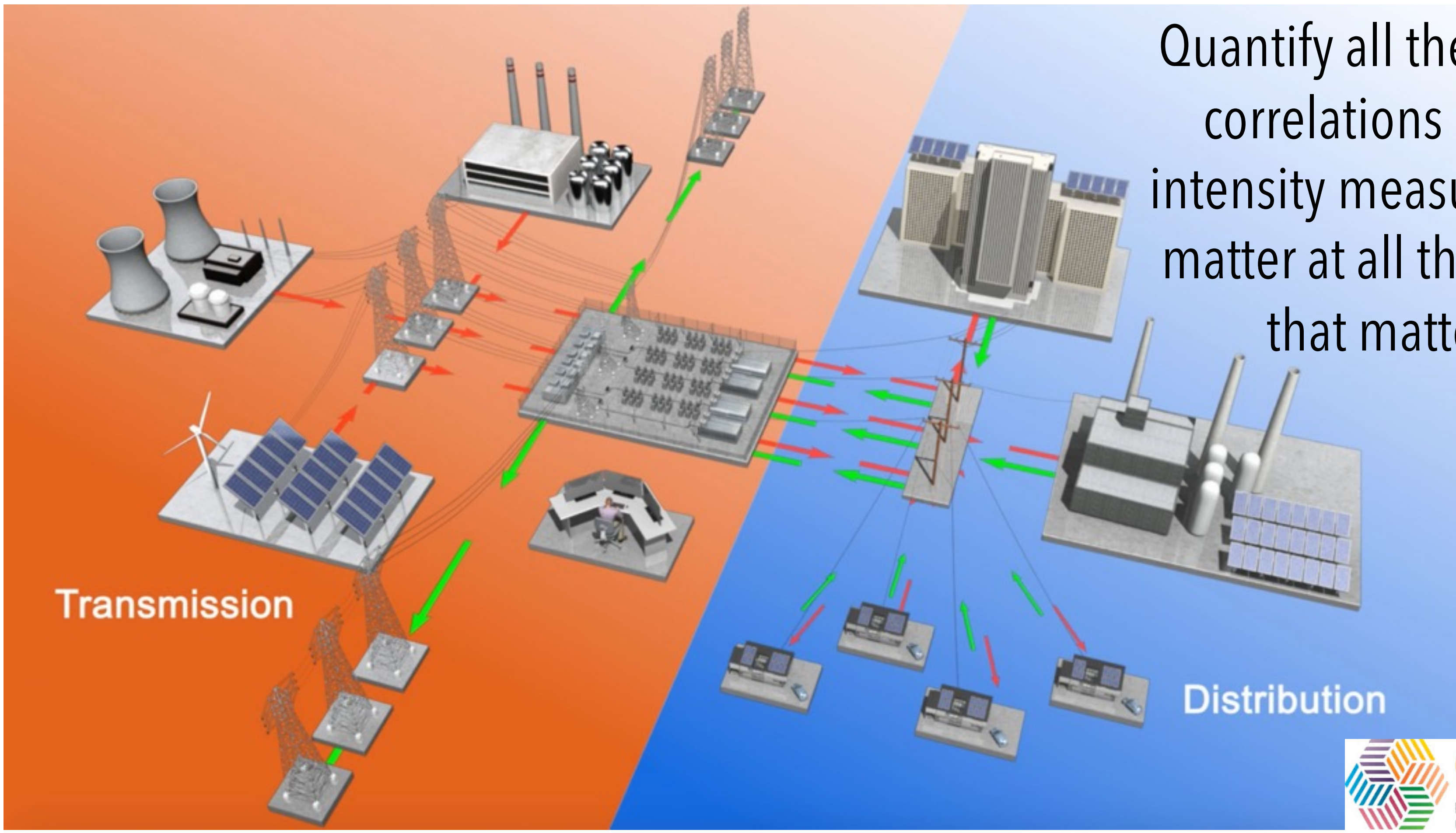
2025 PEER-LBNL Workshop on the Regional Scale Simulated Ground Motion Database (SGMD)
for the San Francisco Bay Area

Simulated Ground Motions in the Development of Machine Learning-Based Risk Estimations

Domniki Asimaki, Zachary E. Ross – California Institute of Technology
Yaozhong Shi, Grigorios (Greg) Lavrentiadis
David McCallen, K. Tsalouchidis -- Lawrence Berkeley National Laboratory
Kamyar Azizzadenesheli - nVIDIA

Imagine a distributed infrastructure system



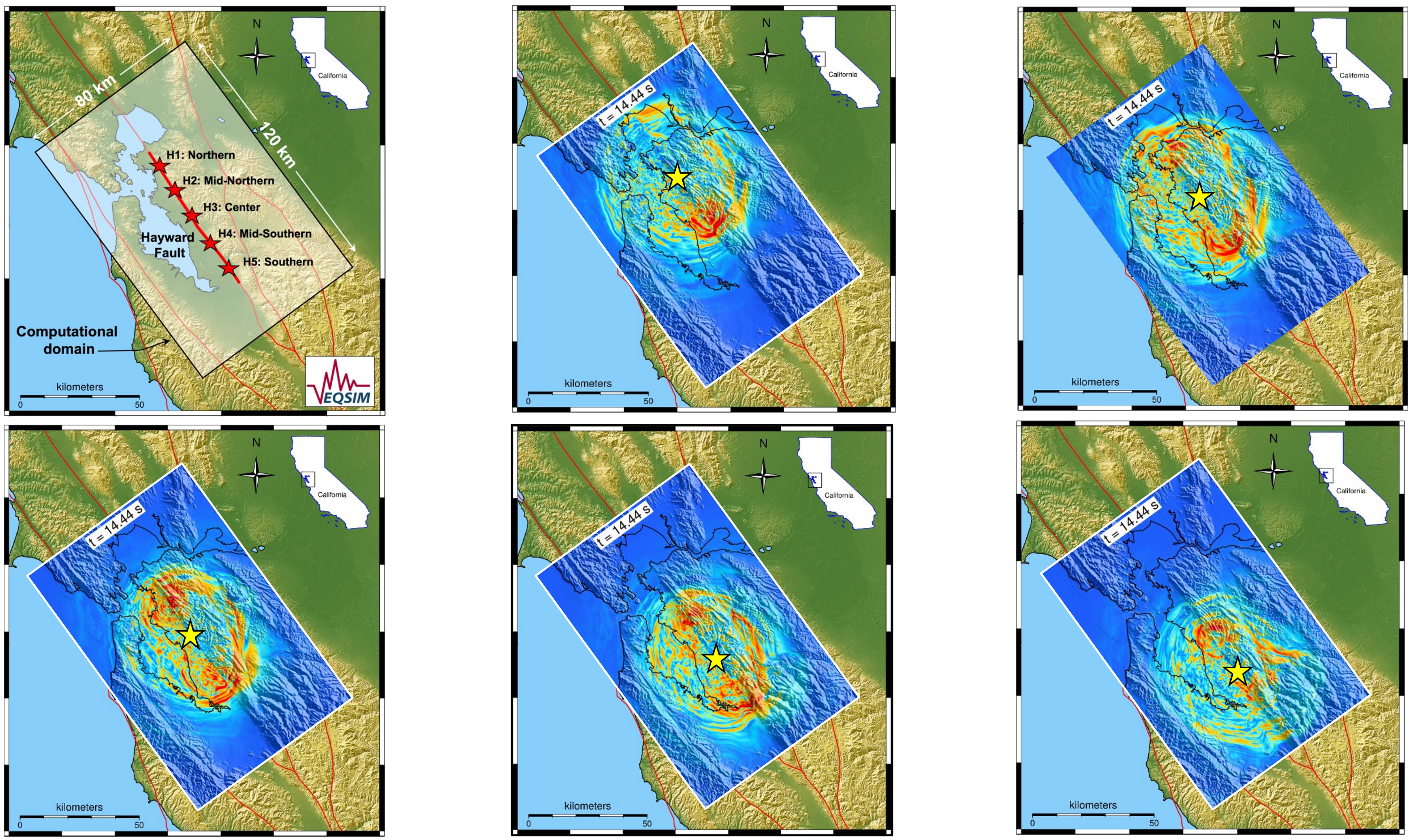
Quantify all the spatial correlations of the intensity measures that matter at all the scales that matter!

GRID MODERNIZATION INITIATIVE
U.S. Department of Energy

The ability to efficiently execute high fidelity simulations enables full exploration of the geophysical parameter space

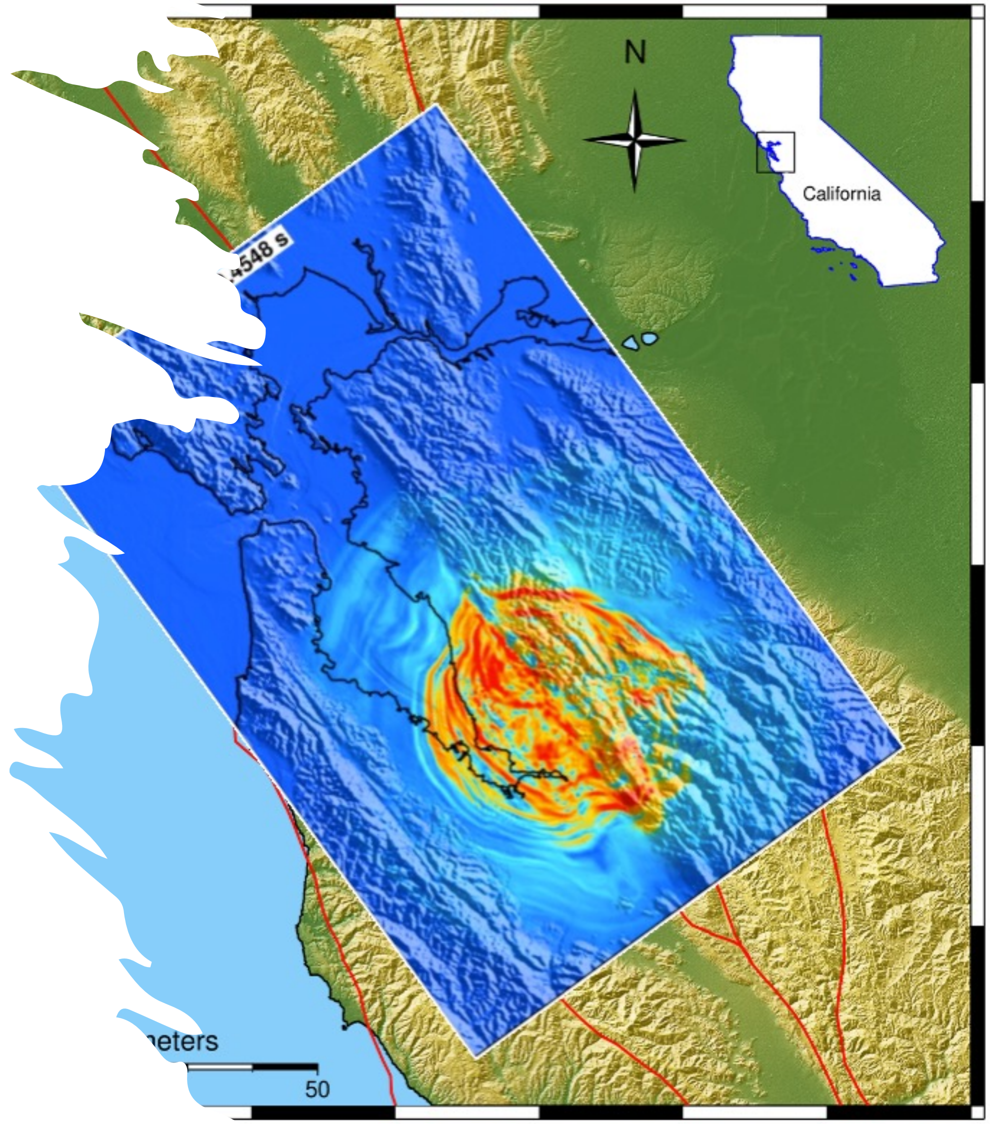
EXASCALE COMPUTING PROJECT

Five Hayward fault M7 rupture realizations



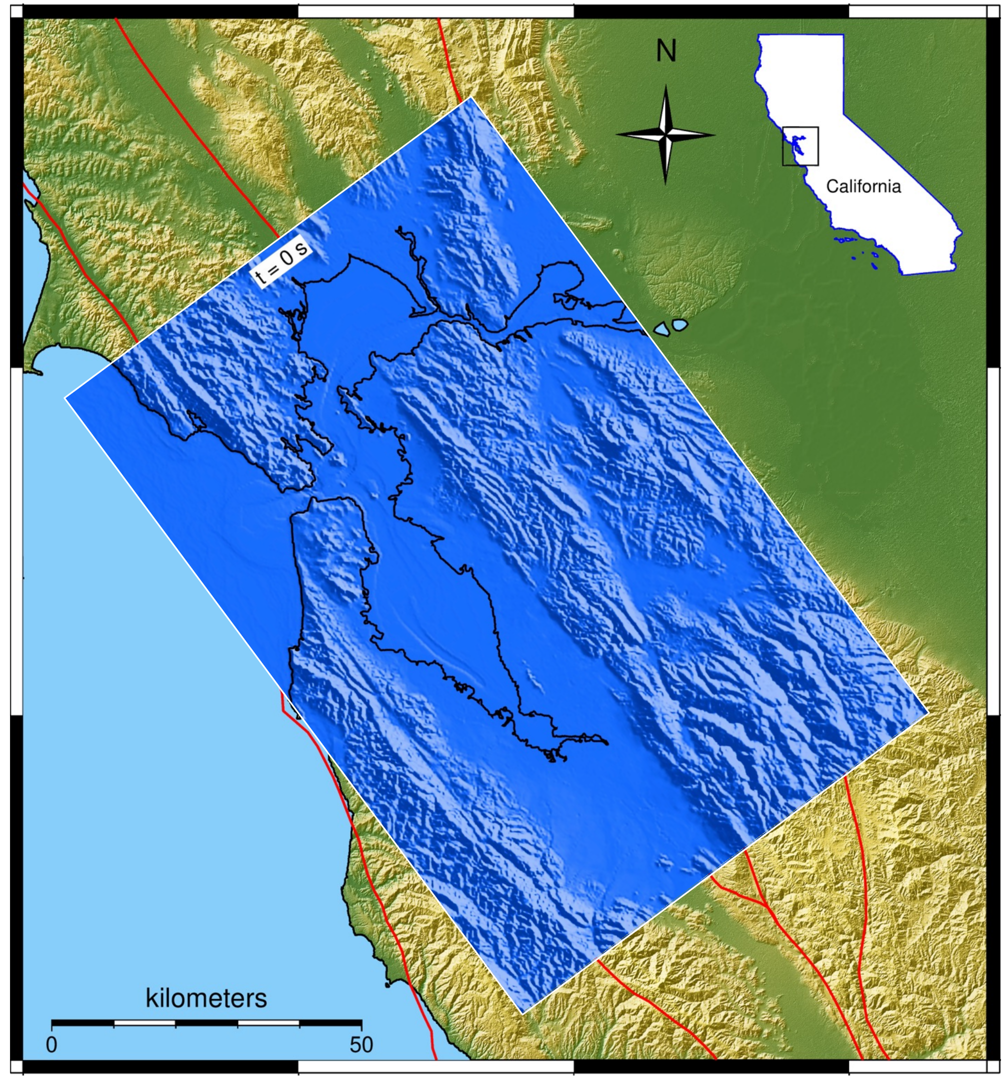
EQSIM

but...



1. The output is as accurate as the input (geology, rheology, hydrology etc, etc…)

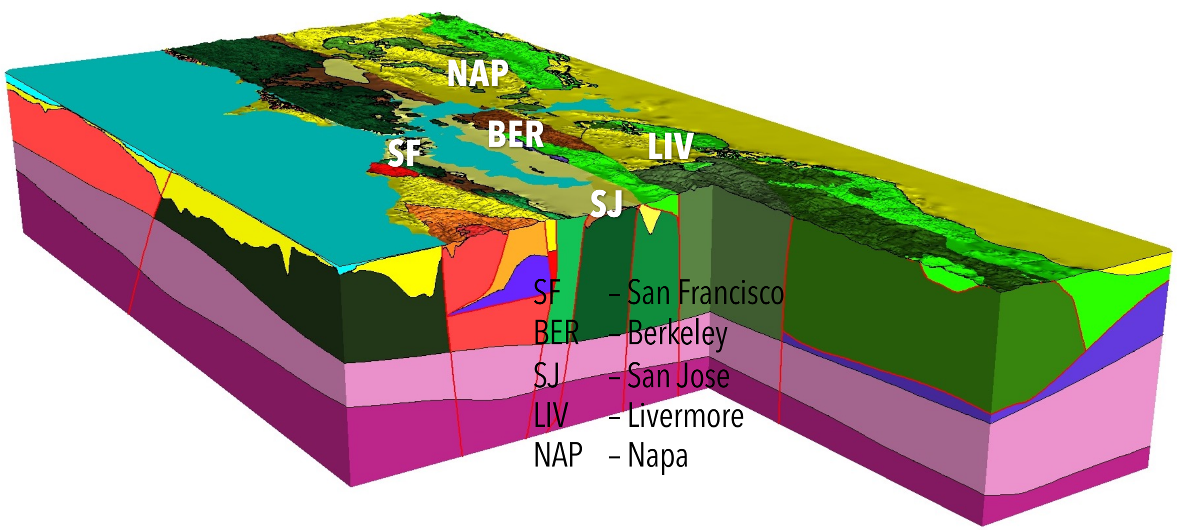
2. The computational cost of physics-based perturbations for risk analyses of distributed systems is high: need sampling of full ground motion distribution (3000 GPU node hours for each M7 simulation).



Now, imagine the distributed system is in the SFBA

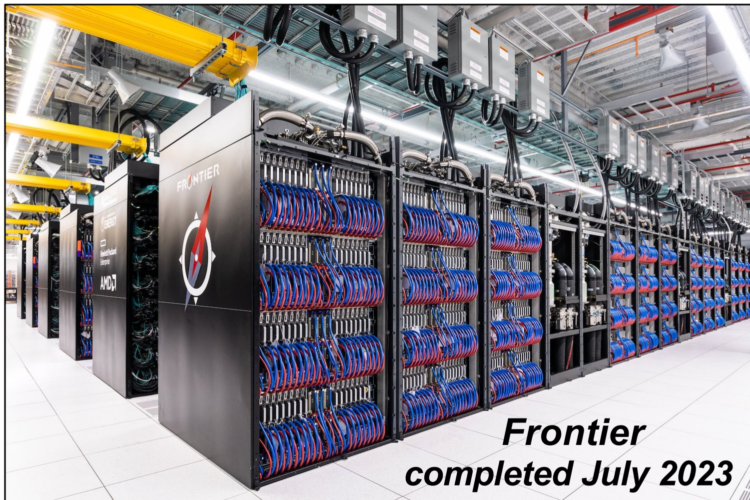
Sample scenarios across the entire hazard curve (it's not only Hayward fault's fault ☺)

We want to include model uncertainties, such as uncertainty in the source, hypocenter, velocity model etc.



DEEPSIMR^3 is developing a <u>hybrid</u> earthquake simulator

Run new supercomputer simulations



9,400 nodes, 1.194 exaflops

T3,6

PROPOSED

HYBRID EARTHQUAKE SIMULATOR

High resolution on demand scenarios with <u>high frequency realism</u>

<u>Scaling</u> constrained by **ECP simulations**

Source-path-site <u>uncertainties</u> constrained by **ground motion observations**

PROS: Ground truth high frequencies and aleatory uncertainties

CONS: Sparse, scarce observations, not available on demand / for large events & near field = **Where it matters!**

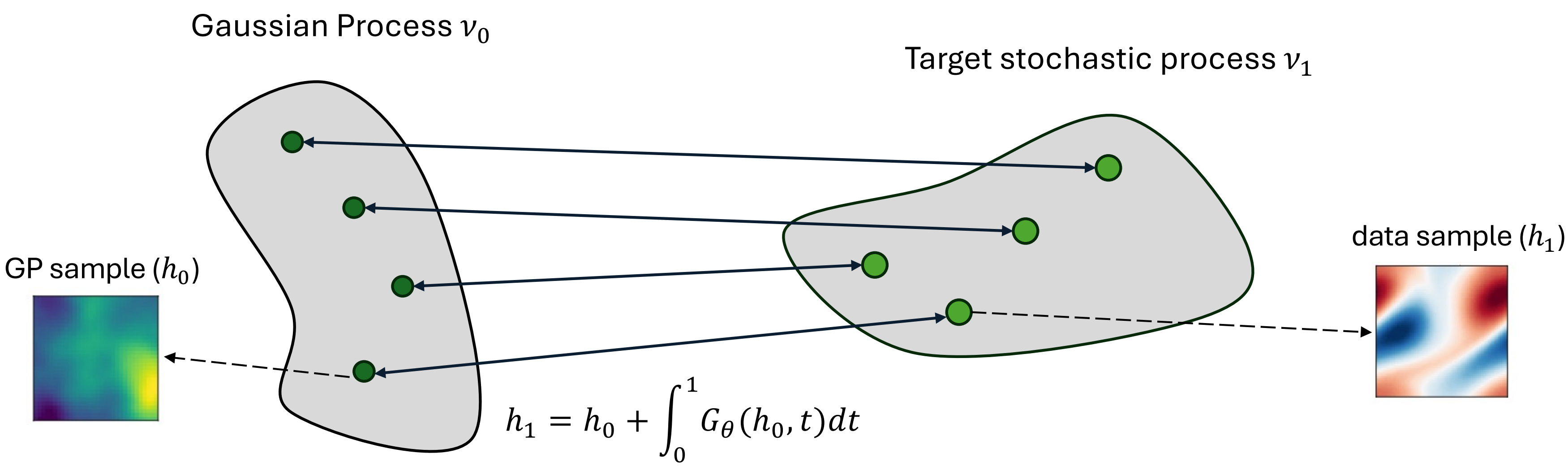
OBSERVATIONS

Flow Matching visualization

Fjelde et al. (2024)

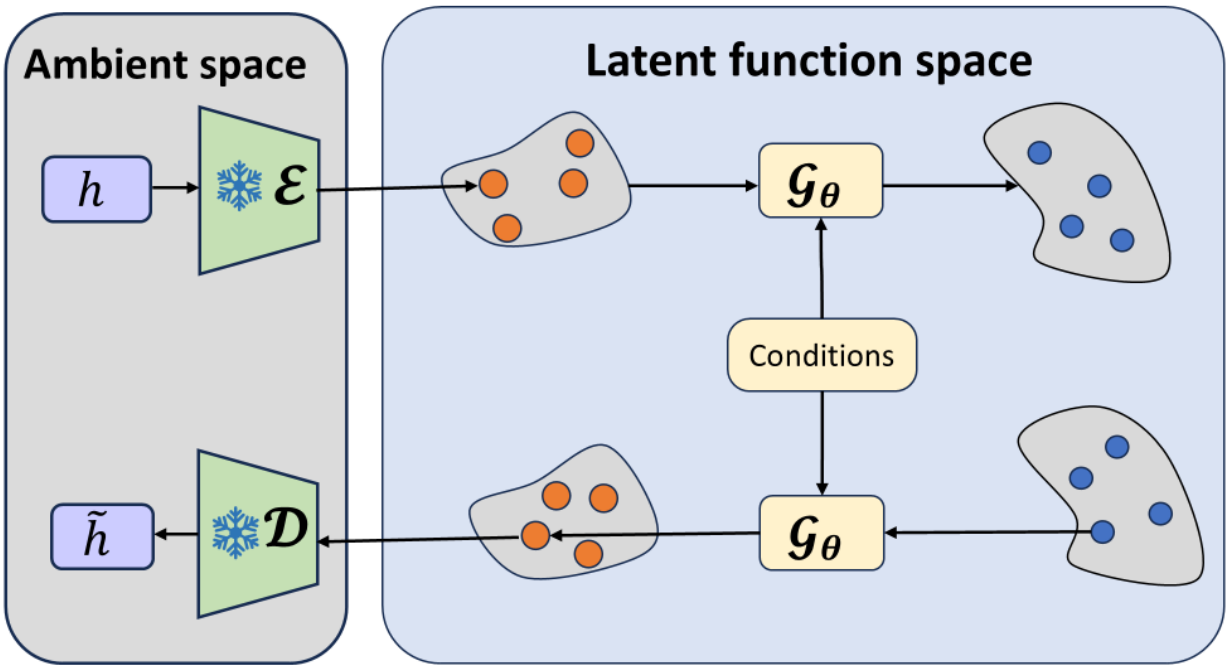
DEEPSIMR^3: Operator Flow Matching (OFM)

Mapping a Gaussian Process to the target stochastic process through flow differential equation with dynamic optimal transport plan.



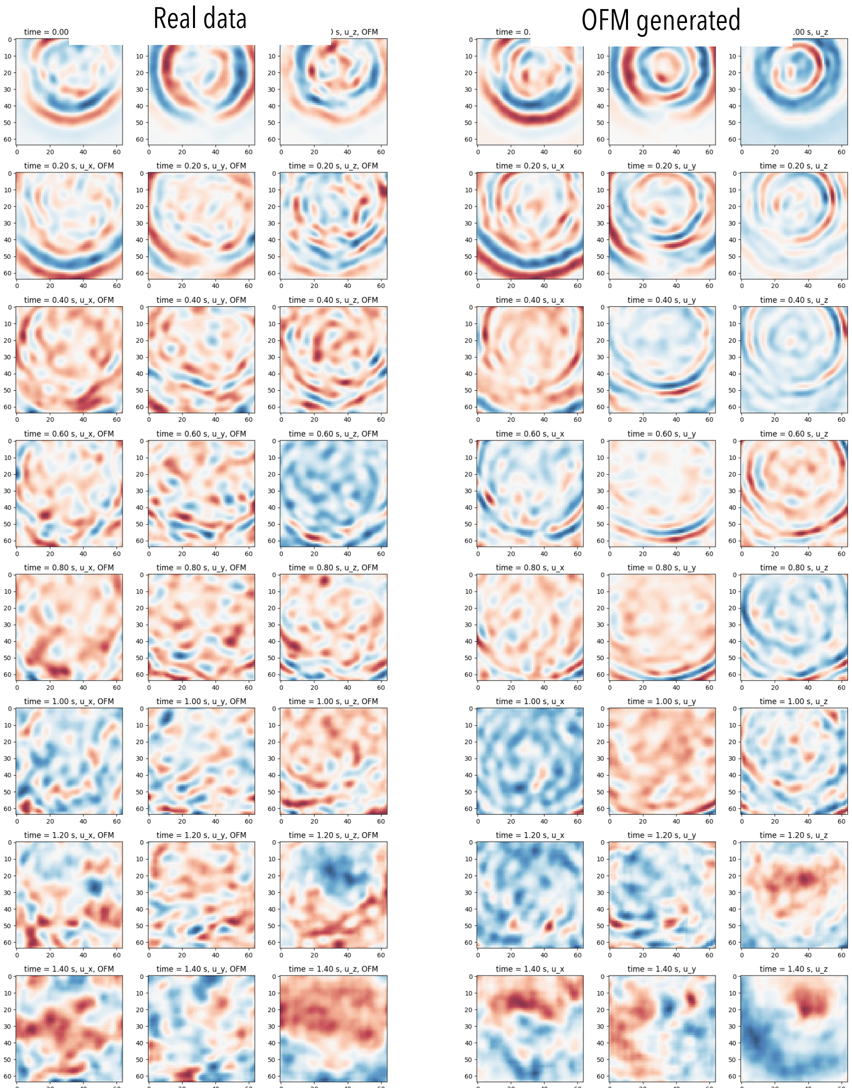
Shi et al. (2025)

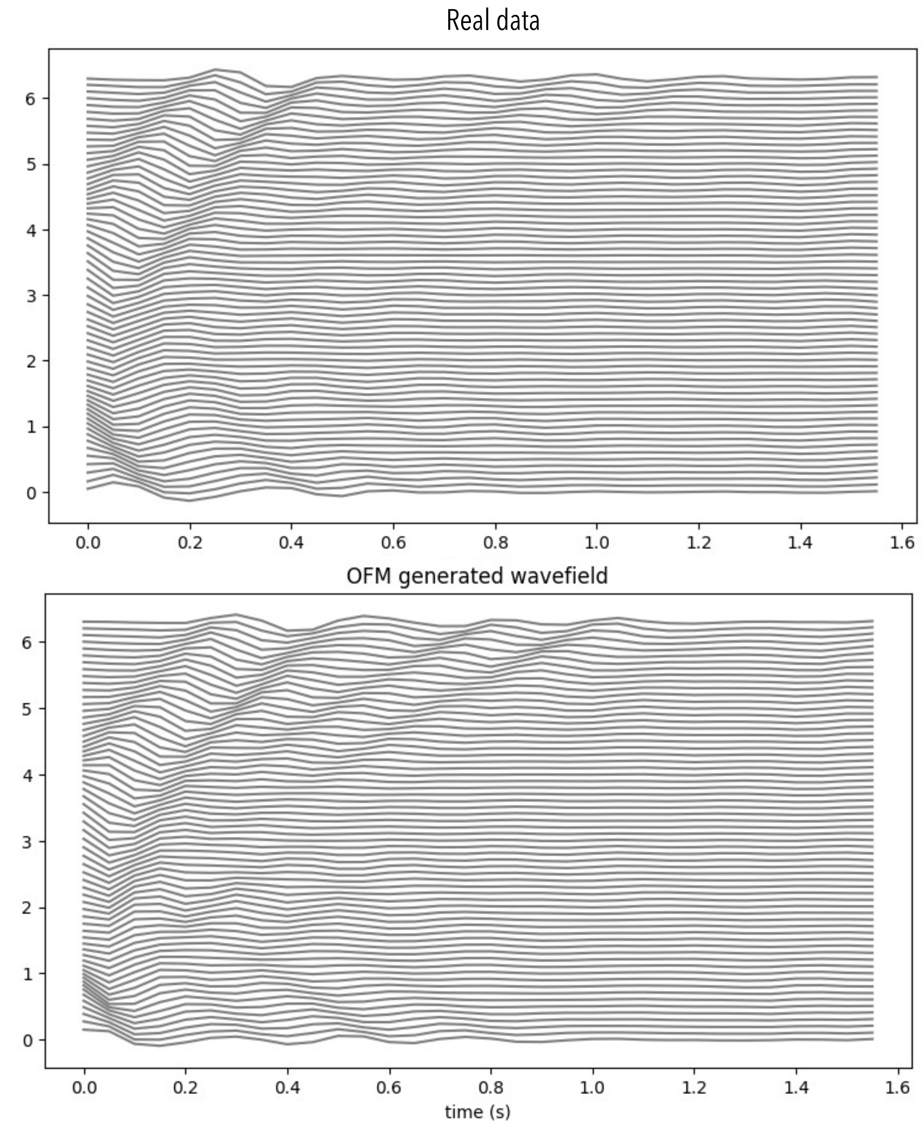
Latent Operator Flow Matching (OFM)



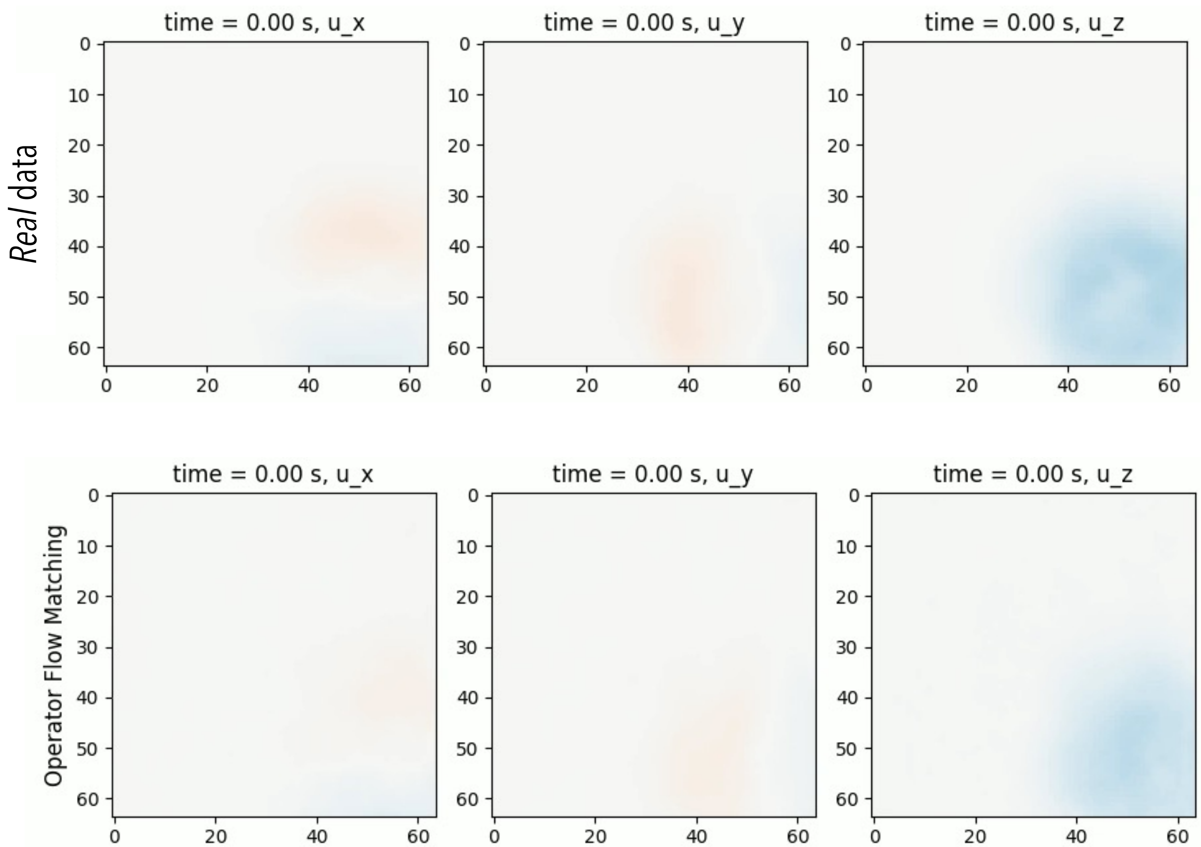
$\mathcal{E}$: **Operator encoder** $\mathcal{D}$: **Operator decoder** $\mathcal{S}$: **super-resolution operator** ❄ : **frozen** $\mathcal{G}_\theta$: **operator flow matching**

3D displacement $\mathbf{\underset{\sim}{u}}$: Point source in random fields



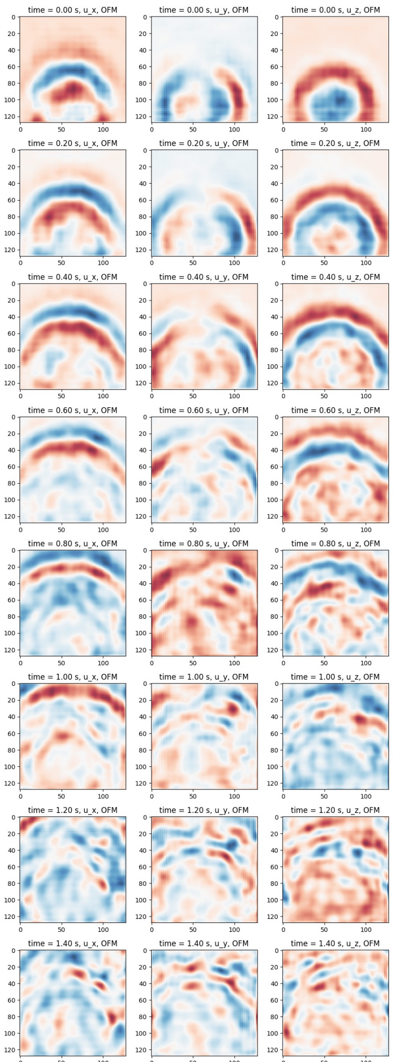


Animation: Latent OFM conditional model vs. data

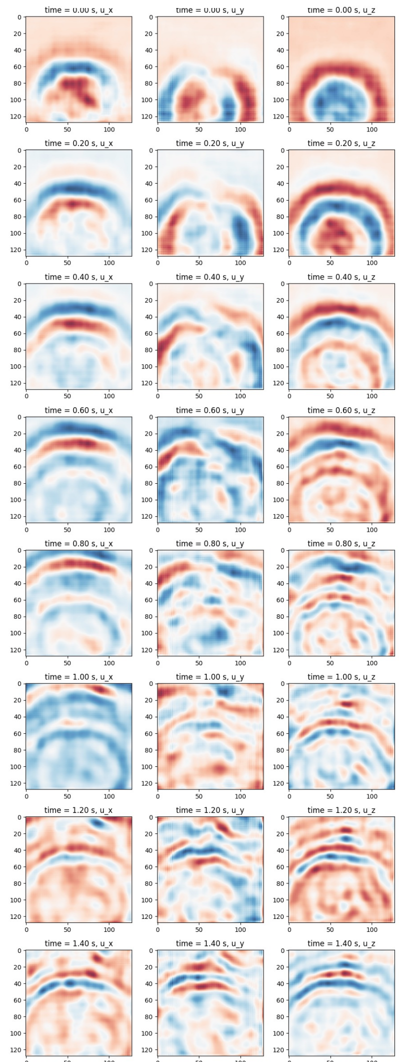


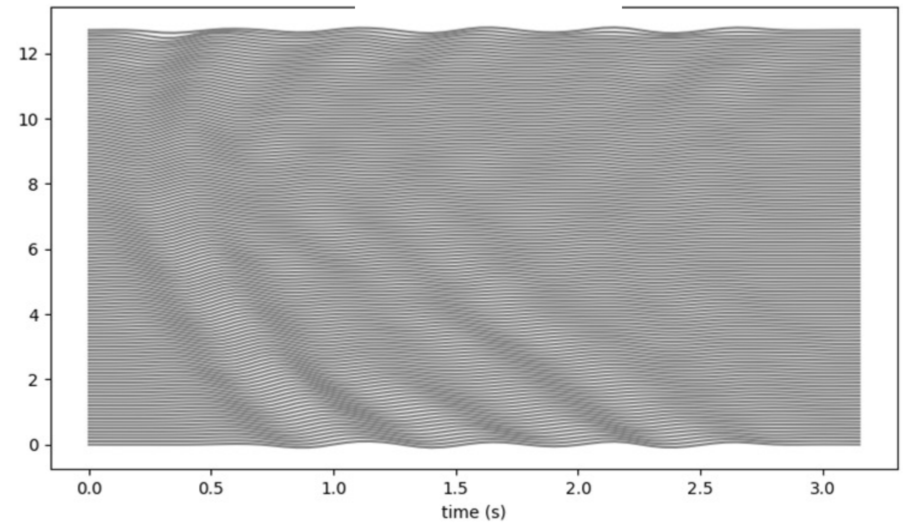
3D displacement $\mathbf{u}$: OFM is resolution invariant

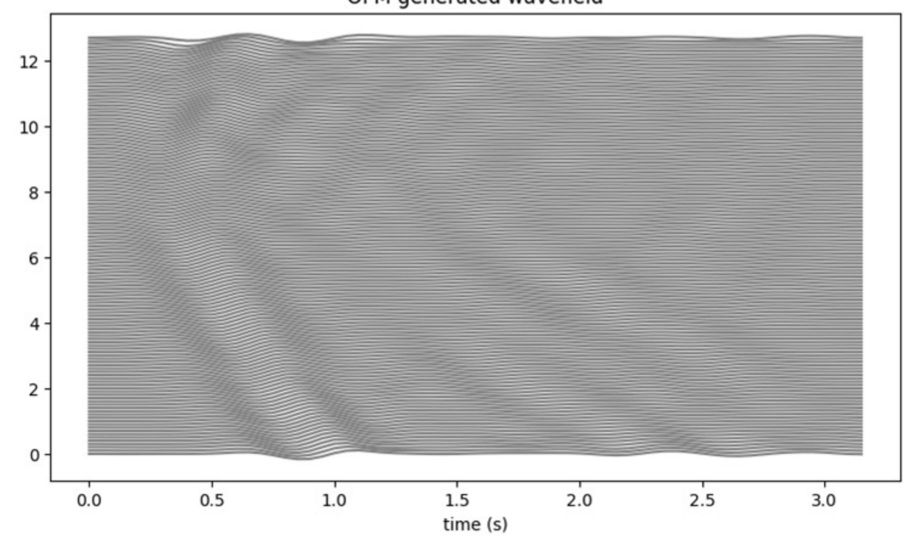
OFM – super resolution



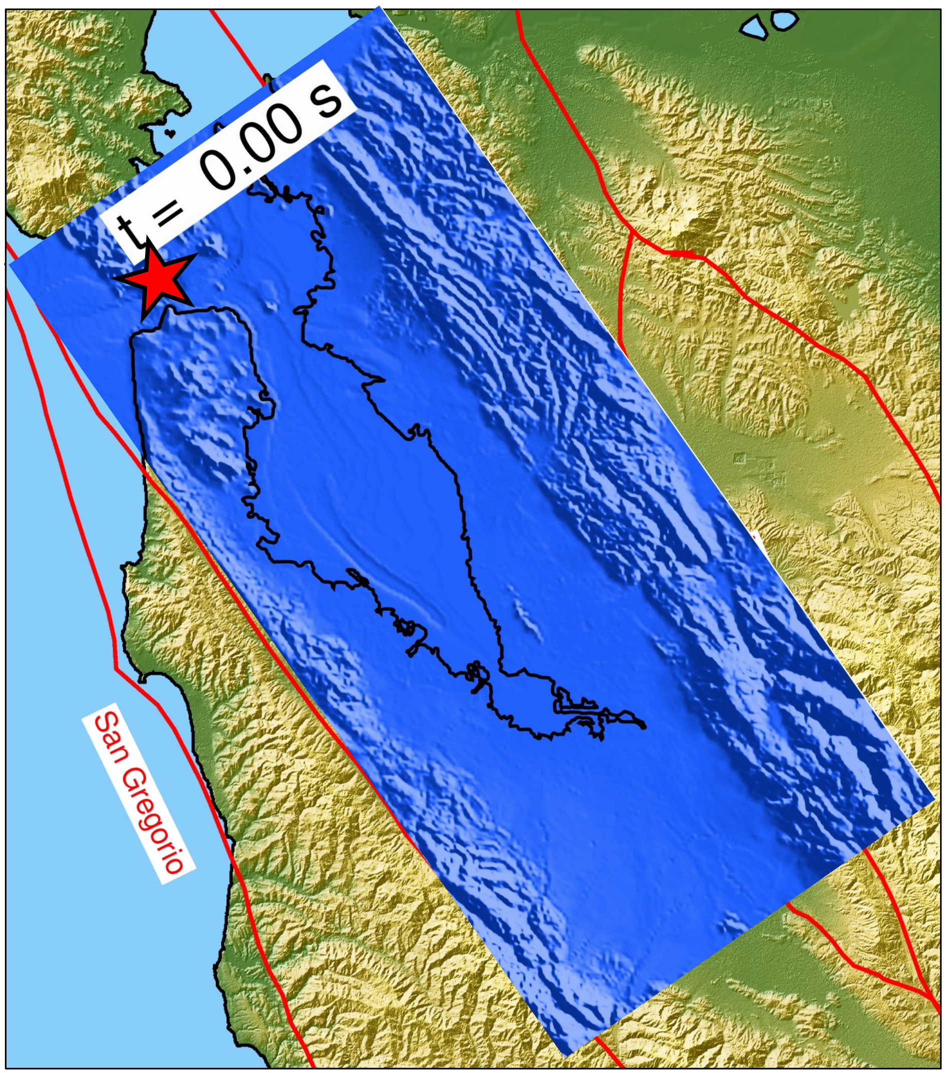
Real data

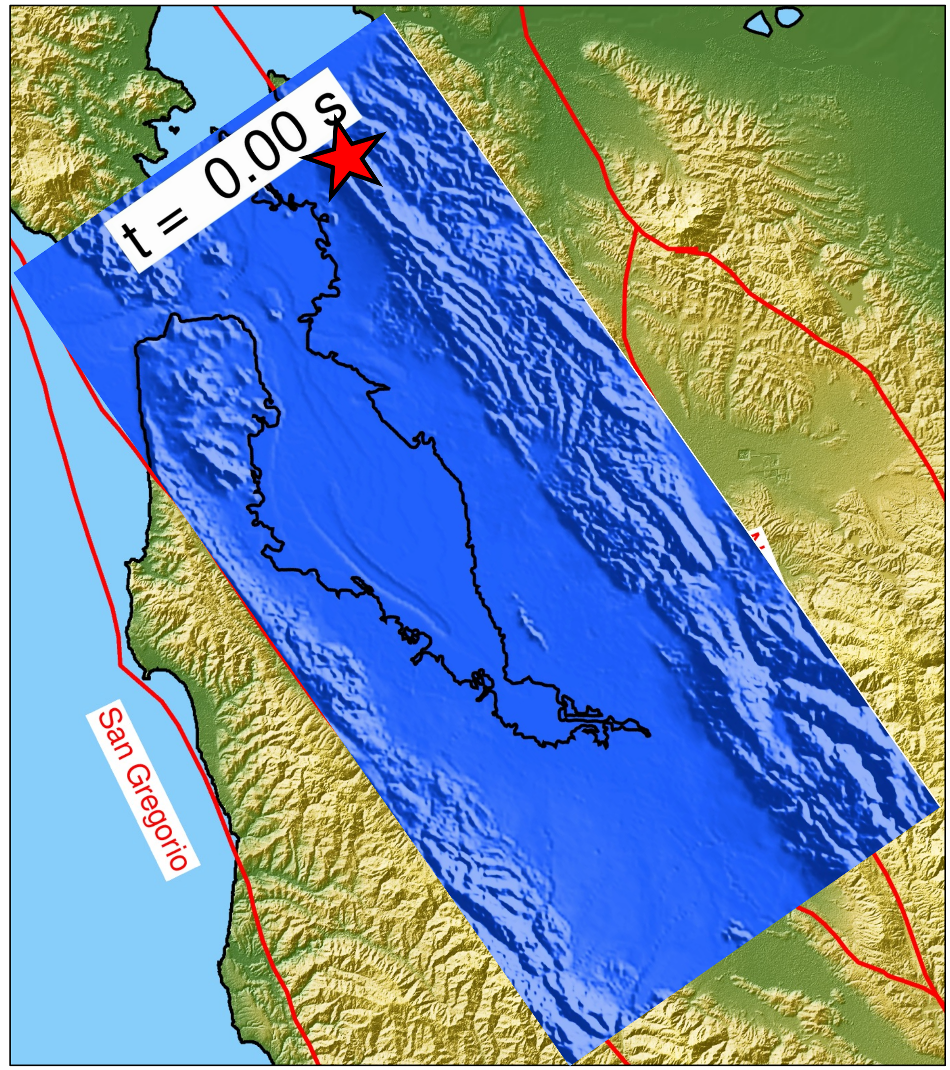


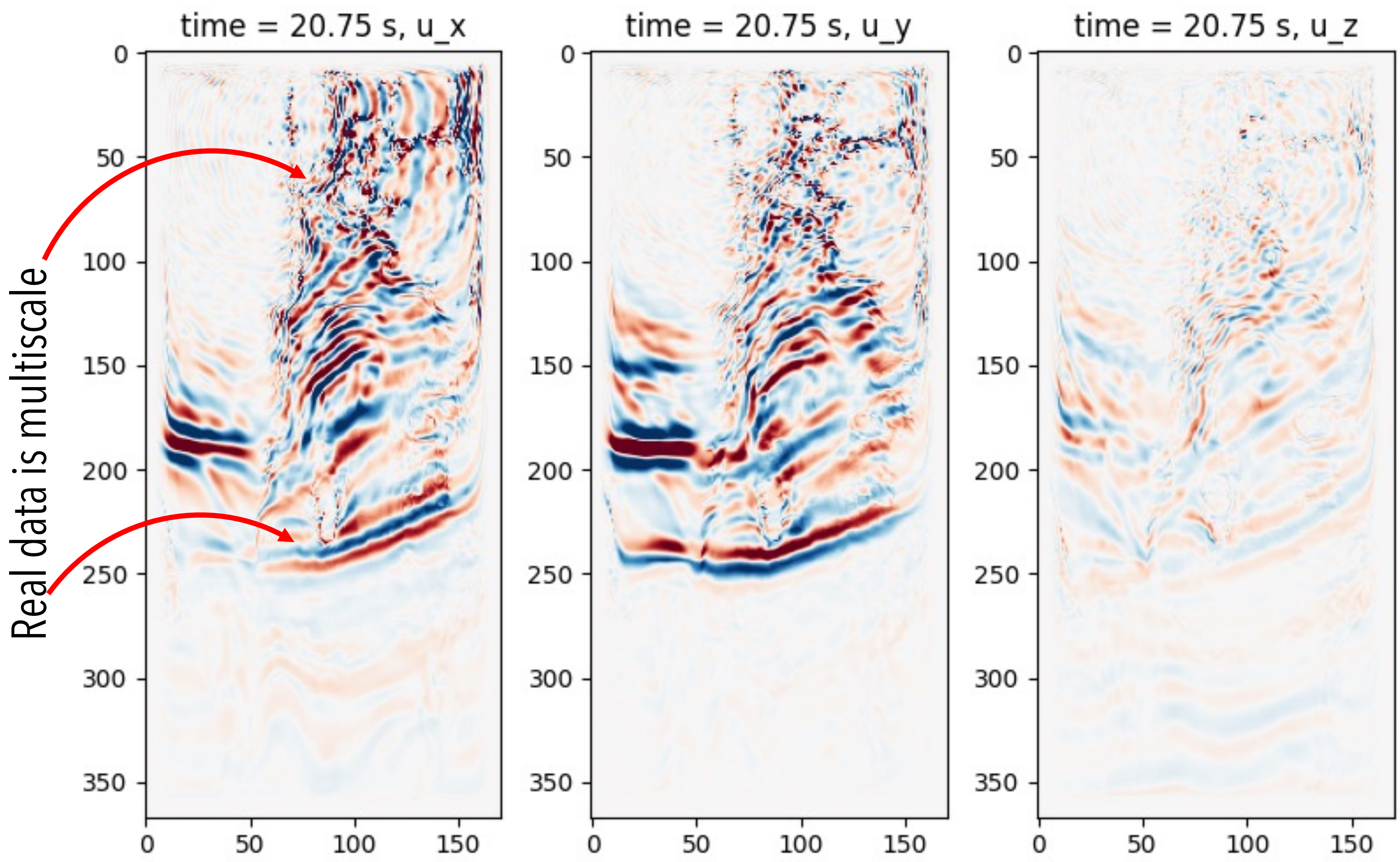




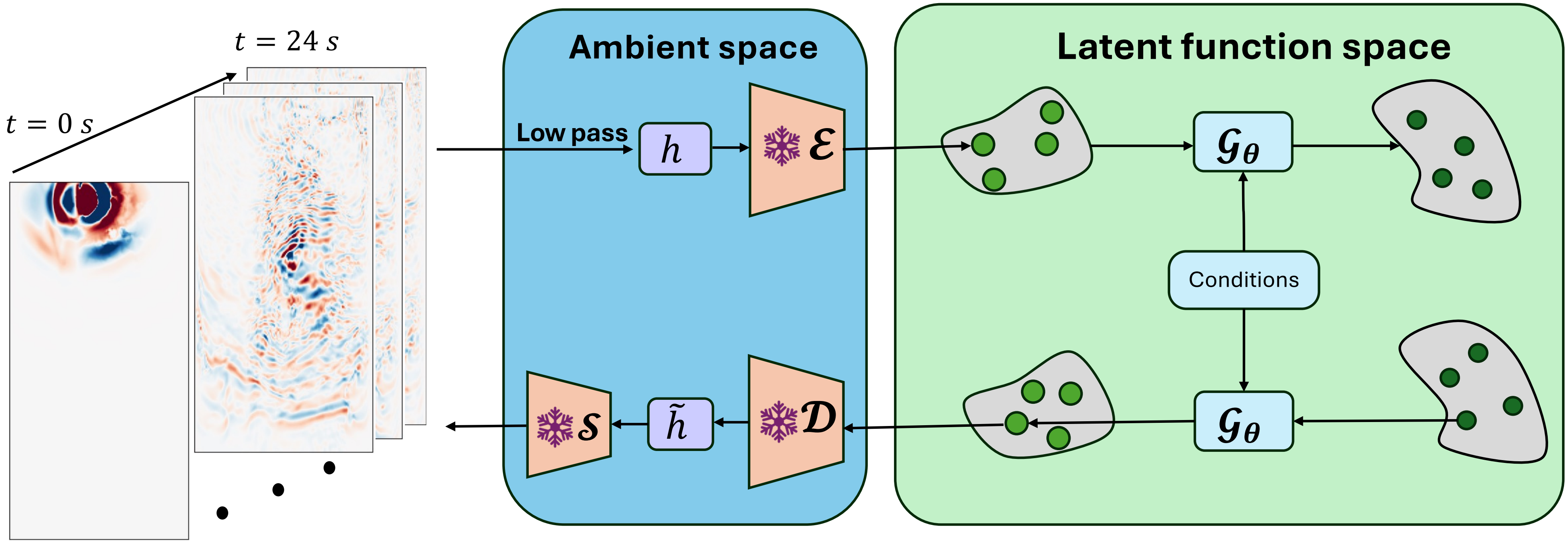
Our goal: LBNL regional model for risk reduction





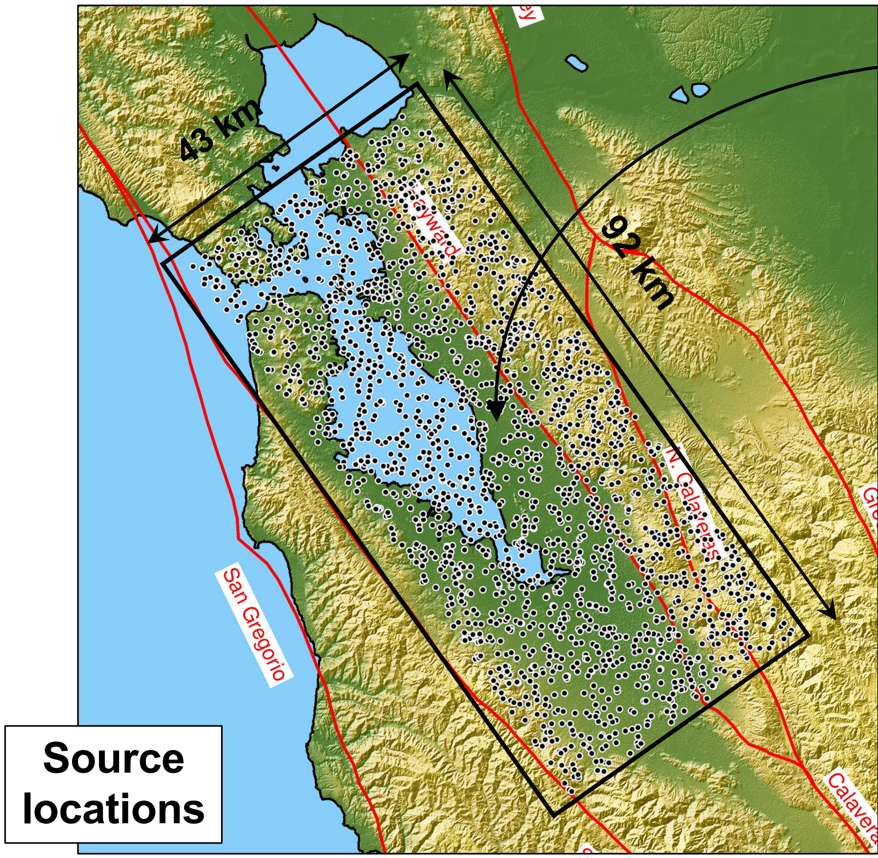

Real data is multiscale
time = 20.75 s, u_x
time = 20.75 s, u_y
time = 20.75 s, u_z

Latent OFM model + super-resolution operator



$\mathcal{E}$: **Operator encoder** $\mathcal{D}$: **Operator decoder** $\mathcal{S}$: **super-resolution operator** ❄ : **frozen** $\mathcal{G}_\theta$: **operator flow matching**

Earthquake simulation dataset overview



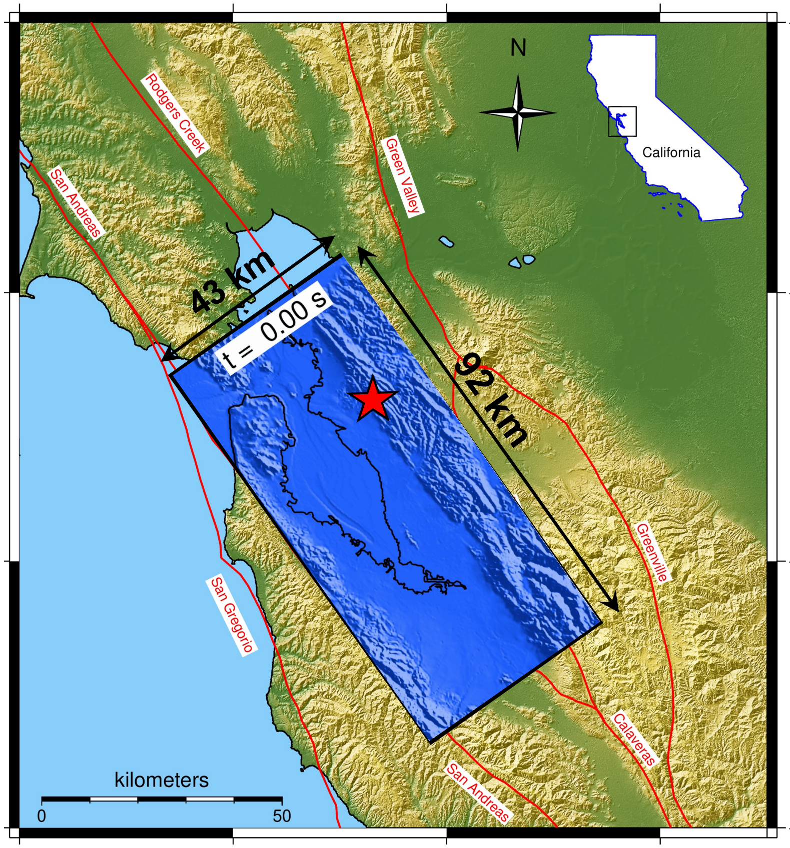
These 2100 locations are the hypocenters of the following earthquake simulations*:*

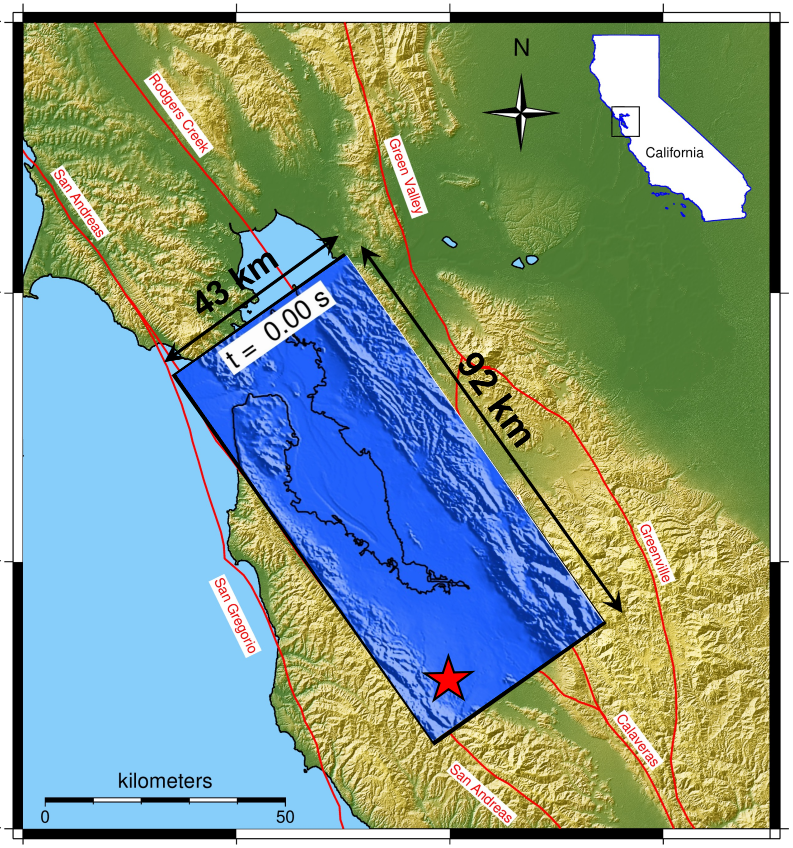
- 2000 simulations – training of ML algorithms
- 100 simulations – validating of ML algorithms
  - Perlmutter, NERSC – f_{max} = 1 Hz
- 40 simulations – further validation, higher frequency resolution
  - Frontier, OLCF – f_{max} = 2.5 Hz

Earthquake simulation characteristics

Vs,min	250 (m/s)
Radiation pattern	dip=80°, strike=145°, rake=180°
Seismic source function	Dirac delta function

Example simulations of two M4.4 earthquake events in the San Francisco Bay Area





★ Hypocenter location, 6km depth

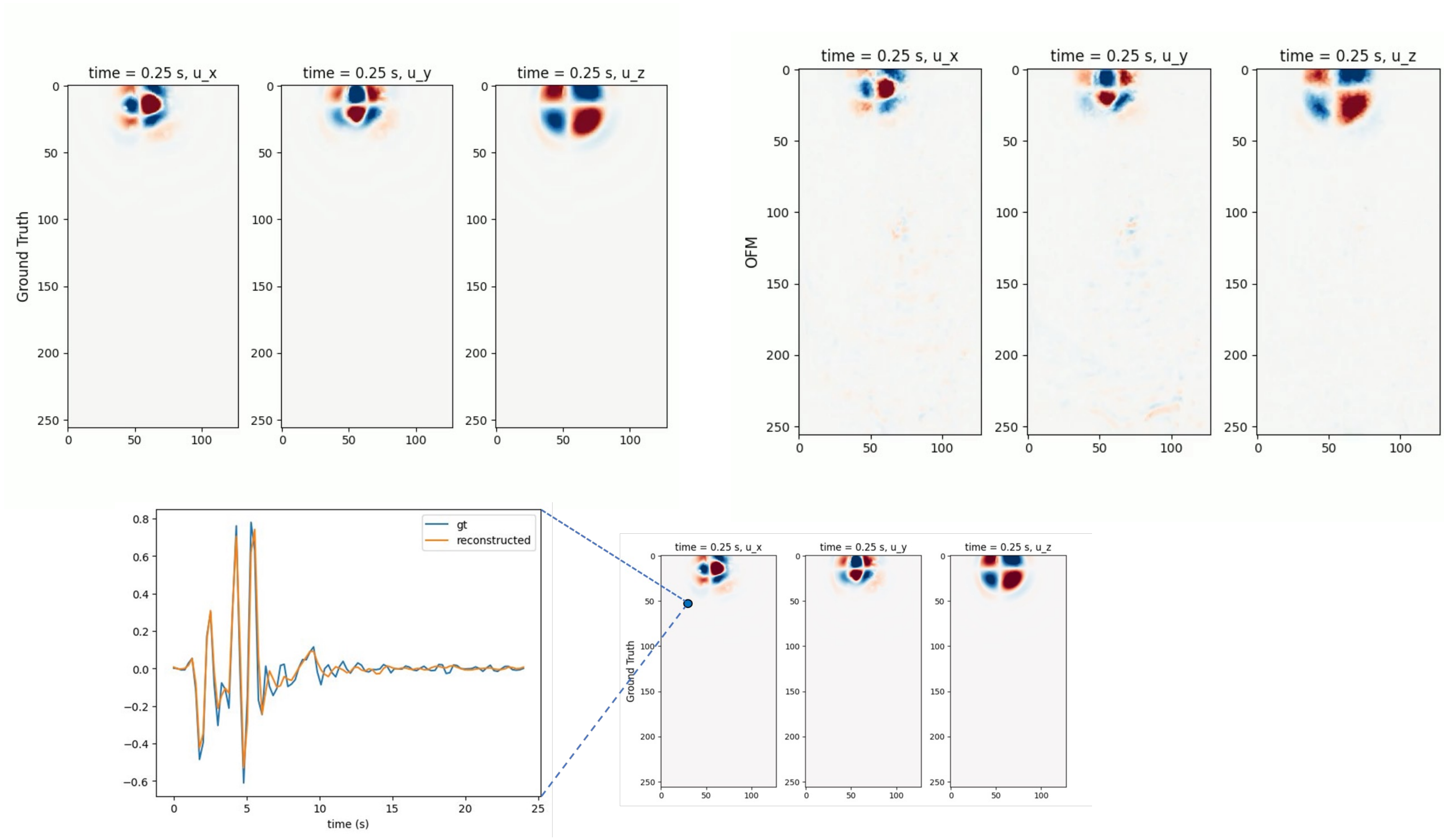
fmax =2.5Hz

Vs,min = 250 m/s

+
2,098 more

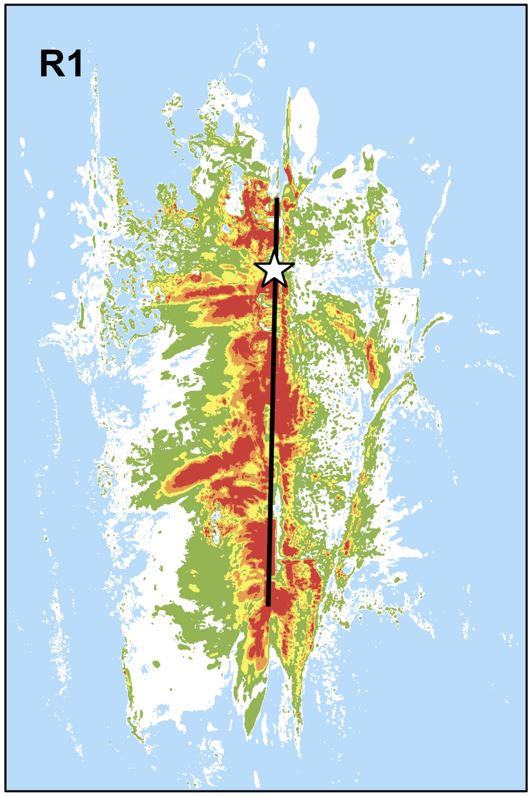
~ 1TB of hdf5, zfp-compressed data

~

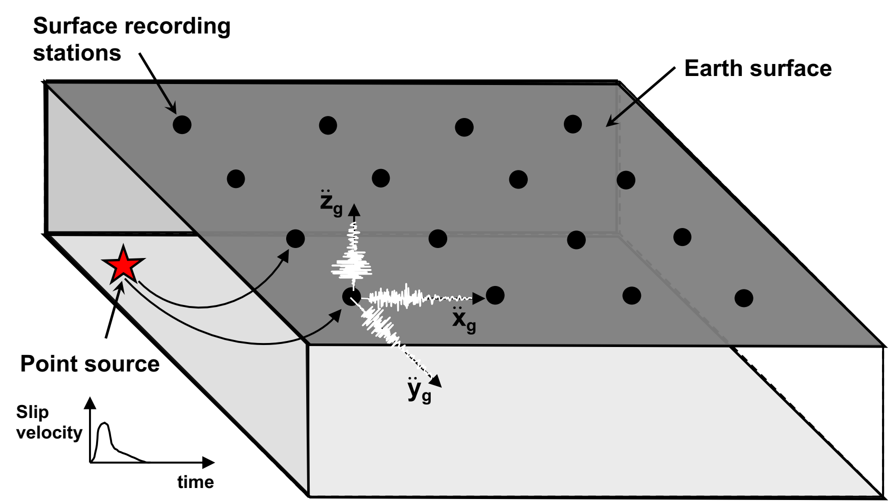


time = 0.25 s, u_x
time = 0.25 s, u_y
time = 0.25 s, u_z
Ground Truth
OFM
gt
reconstructed
time (s)

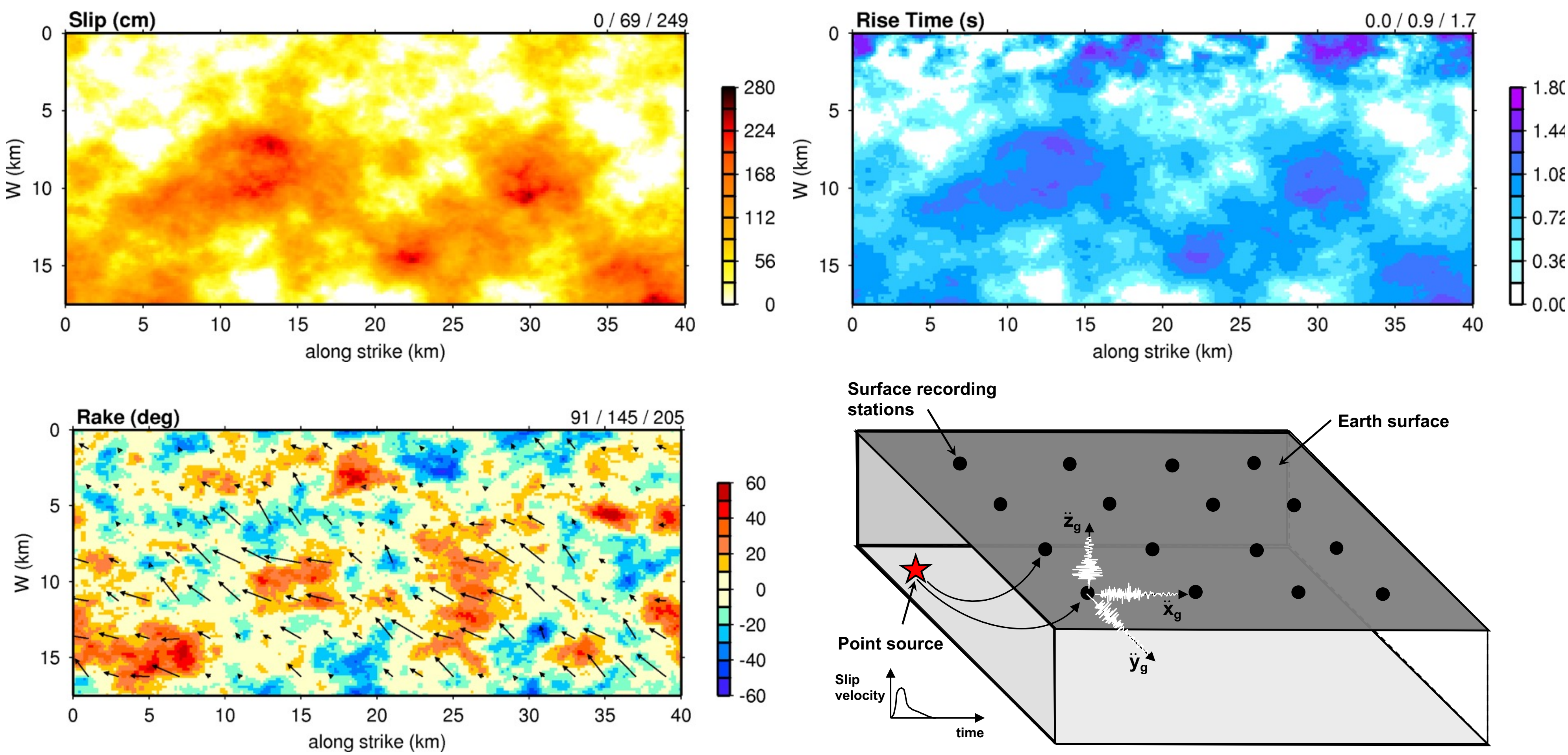
And now what?


R1

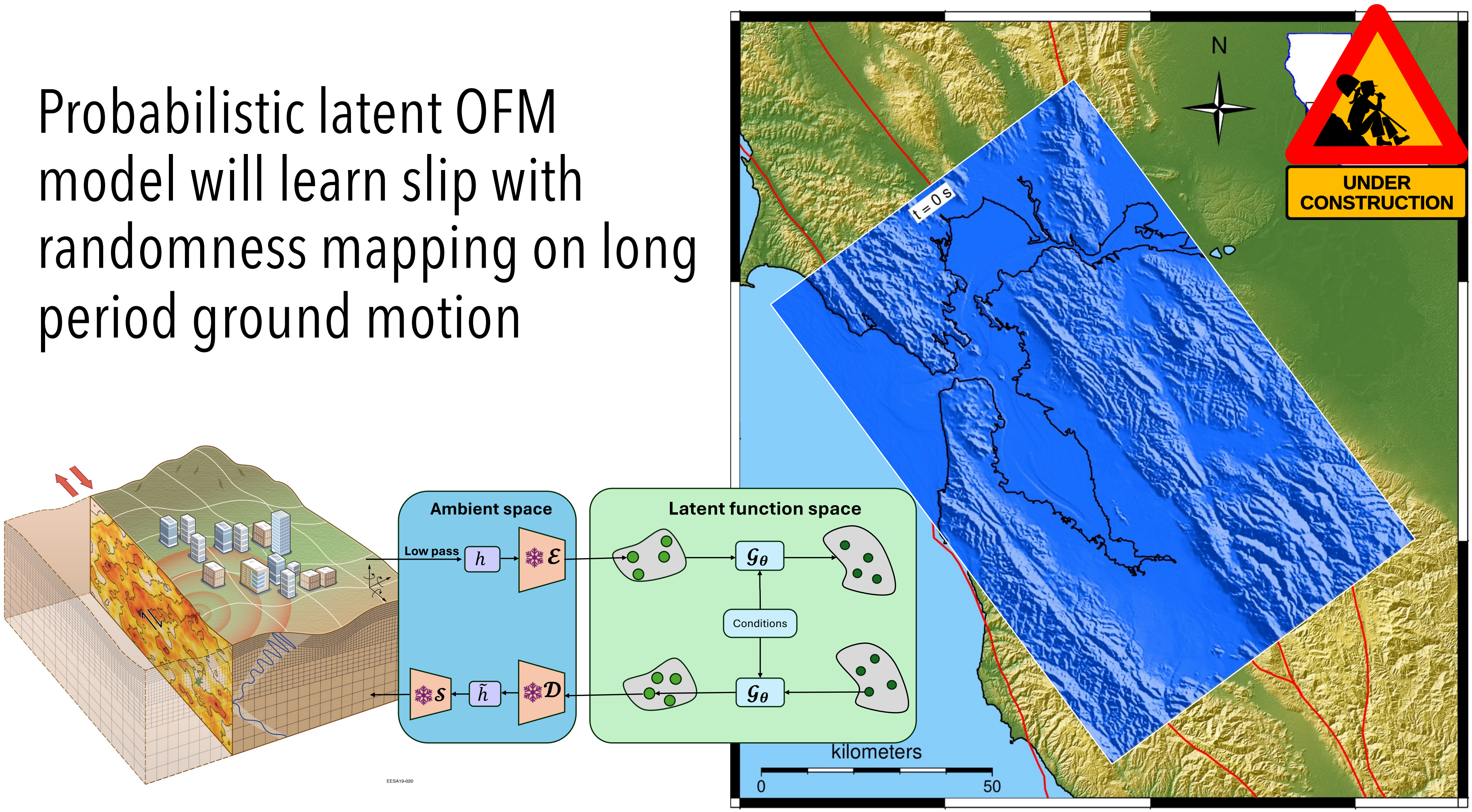
?


Surface recording stations
Earth surface
$\ddot{z}_g$
$\ddot{x}_g$
$\ddot{y}_g$
Point source
Slip velocity
time

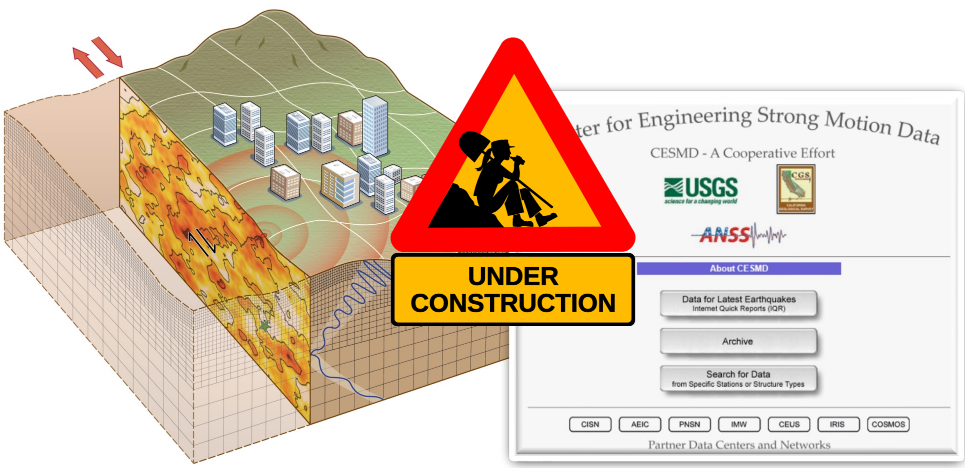
From point to finite sources (Graves & Pitarka, 2016)

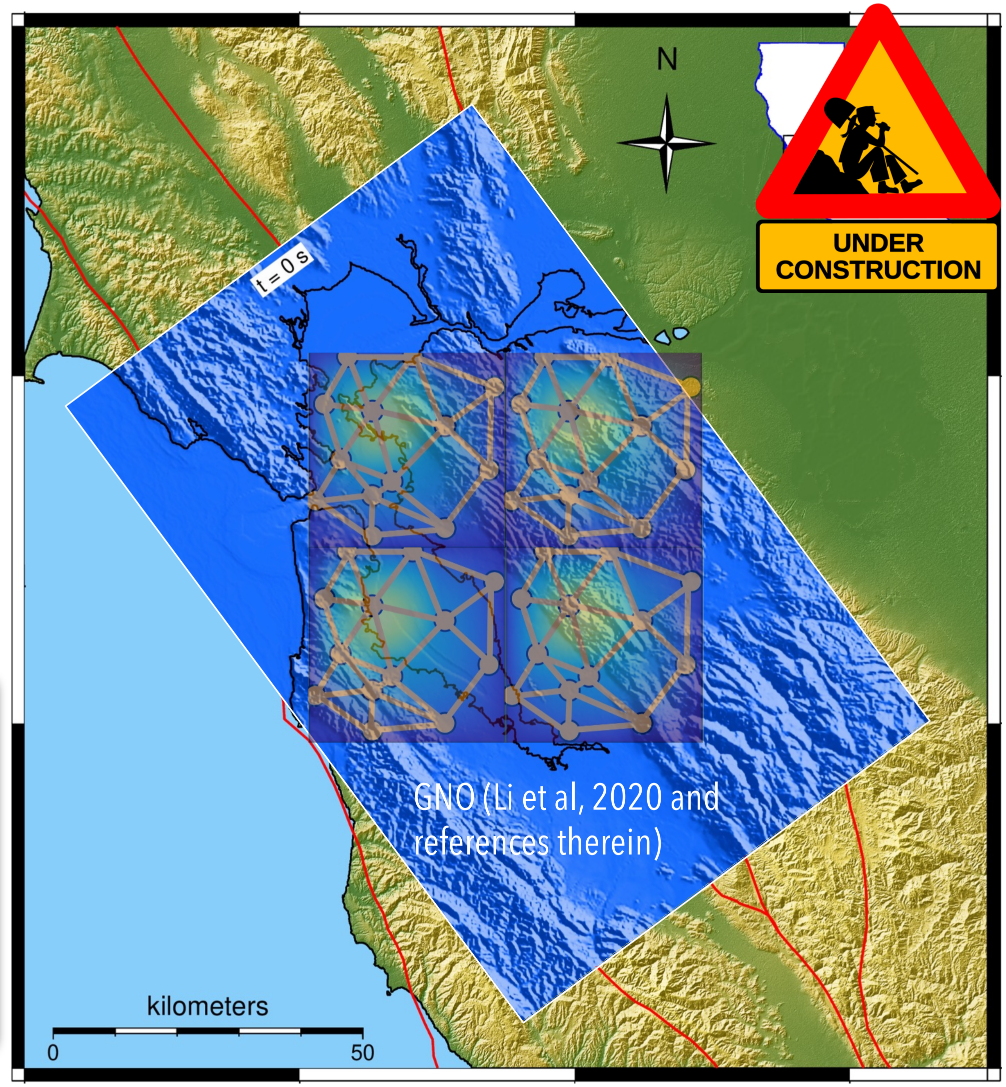


Probabilistic latent OFM model will learn slip with randomness mapping on long period ground motion



[to be named] will fuse long period synthesized ground motions with observed high frequencies and associated spatial uncertainties.





Products of DeepSimR^3

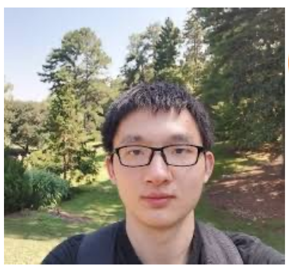
- Shi Y., Azizzadenesheli K., Asimaki D. and Ross Z. (2025). Stochastic Process Learning via Operator Flow Matching, 42nd International Conference on Machine Learning (ICML 2025), July 13–19, Vancouver, Canada (under review)
- Shi Y., Azizzadenesheli K., Zou C., Tsalouchidis K., Lavrentiadis G., McCallen D., Asimaki D. and Ross Z. (2025). Ground-Motion Flow: Latent Operator Flow Matching with Super Resolution Operator for Large-Scale 3D Ground-Motion Synthesis, Science Advances (in preparation)

Presentations on DeepSimR^3

- *Simulated Ground Motions in the Development of Machine Learning Based Risk Estimations*. Asimaki, D., 2025 PEER LBNL Workshop on the Regional Scale Simulated Ground Motion Database (SGMD) for the San Francisco Bay Area, March 24, 2025.
- *Ground-Motion Flow: A Scalable Flow-based Operator Learning Framework for 3D Ground-Motion Synthesis*, Asimaki, D., 2025 Engineering Mechanics Institute 2025 Conference, May 27-30, Anaheim, CA
- *Sensors as supercomputers: Data-driven ground motion models for risk assessment of infrastructure systems*, Asimaki, D., Keynote Lecture, Environmental Seismology: Planning for the Planet's Future, Denver, CO, 14-18 October 2025.

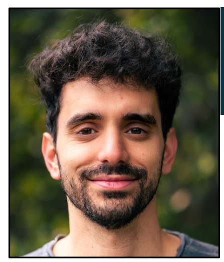
Thank you!

Yaozhong Shi



Machine Learning, Development and Implementation

Konstantinos Tsalouchidis



Simulated ground motions for ML and verification

Kamyar Azizzadenesheli

Machine Learning

Questions?

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Science Foundations for Energy Earthshot under **Award Number DE-SC0024705**.