5th International Bridge Seismic Workshop,
Hunan University, 5-8, June 2026

Unified Transformer-based Ground Motion Prediction Incorporating Finite-Fault Geometry

Xinzheng Lu [1]

Yitian Feng [1], Weiqiang Zhu [2]

1 Tsinghua University, 2 UC Berkeley

Email: luxz@tsinghua.edu.cn

Outline

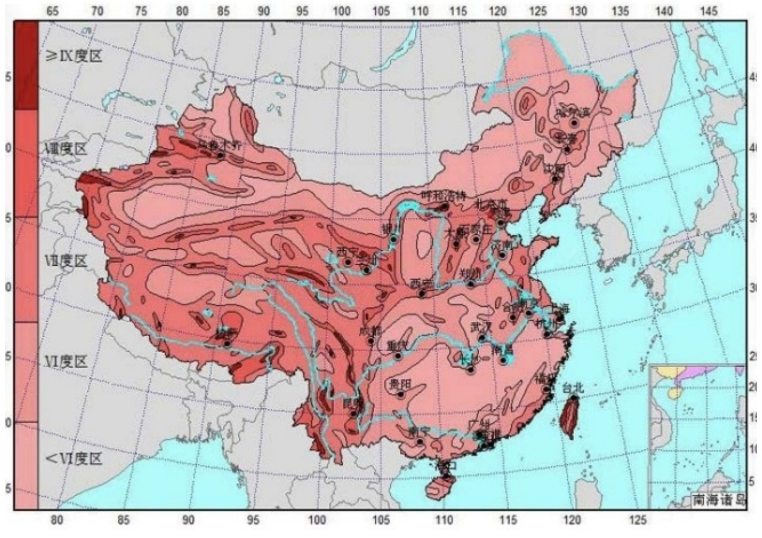
Outline

1.1 Background

- Bridge seismic resilience is essential to lifeline infrastructure, linking ground-motion prediction, emergency response, and post-disaster recovery.

Pre-event

- Seismic fortification standard and bridge design code.

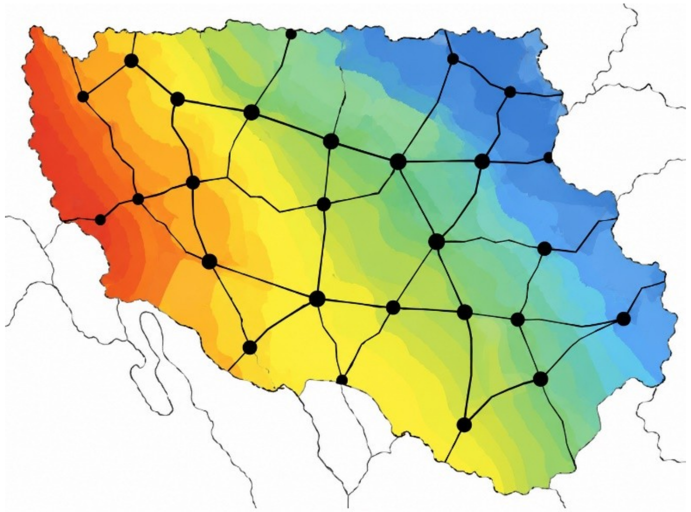


During event

- Real-time decisions such as train braking, traffic control, and bridge closure.

Post-event

- Evaluate regional bridge damage and road-network disruption for reconstruction.

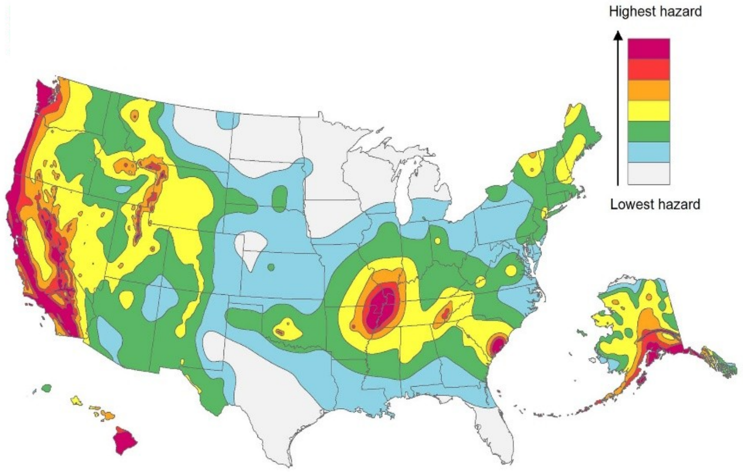


1.1 Background

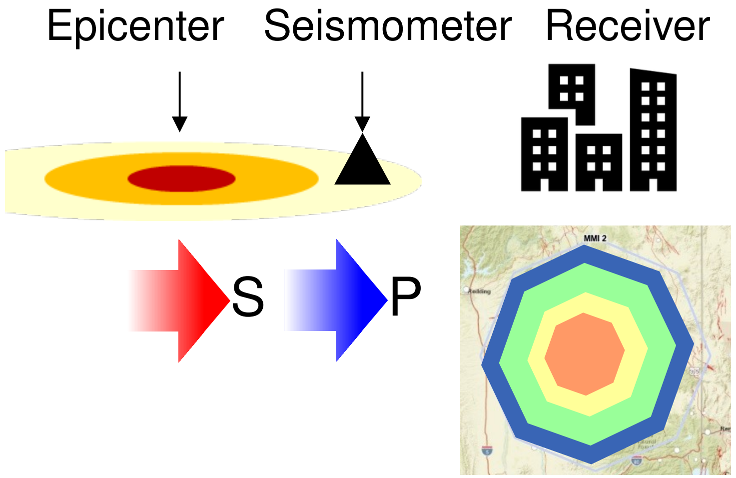
- **Ground Motion Prediction Models (GMM)** is the basis for seismic hazard analysis, early warning, and emergency response

Time of earthquake occurrence	**Pre-event**	**During event**	**Post-event**
	Intensity prediction	Early warning	Interpolation & assessment

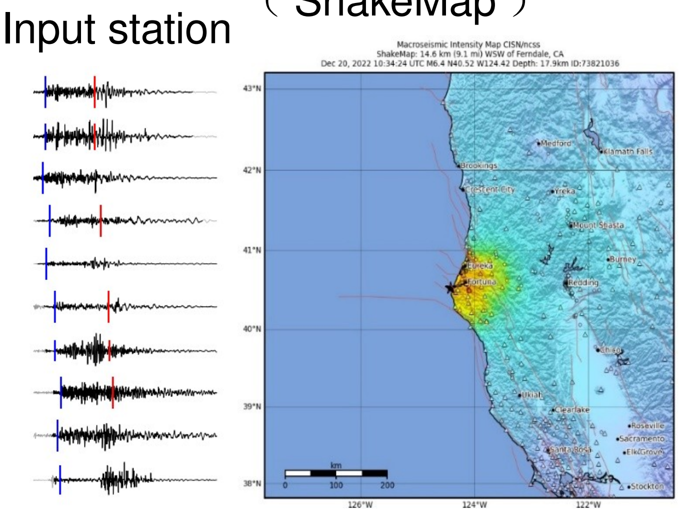
Seismic hazard analysis
（USGS）



Regional GM early warning



Near real-time assessment
（ShakeMap）

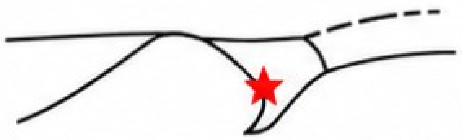


1.2 Pre-event intensity prediction

- **Pre-event intensity prediction** estimates potential ground shaking before an earthquake and provides the basis for seismic hazard analysis.

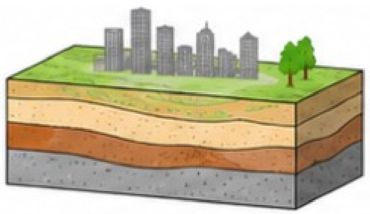
Input

Source scenario



+

Site condition



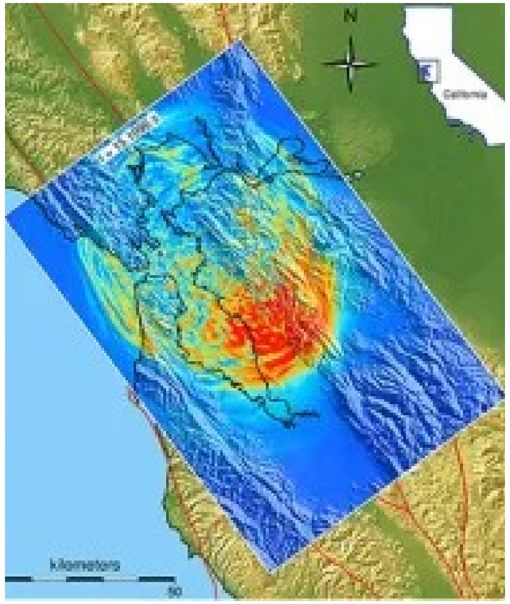
+

Regional geological

Ground motion prediction model

Output

Predicted seismic hazard



EQSIM

1.2 Pre-event intensity prediction

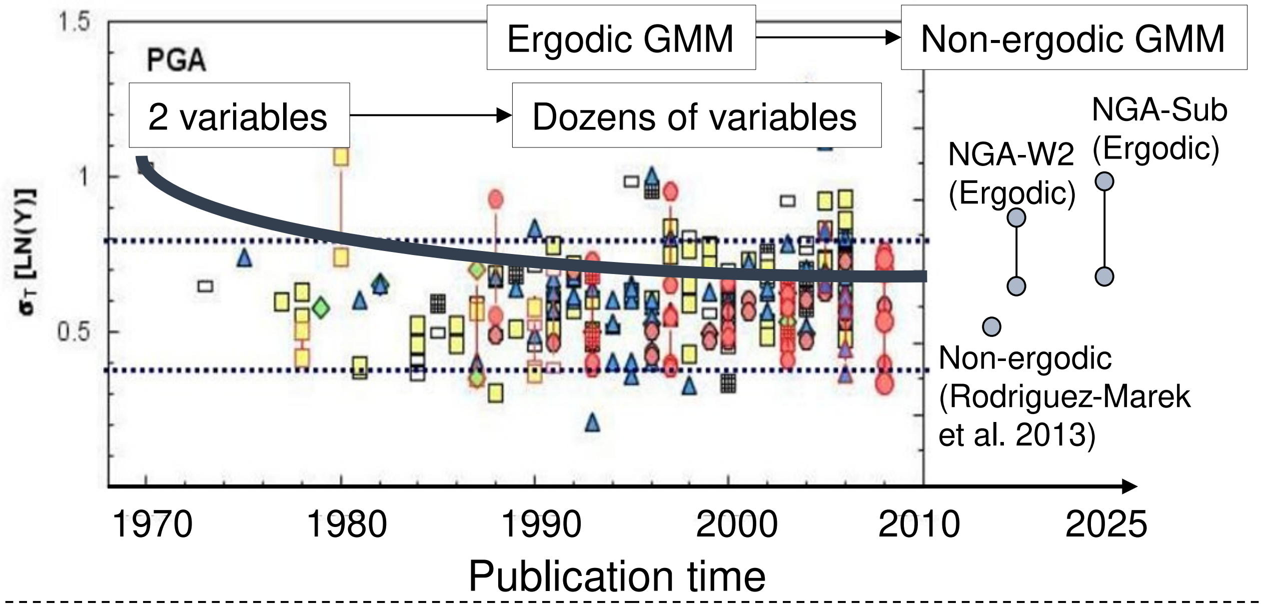
- Traditional methods struggle to balance **accuracy** and **real-time performance**

Empirical methods

Ground motion prediction equations (GMPE)

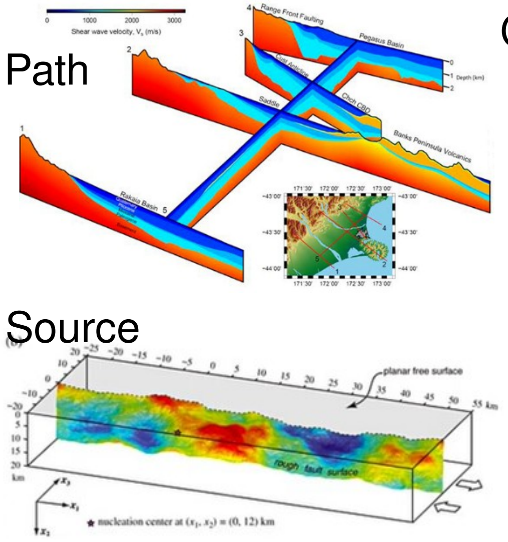
$$\ln IM = f_{median} + f_{non-ergodic} + f_{directivity} + \cdots$$

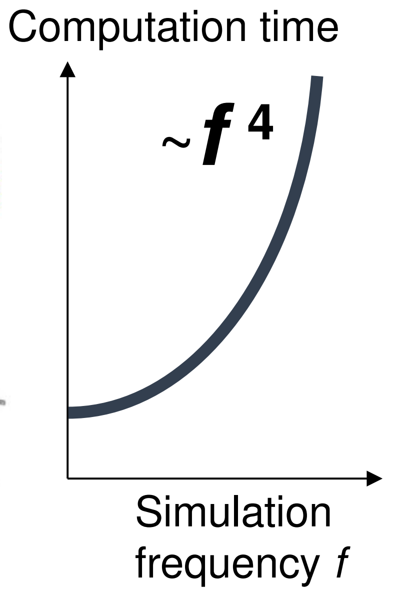
GMPE prediction uncertainty (σ)



Numerical simulations

- Long computation time
- High data storage requirements
- Dependent on source/path/site models

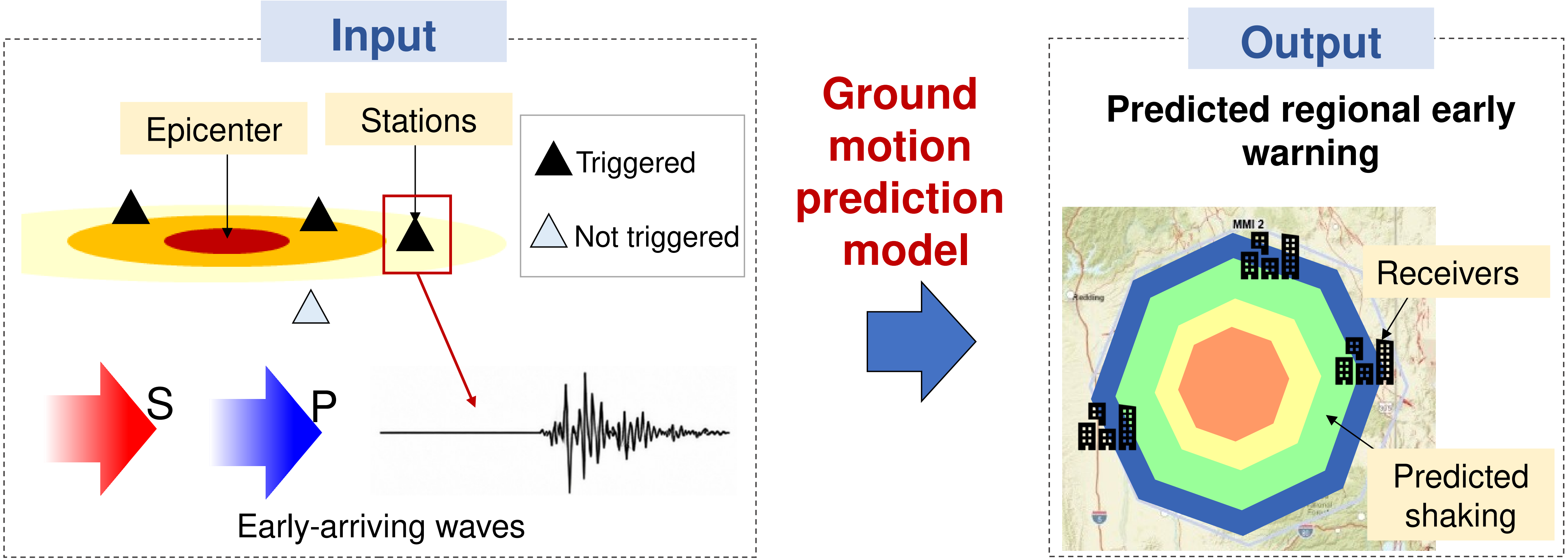




Sigma: Issues, Insights, and Challenges. (Strasser 2009)

1.3 During-event early warning

- **During-event early warning** uses partial real-time observations to rapidly predict ground shaking before strong motion arrives.



1.3 During-event early warning

- Existing EEW methods are limited by **source estimation errors**, **short warning time**, and **dependence on station coverage**

Source-based EEW

Observed waves

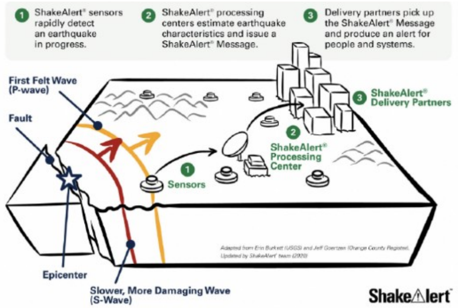
Empirical regression ⬇ Template matching

Source

⬇ **GMPE**

Predicted ground motion

Representative work



ElarmS, FINDER

Propagation-based EEW

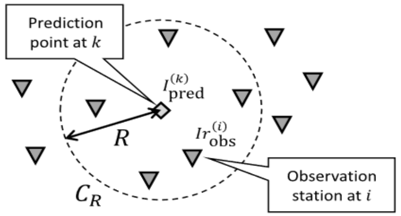
Source (Optional)

Observed waves

⬇

Predicted ground motion

Representative work



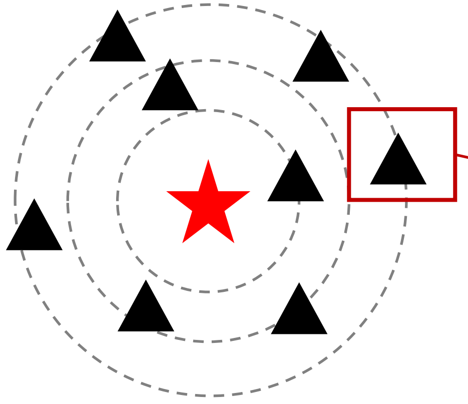
$$I_{\mathrm{pred}}^{(k)} = \max_{i \in C_R} \left\{ Ir_{\mathrm{obs}}^{(i)} - F_0^{(i)} \right\} + F_0^{(k)}$$

PLUM

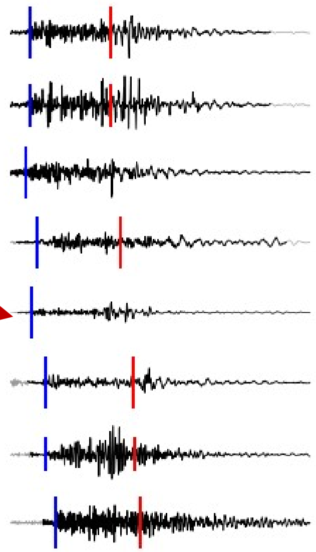
1.4 Post-event interpolation and assessment

- **Post-event interpolation and assessment** reconstruct regional ground-motion fields from station records for emergency response.

Input



Input stations

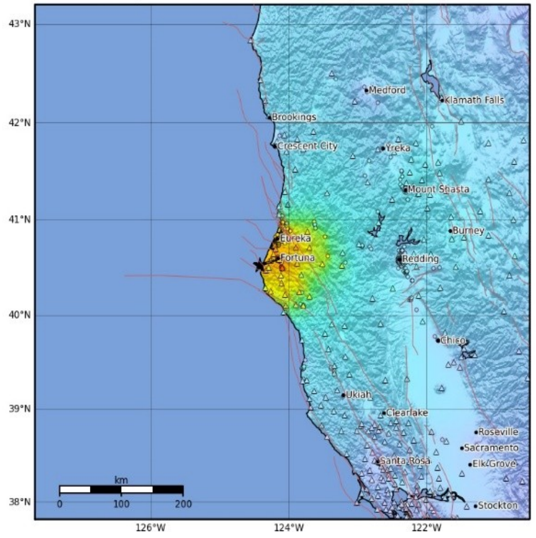


Observed waves

Ground motion prediction model

Output

Shakemap of the entire target region



Courtesy: USGS

1.4 Post-event interpolation and assessment

- Traditional interpolation misses complex source–path–site effects limited by **station coverage and predefined assumptions**

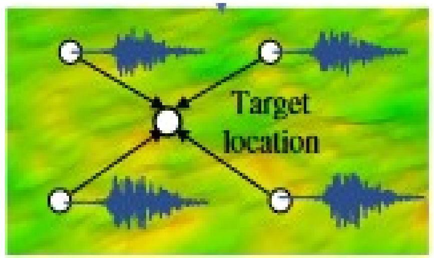
Source-free interpolation

Source Information (unused)

Observed ground motions

Predicted ground motion

Representative work



Kriging
Inverse Distance Weighting
Natural neighbor interpolation

Source-informed interpolation

Source information

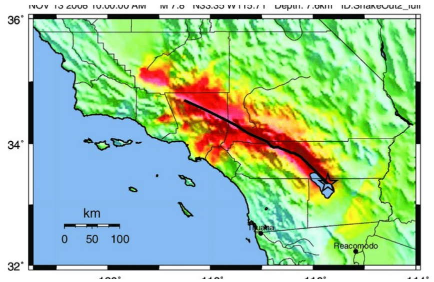
Observed ground motions

GMPE

Condition

Predicted ground motion

Representative work

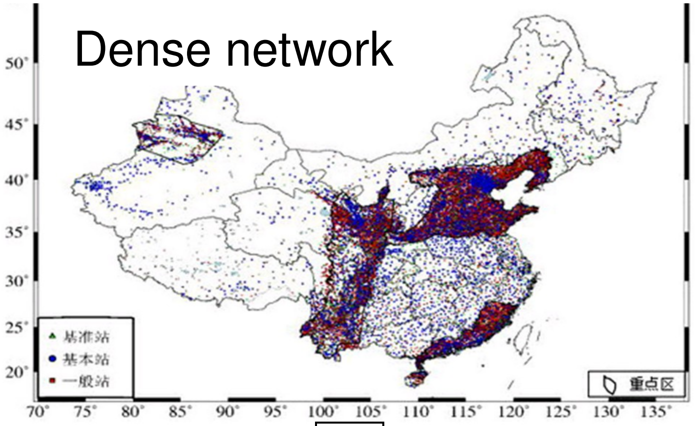


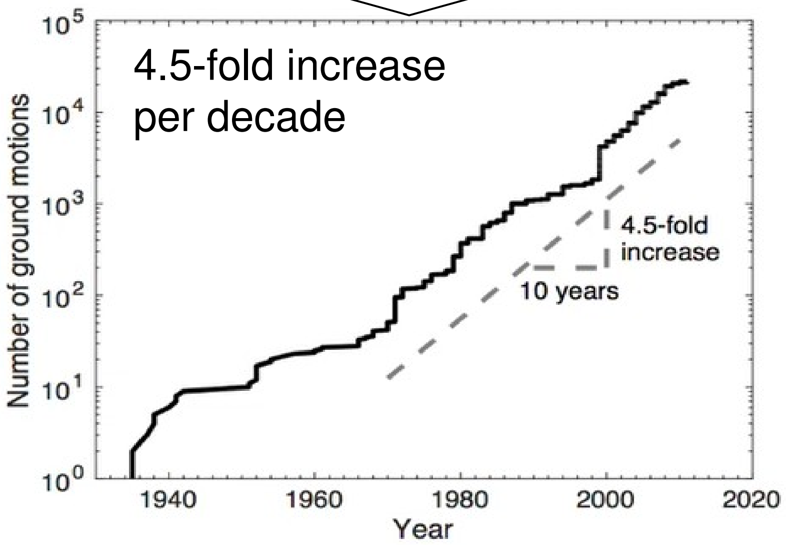
ShakeMap, USGS

1.5 Data-driven research paradigm

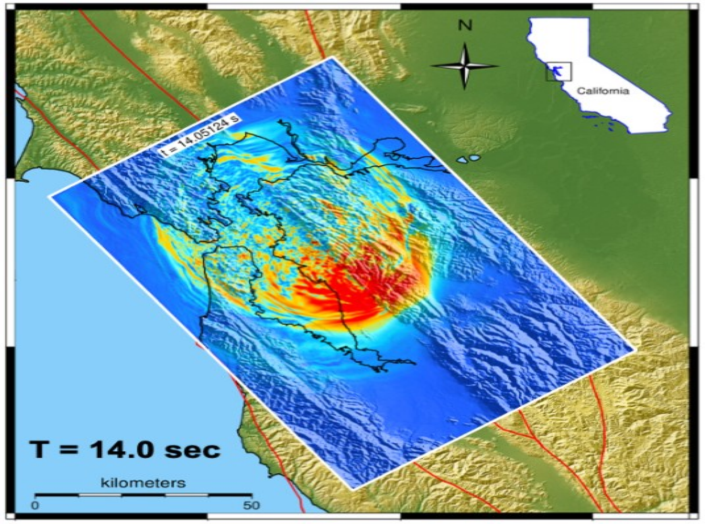
- Rapid growth of mine multi-source seismic data

Seismic network construction





Development of high-performance computing



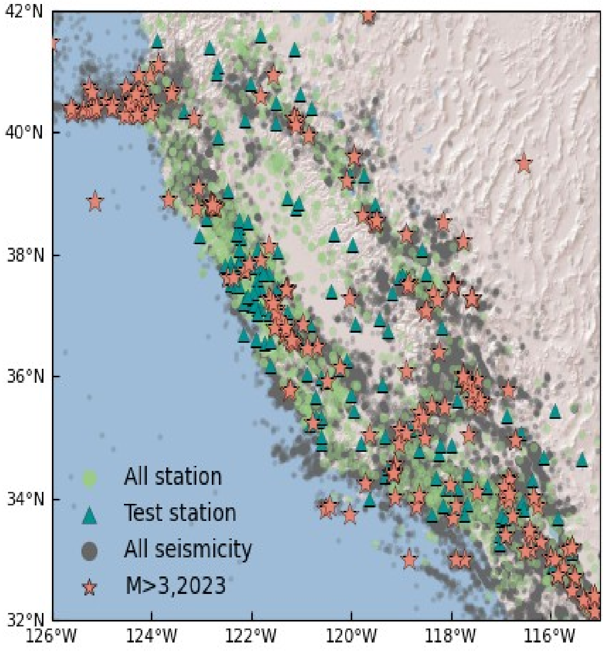
CyberShake
Shakeout
(EQSIM)

1.5 Data-driven research paradigm

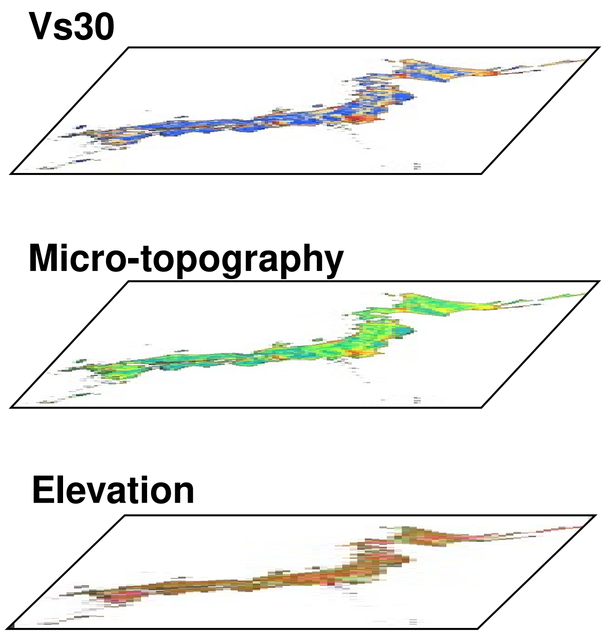
- Develop intelligent methods for predicting ground motions

Multimodal data

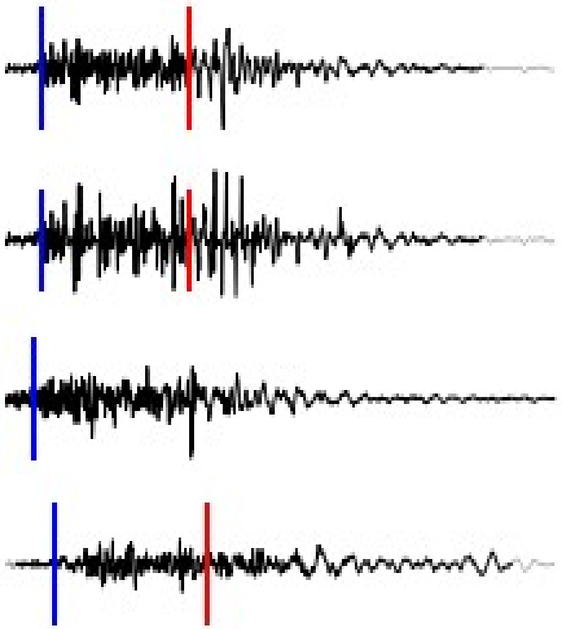
Graph of sources and station



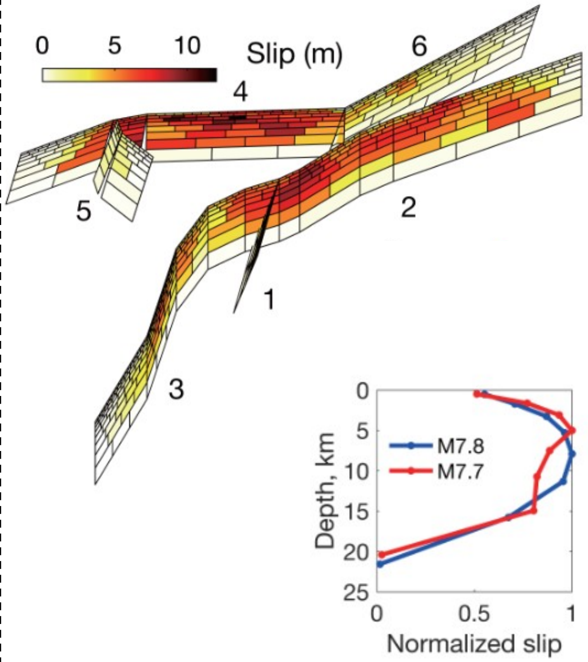
Raster data of site and topography



Waveform data



Unstructured fault data



Outline

1 Background

2 **Unified AI framework for ground motion prediction**

3 Ground motion prediction considering fault segments

4 Conclusion

2.1 Unified AI framework for ground motion prediction

Conventional AI-based ground motion models: **3 tasks**, **3 separate models**

Different information visibility

P(unobserved | observed)

Pre-event	During event	Post-event
Intensity prediction	Early warning	interpolation

Observed

Pre-event	During event	Post-event
• Site • Source	• Site • Fast arrived waves	• Site • Source • Complete waves

Task correlations

Intensity prediction

Initial prediction

Interpolation & assessment

Knowledge transfer

Source-based EEW

Propagation-based EEW

Early warning (EEW)

How to combine different tasks?

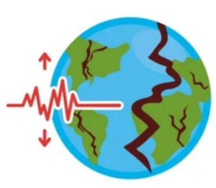
2.1 Unified AI framework for ground motion prediction

Mask Enhances Training

ChatGPT

LLM **Limited Textual Information**

I am eating an ice cream	I am eating an ice cream	I am eating an ice cream	I am eating an ice cream
LLM	LLM	LLM	LLM
predict I	predict eating	predict ice	predict cream

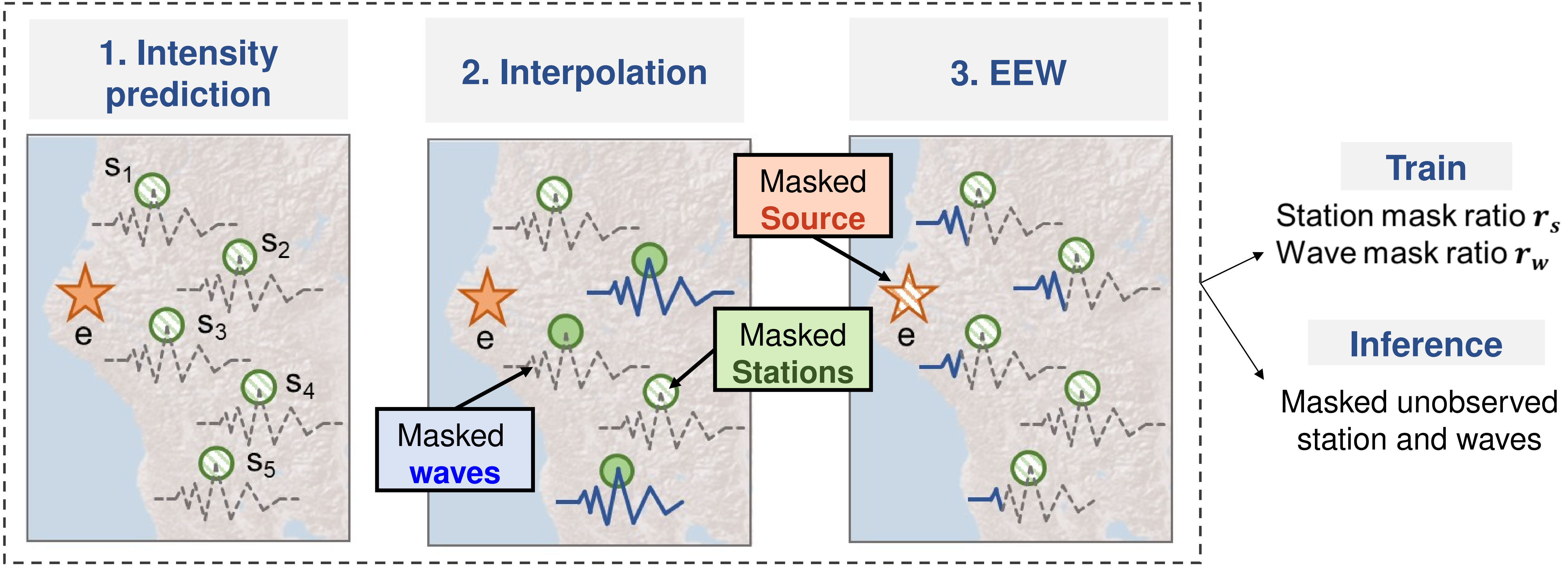


Seismic **Limited Record Information**

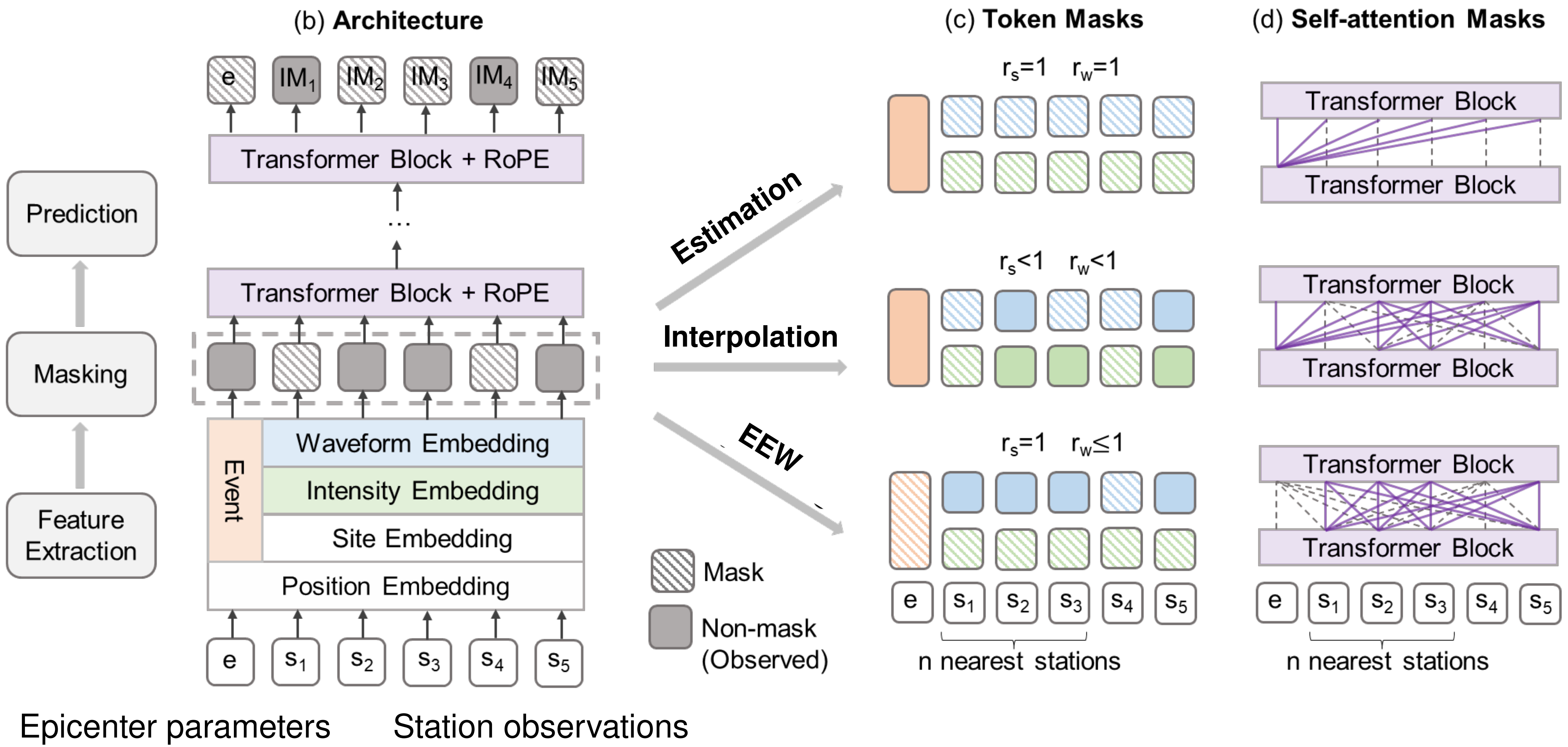
Event	Site	Wave
Event	Site	Wave
Event	Site	Wave

2.2 Unified AI framework: QuakeFormer

- **Spatiotemporal masking mechanism**: task objective is to predict the masked PGA distribution from the unmasked information.



2.2 Unified AI framework: QuakeFormer



A Single Model for Multiple Tasks: Maximizing Data Utility

2.3 Multi-scale absolute position embedding

Ergodic Model vs Non-ergodic on Model

Ergodic Model

Assumes ground motion statistics are same across the world

Non-ergodic Model

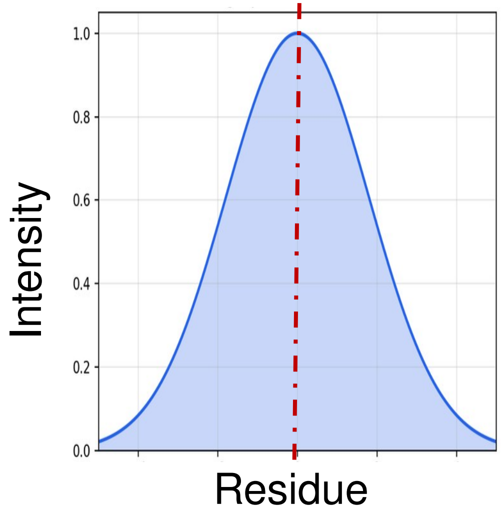
Allow ground motion statistics varied by locations

$$\ln IM = f_{median} + f_{source}(\boldsymbol{x}_e) + f_{path}(\boldsymbol{x}_e, \boldsymbol{x}_s)$$

epicenter coordinate

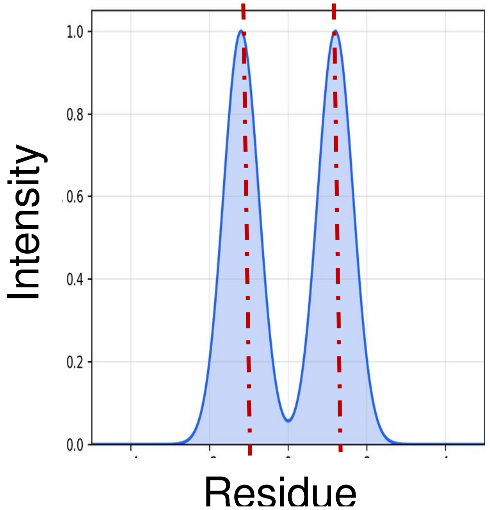
station coordinate

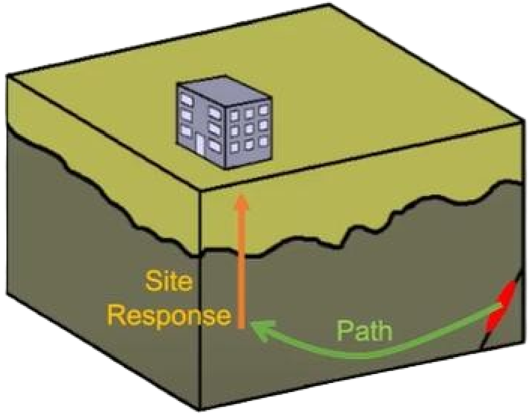
All world



Southern California

Northern California





Source

Non-ergodic model captures location-specific **effects**

2.3 Multi-scale absolute position embedding

Multi-scale absolute position embedding

2-D coordinate

Lat, lon

Neural Network

P

Ground Truth

w/o position embedding
Loss high-frequency spatial features

w position embedding
Clear high-frequency spatial features

2-D coordinate

Lat, lon

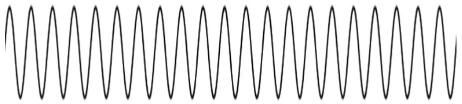
2N-D

Multi-scale position embedding
$\{\cos(2^{L_i}x), \sin(2^{L_i}x)\}$
$i = 0, \ldots, N$

Neural Network

P

$\sin 2^{L_1}\mathbf{x}$  Large-scale feature (North/South California)

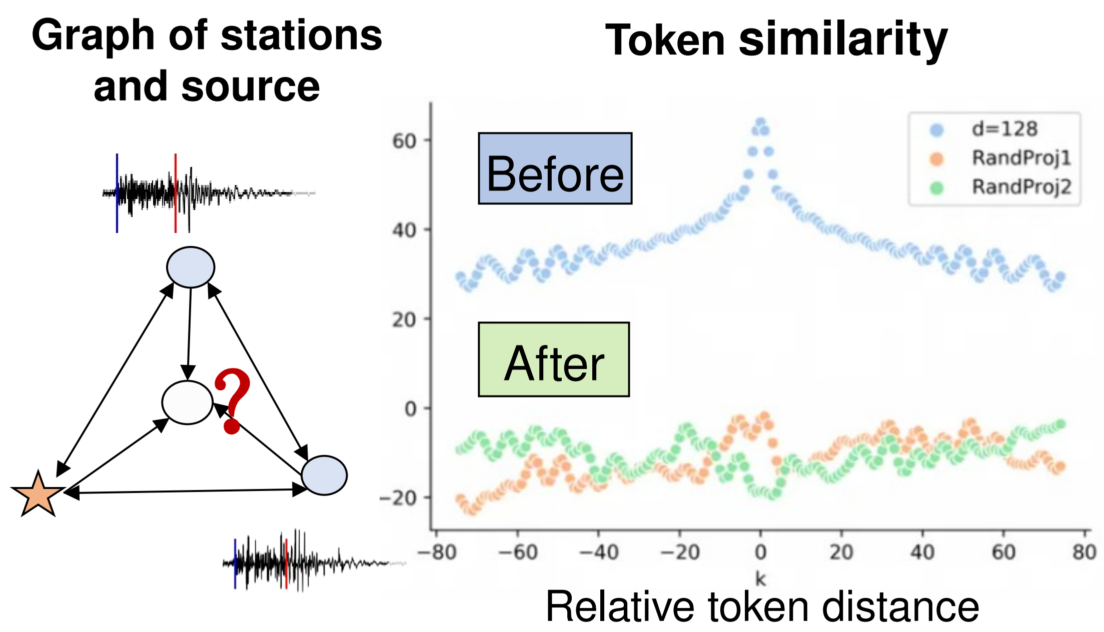
$\sin 2^{L_N}\mathbf{x}$  Small-scale feature (basin, river, rock…)

Multi-scale positional encoding captures high-frequency spatial features

2.4 RoPE Relative position embedding

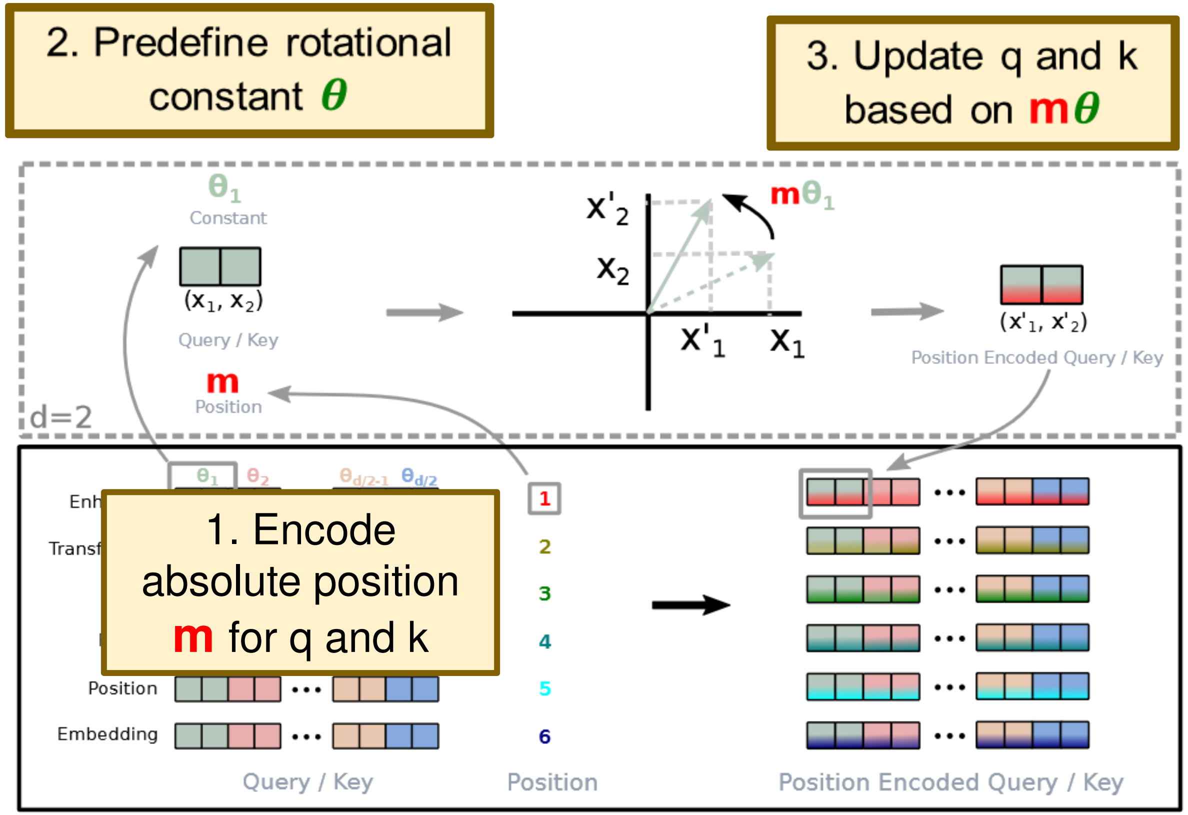
- Relative position embedding: Rotary Positional Embedding (RoPE)

Relative position relationship



After the attention, the correlation with relative position is lost

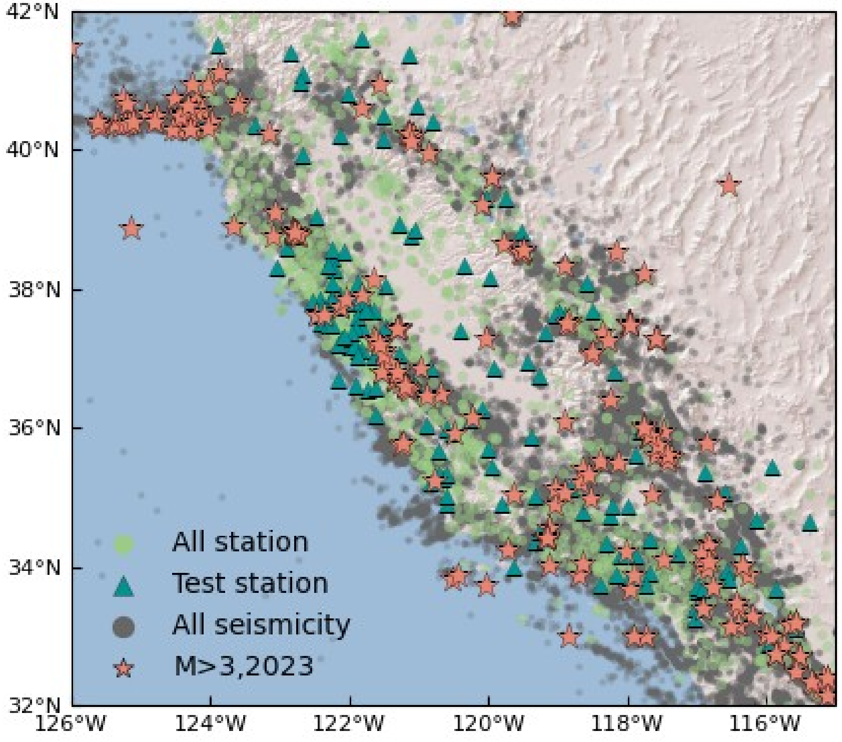
2D location RoPE



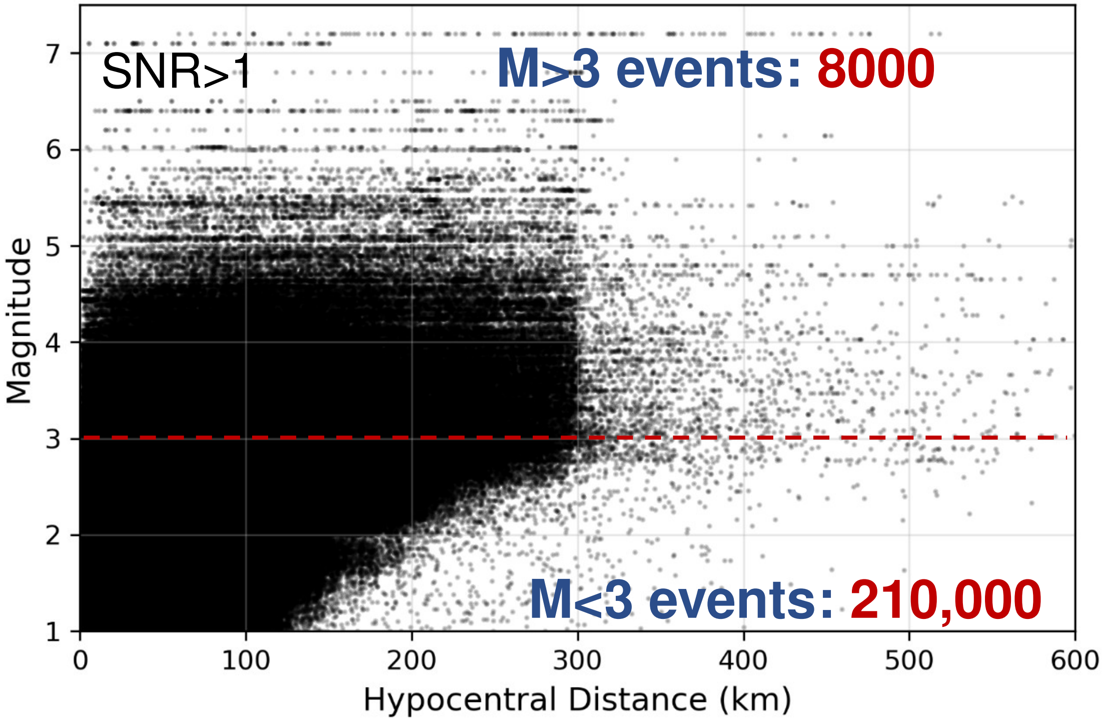
2.5 Training and Validation

- **220,000** earthquake events and **4 million** ground motion records from Northern California (1987-2023) and Southern California (2008-2023)
- Earthquakes range from M1 to M7

Map of events and stations



Magnitude-distance distribution



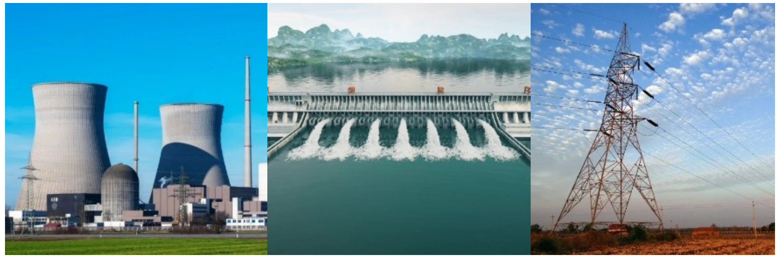
2.5 Training and Validation

- Dataset Separation
  - **Events:** Train (TM), validation and test (TR) by occurrence time
  - **Station:** Train (SM), validation and test (SR) by location

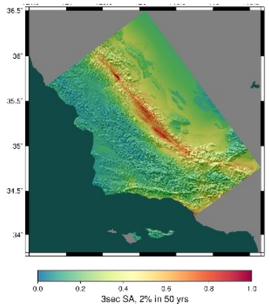
	T_M Before 2023	T_R After 2023
S_M Has historical data	Model development $T_M + S_M$	On-site test $T_R + S_M$
S_R No data		Regional test $T_R + S_R$

Application

On-site hazard and risk analysis for important infrastructure

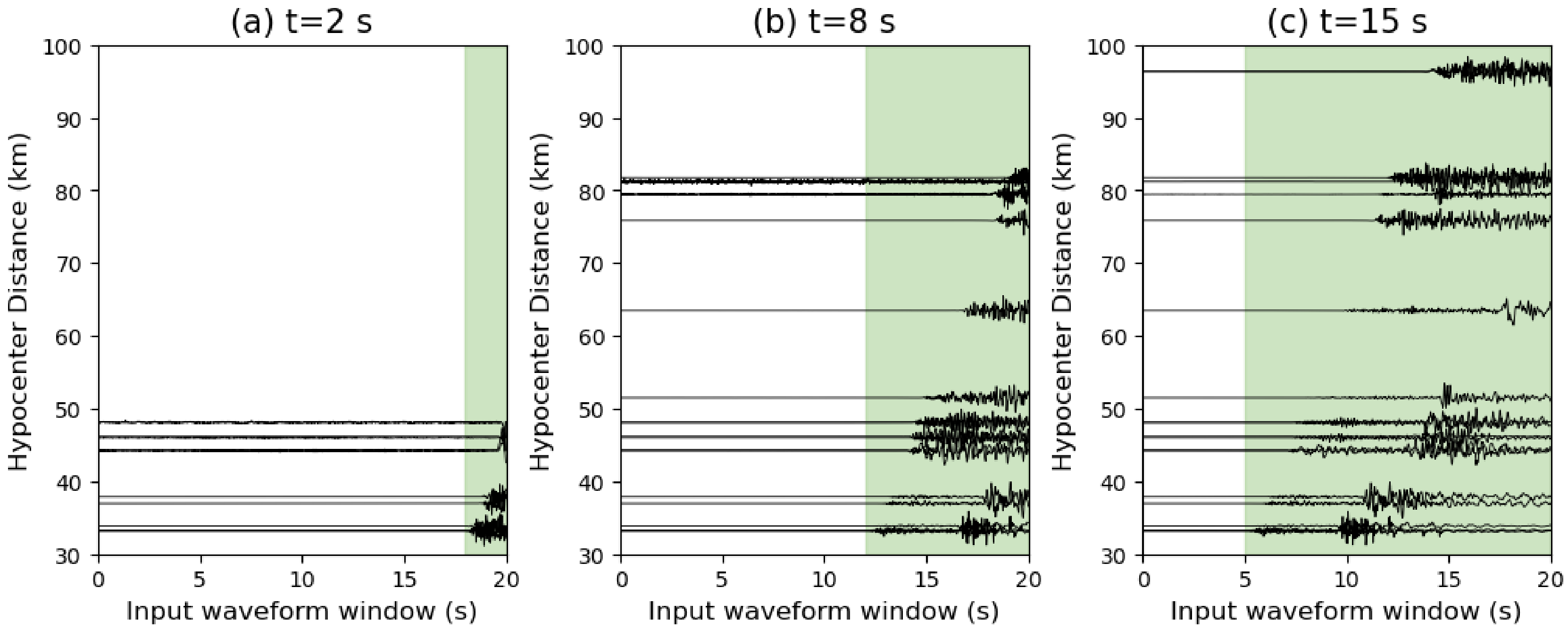


Regional hazard and risk analysis for cities



2.5 Training and Validation

- Training EEW poses a risk of data leakage: the model may infer magnitude from the number of input stations. Therefore, stations are first triggered and then used for prediction.

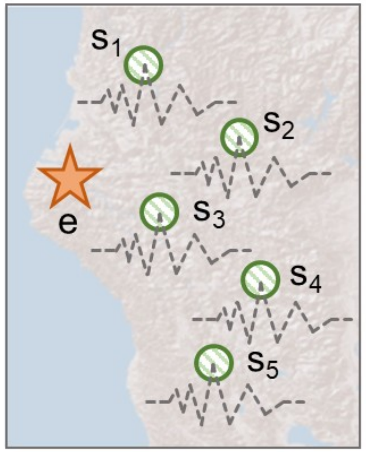


Time after first station trigger

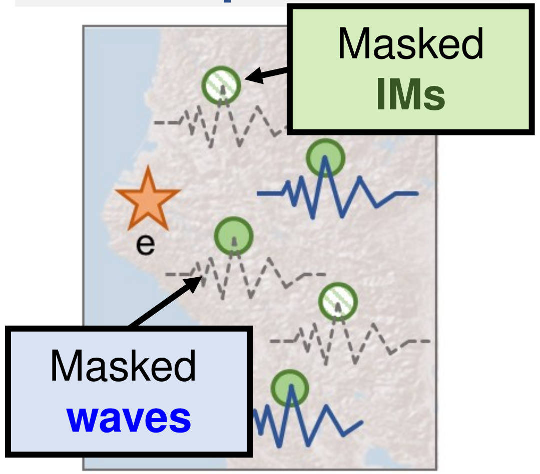
2.6 Results: intensity prediction and interpolation

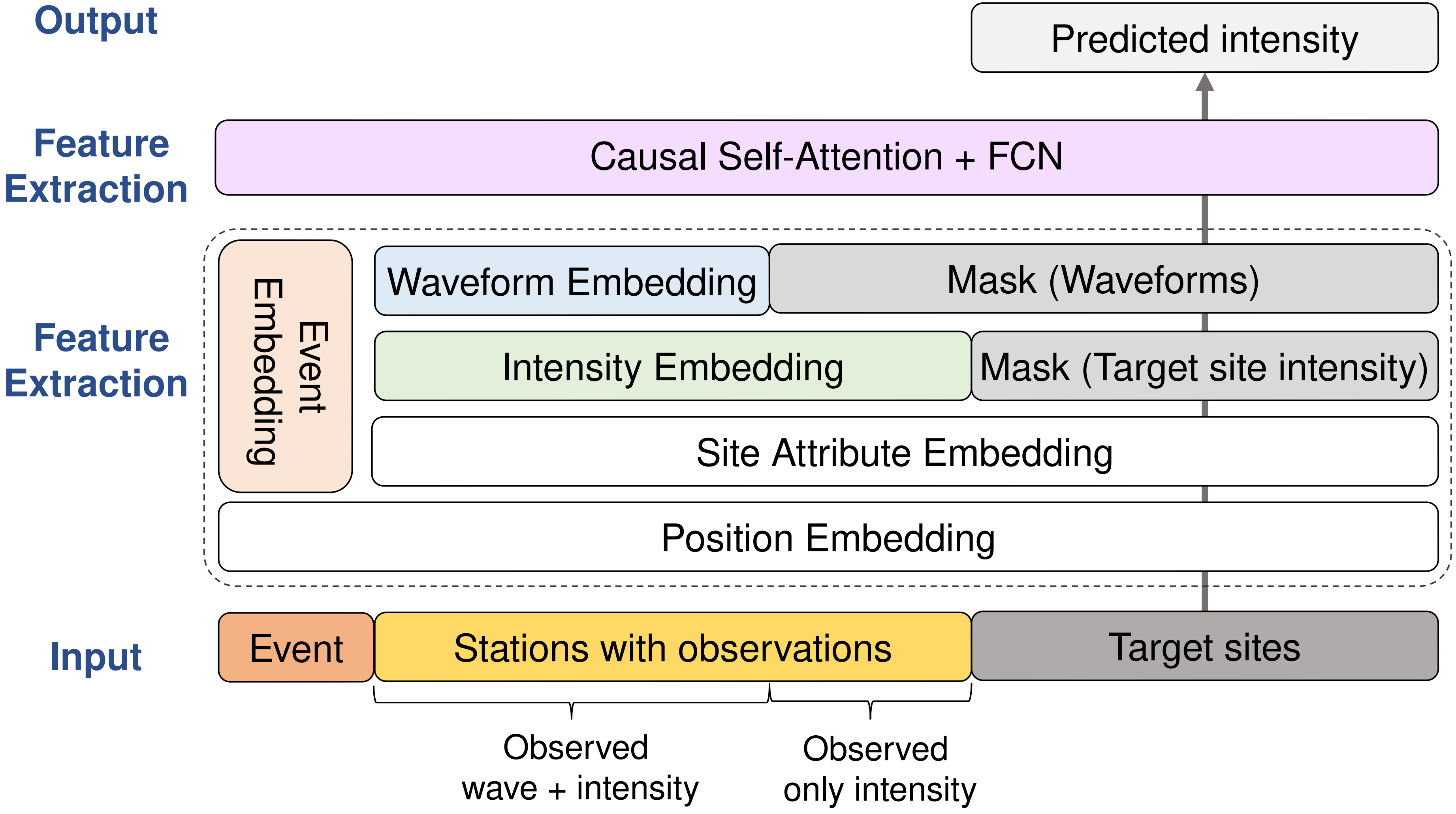
- Inference: Replace unknown intensity and waveform with MASK

Intensity prediction



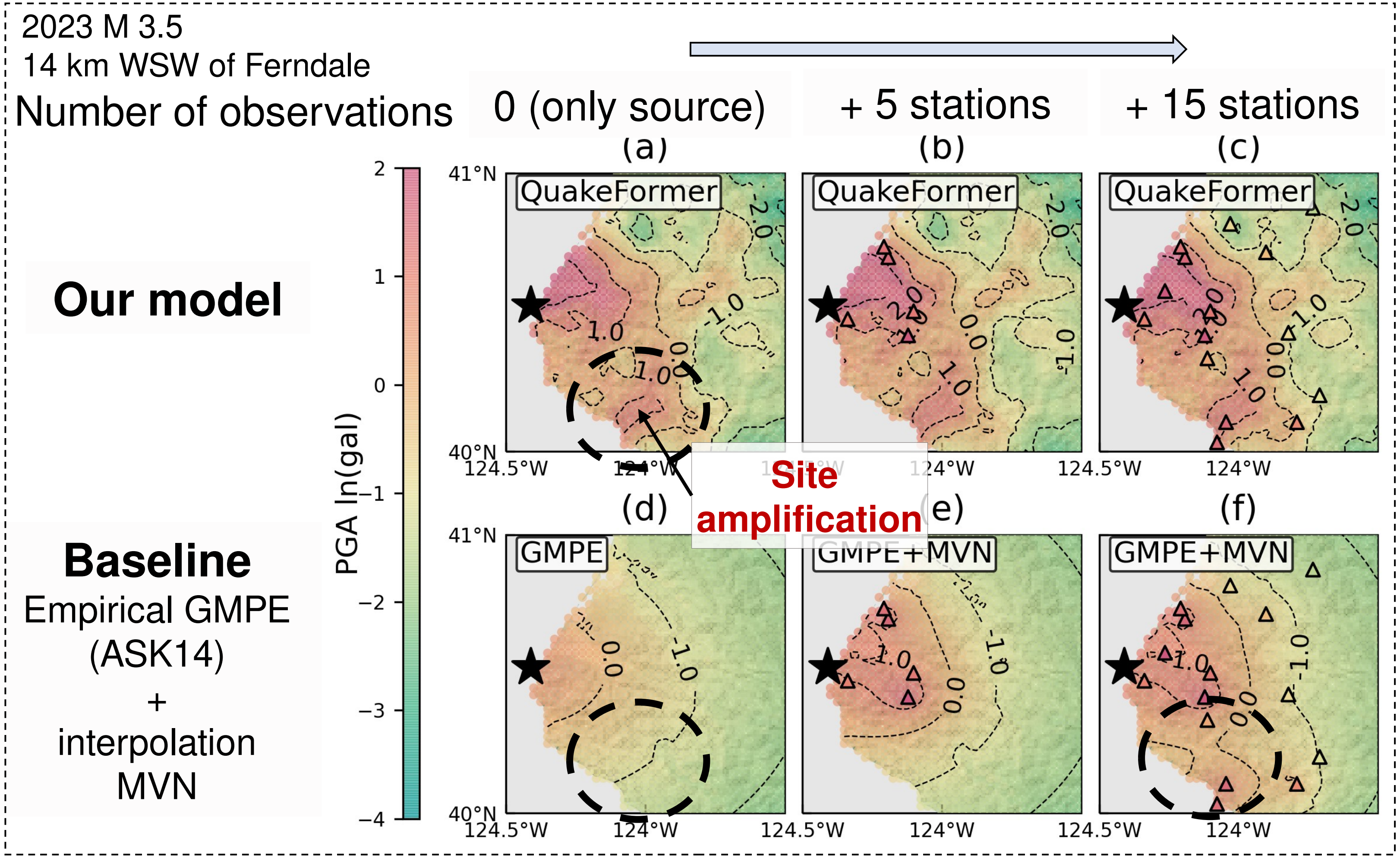
Interpolation

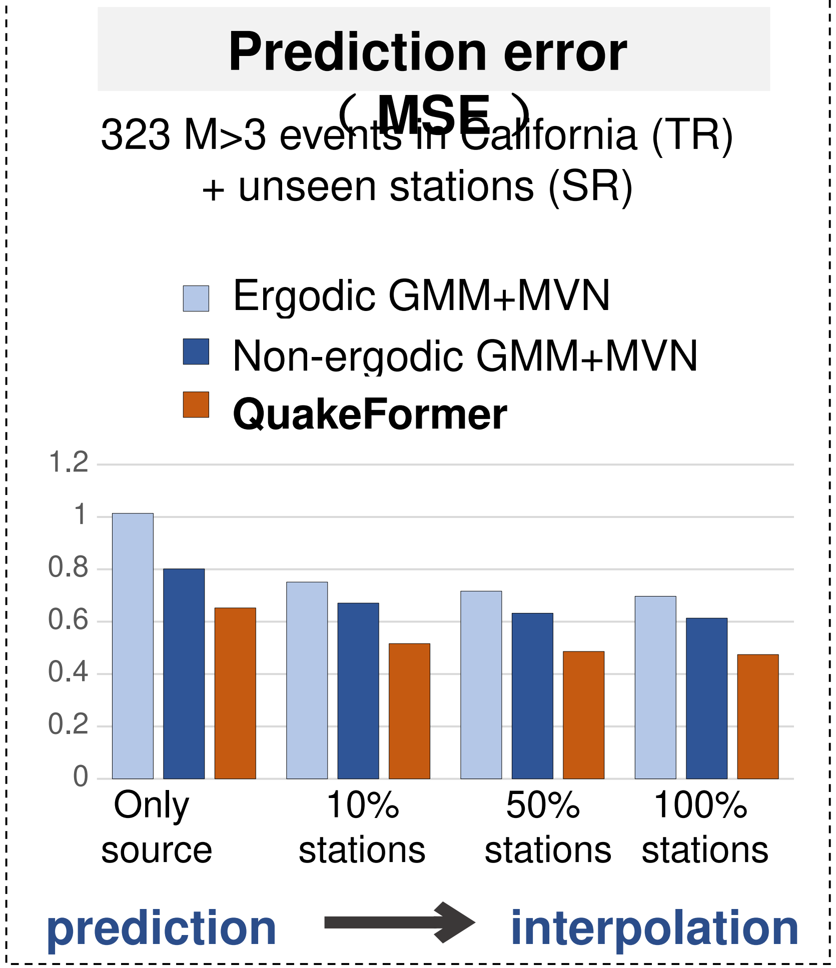




2.6 Results: intensity prediction and interpolation

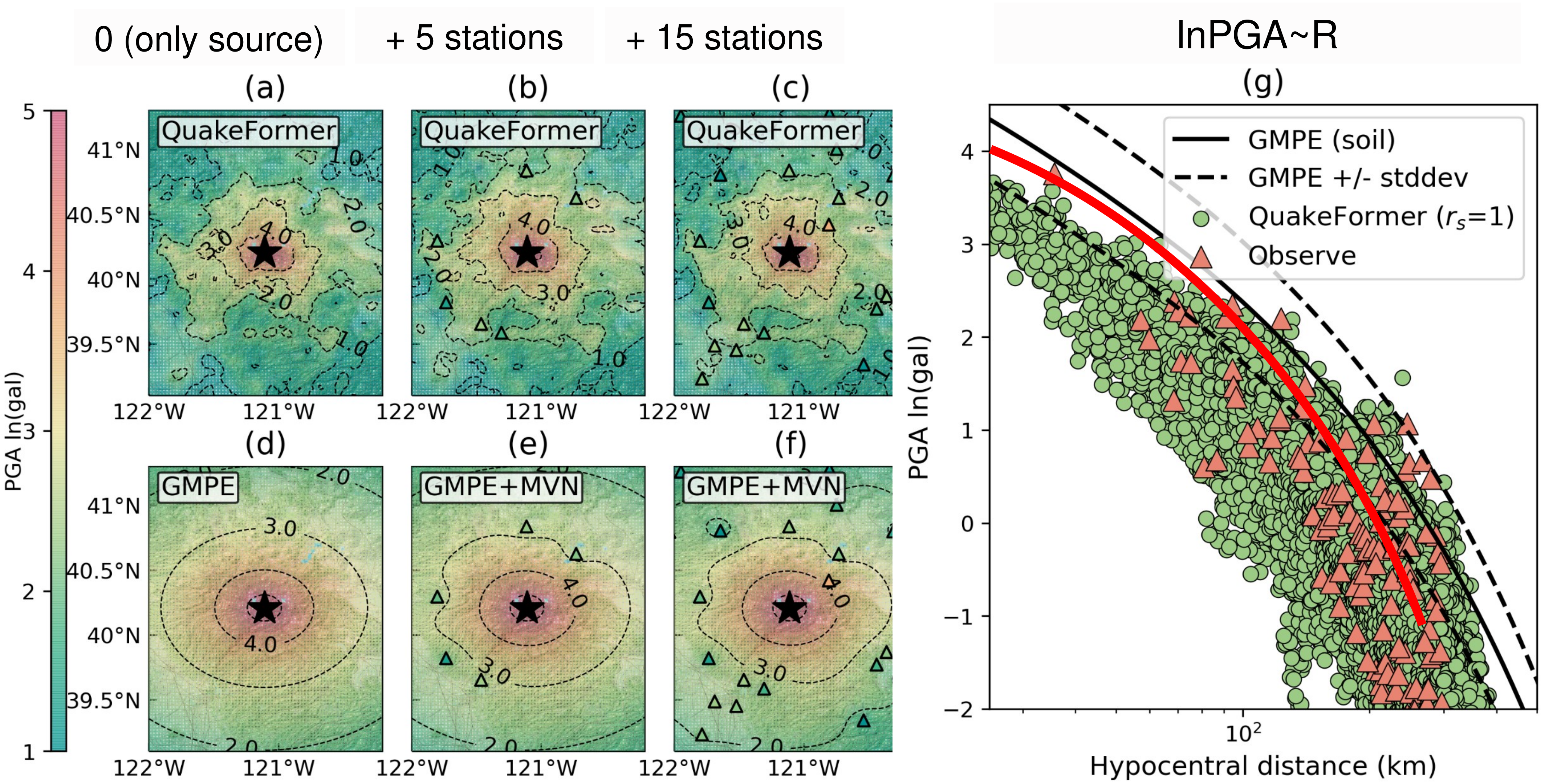
- As more station observations become available, the task naturally transitions from intensity prediction to interpolation.





2.6 Results: intensity prediction and interpolation

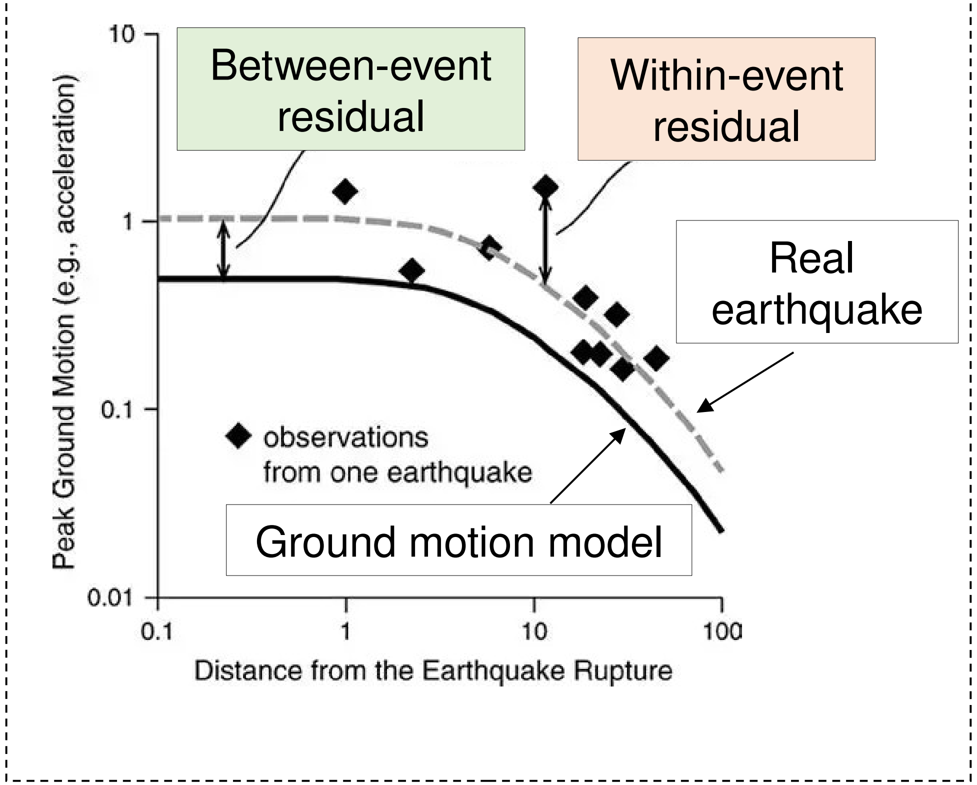
2023 M 5.5 - Lake Almanor Earthquake

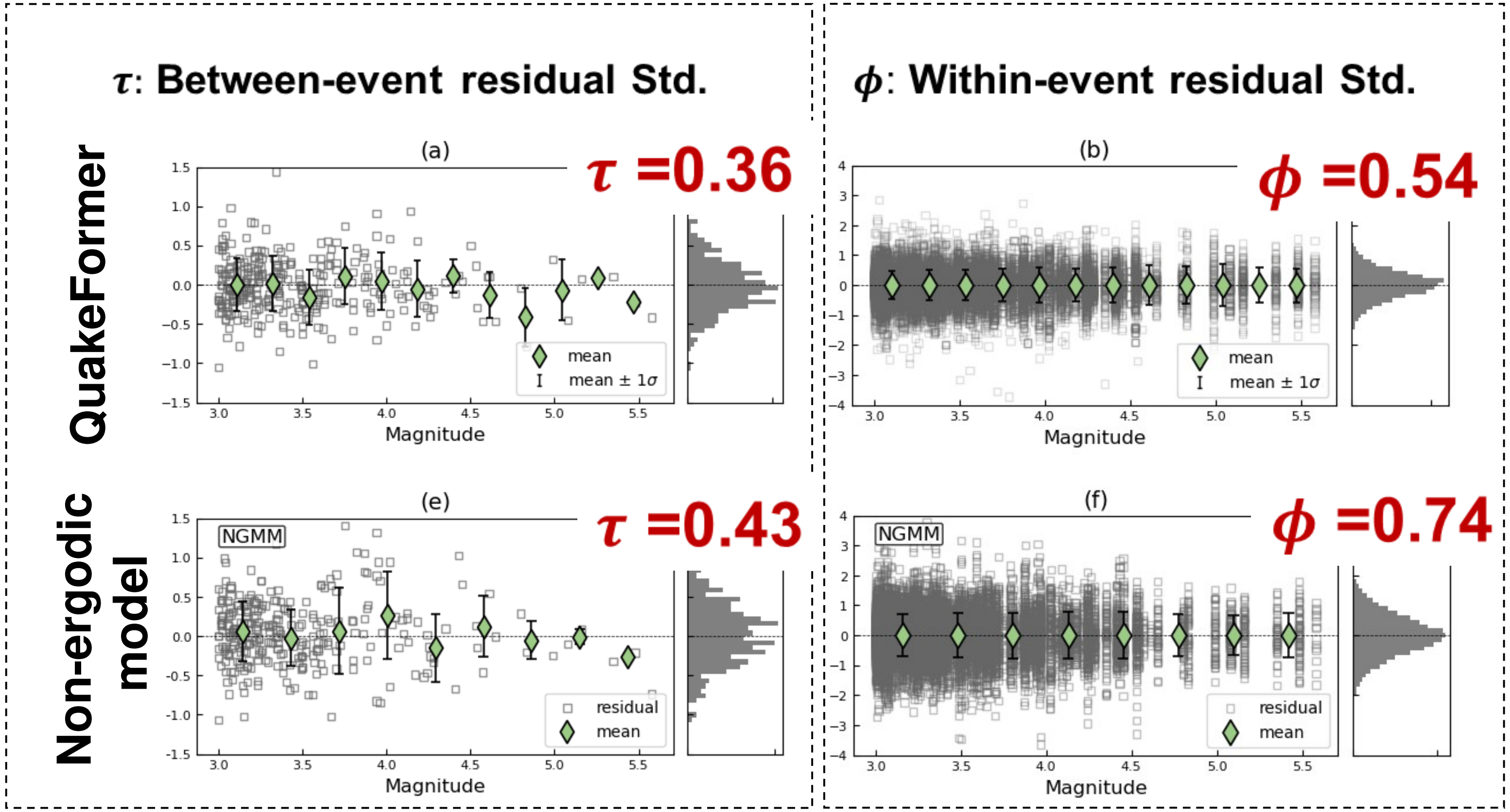


2.6 Results: intensity prediction and interpolation

- Separate the prediction residual
- QuakeFormer has lower between-event residual and within-event residual variation

Residual illustration



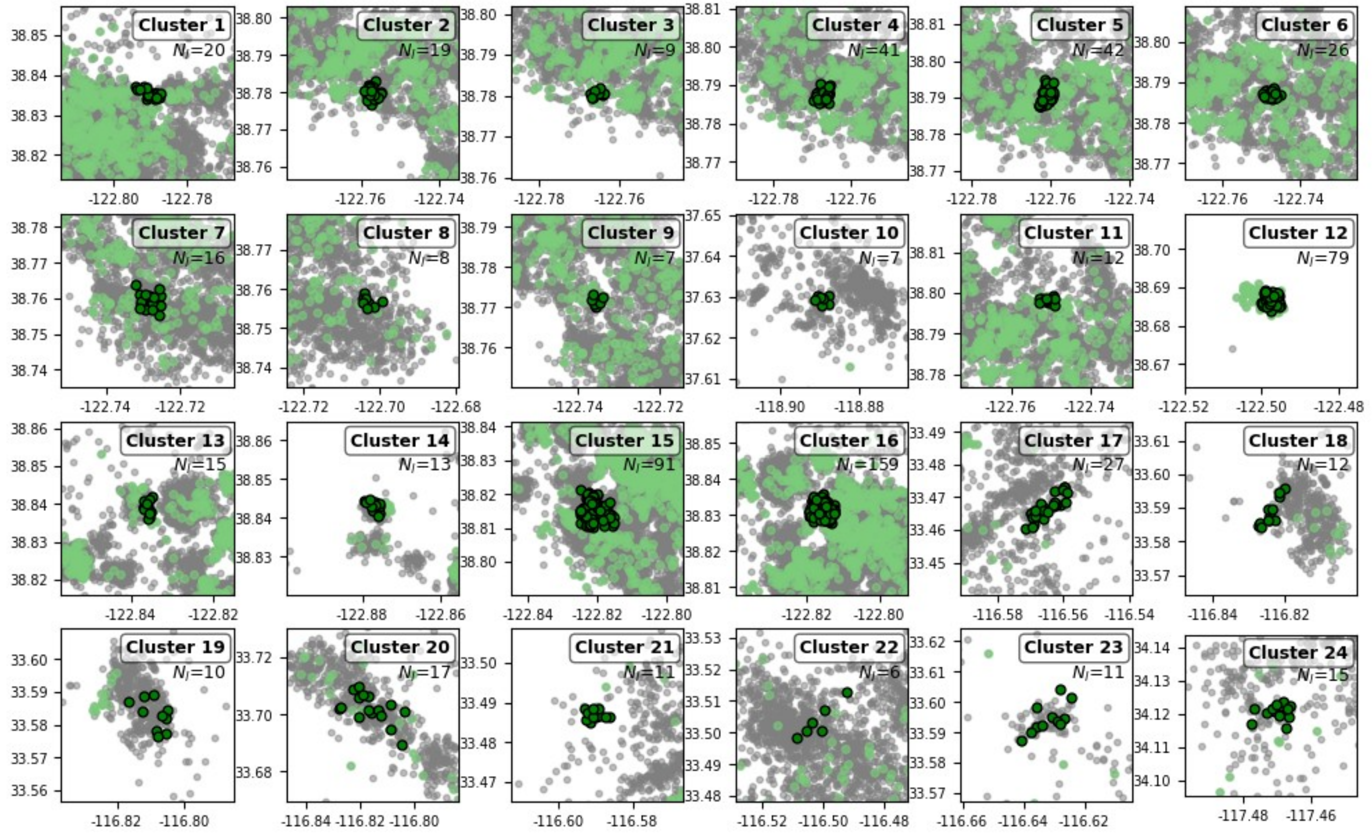


Norman A. Abrahamson, et al. 2014

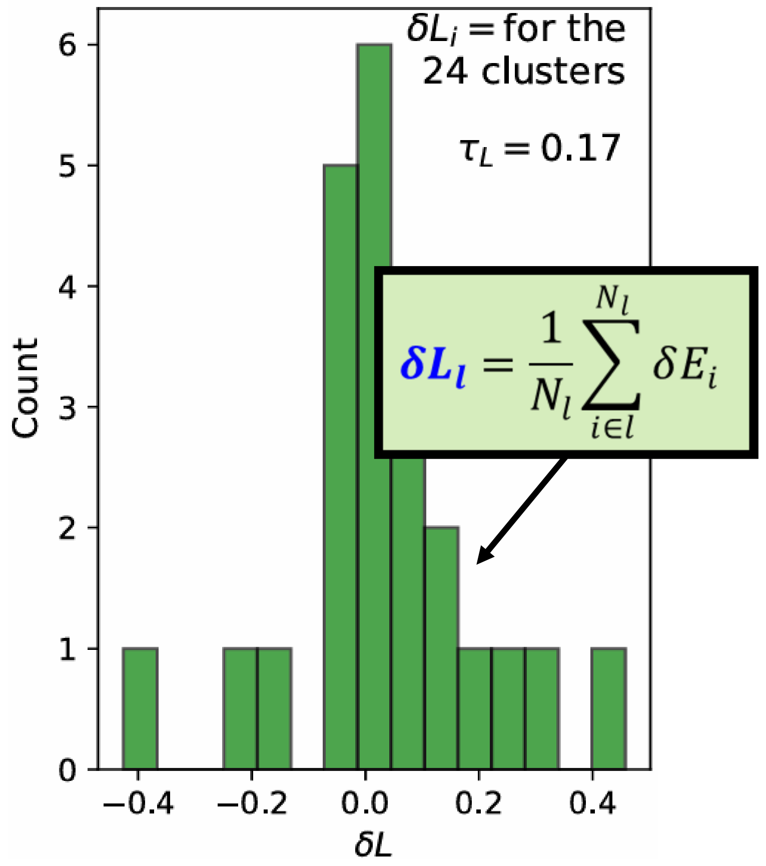
2.6 Results: intensity prediction and interpolation

- Location-based between-event residual (event-term) bias δL_l

Separate 24 earthquake clusters in 2023



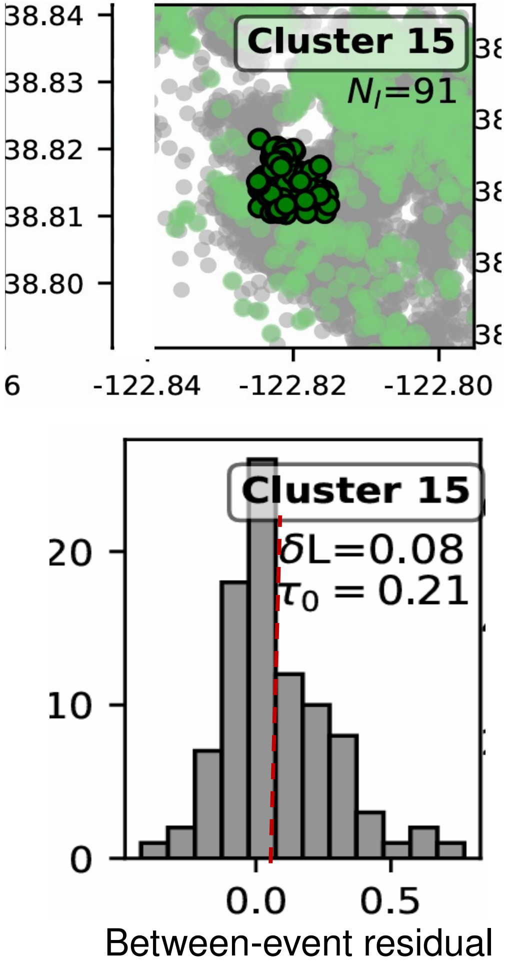
Event term bias hist (δL_l)



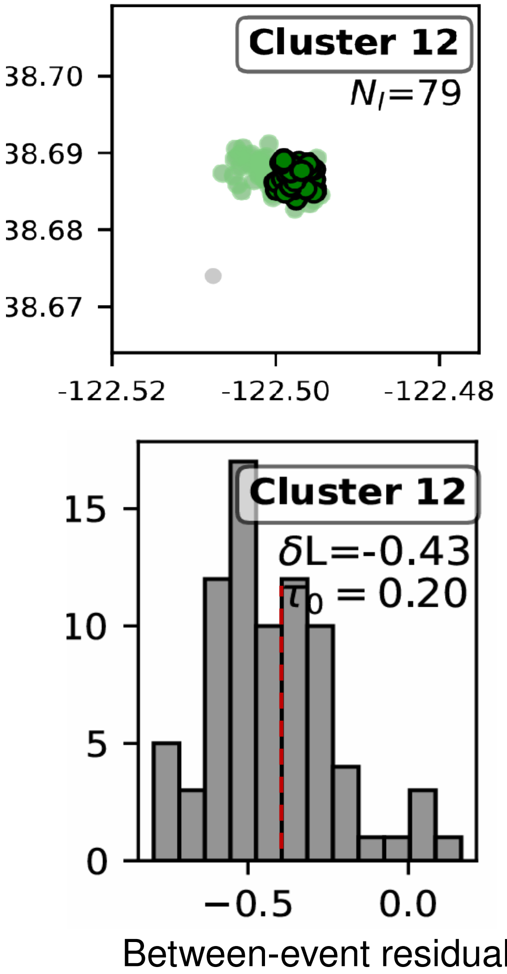
2.6 Results: intensity prediction and interpolation

- Historical events and station observations reduce δL_l (event term bias)

Location with historical events $\delta L_l \approx 0$



Location with historical events large δL_l (−0.43)



2023 events

Event cluster

Historical event

δL_l **:** Mean bias

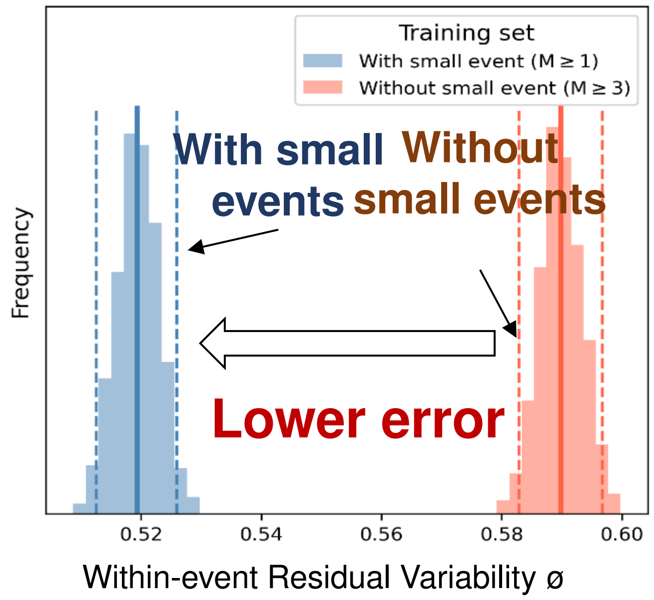
τ_0 **:** Std.

2.6 Results: intensity prediction and interpolation

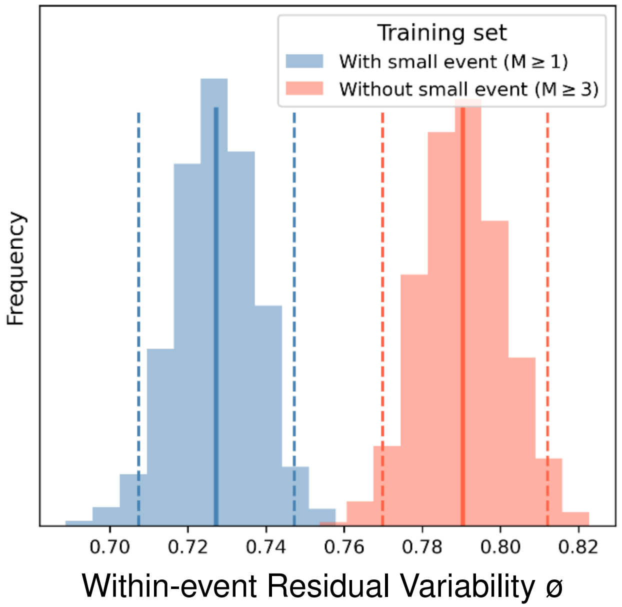
- Training with vs. without small events： M ≥ 1 vs M ≥ 3
  - **Test：** 323 M > 3 events in California (2023)
  - 1000 **bootstrap resamples** to estimate confidence intervals of PGA residual std

Within-event error (path and site uncertainty)

Seen station



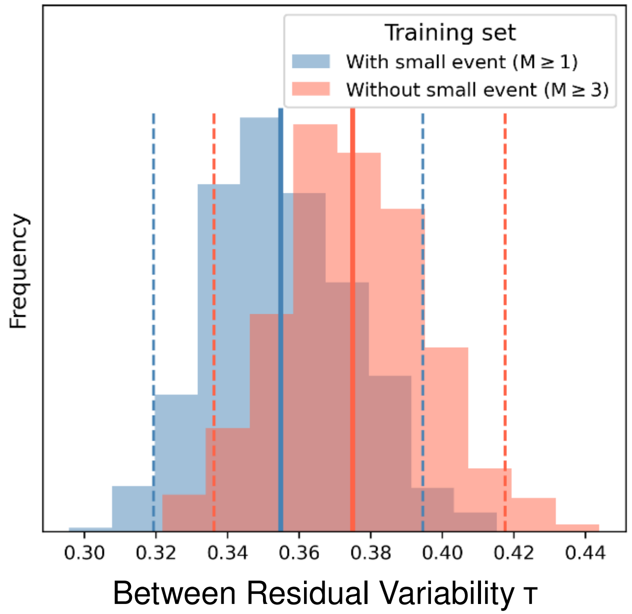
Unseen station



Small events provide dense sampling of site and path effects and therefore help **significantly** reduce within-event error.

Between-event error (source uncertainty)

(a)



Reduce between-event residual

Not significant

2.6 Results: intensity prediction and interpolation

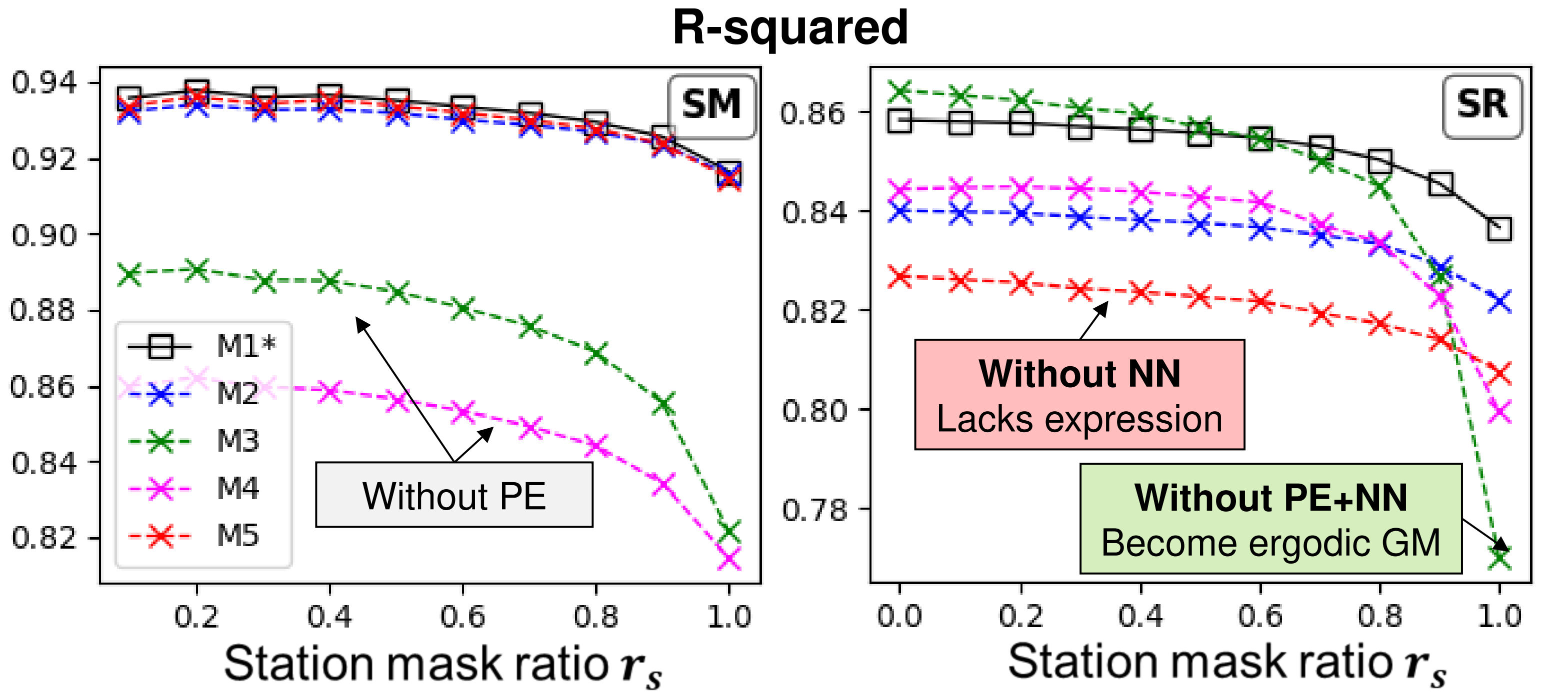
- Absolute and relative position encodings improve model performance
  - PE maps low-dimensional coordinates into a high-dimensional space, easing training and better fitting high-frequency features

Ablation experiment | **Seen station (SM)** | **Unseen station (SR)**

	M1	M2	M3	M4	M5
	✓	✓		✓	
	✓	✓			✓
RoPE	✓		✓	✓	✓

	M1	M2	M3	M4	M5
NN(·)	✓	✓		✓	
$PE(\cdot)$	✓	✓			✓
RoPE	✓		✓	✓	✓

	M1	M2	M3	M4	M5
NN(·)	✓	✓		✓	
$PE(\cdot)$	✓	✓			✓
RoPE	✓		✓	✓	✓



Interpolation → Intensity prediction

2.6 Results: intensity prediction and interpolation

- However, large frequency in absolute PE harms performance, exceeding the spatial resolution supported by station density

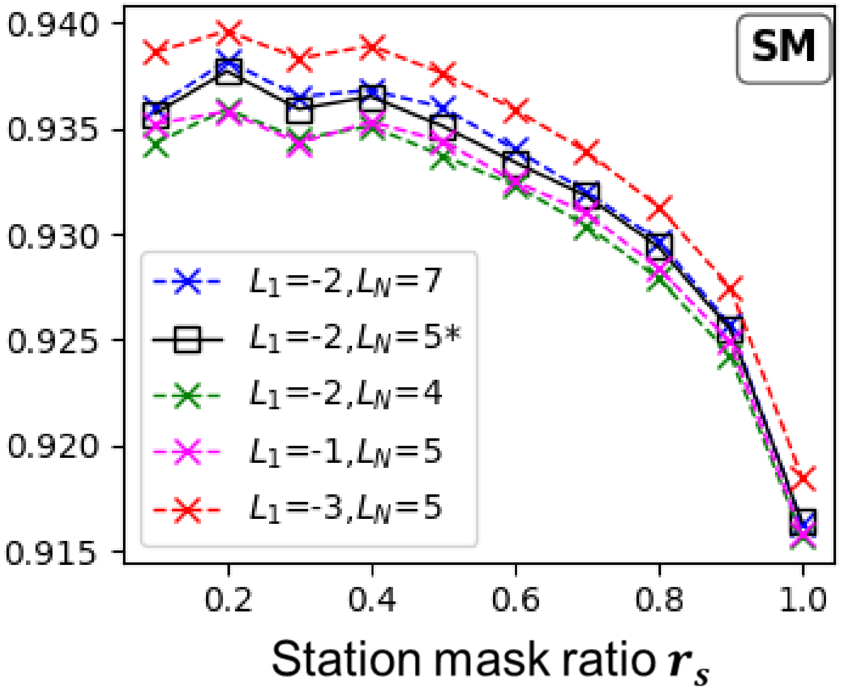
$$PE(\mathbf{x}) = (\sin 2^{L_1}\mathbf{x}, \cos 2^{L_1}\mathbf{x}, \ldots, \sin 2^{L_N}\mathbf{x}, \cos 2^{L_N}\mathbf{x})$$

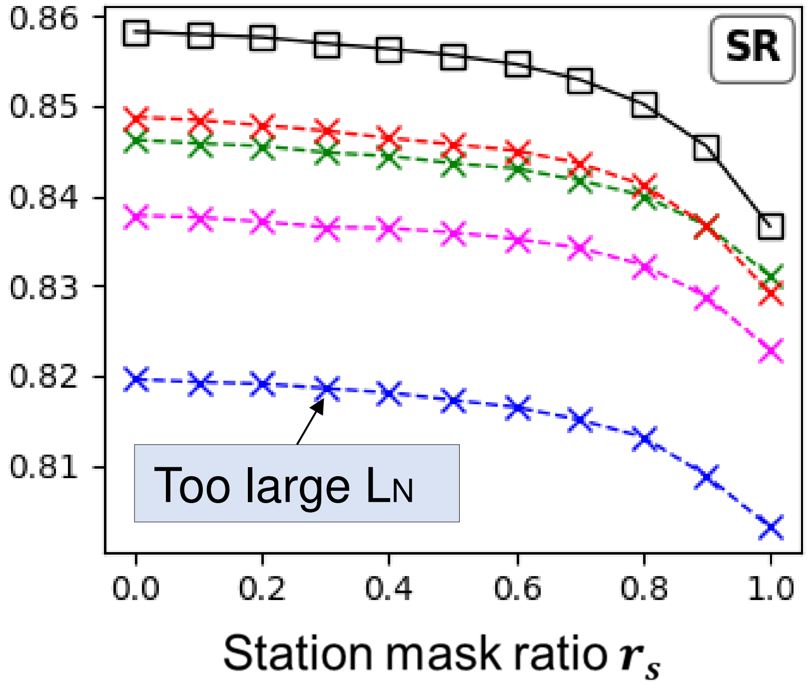
Seen station (SM)

Unseen station (SR)

Application Advice

R-squared





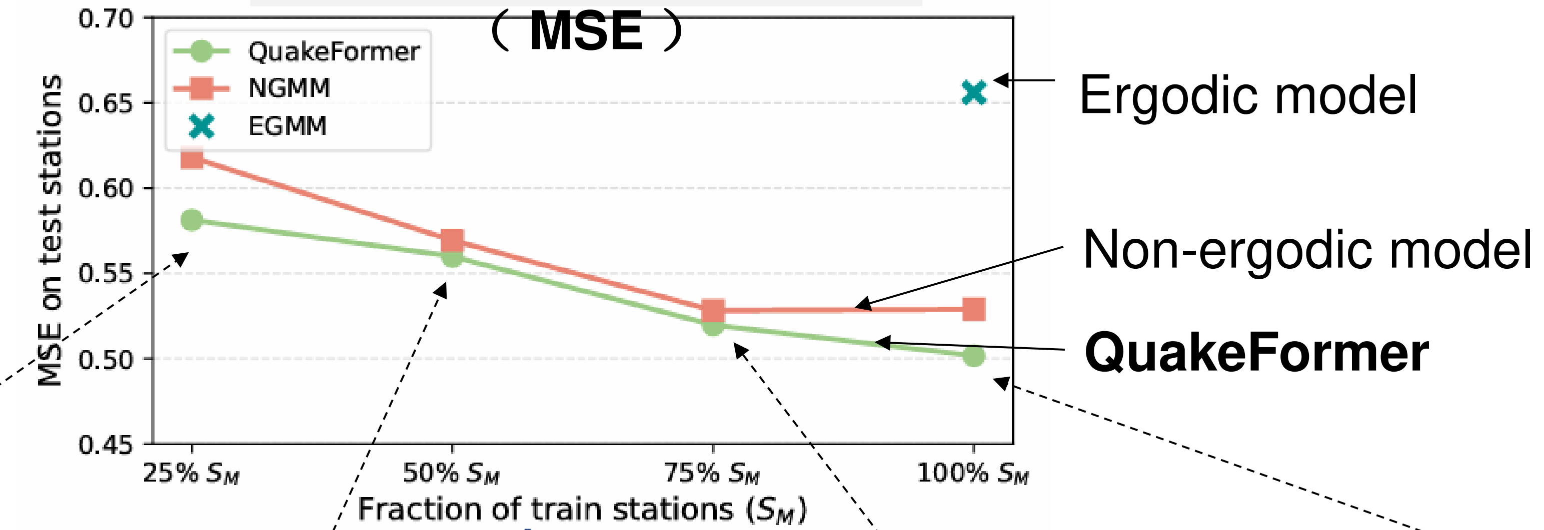
1. Increase seismic station；
2. For on-site prediction, Ln could be large；
3. For regional prediction, carefully choose Ln to balance resolution and stability.

2.6 Results: intensity prediction and interpolation

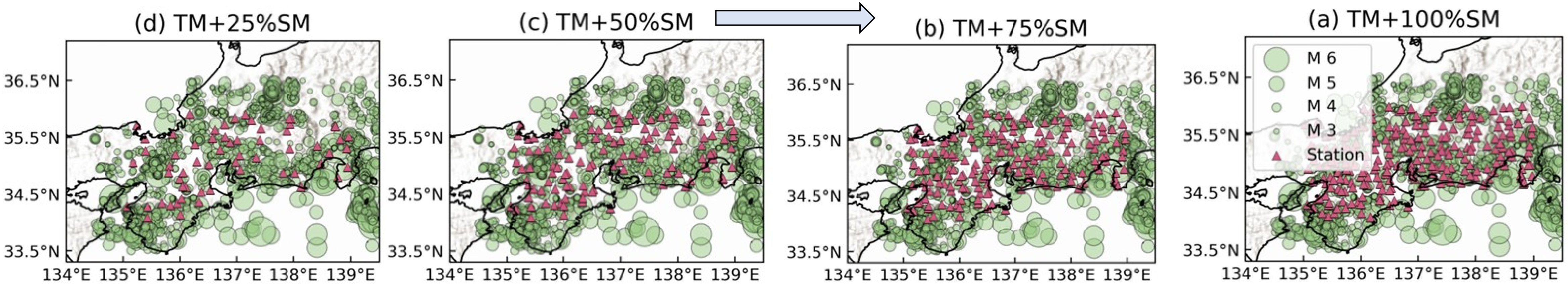
- Dense seismic network vs. sparse seismic network

Trained on Japan K-NET and KiK-NET records (1997–2022)

Prediction error （MSE）



denser



2.6 Results: intensity prediction and interpolation

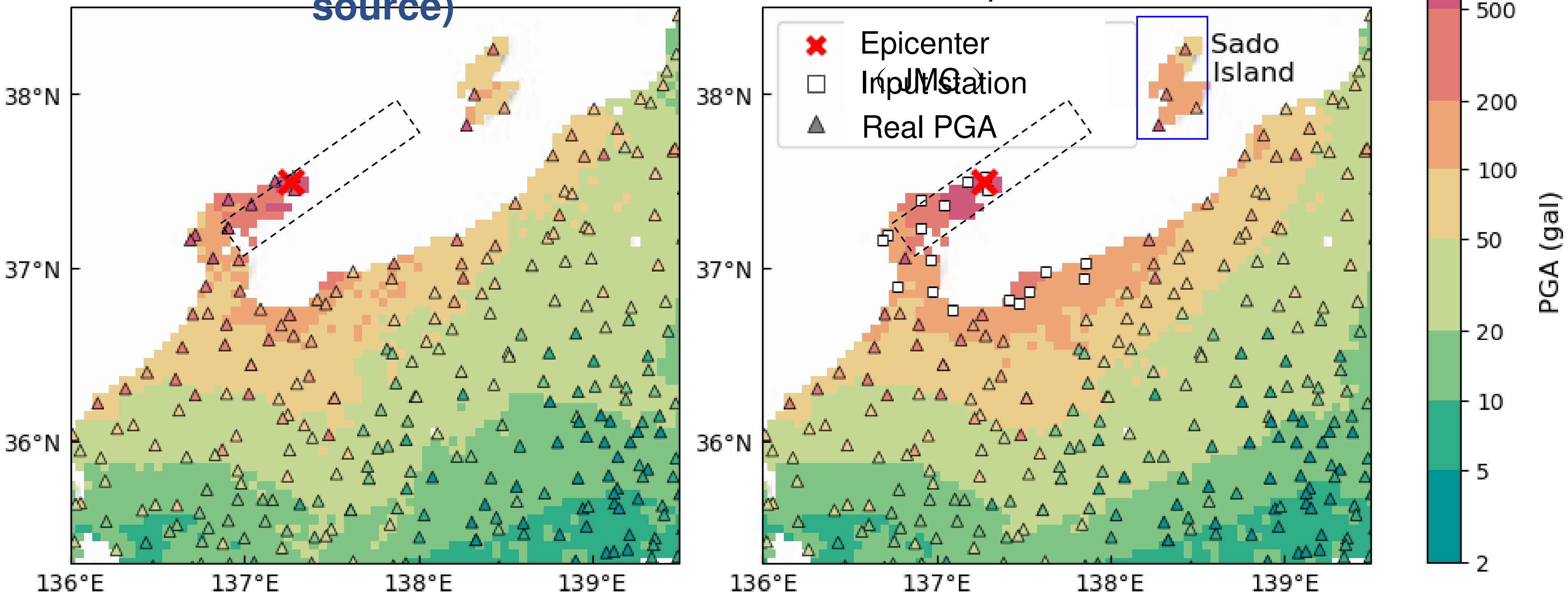
- After inputting the 10 closest stations, near-fault PGAs amplified;
- No inputs exist northeast of the source, yet the Sado Island prediction still matches observations: The model infers finite-fault effects from given observations, rather than simple interpolation.

Number of input stations

0 (only source)

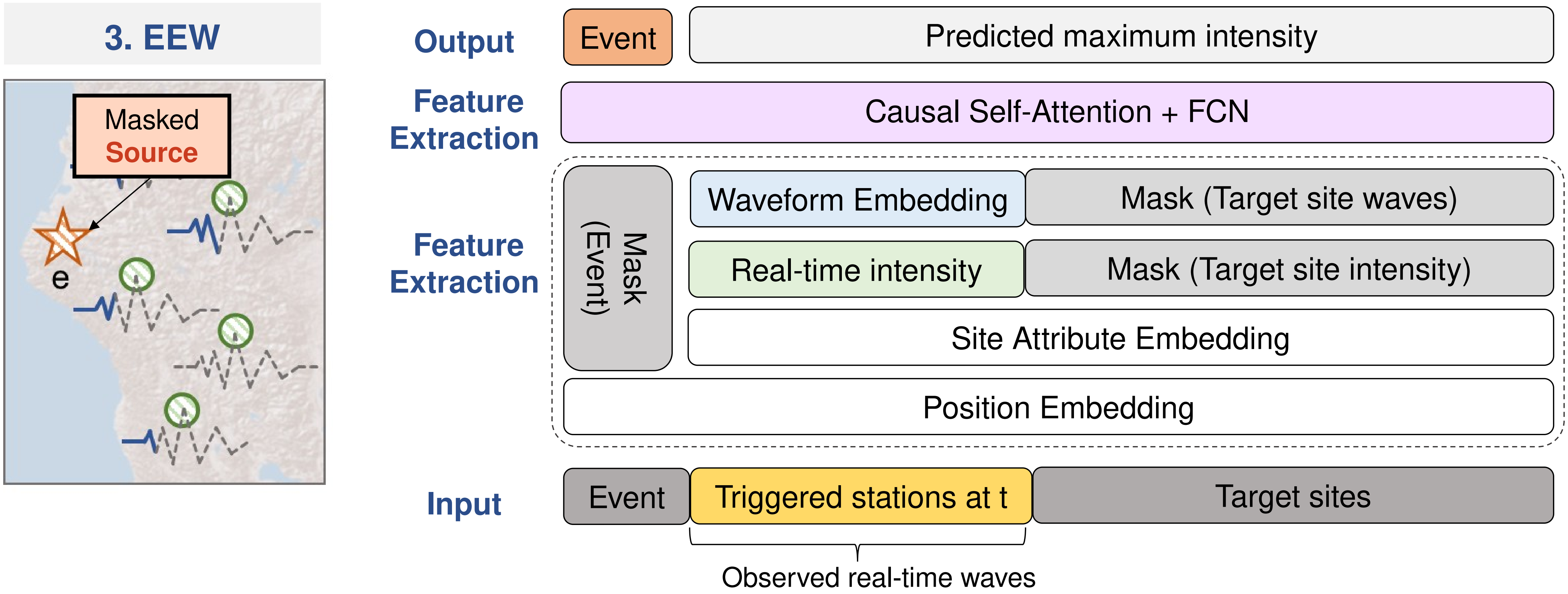
+ 10 stations

2024 Noto Peninsula Earthquake, Japan (M7.6)



2.7 Results: Earthquake early warning

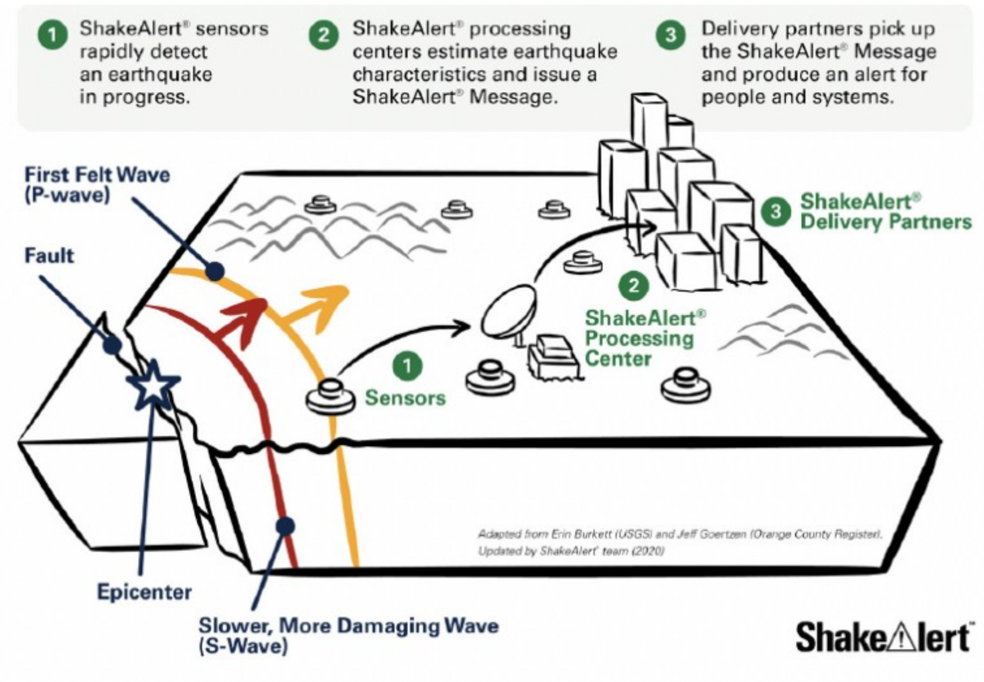
- Inference: At time **t**, the fast P-wave trigger a set of near by stations, while the event and target-site information is replaced with MASK.



2.7 Results: Earthquake early warning

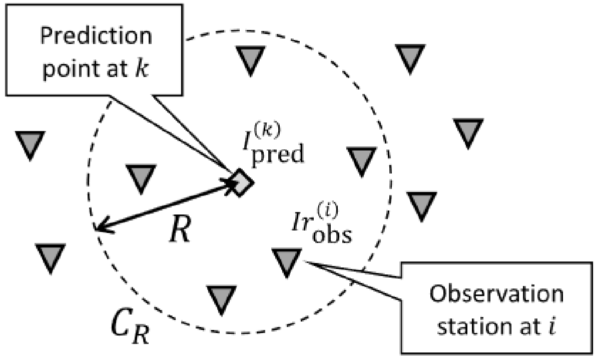
- Existing EEW system

ShakeAlert: Source-based EEW



Widely applied in California region

PLUM: Propagation-based EEW

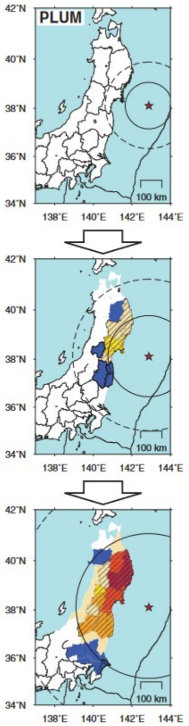


$$I^{(k)}_{\mathrm{pred}} = \max_{i \in C_R} \left\{ Ir^{(i)}_{\mathrm{obs}} - F^{(i)}_0 \right\} + F^{(k)}_0$$

$I^{(k)}_{\mathrm{pred}}$: predicted seismic intensity at k

$Ir^{(i)}_{\mathrm{obs}}$: observed real-time seismic intensity at i

$F^{(i)}_0$ and $F^{(k)}_0$: site amplification factors at i and k (converted into seismic intensity differences)



Widely applied in Japan with dense seismic networks

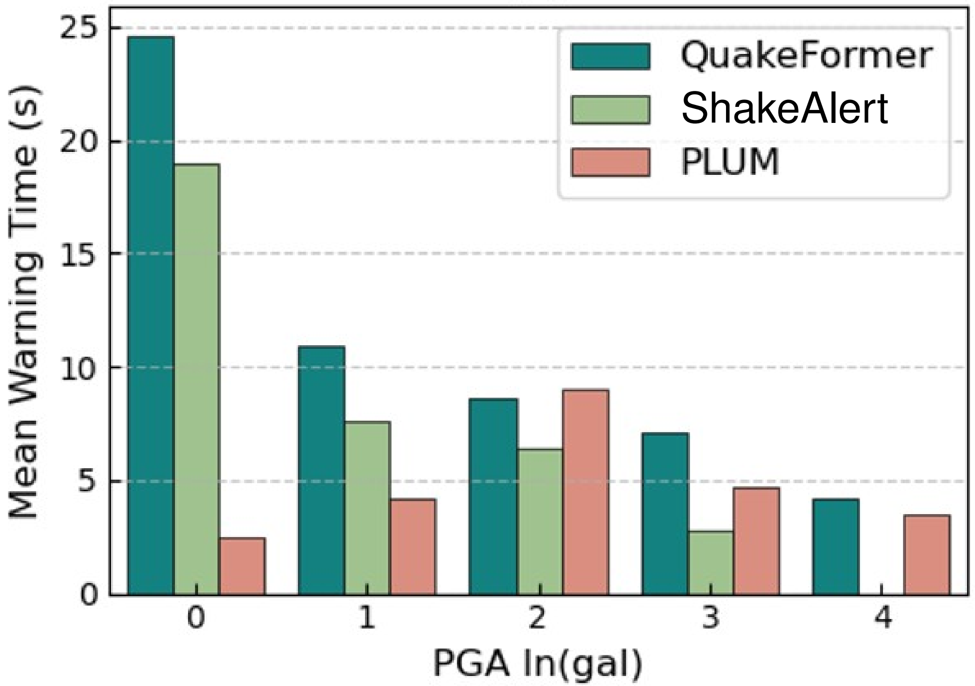
https:// earthquake.usgs.gov/scenarios/catalog

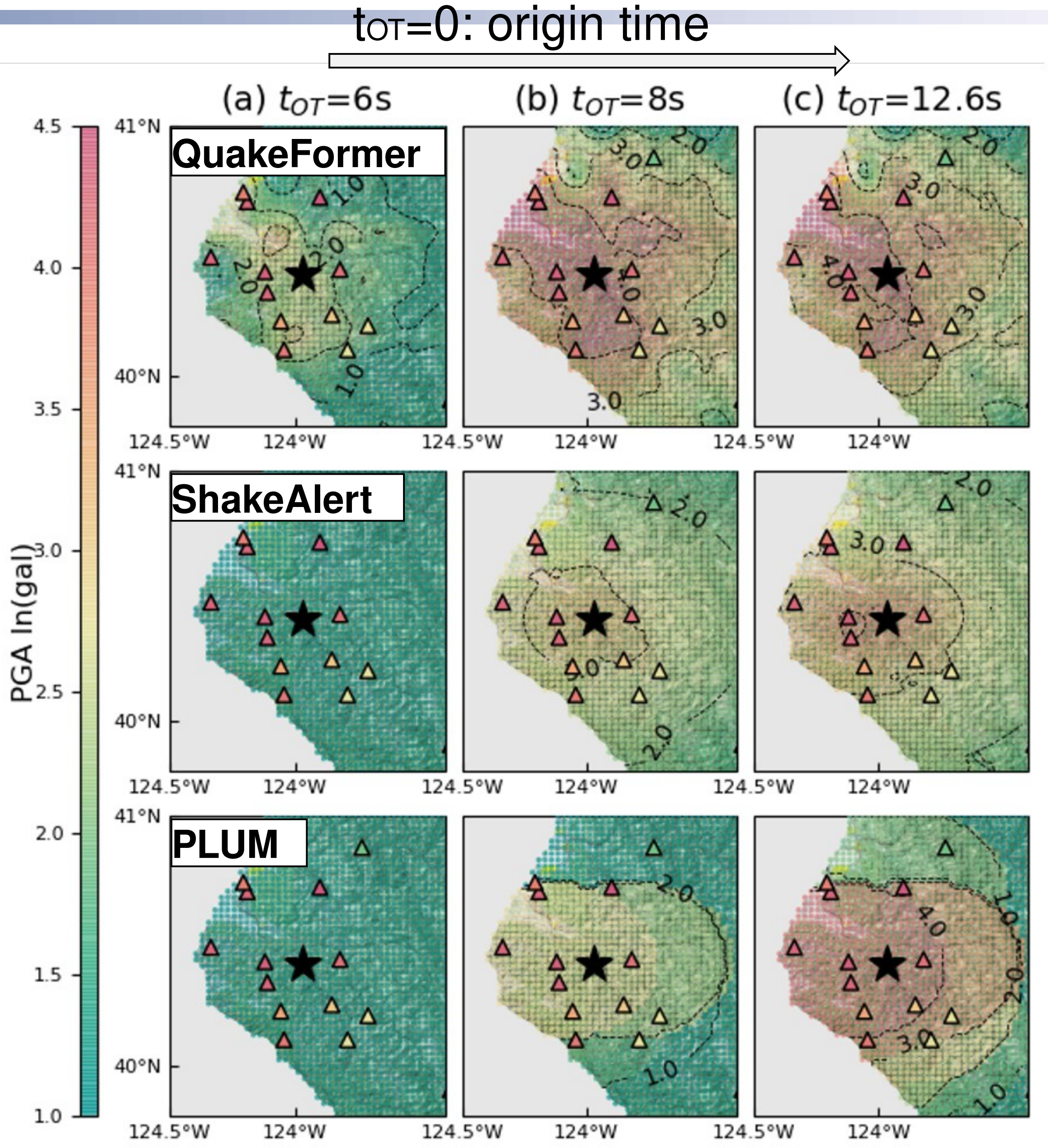
2.7 Results: Earthquake early warning

2023 M 5.4 – 15 km SE of Rio Dell, CA

- t_{OT} = 6 s， QuakeFormer first predicts the extent of shaking
- QuakeFormer provides the longest warning time

Average Warning Time

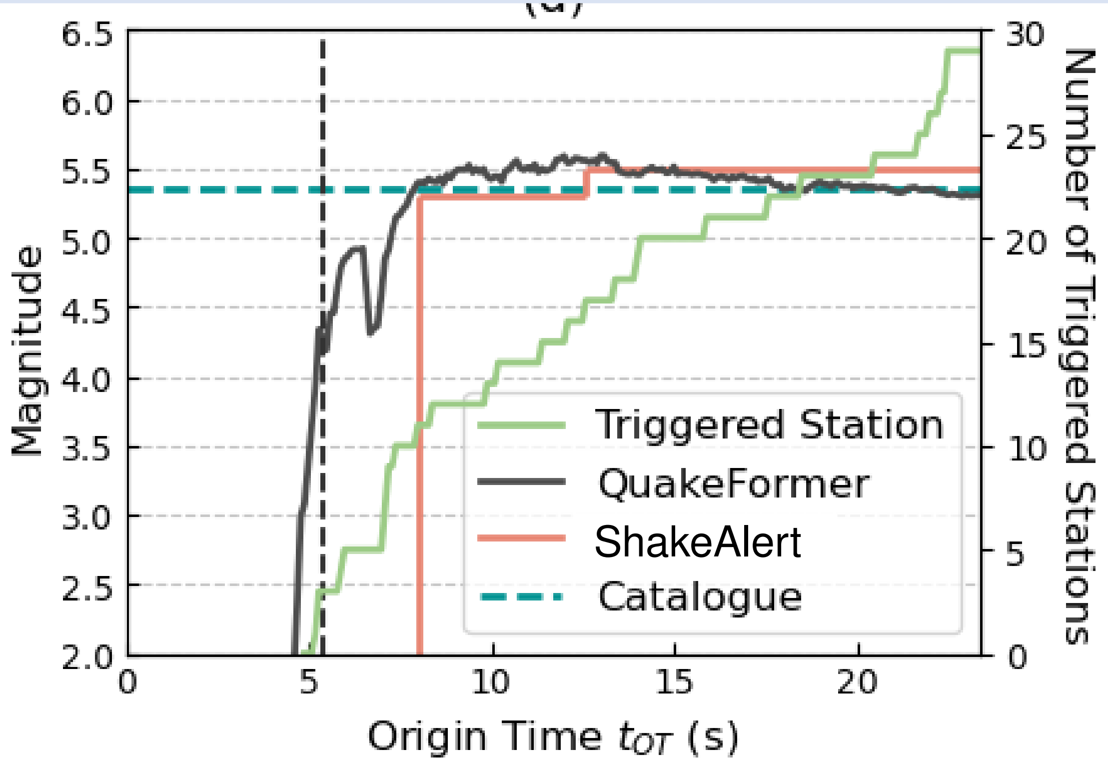




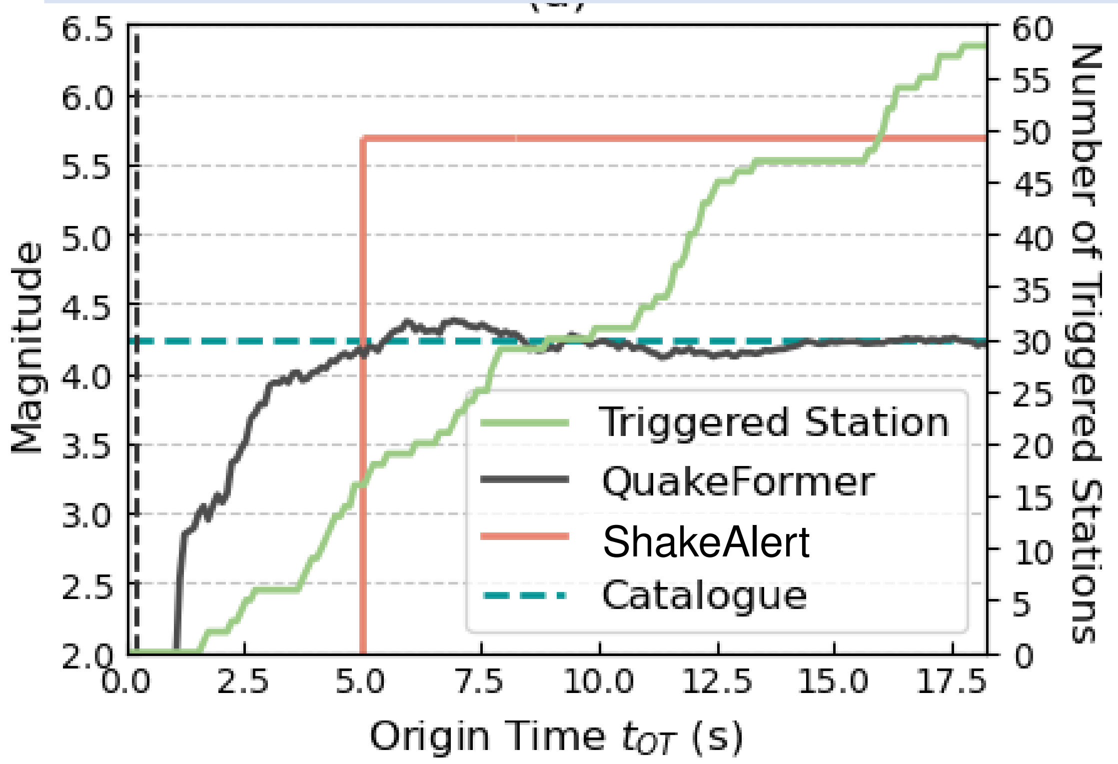
2.7 Results: Earthquake early warning

- Real-time magnitude prediction case：

2023 M 5.4 - 15km SE of Rio Dell, CA



2023 M 4.2 - 5 km SW of Isleton, CA



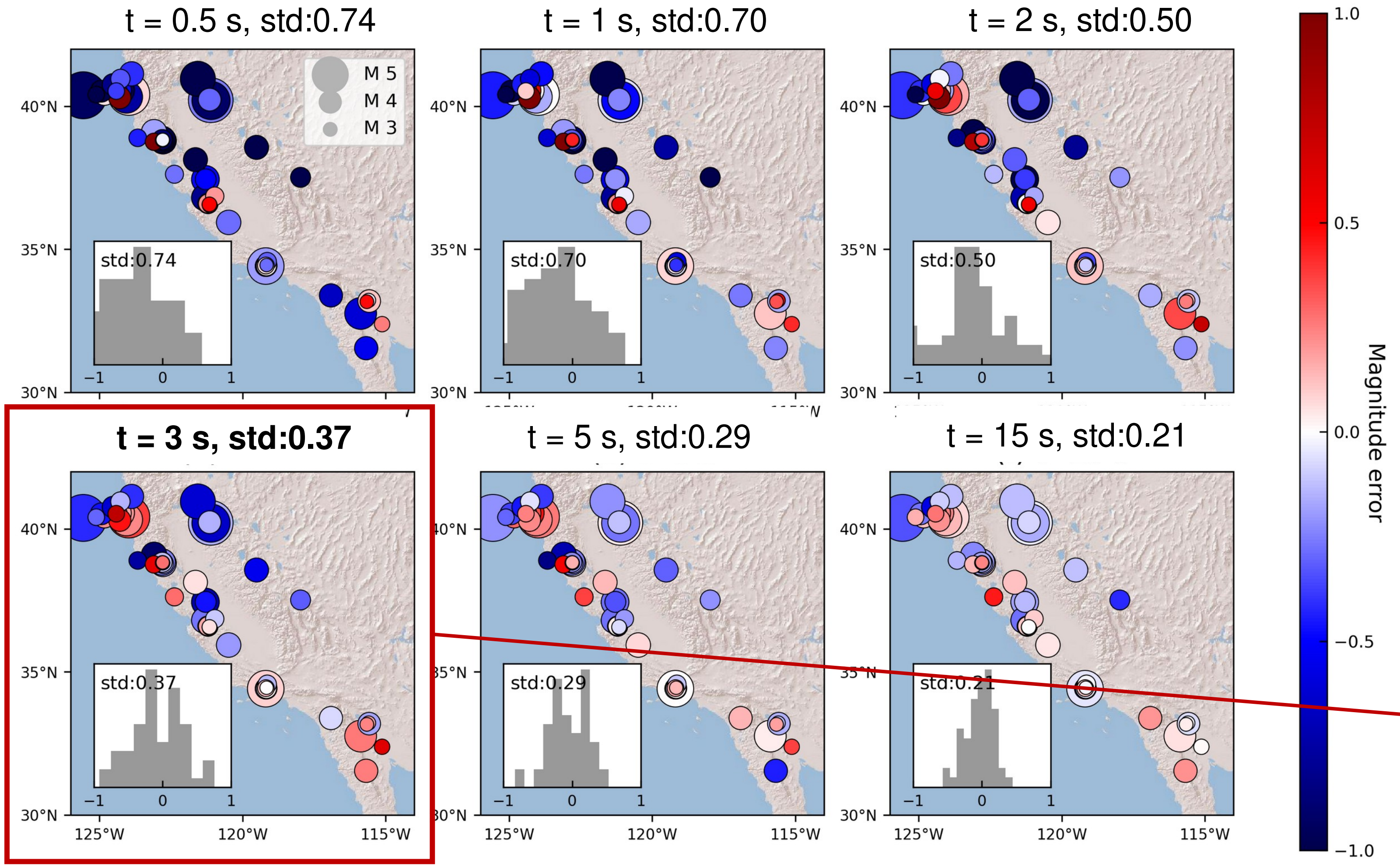
Magnitude prediction is faster and more accurate

2.7 Results: Earthquake early warning

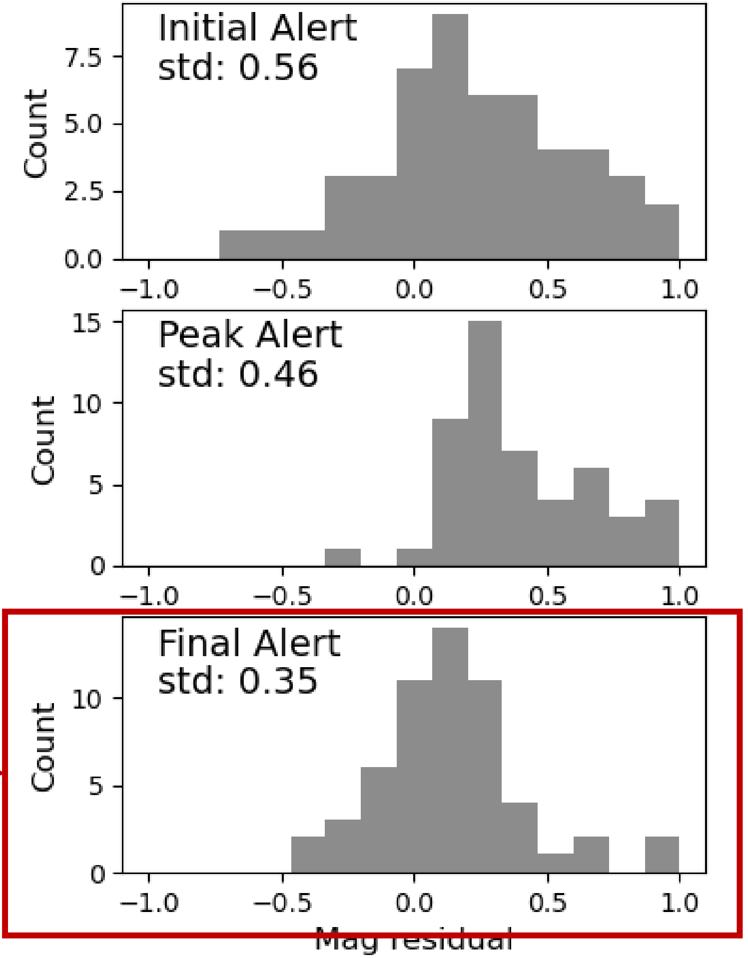
CA 2023
59 M3+ earthquakes

- Rapid Magnitude Prediction with low uncertainty (σ=0.37) at t = 3 s

QuakeFormer



ShakeAlert

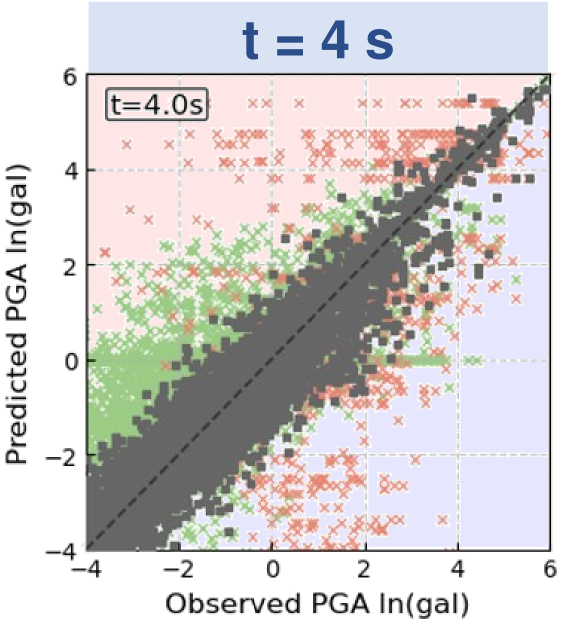


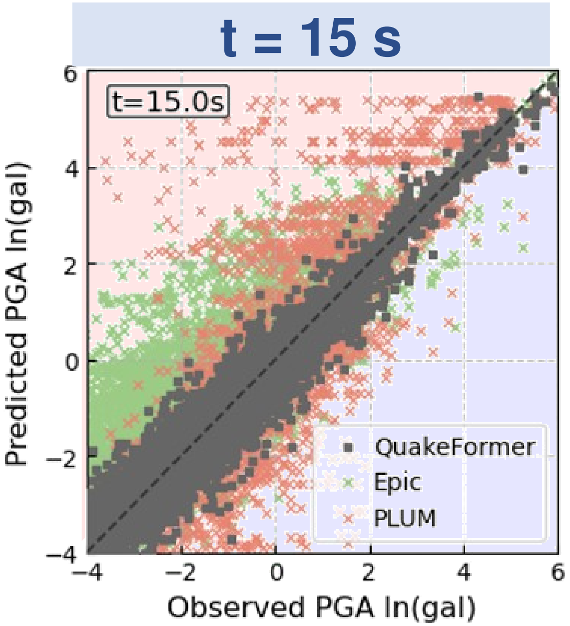
2.7 Results: Earthquake early warning

CA 2023
59 M3+ earthquakes

Time after first station trigger

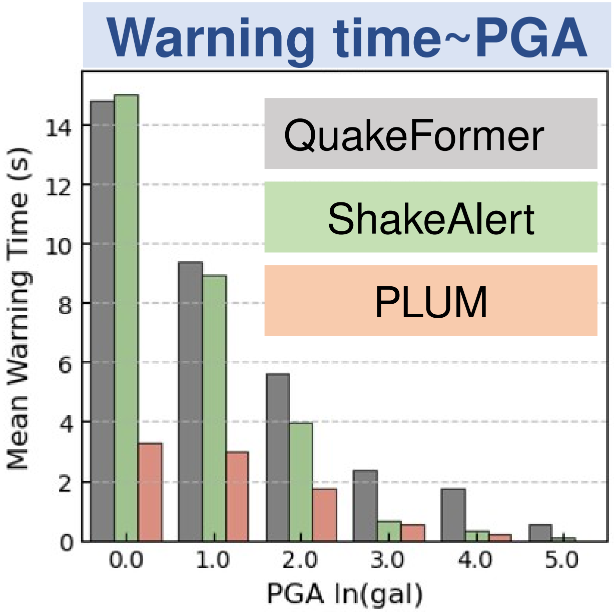
PGA
Prediction

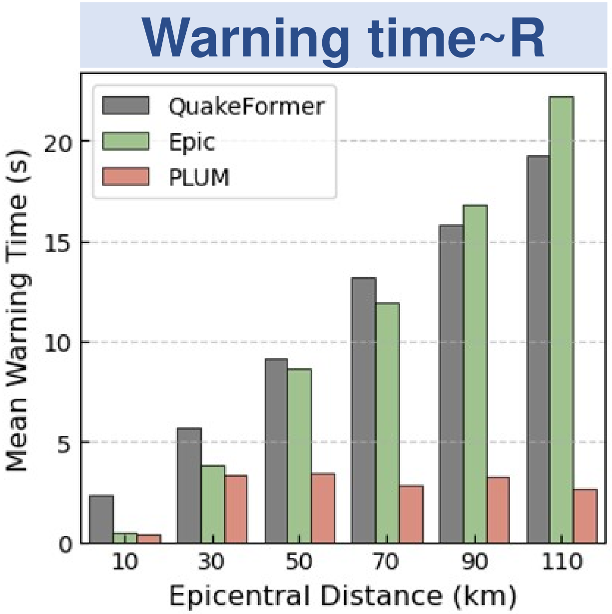




More accurate

Average
Warning Time

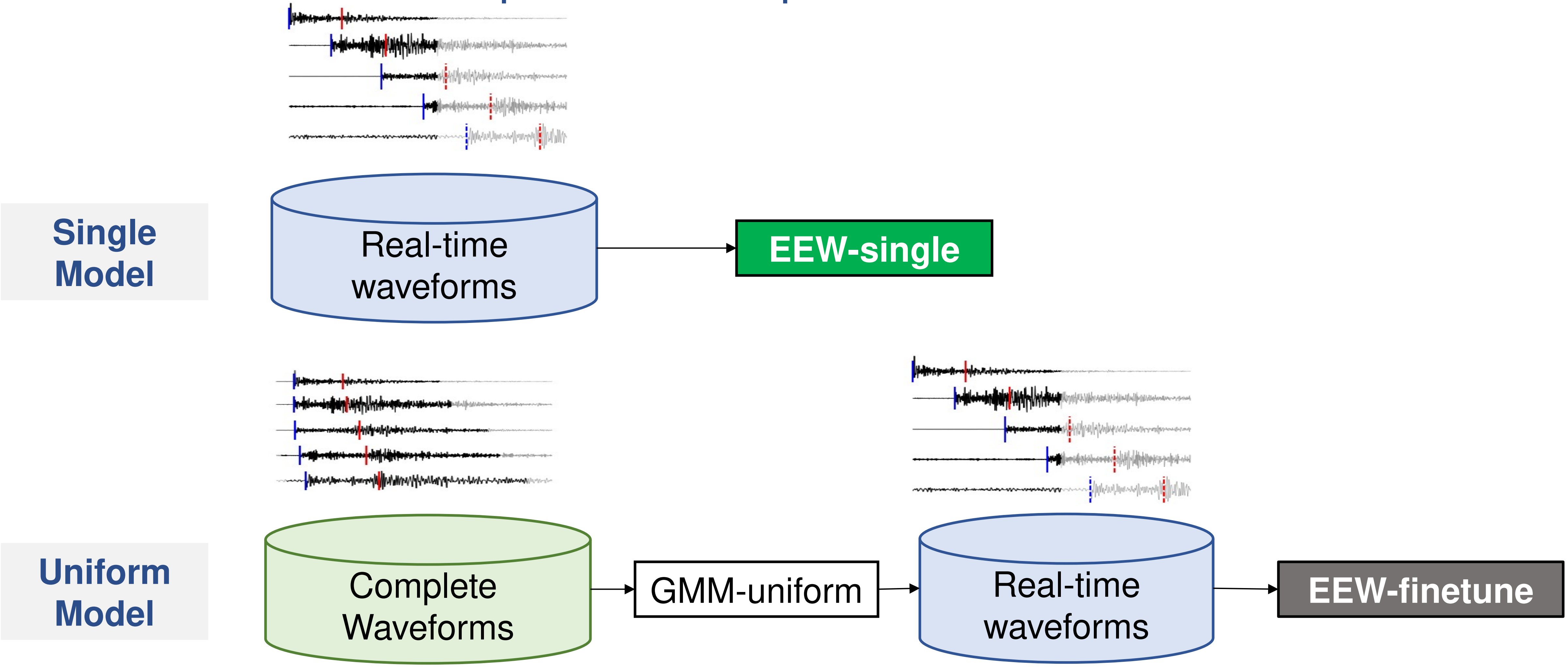




Faster Warning

2.7 Results: Earthquake early warning

- Uniform model improves EEW performance



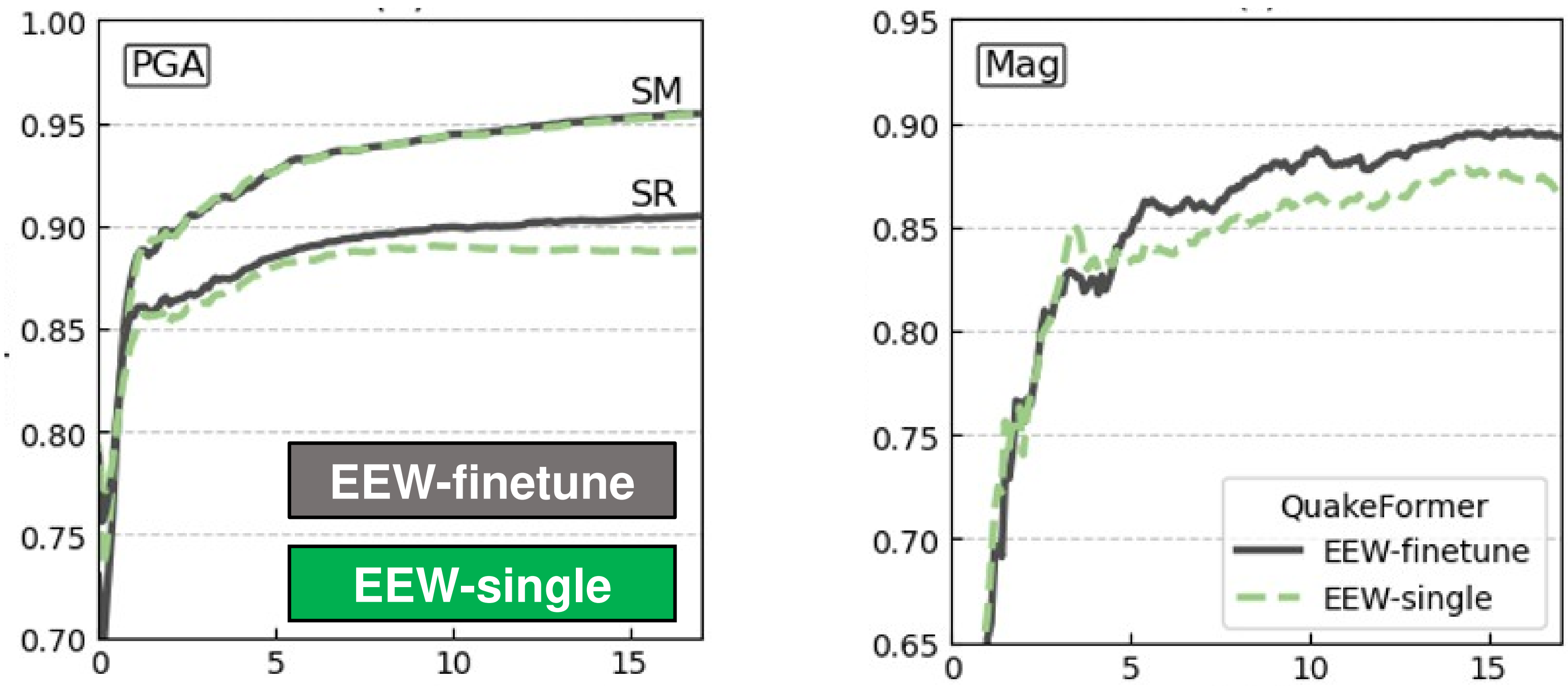
2.7 Results: Earthquake early warning

CA 2023
59 M3+ earthquakes

- Uniform model improves EEW performance

PGA ~ t	**Magnitude ~ t**

R-squared



Time after first station trigger (s) Time after first station trigger (s)

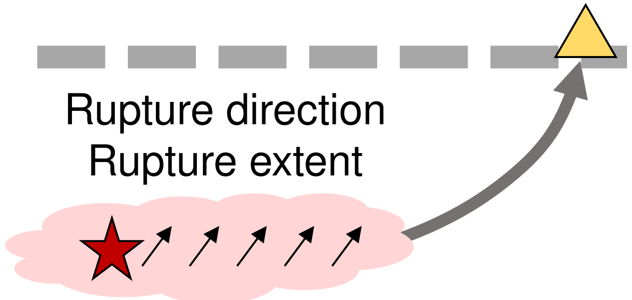
Outline

1. Background
2. Unified AI framework for ground motion prediction
3. **Ground motion prediction considering fault segments**
4. Conclusion

3.1 Ground motion prediction considering fault segments

QuakeFormer
Point-source assumption

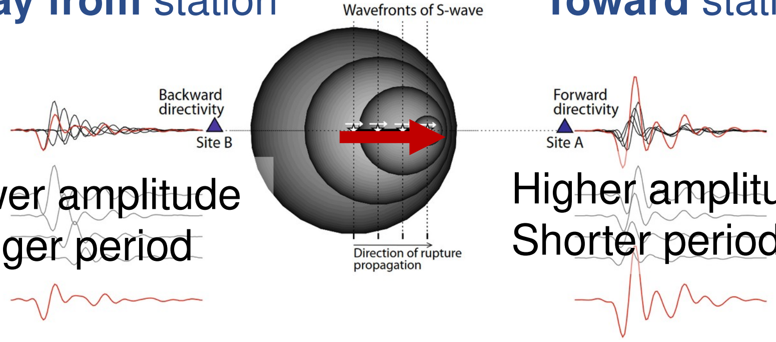
Reality
Finite-fault source



Rupture directivity

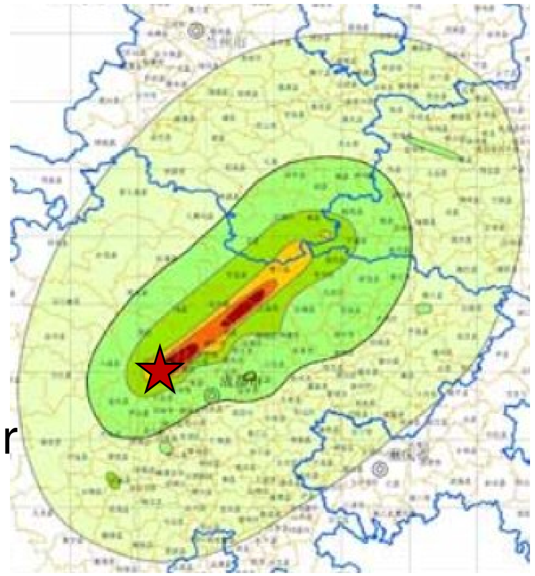
Away from station

Toward station

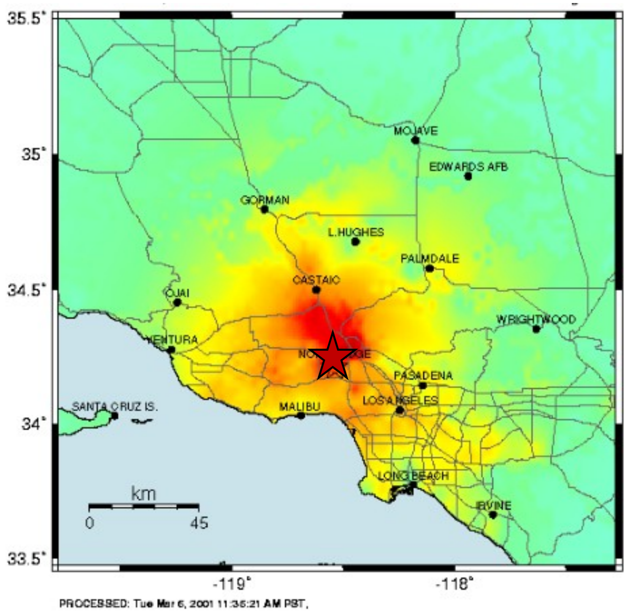


★ Epicenter

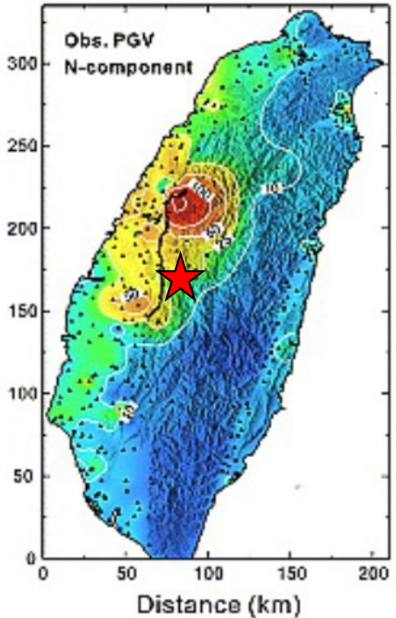
Wenchuan Earthquake



Northridge Earthquake



Chi-Chi Earthquake

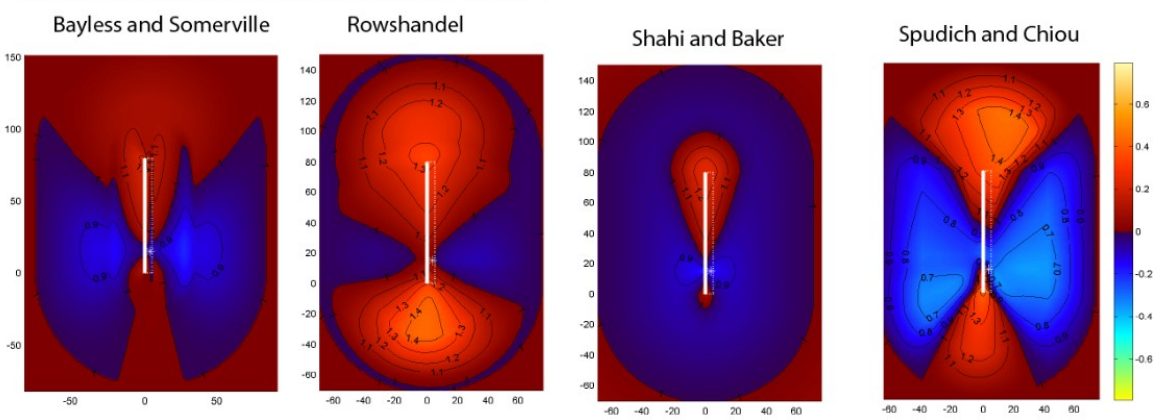


3.2 Motivation

- Towards directivity-based PSHA and post event assessment

Empirical directivity corrections

NGA-West2

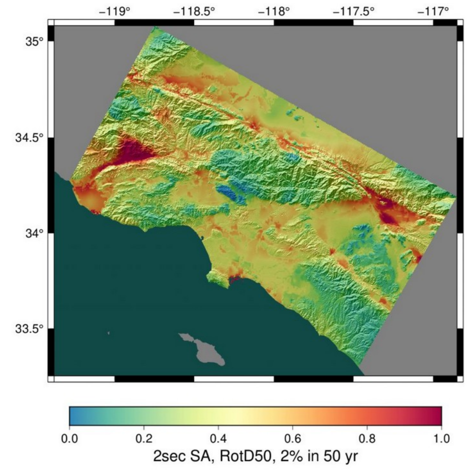


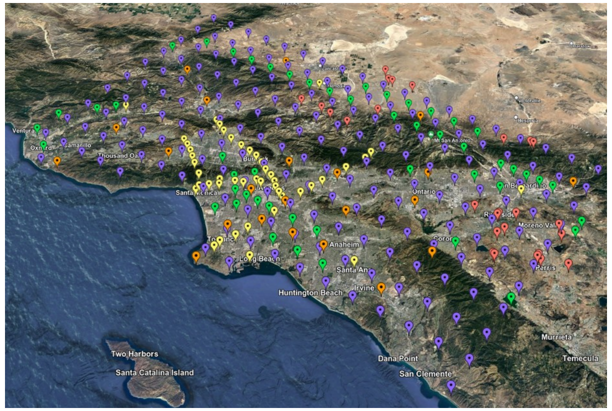
$$\log \mathbf{IM}_{dir} = \log \mathbf{IM} + f_D$$

- Specific assumptions
- Tend to oversimplify multi-segment faults or dipping faults.

Physics-based Simulation

CyberShake (22.12 BB)





- For 1 site, 1.5 TB data need to be staged
- 730 k node-hours for the main study

3.2 Motivation

- Towards directivity-based PSHA and post event assessment

Empirical directivity corrections

Fast but oversimplified

?

Balance

Physics-based Simulation

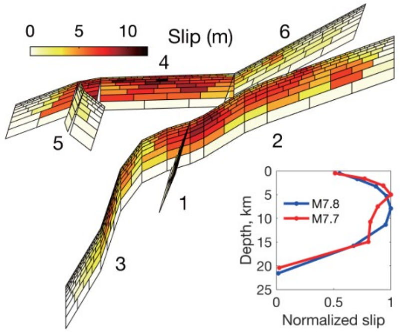
Accurate but expensive

Increasing Synthetic Scenarios

Dataset

Deep learning method

How to represent and encode faults ?



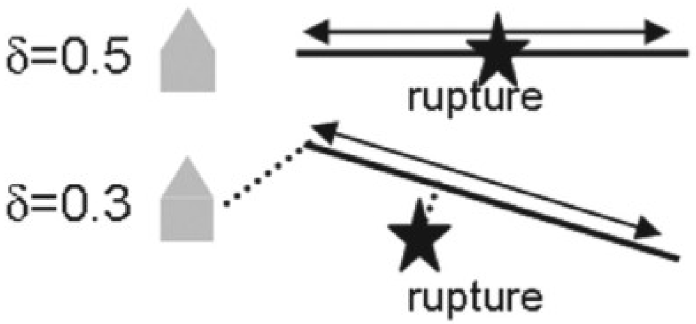
3.3 QuakeF2G

- **Step1**: Separate 3D faults as fault segments. Encode fault sequence.

Existing fault representations in deep learning

By parameters

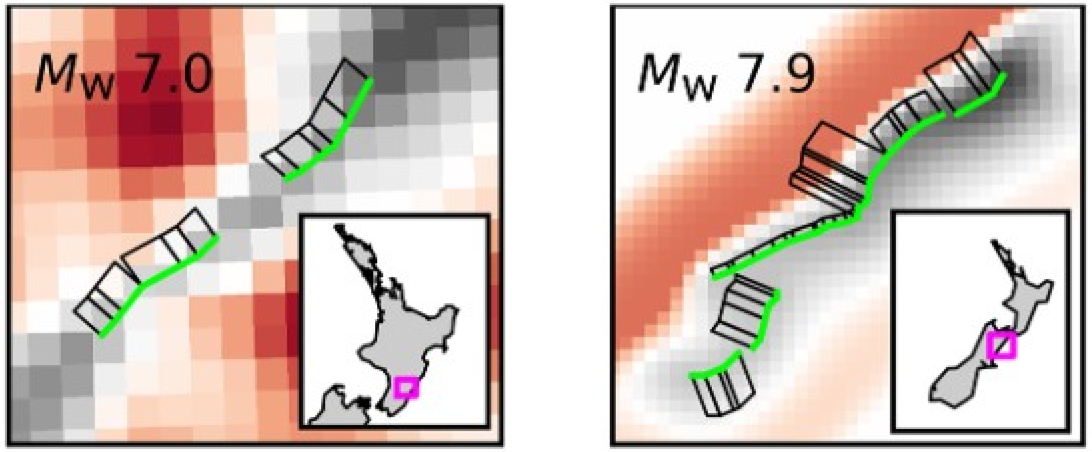
- Rupture length, width, strike, dip, rake
- Rupture ratio



(Bose et al. 2014)

Oversimplified

By image

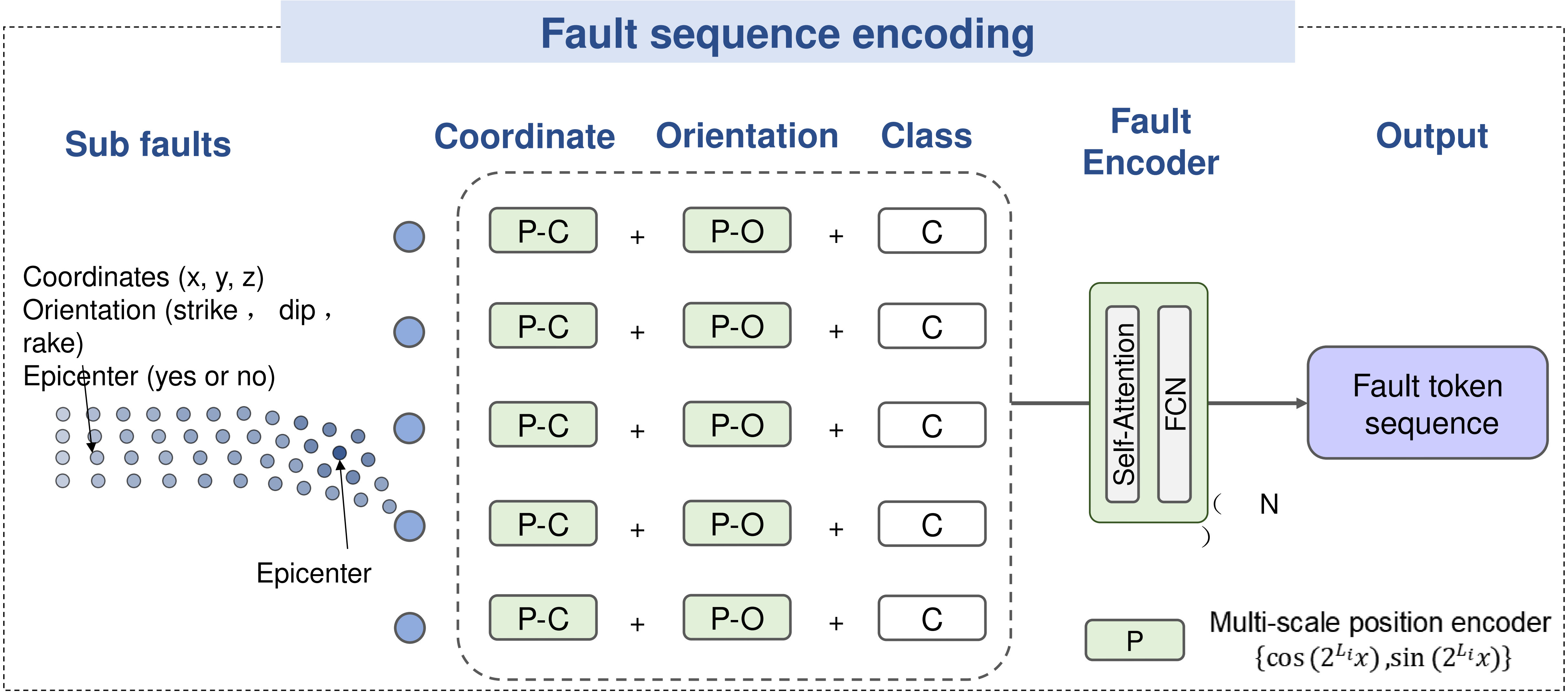


(Weatherill & Lilienkamp, 2024)

Fixed resolution
Cannot represent 3-D faults

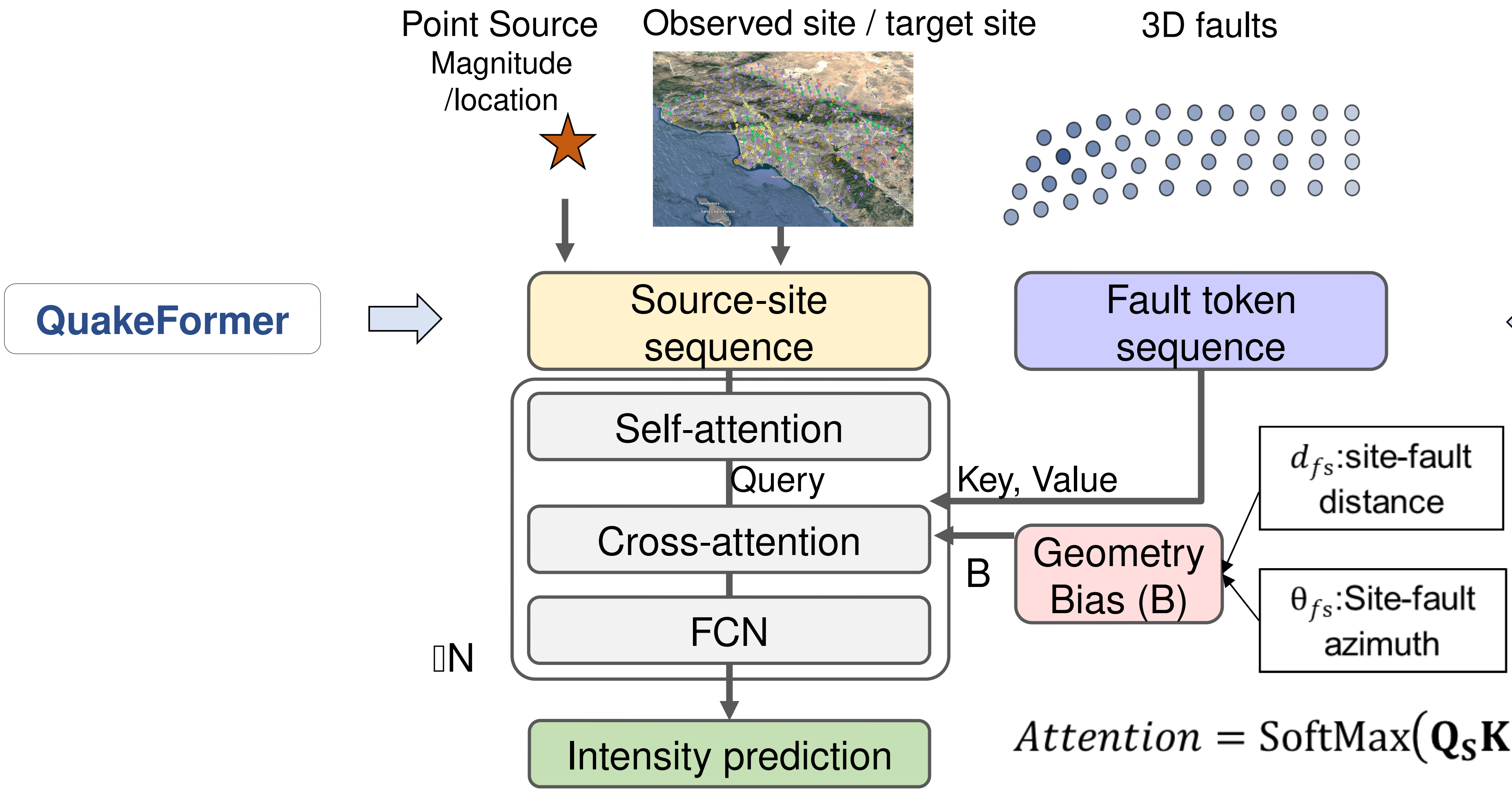
3.3 QuakeF2G

- **Step1**: Separate 3D faults as fault segments. Encode fault sequence.



3.3 QuakeF2G

- **Step 2**: Fault-site information fusion. Add geometry bias to enhance relative position relationship between sites and fault segments



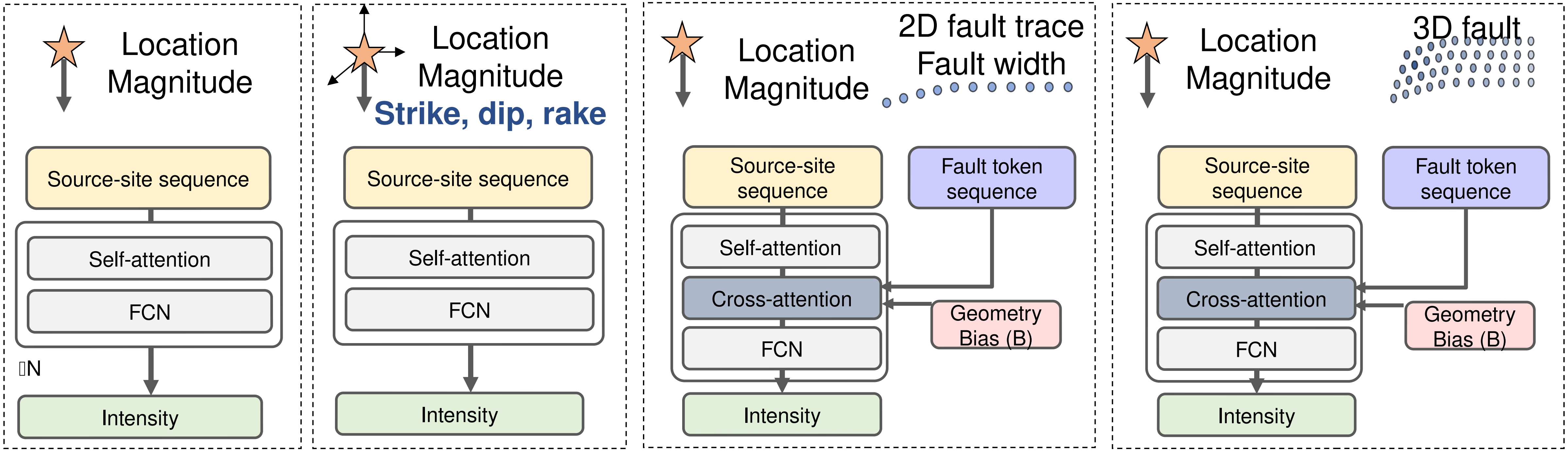
$$Attention = \mathrm{SoftMax}\left(\mathbf{Q_S}\mathbf{K_F^T}/\sqrt{d} + \mathbf{B_{SF}}\right)\mathbf{V_F}$$

3.3 QuakeF2G

- QuakeF2G could be adapted to **multi-level source representations**

Simple → Complex

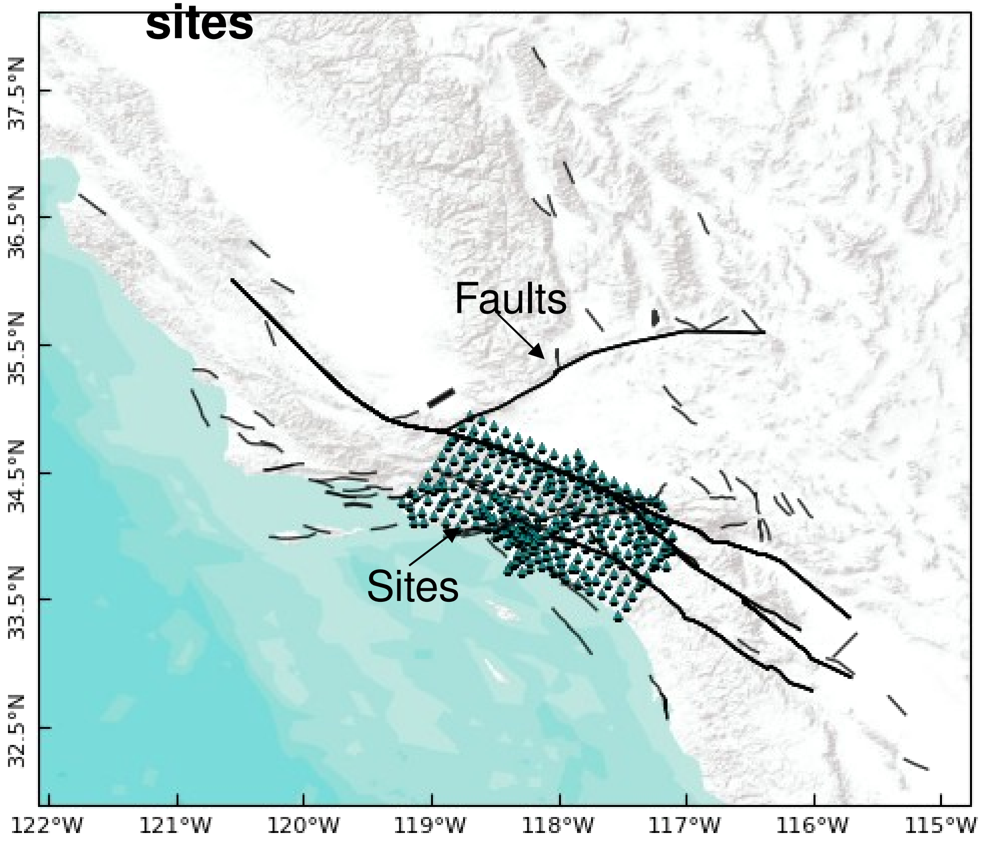
Point Source	Point Source + Focal Mechanism	2D Finite-fault	3D Finite-fault
30 seconds	**30 minutes**	**~1 hour**	**More than 1 hour**

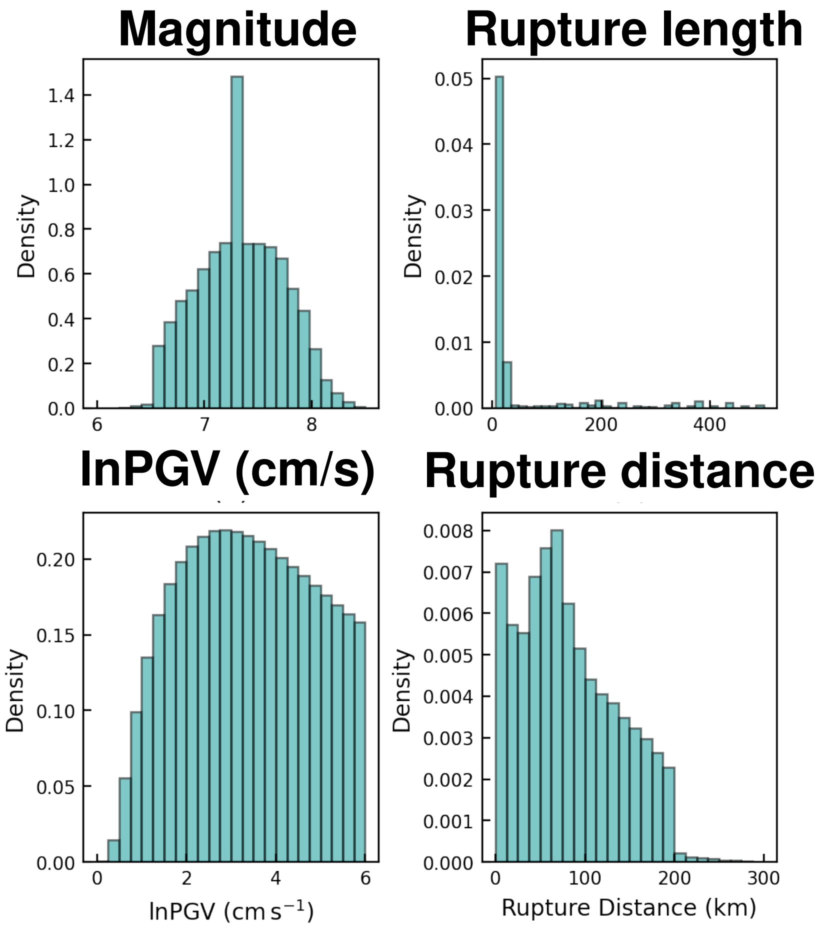


3.4 Dataset

- Current study use CyberShake synthetic waveforms (22.12 BB)

630,000 scenarios， 335 sites

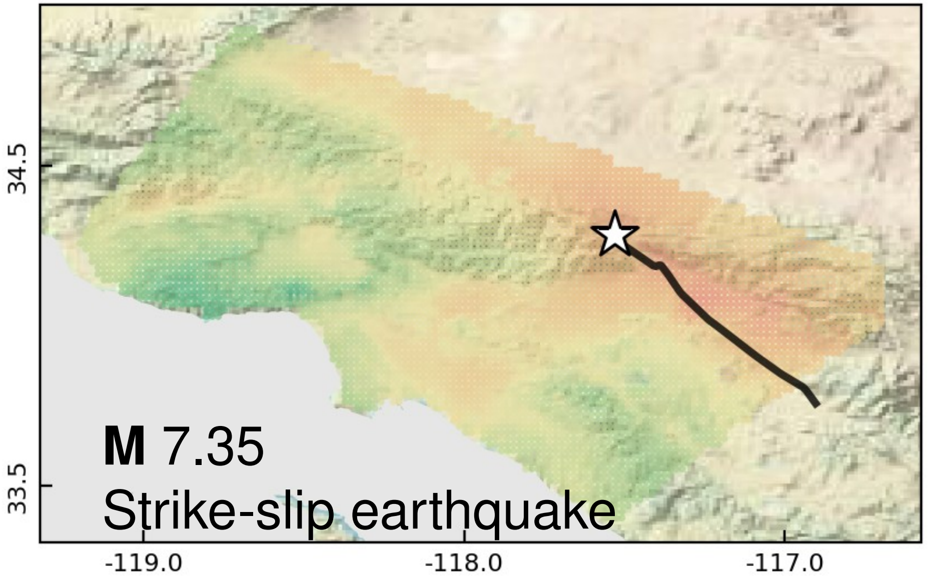




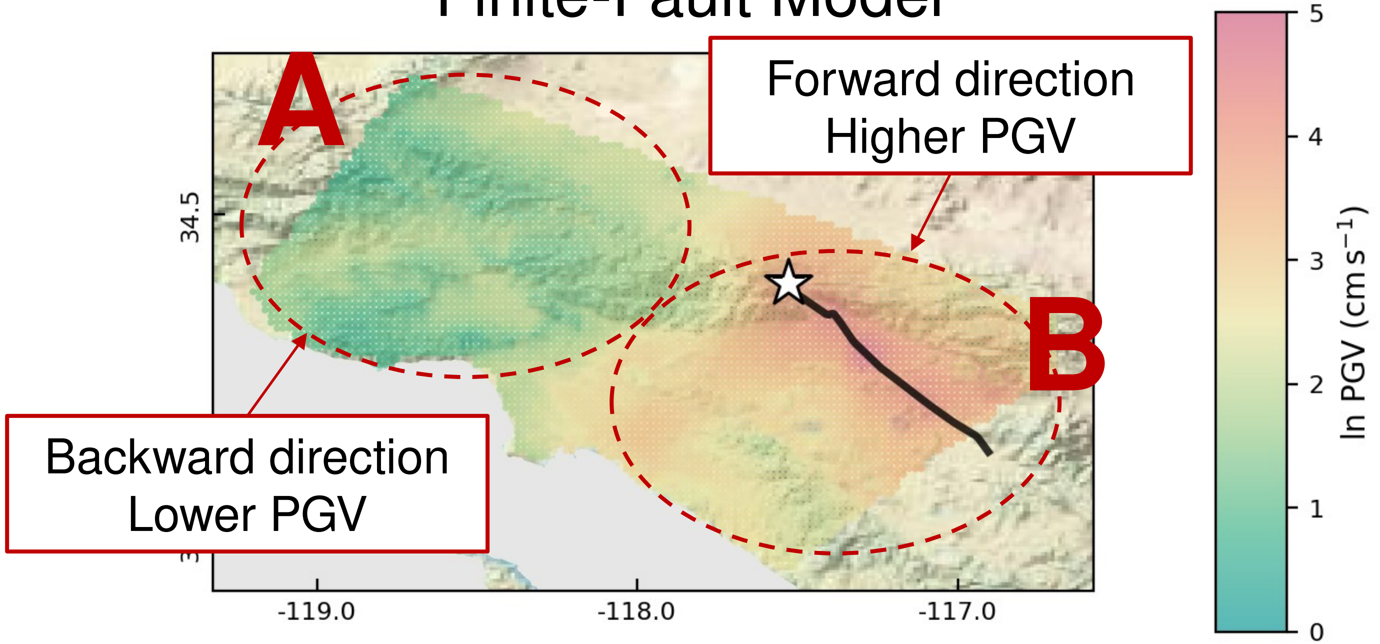
Train: Valid: Test=8:1:1

3.5 CyberShake case: PGV Distribution

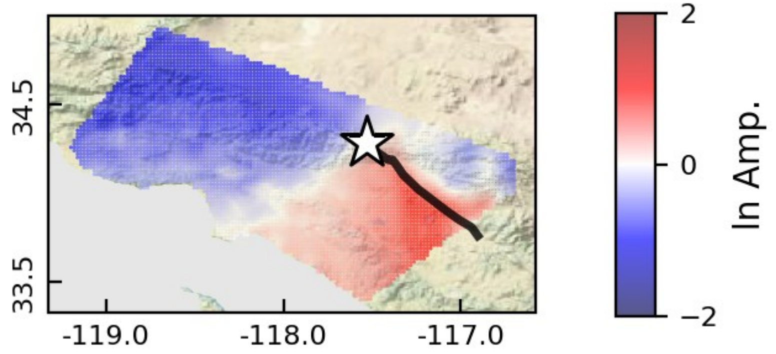
Point Source Model



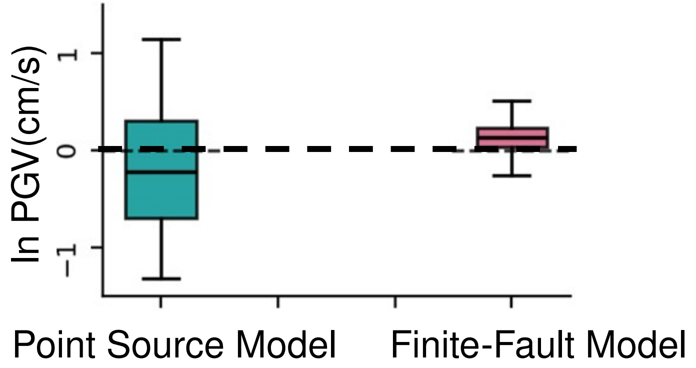
Finite-Fault Model



$$PGV_{\mathrm{FiniteFault}} - PGV_{\mathrm{PointSource}}$$

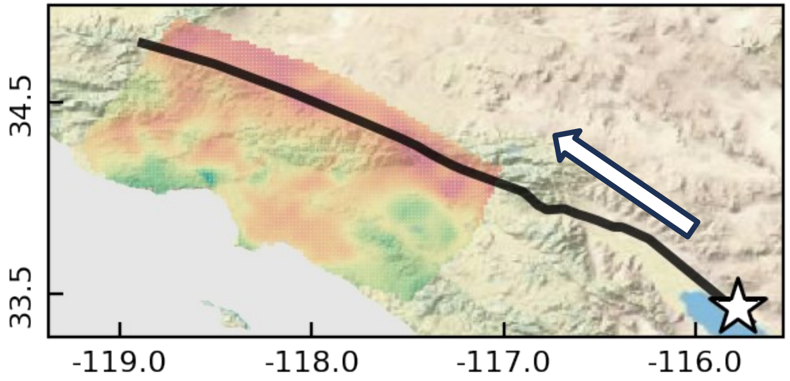


Residuals decrease



3.5 CyberShake case: PGV Distribution

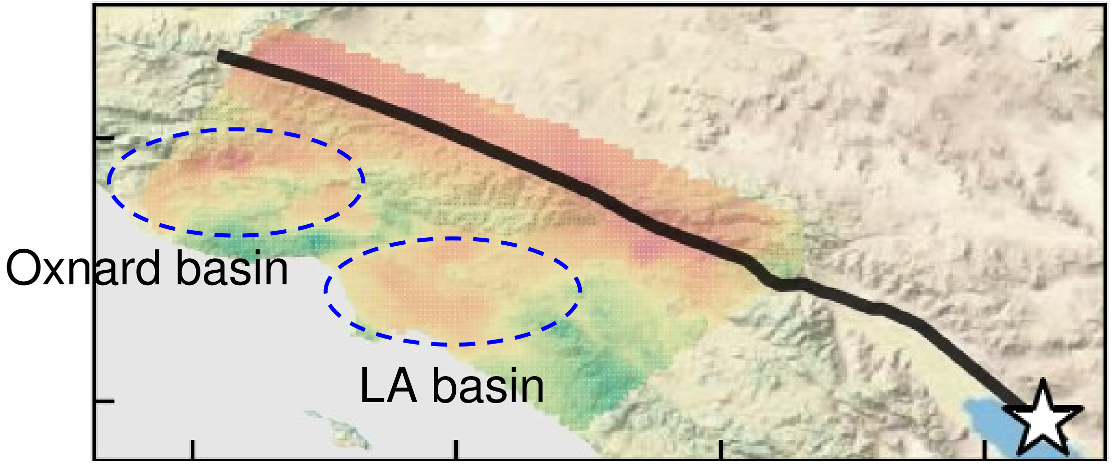
CyberShake BB



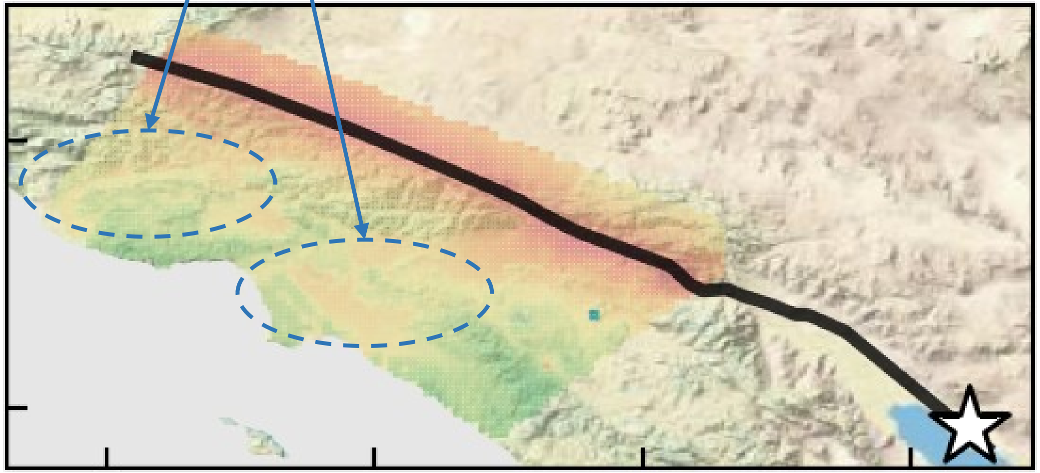
M 7.75
San Andreas Fault

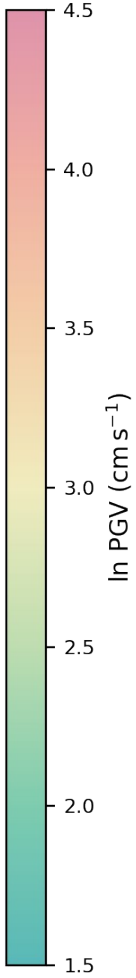
Ignore the coupling between finite-fault directivity and 3D basin effects

QuakeF2G (Finite-Fault)



ASK14+ROW13



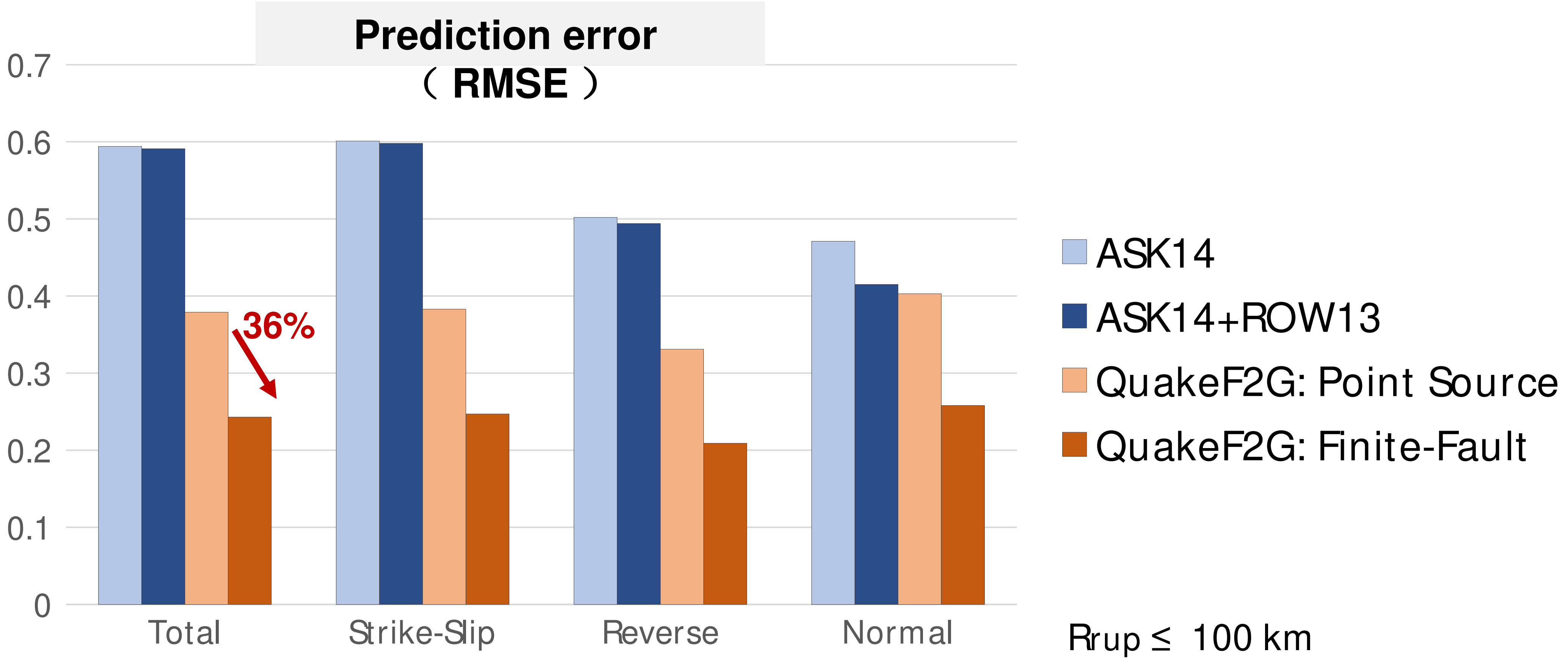


PGV amplified in the forward direction

Underestimate PGV in basins

3.5 CyberShake case: PGV Distribution

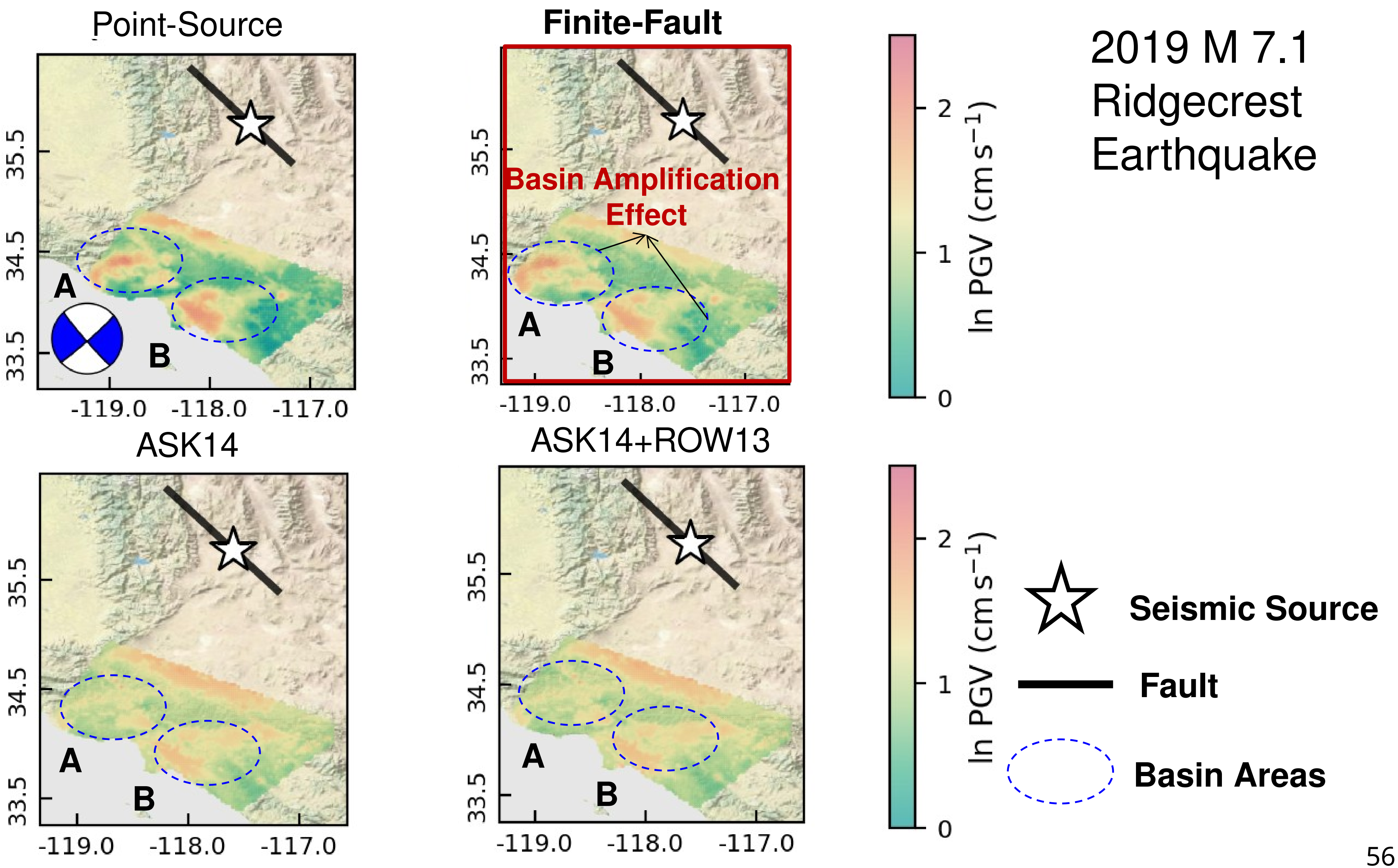
- Much more accurate than existing empirical models
- From QuakeF2G Point Source to Finite-Fault, total RMSE reduces by 36%



3.6 Real-world earthquake

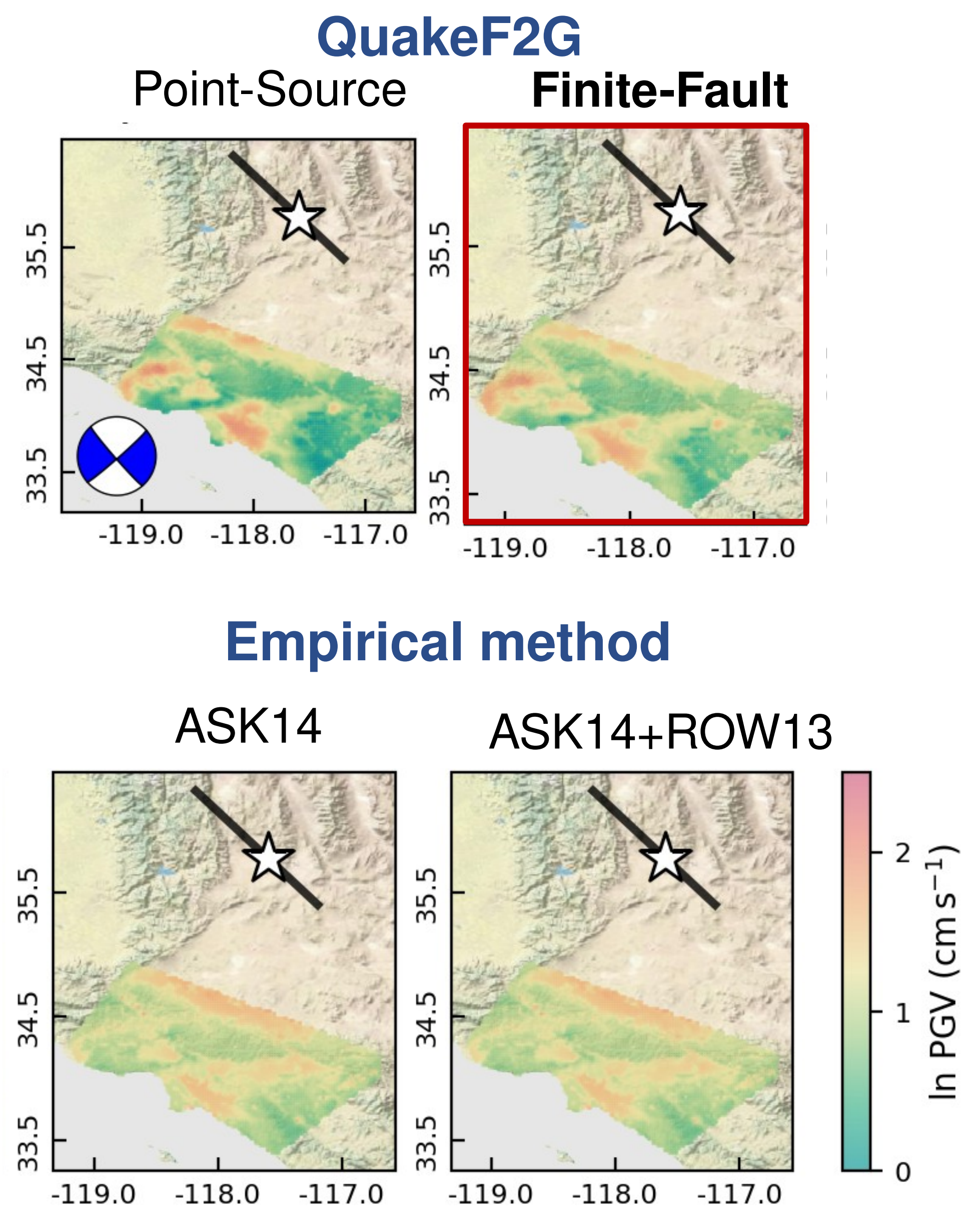
QuakeF2G

Empirical method

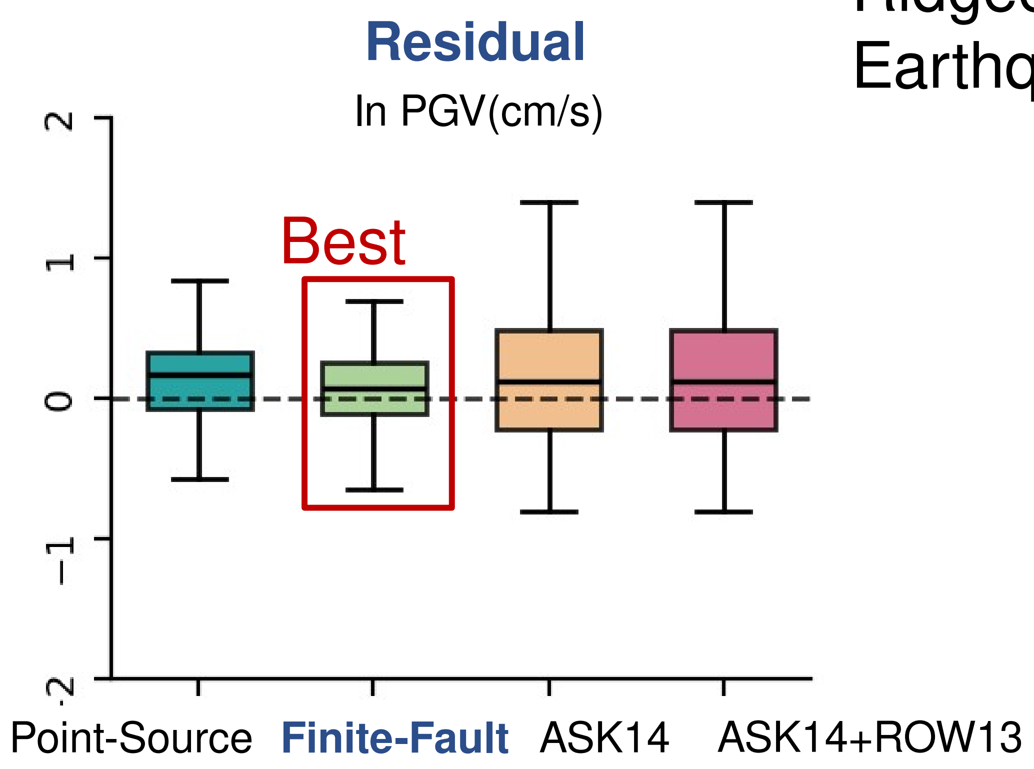


2019 M 7.1 Ridgecrest Earthquake

3.6 Real-world earthquake



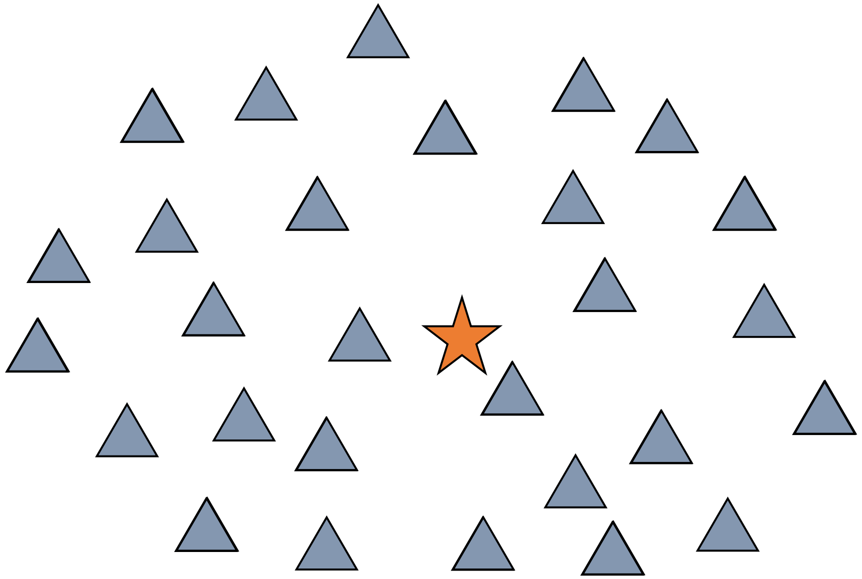
2019 M 7.1 Ridgecrest Earthquake



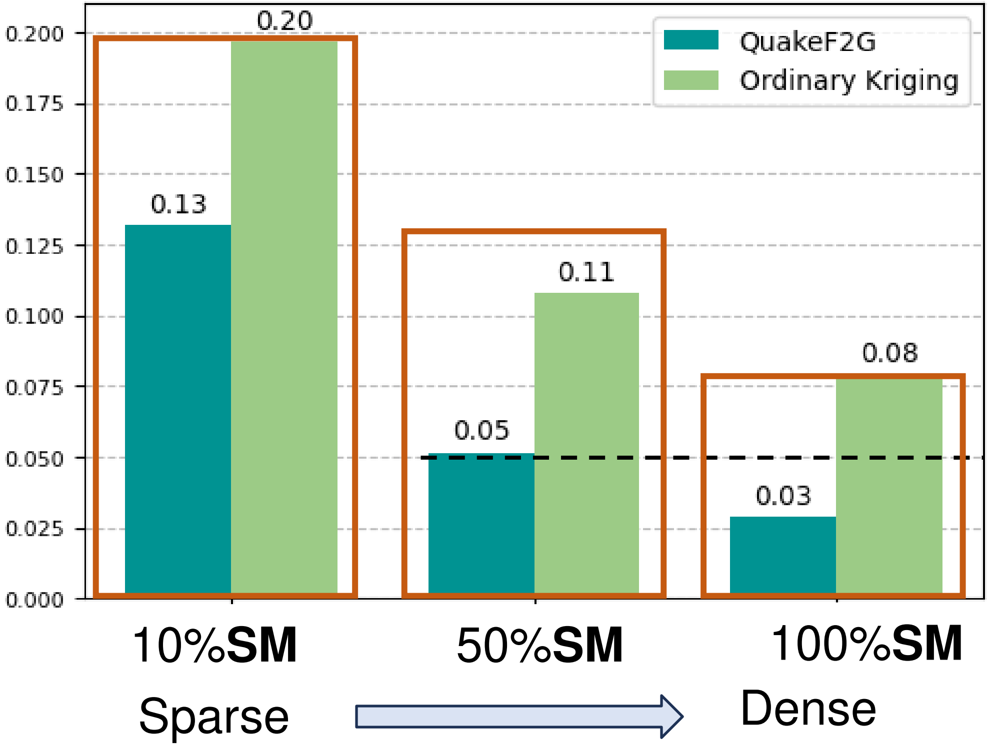
- **Finite-fault information is from the USGS report**
- **Predictions are compared against real recorded ground motions at real stations**

3.6 Real-world earthquake

- **Interpolation**: QuakeFormer outperforms Ordinary Kriging



RMSE



Save synthetic time and storage space

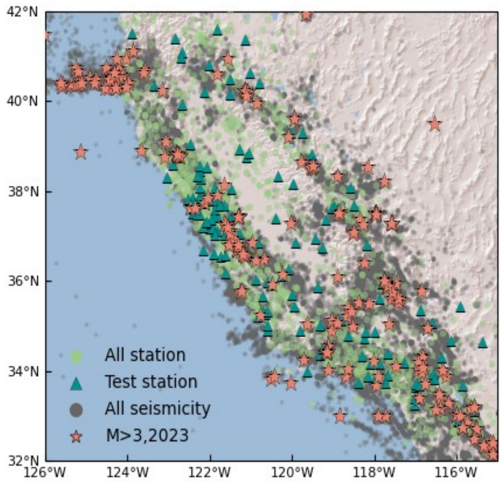
Outline

1. Background
2. Unified AI framework for ground motion prediction
3. Ground motion prediction considering fault segments
4. **Conclusion**

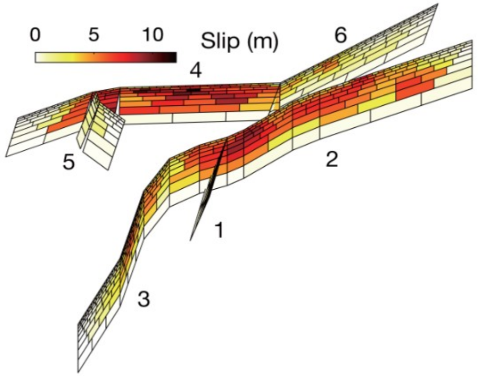
4. Conclusions

Input

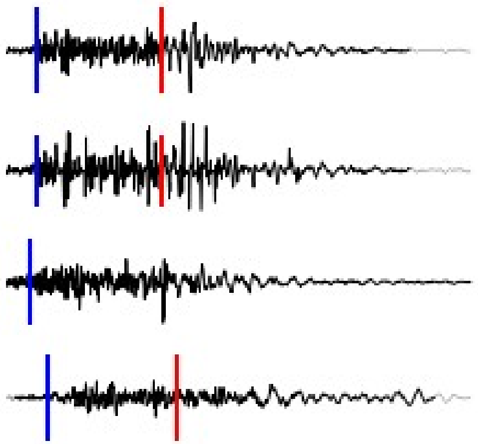
Graph of sources and station (Considered)



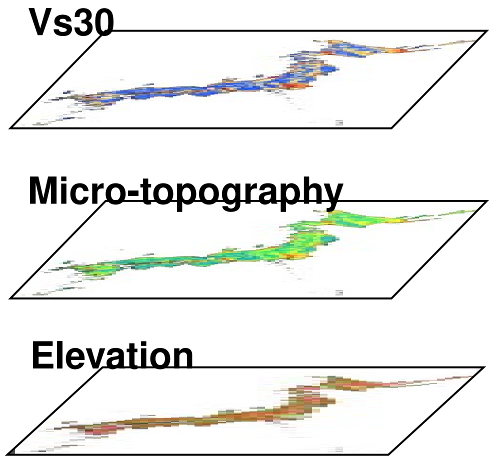
Unstructured fault data (Considered)



Waveform data (Considered)



Raster data of site and topography (to be considered)



Transformer Model

Output

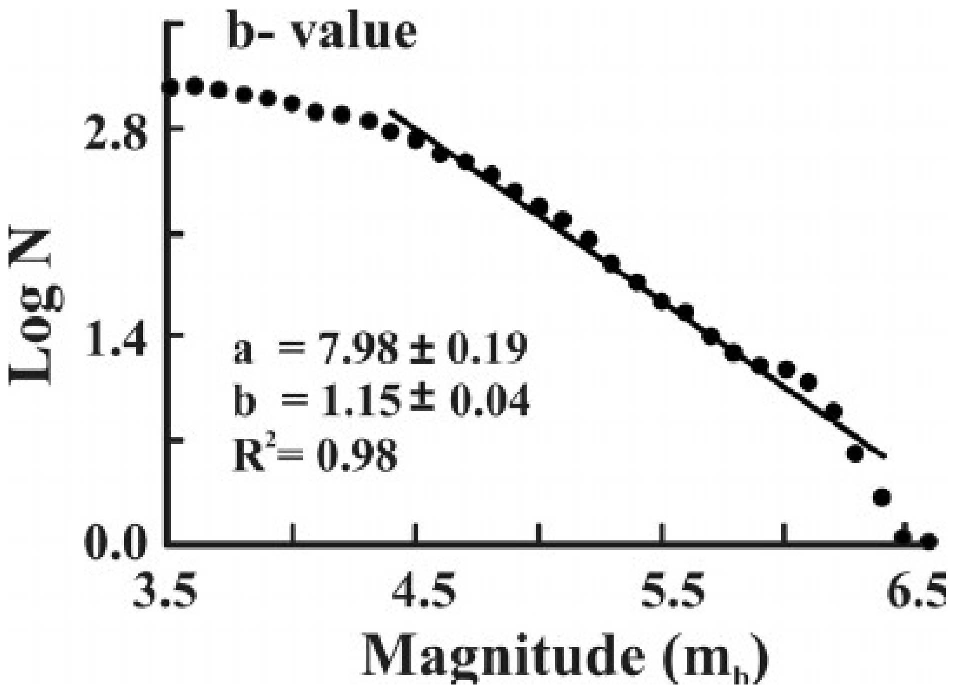
Target Intensity (Solved)

Target Waveform (to be solved)

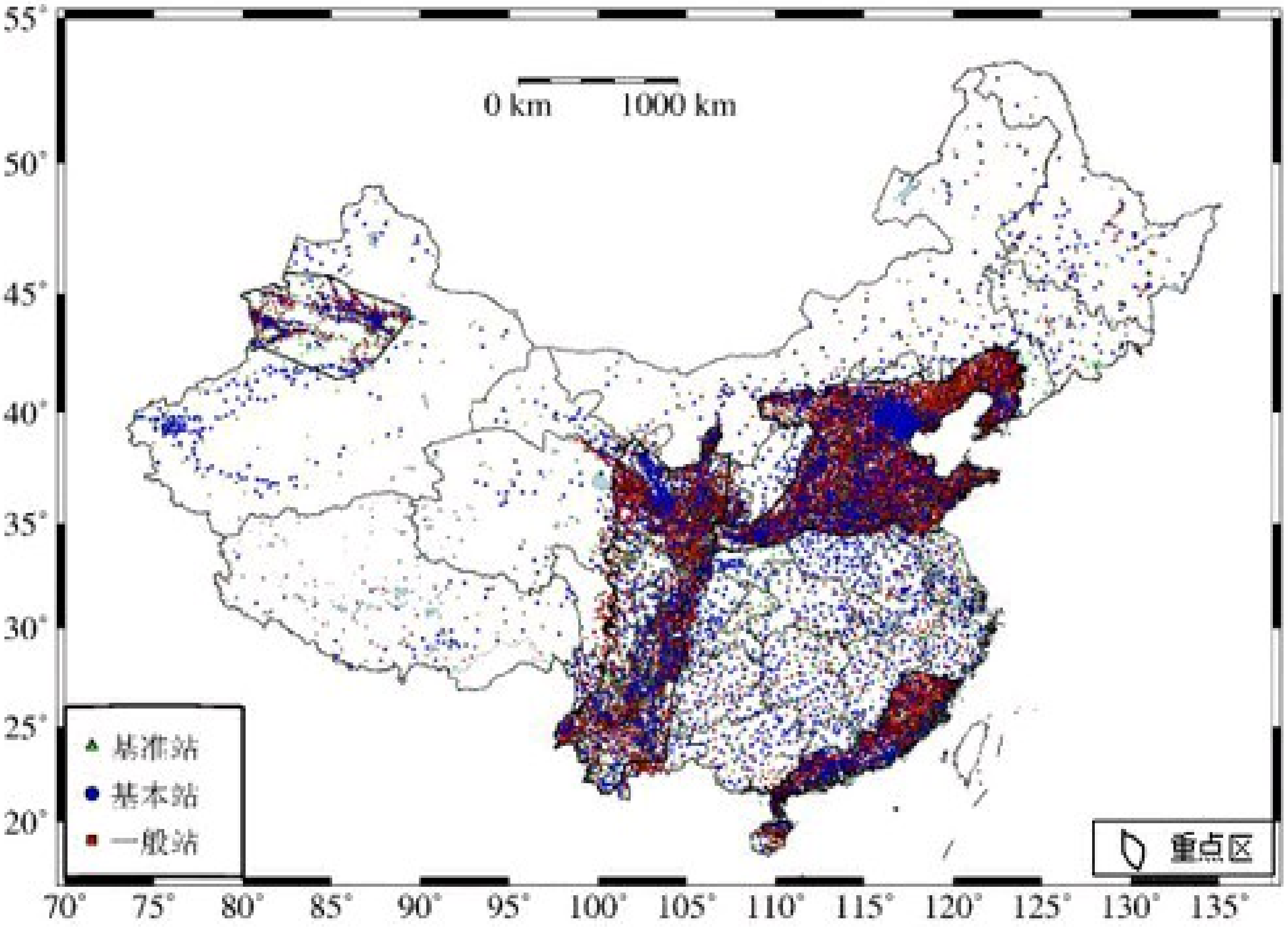
4. Conclusions

Gutenberg–Richter law

$$\log N = a - bM$$

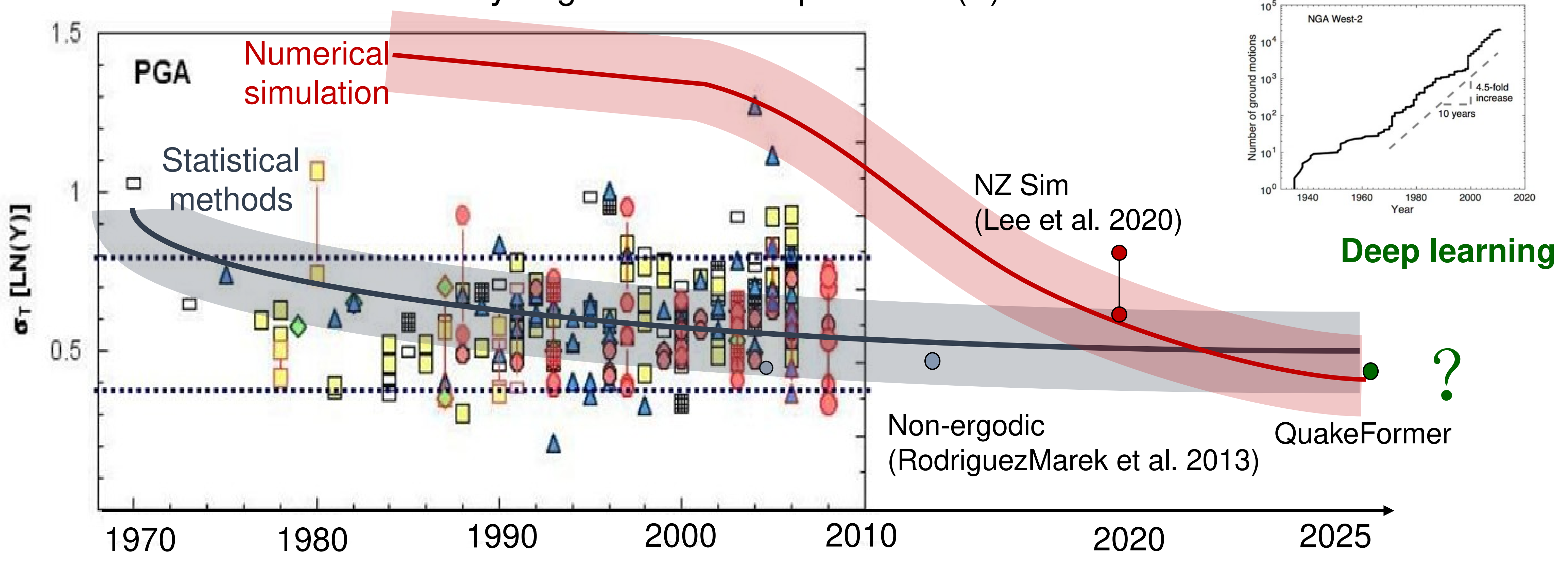


National Seismic Station



Can We Reduce the Uncertainty in Ground Motion Prediction?

Uncertainty in ground motion prediction(σ)

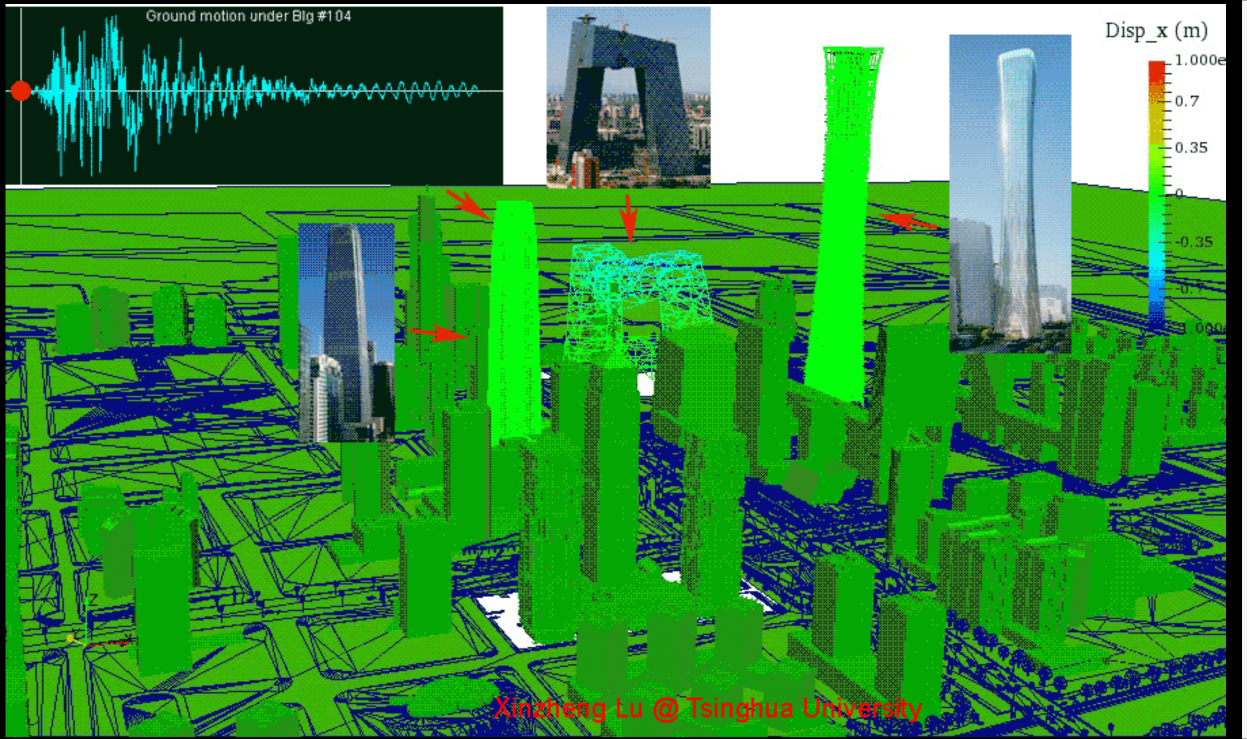


Sigma: Issues, Insights, and Challenges. (Strasser 2009)

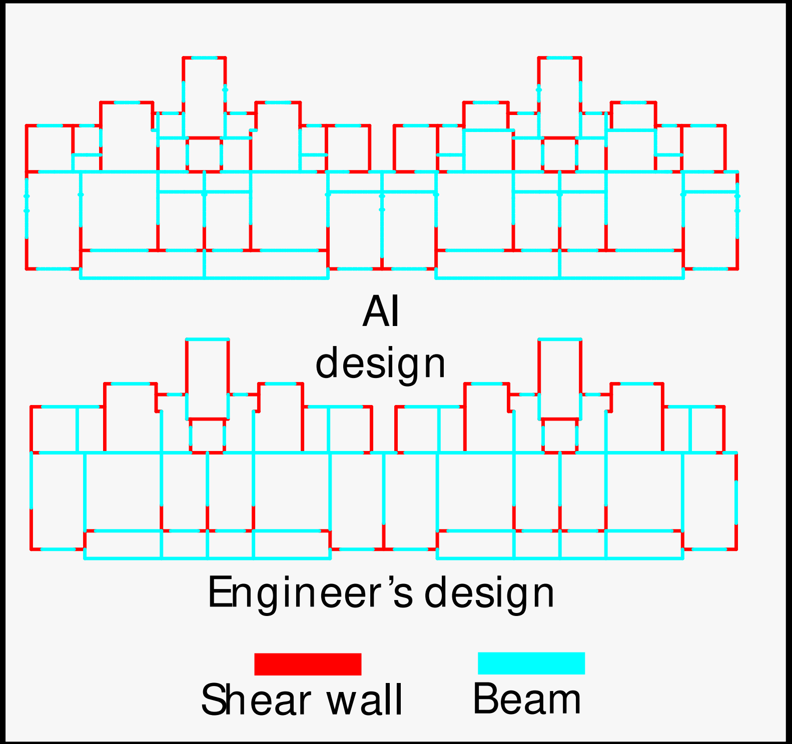
Thank you for your attention!

luxz@tsinghua.edu.cn

Seismic collapse simulation of Shanghai Tower (632 m)



Earthquake scenario simulation of tall buildings at Beijing CBD subjected to the 1679 Sanhe-Pinggu M8 Earthquake



Generative AI design of building structures