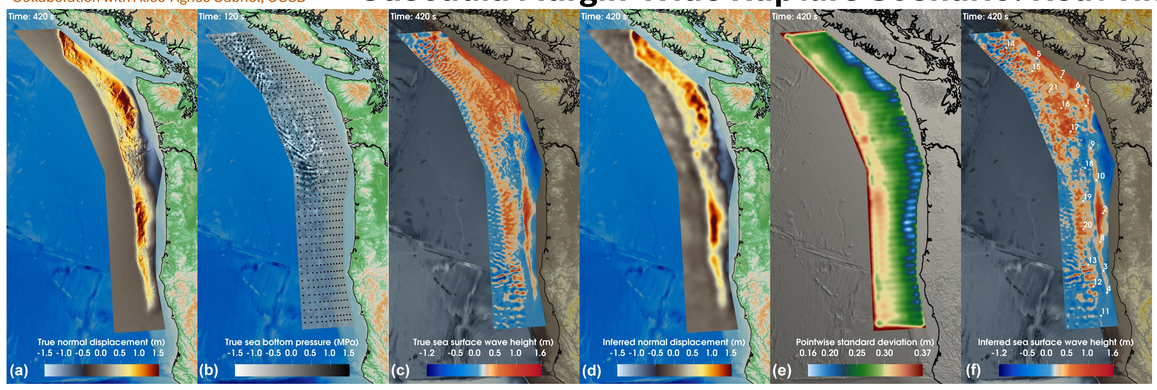
→ Fast matvecs and offline–online decomposition

FFTMatvec[2,3] is a multi-GPU-accelerated algorithm for matrix–vector products with block Toeplitz matrices

[2]Venkat et al. FFT-based GPU algorithms for block-triangular Toeplitz matrices. *SIAM J. Sci. Comput.*, 2025.
[3]Venkat et al. Mixed-precision performance portability of FFT-based GPU algorithms for block-triangular Toeplitz matrices. *SC25 Workshops (P3HPC)*.

Cascadia Margin-Wide Rupture Scenario: Real-Time Inference and Prediction

Collaboration with Alice-Agnes Gabriel, UCSD



Inversion/prediction results

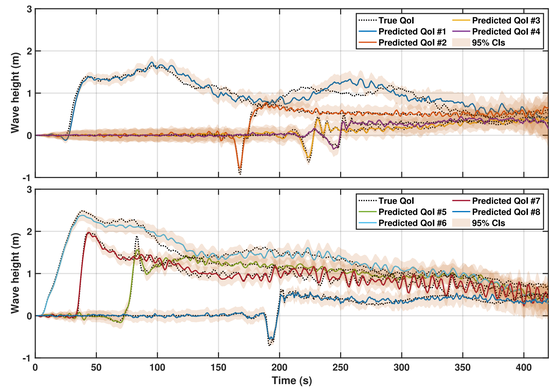
a. True total (normal) seafloor displacement
b. True seafloor pressure field and sensor locations
c. True surface wave height
d. Inferred total (normal) seafloor displacement
e. Pointwise standard deviation of seafloor displacement
f. Predicted surface wave heights and QoI target locations

Animations

[5]Glehman et al. Partial ruptures in 3D Cascadia simulations. *Seismica*, 2025.

[6]Henneking et al. Goal-oriented real-time Bayesian inference. *J. Comput. Phys.*, 2026.

Synthetic, noisy data generated from acoustic–gravity PDE model using source (parameters) from a physics-based magnitude 8.7 dynamic rupture earthquake scenario for a margin-wide rupture in the Cascadia subduction zone.[5]



Real-time prediction of surface wave heights (QoI) with 95% credible intervals (CIs) at selected locations (see plot f). Inferred from noisy, synthetic data of 600 hypothesized seafloor acoustic pressure sensors.