

Operator Learning for Earthquake Location: Methodology and Application to Out-of-Network Events in the Mendocino Triple Junction

Hongyu Sun^{*1} 

ABSTRACT

Accurate location of earthquakes outside the boundaries of seismic networks is a well-known challenge due to limited azimuthal coverage. Offshore earthquakes, particularly along the Mendocino Transform fault, pose considerable risks to nearby communities. However, the lack of permanent stations in marine environments hinders accurate earthquake location. This study investigates the feasibility of locating offshore earthquakes around the Mendocino Triple Junction (MTJ) using advanced deep learning techniques. We propose the location neural operator (LocNO), an innovative operator learning framework that directly estimates earthquake locations from full-waveform data, even under sparse and nonideal network conditions. The approach integrates the Fourier neural operator and graph neural operator to capture the intricate spatiotemporal dependencies of seismic wavefields across stations and to estimate a spatial pseudoprobability density function over earthquake source coordinates (easting, northing, and depth). Trained on historical seismicity and simulated sparse network scenarios, LocNO generalizes across varying network geometries and provides robust estimates for out-of-network events, with the mean absolute errors on the order of 10 km horizontally and 4 km in depth. A case study of the 2017 M_w 5.7 earthquake sequence, 218 km west of Ferndale, California, demonstrates that LocNO yields robust locations for out-of-network earthquakes newly detected by the phase neural operator, even when recorded by only a few stations with large azimuthal gaps. These results open a new avenue for investigating MTJ seismotectonics using deep-learning-enhanced earthquake catalogs.

KEY POINTS

- Sparse or nonideal seismic networks hinder the accurate location of offshore and out-of-network earthquakes.
- Location neural operator (LocNO) improves earthquake location by leveraging advanced operator learning to capture spatiotemporal dependencies.
- LocNO shows promise for improving earthquake monitoring in remote and under-instrumented regions.

INTRODUCTION

The Mendocino Triple Junction (MTJ) is one of the most seismically active regions in North America. This triple junction serves as a tectonic transition zone, linking the convergence of the Cascadia subduction zone with the translation of the San Andreas fault system. These complex interactions generate frequent offshore earthquakes concentrated along the Mendocino Transform fault. However, the lack of permanent seismic stations in the marine environments introduces substantial challenges to accurately locating these events.

Locating out-of-network earthquakes is a long-standing challenge (Ellsworth and Roecker, 1981) due to limited azimuthal coverage, unreliable phase arrival data, and low signal-to-noise ratios (SNRs). These challenges are worsened for offshore earthquakes by the attenuation of seismic signals over long distances and inaccuracies in the regional velocity models. Complex 3D geological structures can perturb ray paths and distort waveform characteristics, making accurate location even more difficult. Traditional methods often struggle with the nonlinear inversion processes required for earthquake location in such scenarios, resulting in high uncertainties, with location errors reaching up to 50 km in some cases (Cabeceles *et al.*, 2020). For example, ShakeAlert, an earthquake early warning system (Burkett *et al.*, 2014; Given *et al.*, 2018;

1. Department of Earth, Environmental and Resource Sciences, The University of Texas at El Paso, El Paso, Texas, U.S.A.,  <https://orcid.org/0000-0003-3828-1859> (HS)

*Corresponding author: hongyu-sun@outlook.com

Cite this article as Sun, H. (2025). Operator Learning for Earthquake Location: Methodology and Application to Out-of-Network Events in the Mendocino Triple Junction, *Bull. Seismol. Soc. Am.* **XX**, 1–20, doi: [10.1785/0120250018](https://doi.org/10.1785/0120250018)

© Seismological Society of America

Kohler *et al.*, 2020), has located many offshore earthquakes near the MTJ with errors exceeding 50 km compared to U.S. Geological Survey (USGS) locations (Williamson *et al.*, 2023). This limitation creates significant uncertainties in the earthquake locations, hindering the development of accurate earthquake catalogs essential for seismic hazard assessment, early warning systems, and seismotectonic research.

Several approaches have been proposed to address these challenges and locate out-of-network earthquakes. Rubinstein and Beroza (2007) developed a relocation technique to improve earthquake locations for sparsely recorded earthquakes using full-waveform information. Array methods have shown promise in locating offshore and out-of-network earthquakes in earthquake early warning systems (Eisermann *et al.*, 2018; Nof *et al.*, 2019; Netanel *et al.*, 2021; Jung *et al.*, 2023; Ziv *et al.*, 2024). In addition, Bayesian frameworks have been applied to improve earthquake location in non-optimal network configurations. For example, Zollo *et al.* (2021) combined arrival times, amplitude ratios, and back azimuths to detect offshore seismicity. Williamson *et al.* (2023) proposed the bEPIC algorithm, which integrates historical seismicity into ShakeAlert's location procedure using a Bayesian framework. By incorporating prior seismic activity to downweight high-error solutions in regions with little or no past seismicity, bEPIC significantly reduced mean location errors offshore northern California from 58 to 14 km, especially improving performance for offshore and out-of-network events.

The emergence of machine learning in seismology offers additional opportunities to advance earthquake location. Some studies have shown that deep learning models can estimate earthquake locations from single-station waveforms, challenging the conventional notion that a minimum of three seismic stations is necessary for triangulating and locating a seismic source (Perol *et al.*, 2018; Mousavi and Beroza, 2020; Elsayed *et al.*, 2023; Castro *et al.*, 2024). Other studies follow the general use of earthquake location with multiple stations (Zhang *et al.*, 2020; Chen *et al.*, 2022; McBrearty and Beroza, 2022, 2025; Zhu, Tai, *et al.*, 2022; Kuang *et al.*, 2024; Si *et al.*, 2024). By incorporating geographic locations of stations as node features, graph neural networks (GNNs; Gilmer *et al.*, 2017) have been applied to explore spatiotemporal relationships within seismic networks (Van Den Ende and Ampuero, 2020; Zhang *et al.*, 2022), enabling more accurate earthquake location with adaptability to diverse station geometries using one trained GNN model. However, these methods are often optimized for well-instrumented seismic networks and may underperform for offshore or sparsely instrumented regions.

To overcome these limitations, this study investigates the feasibility of using deep learning to locate out-of-network earthquakes with sparse data, addressing the inherent challenges of locating offshore earthquakes with limited azimuthal coverage. We propose the location neural operator (LocNO),

an operator learning framework (Kovachki *et al.*, 2023) designed to locate earthquakes directly from the full-waveform data recorded by seismic networks with arbitrary station geometries. Recent advancements in operator learning offer promising alternatives for processing the seismic data (Sun *et al.*, 2022, 2023; Aquib and Mai, 2024). Building on this progress, our approach adapts and extends classic operator learning frameworks (Li *et al.*, 2020, 2021) to enhance offshore earthquake location using the full-waveform seismic datasets from onshore stations.

To improve the offshore earthquake location accuracy in the MTJ region, we develop a training dataset by collecting historical seismicity data from the region and simulating sparse and nonideal station coverage scenarios. We randomly remove stations from cataloged earthquakes and generate an artificial out-of-network earthquake dataset for training purposes. By leveraging the spatiotemporal information of the seismic data recorded across multiple stations, LocNO is trained to directly estimate event locations, even under nonideal network conditions. The performance of the well-trained LocNO model demonstrates its capability to accurately locate offshore earthquakes beyond the boundaries of the onshore seismic network. LocNO provides locations for newly detected out-of-network earthquakes in the MTJ region that are recorded by only a few stations in the phase neural operator (PhaseNO)-detected catalog (Sun *et al.*, 2023), a setting that is challenging for travel-time-based location methods, as demonstrated by its application to an offshore sequence: the 22 September 2017 M_w 5.7 event located 218 km west of Ferndale, California. Incorporating LocNO into the standard seismic monitoring workflow has the potential to improve offshore location accuracy, thereby advancing our understanding of the seismotectonics in the MTJ region.

METHOD: LOCNO

We begin by introducing the fundamental principles underlying neural operators, which form the basis of the LocNO implemented in our public repository (Sun, 2025). For clarity, all notations are summarized in Table 1. Neural operators extend traditional neural networks by enabling mappings between functions defined over infinite-dimensional spaces. These models adhere to a version of the universal approximation theorem specifically formulated for operators (Kovachki *et al.*, 2023). One of the key applications of neural operators is solving partial differential equations (PDEs), which inspires their foundational design as approximations of inverse differential operators, typically represented by integral operators:

$$u(\mathbf{x}) = (\kappa * v)(\mathbf{x}) = \int \kappa(\mathbf{x}, \mathbf{y}) v(\mathbf{y}) d\mathbf{y}, \quad (1)$$

in which v and u are input and output functions of one integral operator layer $\mathcal{K}$, and κ denotes the integration kernel. Here, $\mathbf{x}$

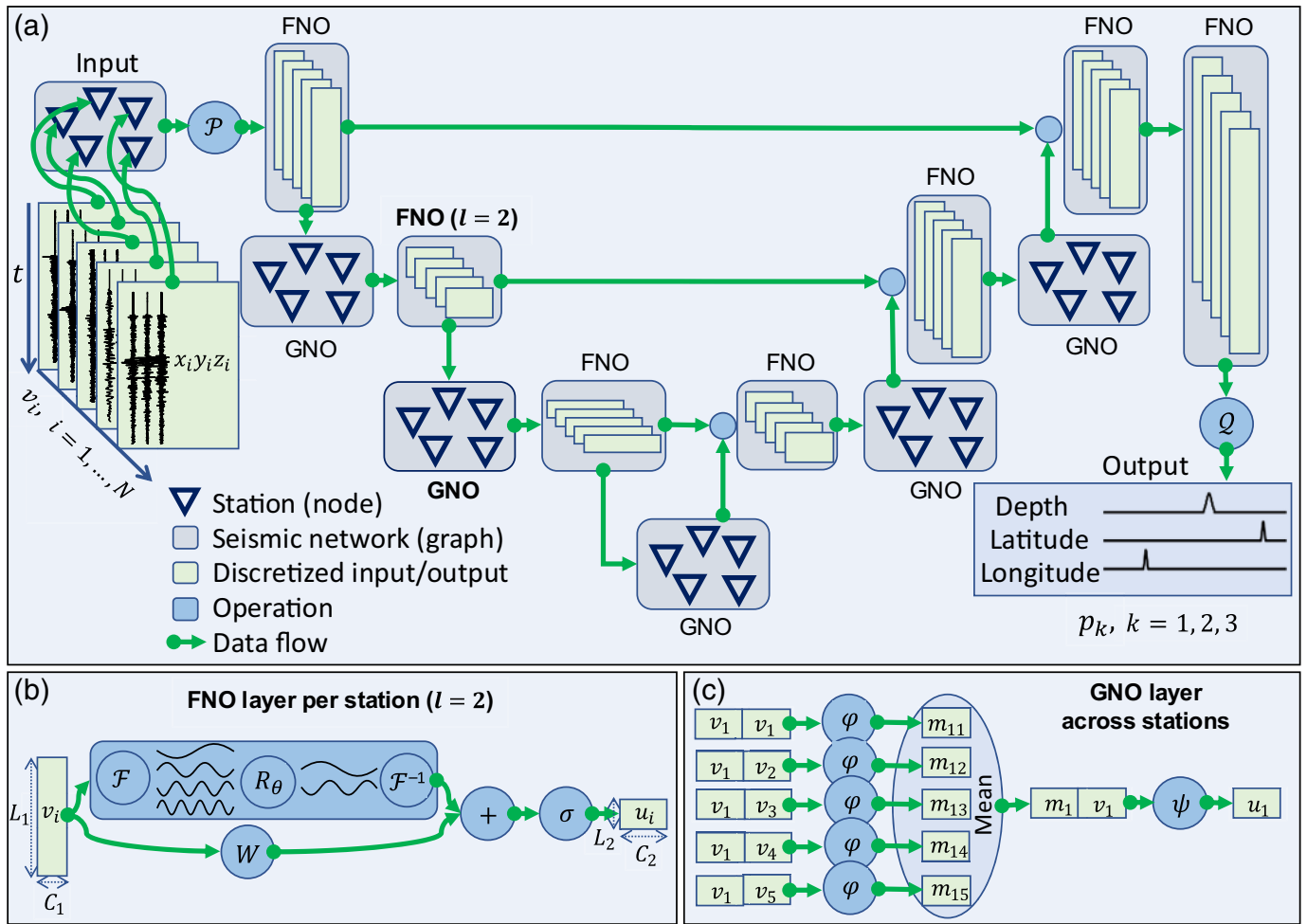
TABLE 1
Summary of Notations

Notation	Description
General Variables	
f	Input seismic waveform function
t	Time coordinate in the waveform input
C, L, N	Number of channels, discrete grid points, and stations, respectively
$i = 1, \dots, N$	Index of seismic station
$k = 1, 2, 3$	Index of spatial direction: $k = 1$ (easting), $k = 2$ (northing), and $k = 3$ (depth)
κ	Kernel for integral operator layer
v	Input of an operator layer
u	Output of an operator layer via integration with κ
σ	Activation function (e.g., GELU or sigmoid function for converting logits into probabilities)
$\mathcal{P}$	Lifting map to project f into a high-dimensional latent space
$\mathcal{U}$	U-shaped neural operator composed of FNO and GNO layers
h	Extracted spatiotemporal features via U-shaped neural operator $\mathcal{U}$
Fourier Neural Operator (FNO)	
$\mathcal{F}, \mathcal{F}^{-1}$	Fourier and inverse Fourier transforms
R_θ	Learnable spectral kernel in frequency domain parameterized by θ
W	Local linear transformation
M_l	Number of Fourier modes retained at FNO layer l
C_l, L_l	Number of channels and discretization points at FNO layer l
Graph Neural Operator (GNO)	
$\mathcal{N}(\mathbf{x}_i)$	Neighborhood of node $\mathbf{x}_i$
φ	Message function (e.g., two-layer MLP)
$m_{ij} = m(\mathbf{x}_i, \mathbf{x}_j)$	The edge features (message) computed using $\varphi(v(\mathbf{x}_i), v(\mathbf{x}_j))$ between the node $\mathbf{x}_i$ and $\mathbf{x}_j$, in which $\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)$ is a neighboring node of the station i
$m_i = m(\mathbf{x}_i)$	The aggregated message at station i
ψ	Update function (e.g., two-layer MLP)
D	Radius threshold for defining graph edges
(a_0, b_0, c_0)	Geographical origin (longitude, latitude, and depth) used as the reference point for normalization
(a_i, b_i, c_i)	Geographical location of station i (longitude, latitude, and depth) in physical coordinates
(A, B, C)	Physical range in longitude, latitude, and depth used for normalization
$\mathbf{x}_i = (x_i, y_i, z_i)$	Input spatial coordinates of station i , normalized to the range $[0, 1]$
Decoder and Output	
$g_i(t) \in \mathbb{R}^{C \times L}$	Spatiotemporal feature embedding at station i
$s_i(n) \in \mathbb{R}^{3 \times L}$	Directional score from station i
$s(n) \in \mathbb{R}^{3 \times L}$	Aggregated directional score over space
$\mathcal{Q}$	Station wise decoder mapping $g_i(t)$ to $s_i(n)$
n_k^ℓ	The ℓ th discrete grid point along the k th spatial direction
$s_k^\ell = s(n_k^\ell)$	Logit score at discrete grid point ℓ in the spatial direction k
$p_k(n_k^\ell)$	Output probability at discrete grid point ℓ in the spatial direction k
$\mathcal{A}$	Probability aggregator across stations
α_i	Attention weight for station i , satisfying $\sum_i \alpha_i = 1$
Loss Function	
$p_k^*(n_k^\ell)$	Ground-truth probability at discrete grid point ℓ in the spatial direction k
w_k	Loss weight for direction k
$\mathcal{L}_{\text{LocNO}}$	Total loss defined as weighted binary cross-entropy across all directions and grid points

and $\mathbf{y}$ denote coordinates (e.g., space, time, or other domain variables) in the output and input domains, respectively.

Our approach extends the applicability of neural operators beyond their traditional scope of solving PDEs by extracting features from seismic wavefields and predicting a pseudo-probabilistic function indicating the earthquake location. For seismic wavefields, both $\mathbf{x}$ and $\mathbf{y}$ are 4D vectors representing spatiotemporal sampling points, and $\kappa(\mathbf{x}, \mathbf{y})$ is a learnable spatiotemporal kernel that governs the integration over

neighboring points. Each spatiotemporal point $\mathbf{x} \in \mathbb{R}^4$ consists of three spatial coordinates and one time coordinate. For clarity, we treat time t separately from $\mathbf{x}$. At the i th station (in which $i = 1, \dots, N$), the input waveform is expressed as $v_i(t) = v(\mathbf{x}_i, t)$, in which $\mathbf{x}_i$ represents the 3D spatial location of the station. In practice, $v_i(t)$ is not only the three-component ground-motion time series but also includes the spatial coordinates $\mathbf{x}_i$ concatenated as three additional channels, so that station location is explicitly embedded into the input



representation. This time-dependent function serves as the input representation for station i in a neural operator layer. The corresponding learned feature representation after a neural operator layer is denoted as $u_i(t) = u(\mathbf{x}_i, t)$.

Various methods can be employed to parameterize the kernel function κ and implement $\mathcal{K}$ (Kovachki et al., 2023). Seismic data collected by a network of multiple stations can be viewed as a function discretized regularly in time at each station and irregularly in space due to the nonuniform distribution of station locations. Thus, we use graph neural operators (GNOs) to represent the spatially irregular discretization and Fourier neural operators (FNOs) to handle the temporally regular discretization. Moreover, applying a kernel integral fully over the 4D spatiotemporal domain is computationally inefficient and unnecessary because only a sparse set of seismic stations provides observed 1D time-series data. Therefore, we decouple temporal and spatial feature extraction: temporal features are modeled using 1D FNOs applied at individual stations, whereas spatial dependencies across stations are captured using GNOs, where the temporal signal at each station is considered as the node features in the computational graph. Consequently, LocNO integrates both FNO and GNO as the foundation of its architecture (Fig. 1). In addition, FNO and GNO layers are interleaved, allowing

Figure 1. Location neural operator (LocNO) architecture: (a) overview of the LocNO model. The feature update process in (b) the second Fourier neural operator (FNO) layer ($l = 2$), and (c) the following graph neural operator (GNO) layer. $\mathcal{P}$, lifting map; $\mathcal{Q}$, decoder map; $\mathcal{F}$, Fourier transform; $\mathcal{F}^{-1}$, inverse Fourier transform; R_θ , learnable spectral kernel; W , local linear transformation; σ , activation function; ϕ , message function; ψ , update function; v_i , input feature at station i to one layer; u_i , output feature at station i from one layer; $L_l \times C_l$, feature dimension at the l th FNO layer; m_{ij} , edge feature between stations i and j ; m_i , aggregated message feature at station i ; $\mathbf{x}_i = (x_i, y_i, z_i)$, spatial coordinates of station i ; and p_k , pseudo-probability distribution of earthquake location. The model combines 1D FNOs to extract temporal features at each station and GNOs to capture spatial dependencies across stations. For clarity, only five stations are illustrated ($N = 5$). See Table 1 and main text for full descriptions. The color version of this figure is available only in the electronic edition.

temporal and spatial features to be progressively exchanged and jointly encoded through successive layers.

FNO

The FNO computes the kernel κ in the frequency domain by transforming the input into the frequency domain with fast Fourier transforms (FFT) (Li et al., 2021). Because the FFT requires regular spacing, the FNO is best suited for inputs

defined on regular grids, such as time series recorded at individual seismic stations. According to the convolution theorem:

$$u_i(t) = \mathcal{F}^{-1}(R_\theta \mathcal{F}(v_i))(t), \quad (2)$$

in which $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the FFT and inverse Fourier transform, respectively, and R_θ is the Fourier-domain representation of the kernel, learned via parameters θ . During training, the model effectively learns the frequency spectrum of $u_i(t)$ through the multiplication in the Fourier domain. However, learning high-frequency components can be more challenging because high-frequency signals are often noisy. Therefore, in practice, only the first M lowest-frequency modes are retained and learned, whereas high-frequency modes are truncated (Li *et al.*, 2021). This strategy helps reduce overfitting risk, improve numerical stability, lower training difficulty, and decrease model complexity. Incorporating a local activation function σ , the output of one FNO layer becomes

$$u_i(t) = \sigma(Wv_i(t) + \mathcal{F}^{-1}(R_\theta \mathcal{F}(v_i))(t)), \quad (3)$$

with W being a pointwise linear transformation. Each FNO layer includes two branches (Fig. 1): one performs the Fourier-based global convolution, whereas the other applies the local linear transformation W . The outputs are combined before the activation function. The purpose of introducing a nonlinear activation function after each integral operator is to enable the network to approximate complex, nonlinear mappings. Without such nonlinearities, the composition of multiple integral operators would collapse into a single linear operator, severely limiting the model's expressive power. The activation function σ used in LocNO is the Gaussian error linear unit (GELU; Hendrycks and Gimpel, 2016).

LocNO is structured as a U-shaped neural operator, in which changes in feature dimensions are governed by the FNO layers applied at each spatially sampled station. Each FNO layer models the long-range temporal dependencies present in seismic waveforms by performing 1D global convolution along the time dimension in the frequency domain. Specifically, the input sequence at each station is first transformed via the FFT, followed by a complex-valued multiplication with a learnable spectral kernel R_θ , and then mapped back via the inverse FFT. Operating in the Fourier domain allows the network to capture global temporal patterns efficiently, enabling it to integrate full-waveform information that is critical for improving the robustness and accuracy of hypocenter estimation. At station i , the learned feature representation u_i at layer l has shape $C_l \times L_l$, in which C_l denotes the number of channels and L_l represents the number of discretization points along the primary axis (e.g., input time). LocNO consists of seven FNO layers in total. The value of L_7 is related to the number of spatial bins in the predicted probability distribution and thus constrains the spatial resolution of the estimated source

location along the easting, northing, and depth directions. Unless otherwise specified, we adopt the following parameter settings for the feature dimensions and the number of retained Fourier modes M_l across FNO layers:

$$\{L_l\}_{l=1}^7 = \{3000, 750, 200, 750, 3000, 3000, 6000\}, \quad (4)$$

$$\{C_l\}_{l=1}^7 = \{48, 96, 192, 96, 48, 48, 48\}, \quad (5)$$

$$\{M_l\}_{l=1}^7 = \{24, 12, 8, 8, 12, 24, 24\}. \quad (6)$$

In this U-shaped neural operator architecture, the number of Fourier modes M_l is progressively reduced in proportion to the downsampling of the feature dimension along the primary axis. Concurrently, the number of channels is increased to compensate for the loss of feature resolution and to enrich the feature representation in deeper layers (Fig. 1).

GNO

The kernel integration κ in the spatial domain is computed through message-passing graph neural networks (GNNs) (Li *et al.*, 2020; McBrearty and Beroza, 2023). In our method, we treat the spatially irregular structure of the seismic data as a graph: seismic stations are nodes, and their connectivity forms the graph edges. Given node features $v(\mathbf{x}_i)$ computed by an FNO layer at individual stations, a GNO layer updates the value $v(\mathbf{x}_i)$ of node $\mathbf{x}_i$ to a new representation $u(\mathbf{x}_i)$ through an averaging aggregation:

$$u(\mathbf{x}_i) = \psi\left(v(\mathbf{x}_i), \frac{1}{|\mathcal{N}(\mathbf{x}_i)|} \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} \varphi(v(\mathbf{x}_i), v(\mathbf{x}_j))\right), \quad (7)$$

in which $\mathcal{N}(\mathbf{x}_i)$ denotes the set of neighboring nodes of $\mathbf{x}_i$. The edge features $m_{ij} = m(\mathbf{x}_i, \mathbf{x}_j)$ are computed using $\varphi(v(\mathbf{x}_i), v(\mathbf{x}_j))$, in which φ is a differentiable function that maps the node features from a pair of connected nodes concatenated along the channel axis to a latent edge representation. In LocNO, φ is implemented as a two-layer multilayer perceptron (MLP) with a hidden layer size of $4C_l$, in which C_l is the channel dimension of the node features produced by the l th FNO layer preceding the l th GNO layer (Fig. 1).

Once edge features are generated, each node aggregates messages from its neighbors by averaging, and then concatenates the aggregated message $m_i = m(\mathbf{x}_i)$ with its own original features $v(\mathbf{x}_i)$ along the channel axis. This concatenated tensor is passed through another differentiable function, $\psi(m(\mathbf{x}_i), v(\mathbf{x}_i))$, also implemented as a two-layer MLP with the same architecture as φ . The resulting $u(\mathbf{x}_i)$ is the updated node representation. This message-passing paradigm allows information to be exchanged across neighboring nodes and supports the learning of spatial correlations among seismic stations (Fig. 1).

GNO adopts a message-passing framework similar to GNNs, but is specifically designed for operator learning, enabling the approximation of the kernel integration and the generalization across different discretizations. In GNO, the graph is constructed based on a physical distance threshold D defined in the input physical space, to approximate integration over that domain. For any given node, its neighborhood set is composed of all discretized points located within a radius D centered at that node. In the context of spatial domains such as seismic networks, this means that the number of neighbors corresponds to the number of seismic stations located within that physical radius (geographic distance). Unlike radius-based graphs, k-NN graphs (Franceschi *et al.*, 2019) are not suitable for GNO layers because they may restrict connectivity to local clusters, which undermines the uniform spatial coverage needed to accurately approximate kernel integrals.

Within a specified threshold D , nodes are fully connected to form a graph, including self-loops where each node is connected to itself. This threshold serves as a tunable hyperparameter before training. Although seismic stations are irregularly distributed in physical space, increasing the distance allows more stations to be included within a single graph, which consequently increases computational cost due to the increased number of node-to-node communications. In this study, we use a distance threshold of 300 km to account for the long-range wave propagation from offshore earthquakes to onshore stations. This ensures that all stations can effectively exchange information to accurately determine hypocenter locations.

At each GNO layer, we concatenate the normalized node coordinates with the output from the preceding FNO layers to form the node attributes $v(\mathbf{x}_i)$. The geographical locations of all stations in the study area are normalized to the range $[0, 1]$ before concatenation, assuming a geographical origin of (a_0, b_0, c_0) and a physical range of (A, B, C) , representing longitude, latitude, and depth, respectively. If the geographic location and depth of one station is (a_i, b_i, c_i) , the normalized node coordinates are computed as

$$x_i = \frac{a_i - a_0}{A}, \quad y_i = \frac{b_i - b_0}{B}, \quad z_i = \frac{c_i - c_0}{C}. \quad (8)$$

We use three channels to encode the node location information in both the input layer and the GNO layers (Fig. 1). In each channel, the corresponding normalized coordinate is repeated along the primary discretization dimension, allowing it to be concatenated with either the waveform input or the FNO output features. This strategy ensures consistent spatial encoding for all stations in the study area and enables the model to learn spatial relationships among seismic stations, which is crucial for accurate hypocenter location.

LOCNO

The LocNO is a hybrid architecture that combines the neural operator $\mathcal{U}$ to extract the spatiotemporal features from seismic

wavefields with a decoding module $\mathcal{Q}$ to estimate earthquake locations (Fig. 1). The architecture of $\mathcal{U}$ integrates seven 1D FNO layers to extract temporal features at each station and five GNO layers to capture spatial dependencies across stations. The feature dimensions change across FNO layers but remain constant across GNO layers. These layers are sequentially connected; however, the input to deeper FNO layers is formed by concatenating the features from earlier FNO layers with the output of the corresponding deeper GNO layers, thereby introducing skip connections. Such skip connections help preserve low-level temporal information and facilitate gradient flow during training (He *et al.*, 2016). The model is capable of processing seismic data with an arbitrary number of stations and flexible station geometries. For clarity, Figure 1 illustrates only five stations. The bottom panels in Figure 1 depict the feature update process in the second FNO layer ($l = 2$) and the second GNO layer. In this figure, the five stations ($N = 5$) are fully connected in the computational graph of the GNO layer; however, the actual graph connectivity is determined by the interstation distances and the threshold D . The message aggregation and update process is illustrated for the station $i = 1$, but the same procedure is applied to all other stations. The output is a pseudoprobability distribution of the earthquake source location along the easting, northing, and depth directions, with the index of the highest probability corresponding to the ground truth in each direction.

After discretization, the input function f is represented as a tensor of shape $C \times L \times N$, in which C is the number of channels, L is the number of temporal samples, and N is the number of stations. Each station provides three-channel waveforms, either recorded along three orthogonal directions or repeated single-channel measurements. In addition, the spatial coordinates $\mathbf{x}_i = (x_i, y_i, z_i)$ of each station are repeated L times and concatenated to the time series as three additional channels, encoding positional information directly into the temporal feature input $v_i(t)$. The spatiotemporal feature $h \in \mathbb{R}^{C_7 \times L_7 \times N}$ extracted by the U-shaped neural operator $\mathcal{U}$ is given by

$$h = \mathcal{U}[\mathcal{P}(f)], \quad (9)$$

in which $\mathcal{P}$ is a lifting operator that maps the input function f into a higher-dimensional representation $\mathcal{P}(f) \in \mathbb{R}^{C_1 \times L_1 \times N}$. $\mathcal{P}$ in LocNO is a single fully connected layer, which maps the discretized input function to a latent space with C_1 channels. This lifting operation is applied locally at each spatial or temporal location, with shared parameters across the entire input domain. Such parameter sharing ensures that $\mathcal{P}$ behaves as a proper operator, independent of the discretization of the input function.

The architecture of $\mathcal{U}$ integrates FNO and GNO to leverage their complementary strengths. Similar to PhaseNO (Sun *et al.*, 2023), the FNO component establishes a U-shaped architecture, whereas the GNO facilitates information exchange among multiple stations. To enhance the exchange of features between the temporal and spatial domains, a GNO layer is strategically placed between two FNO layers.

We treat the output of the neural operator $h_i \in \mathbb{R}^{C_7 \times L_7}$ at the station i as a latent embedding that implicitly captures the station's spatiotemporal response to a seismic event. The following layer of the model is a shared decoder $\mathcal{Q}$ which maps the station-level latent features to directional score vectors over discretized spatial coordinates. Specifically, the latent embedding h_i is decoded into a three-channel output corresponding to the easting, northing, and depth directions:

$$s_i(n) = \mathcal{Q}(h_i) \in \mathbb{R}^{3 \times L}, \quad (10)$$

in which each channel of $s_i(n)$ represents a discrete score vector over one spatial axis, defined on L grid points $\{n_k^\ell\}_{\ell=1}^L$ for direction $k = 1, 2, 3$. $\mathcal{Q}$ in LocNO is implemented as a two-layer fully connected network with an input width (channel) of 48, a hidden layer of size 96, and an output layer with three channels.

To obtain the score vector $s(n)$ for all stations, we aggregate the station-wise estimates $\{s_i(n)\}_{i=1}^N$ using a structured attention pooling operator $\mathcal{A}$. Specifically, for each station i , a scalar attention weight is computed by applying a convolutional neural network (CNN) with a kernel size of one over the score $s_i(n)$. The resulting attention logits are averaged over the feature dimensions to yield a single scalar per station, and then normalized across stations via a softmax function to produce the attention weights. The total score is then given by

$$s(n) = \mathcal{A}(\{s_i(n)\}_{i=1}^N) = \sum_{i=1}^N \alpha_i s_i(n) \in \mathbb{R}^{3 \times L}, \quad (11)$$

in which $\alpha_i \in [0, 1]$ is the normalized attention weight assigned to station i , satisfying $\sum_{i=1}^N \alpha_i = 1$. As a special case, uniform averaging corresponds to setting $\alpha_i = \frac{1}{N}$ for all i , which is equivalent to aggregating the station-level predictions via mean pooling.

Finally, for each direction k , the score $s_k^\ell = s(n_k^\ell)$ is transformed into a pseudoprobability distribution by applying the sigmoid function σ independently to each entry

$$p_k(n_k^\ell) = \sigma(s_k^\ell) = \frac{1}{1 + e^{-s_k^\ell}} \quad \text{for } \ell = 1, \dots, L. \quad (12)$$

The loss function is defined as the sum of binary cross-entropy losses between the predicted and true probability distributions, applied independently to each spatial direction k and grid point ℓ . The ground-truth distribution in each direction $p_k^*(n_k) \in [0, 1]$ is modeled as a truncated Gaussian centered at the true source coordinate, allowing the network to learn a continuous approximation over the discretized spatial grid. The loss is defined as

$$\mathcal{L}_{\text{LocNO}} = \sum_{k=1}^3 w_k \sum_{\ell=1}^L [-p_k^*(n_k^\ell) \log p_k(n_k^\ell) - (1 - p_k^*(n_k^\ell)) \log(1 - p_k(n_k^\ell))], \quad (13)$$

in which w_k is a user-defined weight for direction k . This formulation treats each grid point as an independent binary classification task and supports soft target distributions, encouraging the model to produce accurate pseudoprobability densities over the discretized spatial grid in each direction.

The LocNO model is implemented using PyTorch, with custom modules for neural operator layers. Optimization is performed using the Adam optimizer with an initial learning rate of 1×10^{-4} . Adam adaptively adjusts learning rates for individual parameters using estimates of the first and second moments of the gradients, which improves convergence stability and efficiency. To further refine training, the learning rate is reduced by a factor of 0.1 when the validation loss does not improve for 10 consecutive epochs, where one epoch refers to a complete pass through the entire training dataset. This scheduler helps the model escape flat regions or suboptimal local minima during training. Model training and evaluation are conducted on a high-performance computing cluster equipped with NVIDIA graphics processing units, which enables efficient processing of large-scale seismic datasets.

DATA

To test the algorithm and to locate offshore earthquakes in the MTJ region, we use a dataset composed of earthquake waveforms and event catalogs around the MTJ region spanning three decades (Zhu *et al.*, 2025), sourced from the Northern California Earthquake Data Center (NCEDC). Many downloaded earthquake data based on the NCEDC catalog are in-network events, where seismic stations are distributed around the earthquake's location. The available out-of-network earthquakes are insufficient in quantity to create a large and diverse training dataset for deep learning models.

To enable the LocNO model to handle offshore and out-of-network earthquakes, we adopt a strategy to artificially generate additional out-of-network events from the existing in-network events. This is achieved by systematically removing selected seismic stations from the network surrounding an in-network earthquake. By strategically reducing the station coverage in specific areas, we simulate scenarios where the earthquake occurs outside the dense monitoring regions, effectively creating synthetic out-of-network events. These generated samples are then incorporated into the training dataset, enhancing the model's ability to generalize to real-world cases of out-of-network earthquakes.

Figure 2 illustrates the method for constructing an out-of-network earthquake sample from an in-network event. The process begins by positioning the earthquake's location at the center of an easting–northing coordinate system. A series of straight lines with slopes ranging from -1 to 1 is systematically drawn through this central point. For each line, the seismic stations are counted and grouped based on their positions relative to the line (either above or below). The line that minimizes the total number of seismic stations on one side is identified.

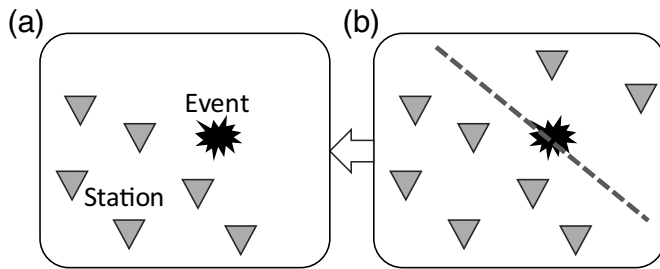


Figure 2. Diagram illustrating the procedure for generating (a) an out-of-network earthquake sample from (b) an in-network event. Starting at the earthquake's location in an easting–northing coordinate system, draw a straight line and count the seismic stations on each side. Repeat the process with lines of varying slopes from -1 to 1 . Choose the line where one side has the fewest seismic stations and remove those stations. The remaining stations, all located on one side of the earthquake's position, form the out-of-network earthquake sample.

Stations on the less-populated side are excluded from further analysis. This approach ensures that the remaining seismic stations are all located on one side of the earthquake's position, effectively isolating the event from a portion of the seismic network and generating a well-defined out-of-network earthquake sample. Furthermore, we include the resulting earthquake sample in the training dataset only if the out-of-network earthquake is recorded by at least four stations.

Figure 3 shows the map of all out-of-network earthquakes and the remaining stations in the resulting dataset. All stations

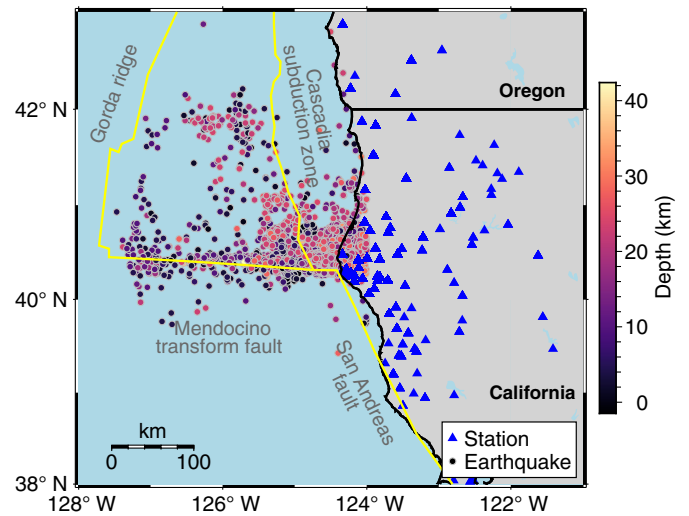


Figure 3. Earthquakes near the Mendocino Triple Junction (MTJ) region from 1987 to 2021. Plate boundaries are drawn following Bird (2003). The color version of this figure is available only in the electronic edition.

are onshore, whereas most earthquakes are offshore and classified as out-of-network earthquakes for these onshore stations. The dataset was split chronologically, with the events occurring before 2016 used for training, the events from 2016 for validation, and those from 2017 to 2021 for testing (Fig. 4). This resulted in 2486 training samples, 209 validation samples, and 911 test samples. We remove microseismicity below a magnitude of one from this dataset due to the large

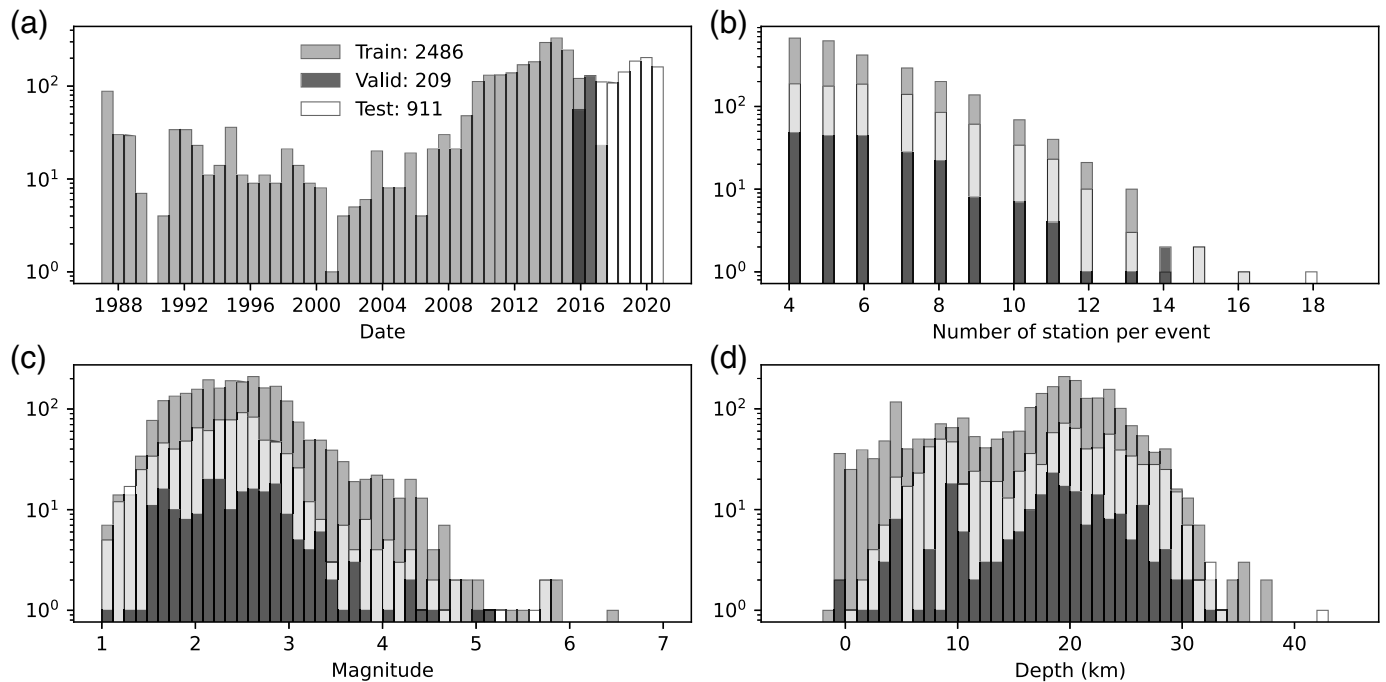
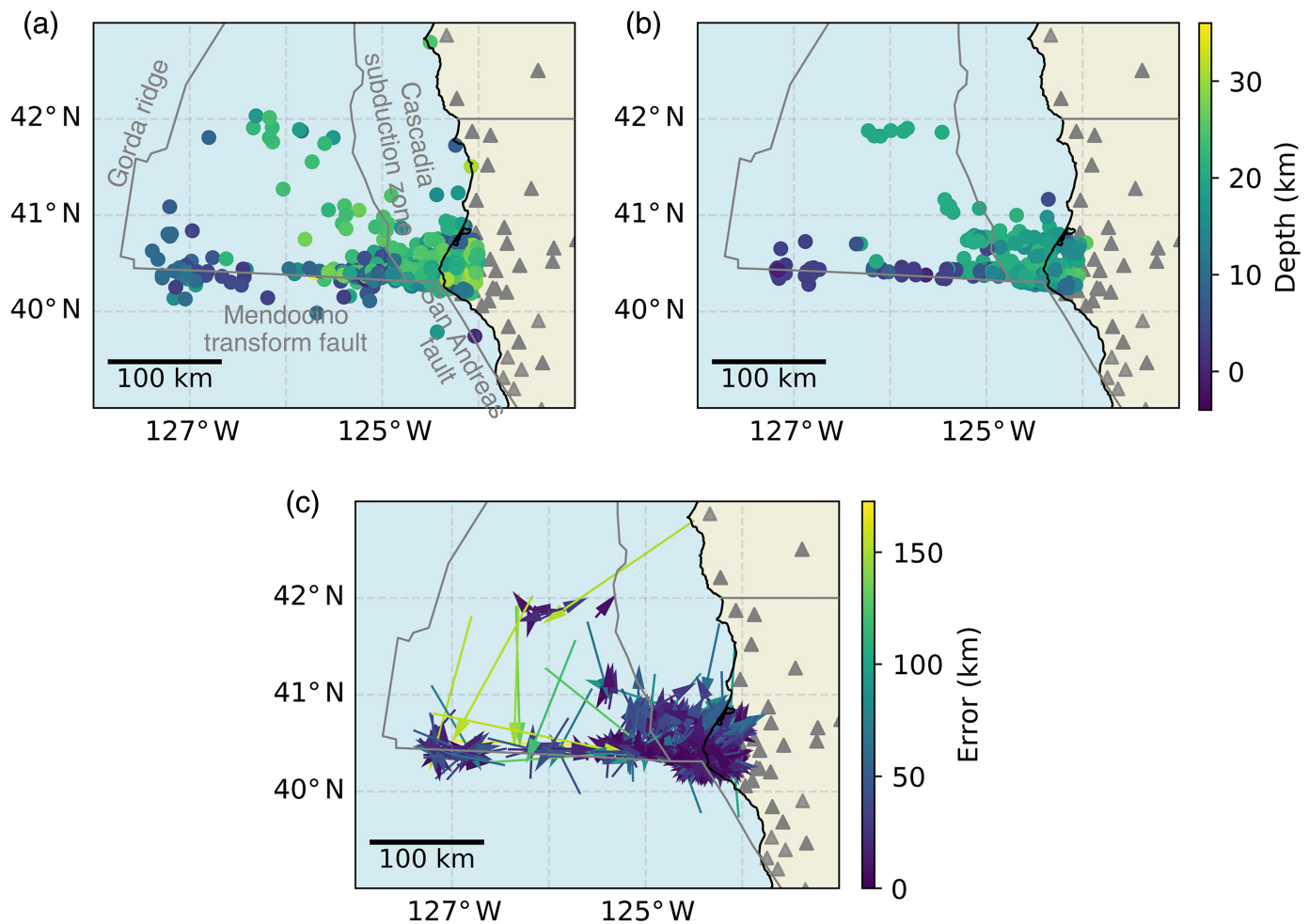


Figure 4. (a–d) Dataset of out-of-network earthquakes in the MTJ region for LocNO training and evaluation: We split the dataset based on their time:

earthquakes before 2016, in 2016, and from 2017 to 2021, respectively, into the training, validation, and test datasets.



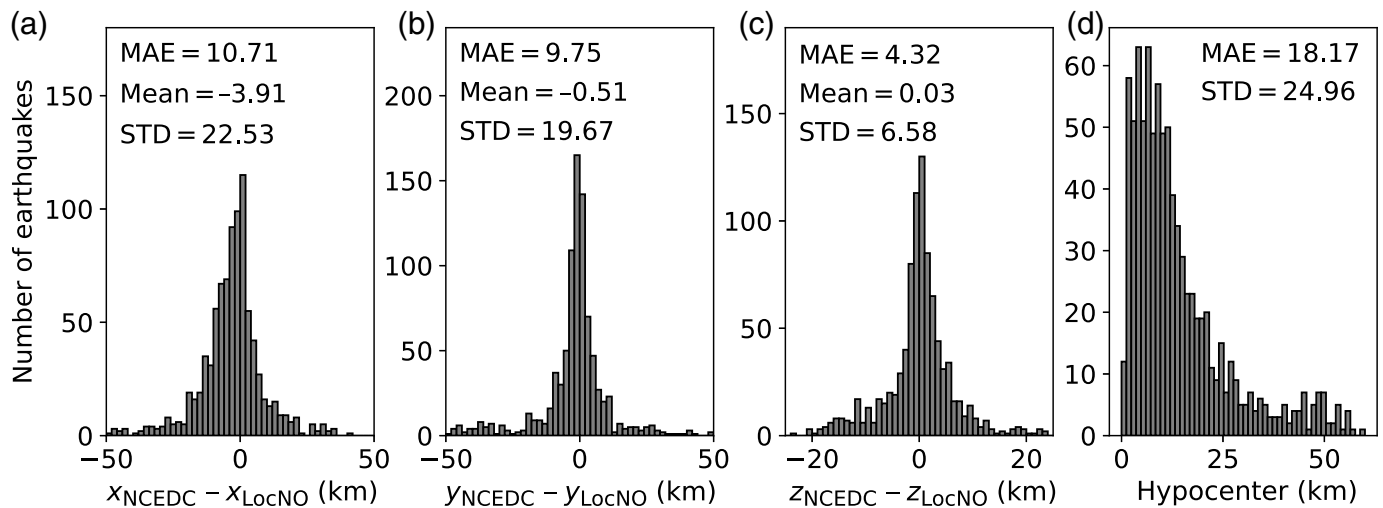
offset between earthquake hypocenters and onshore stations. Both the training and test datasets cover a wide range of depth from surface to 40 km and magnitudes from 1 to 6 (Fig. 4). Most events have magnitudes between 1 and 5, at depths of 0–30 km, and were recorded by 4–13 stations.

LocNO is designed to be used in conjunction with external earthquake detection algorithms as part of a seismic monitoring workflow. Once an event is detected, LocNO is applied to estimate its location using waveforms from the stations where seismic signals have been identified. We use a waveform length of 30 s for all input stations, unless otherwise noted, where we study the impact of alternative lengths (5, 10, and 20 s) on performance. Because the input is event-based, the *P*-wave arrival exhibits moveout and therefore varies across stations. To prevent the model from relying on the absolute arrival time for location, we randomly place the earliest *P*-wave arrival within a window between 1 and 5 s from the beginning of the input waveform during training. During testing, the earliest arrival time is fixed at 1 s. In addition, in practical applications, it is common that a small subset of stations in a network may record clipped, low SNR, or otherwise unusable data. To improve the model's robustness, we explicitly simulated this condition during training by randomly including virtual stations that contain only noise in each training sample. This

Figure 5. Comparison of earthquake locations between (a) the Northern California Earthquake Data Center (NCEDC) catalog and (b) the LocNO prediction in the out-of-network earthquake test dataset. (c) Epicentral location errors illustrated with arrows pointing from catalog locations to the corresponding LocNO-predicted locations. The color version of this figure is available only in the electronic edition.

approach allows LocNO to learn to downweight or ignore stations with uninformative or misleading input. To simulate real-world scenarios where not all stations record high-quality earthquake signals, we augmented the dataset with up to two virtual stations placed randomly within the boundary of stations. Noise waveforms were randomly selected from the Stanford EArthquake Dataset (Mousavi *et al.*, 2019), ensuring sample-specific variability. Note that in practical applications, the majority of associated stations for a detected earthquake are expected to contain meaningful signals. Therefore, it is not necessary to add many virtual stations, particularly in this dataset where many samples include only four stations (Fig. 4). All waveforms were preprocessed with trend removal, band-pass filtering (1–10 Hz), and normalization.

The study area for the MTJ region (Fig. 3) is defined as a geographical domain spanning 8° in both longitude and latitude. The longitude range extends from -128° to -120° , and



the latitude range covers from 35° to 43° . In addition, the depth ranges from -5 to 45 km, including the depth of stations above the sea level and the maximum depth of earthquakes in the dataset. The domain's coordinates in the three directions were independently normalized to $[0,1]$ for compatibility with the learning process. Normalized positions of stations (x_i, y_i, z_i) are included in the node attributes along with three-component waveform data.

The sample label is generated based on catalog locations and consists of three channels, each representing a 1D probability function. These channels indicate the likelihood of the earthquake's occurrence along the easting, northing, and depth directions. Each channel is modeled as a truncated Gaussian function centered on the true location, where the real position has a maximum probability of one, gradually decaying to zero. For longitude and latitude, the probability function spans a width equal to $1/30$ of the total range, whereas for depth, it spans a wider width of $1/12$ of the total range. This broader Gaussian function for depth accounts for the greater uncertainty in determining earthquake depth using surface observations, particularly for out-of-network events.

RESULTS

Locating offshore earthquakes in the MTJ region

The LocNO model was evaluated after being trained for 20 epochs. Figure 5 compares the earthquake locations determined by the trained LocNO model with those from the NCEDC catalog using the out-of-network earthquakes in the test dataset. The results demonstrate that LocNO provides location estimates aligned with the NCEDC catalog. Notably, the model achieves higher accuracy for events near the Mendocino transform fault, showcasing its effectiveness in this tectonically active region.

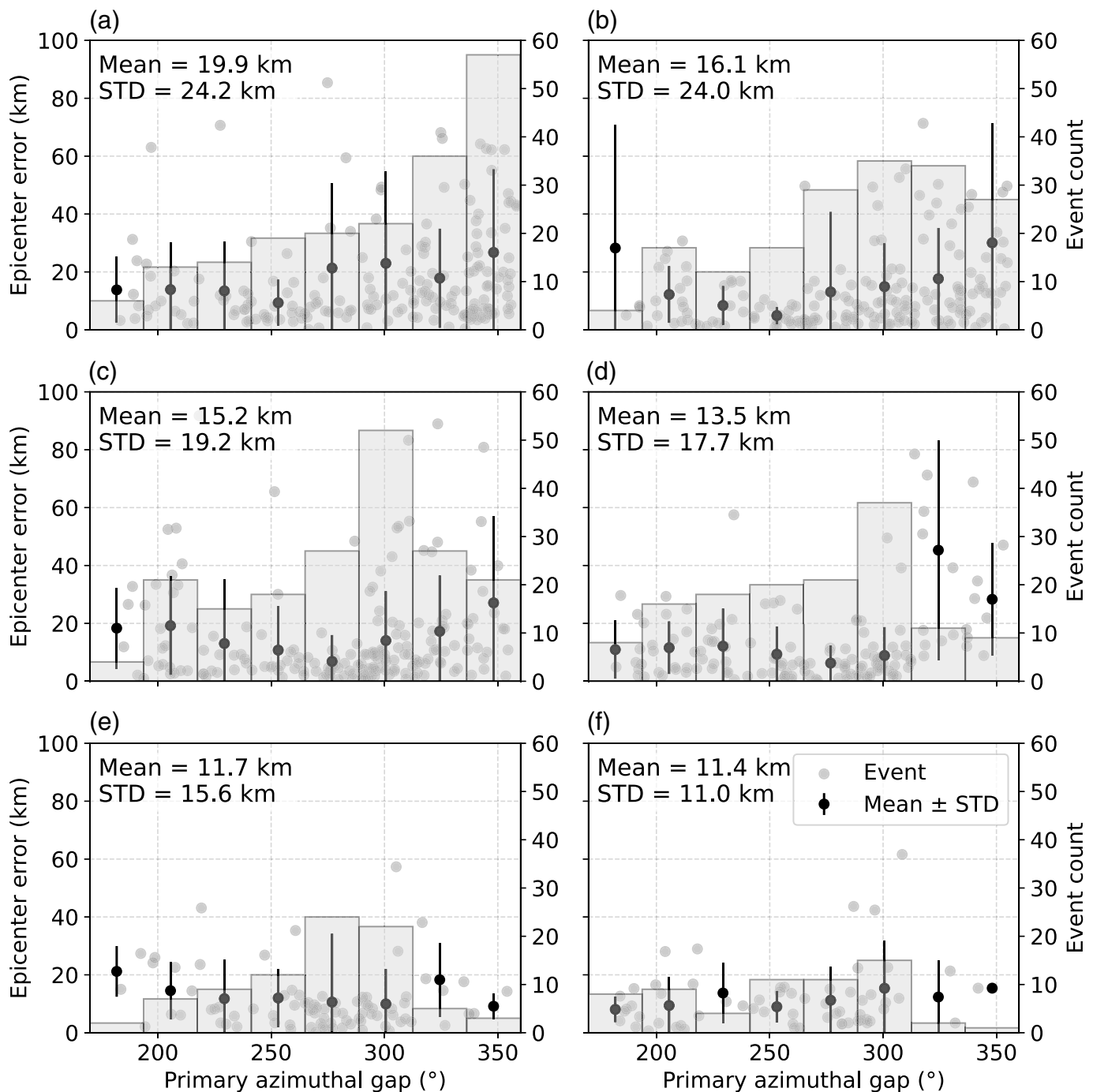
However, larger discrepancies are observed for some events located farther from the Mendocino transform fault. These larger differences can be attributed to the limited number of prior seismic events in those areas available for training. As

Figure 6. Histograms of location errors between LocNO and the NCEDC catalog for out-of-network earthquakes in the test dataset, shown separately for (a) easting (x), (b) northing (y), (c) depth (z), and (d) total hypocenter error. Evaluation metrics include mean absolute error (MAE), mean error (Mean), and standard deviation (STD), all in kilometers (km).

a supervised learning method, LocNO learns spatiotemporal features by capturing statistical relationships among multistation waveform patterns. Its generalization ability is, therefore, the strongest in regions where the training data are sufficiently dense and representative. In this study, the training catalog contains relatively few well-recorded events in the offshore region far from the fault, which limits the model's ability to learn reliable waveform–location relationships for those areas.

Figure 6 shows a quantitative comparison of location errors in easting, northing, and depth, evaluated using the metrics of mean absolute error (MAE), mean error (Mean), and standard deviation (STD). Location errors in degrees were converted to kilometers using a conversion factor of 111 km° for latitude and $\sim 85 \text{ km}^{\circ}$ for longitude, consistent with the geographic position near the Mendocino transform fault. The MAE of the location errors is 10.71 km in easting, 9.75 km in northing, and 4.32 km in depth. LocNO provides location estimates within 20 km of the NCEDC catalog for 75% of the events in the test dataset (Fig. 6). These location errors are reasonable given that most events are located using only 4–6 stations (Fig. 4). The distribution of seismic stations, which is sparse along the easting direction but relatively uniform along the northing direction (Fig. 3), leads to larger location errors in easting than northing.

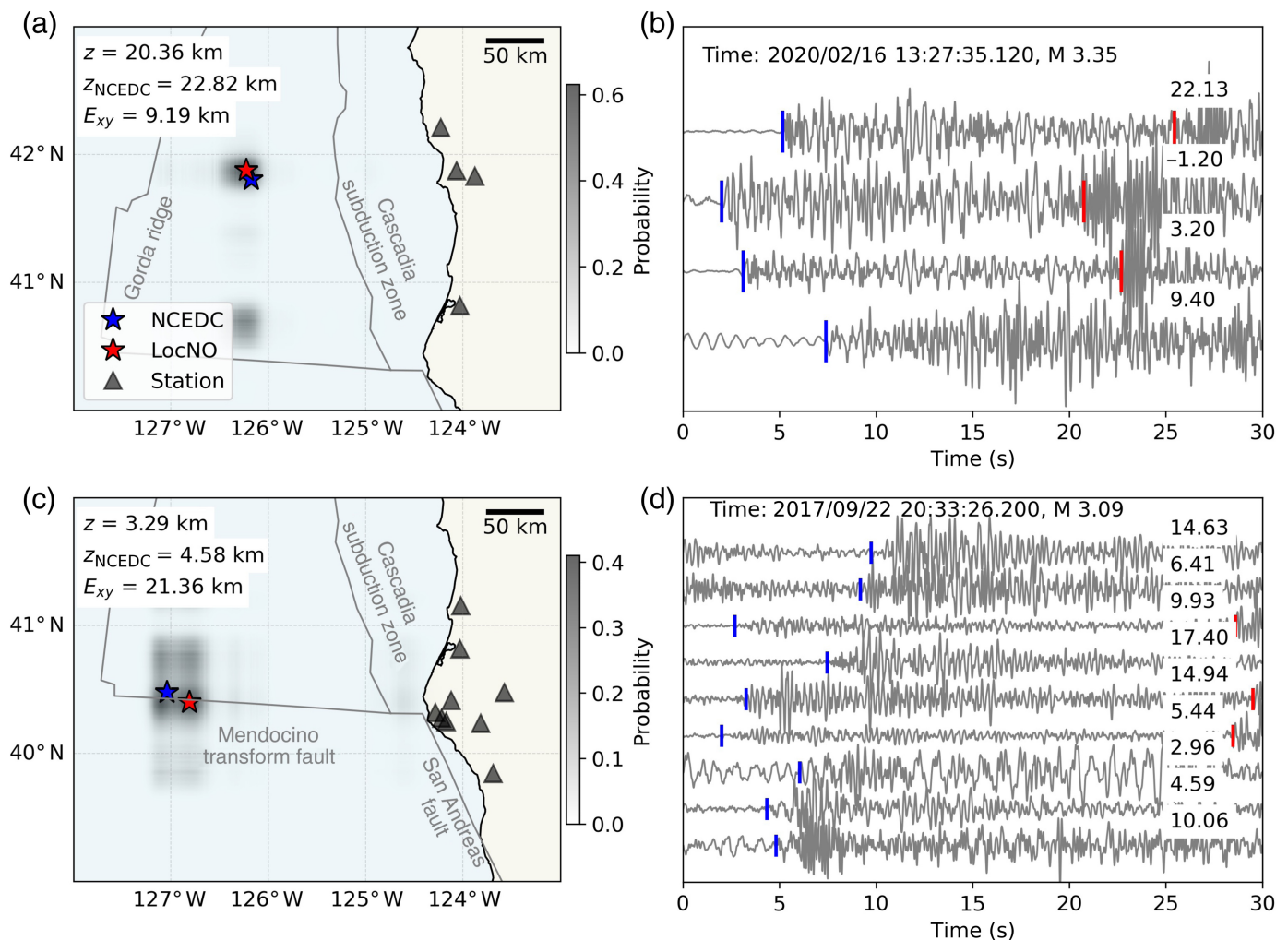
Figure 7 shows the epicenter errors between LocNO predictions and the NCEDC catalog for events located using 4–9 stations in the test dataset. Each panel corresponds to a fixed number of stations and presents epicenter errors as a function of the primary azimuthal gap. The minimum primary azimuthal gap is 180° for all samples due to the inclusion of out-of-network events. We bin the data into azimuthal gap ranges



from 180° to 360° and compute the mean error (Mean) and standard deviation (STD) of epicenter errors in each bin. These statistics are most informative when sufficient samples exist in a bin. As described more fully subsequently, this figure demonstrates that increasing the number of stations and reducing the azimuthal gap improves the earthquake locations.

LocNO demonstrates the ability to handle out-of-network scenarios and maintains reliable performance even as the azimuthal gap increases. Although epicenter errors tend to increase slightly with larger azimuthal gaps, particularly when only a few stations are available, the overall performance

Figure 7. Epicenter errors as a function of the primary azimuthal gap, comparing LocNO predictions with the NCEDC catalog for events located using (a) 4, (b) 5, (c) 6, (d) 7, (e) 8, and (f) 9 stations in the test dataset. Histograms show the number of samples in each range of the primary azimuthal gap. The mean (Mean) and standard deviation (STD) of epicenter errors within each bin are labeled on the histograms. The overall Mean and STD for all samples in each panel are noted at the top of the corresponding subplot.

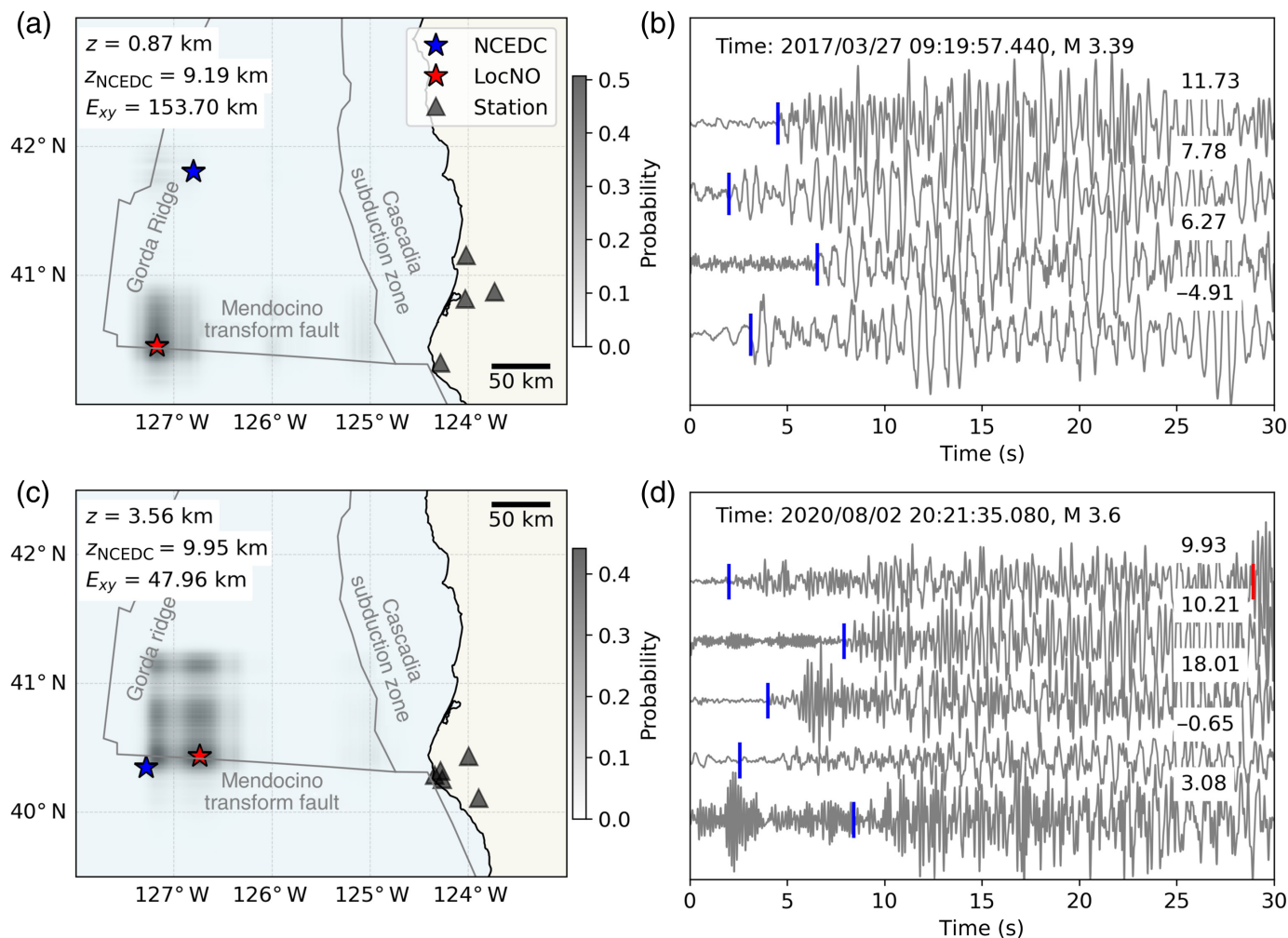


improves with more stations. Specifically: (1) for events located with nine stations, the overall mean and standard deviation of epicenter errors are 11.4 and 11.0 km, respectively, which are nearly half the values observed with only four stations; and (2) the number of outliers, defined as samples with unusually large epicenter errors relative to their bin mean, declines because the number of stations increases.

Figure 8 shows two examples of offshore earthquakes that are relatively well located by LocNO, with their origin times, magnitudes, and depths annotated on the figure. The seismic stations used by LocNO for locating these earthquakes are also highlighted. These examples demonstrate that, despite the challenges of having only several stations, large distances from the earthquake origin, and poor geometric coverage, the trained LocNO model accurately estimates the earthquake locations. This performance is achieved by effectively utilizing seismic records with relatively high SNR, which was estimated as the ratio of standard deviations between two five-second windows following and preceding the P -wave arrival. In contrast, offshore events with low SNR and clustered station configurations are poorly located by LocNO under these challenging monitoring conditions (Fig. 9).

Figure 8. Examples of offshore earthquakes that are relatively well located by LocNO. The map annotates the epicentral error E_{xy} , the catalog depth z_{NCEDC} , and the LocNO-predicted depth z . For each event (row), the panels (a,c) show the earthquake and station locations, and the panels (b, d) present the waveform inputs to LocNO. The origin time and magnitude are labeled above the waveforms. The signal-to-noise ratio (SNR) (in dB) is labeled on each trace. The P -wave (blue) and S -wave (red) arrival times used for earthquake locations in the NCEDC catalog are labeled on the waveforms. The color version of this figure is available only in the electronic edition.

The NCEDC catalog locations used as a reference point in these comparisons are also subject to significant uncertainties, especially for offshore events with poor station coverage. Figures 8 and 9 include the full spatial probability maps predicted by LocNO, which provide a more informative representation of location uncertainty. These maps allow readers to evaluate whether the catalog locations fall within high-probability regions of the LocNO's output. When comparing with NCEDC locations, we should keep in mind that both LocNO and the catalog locations are subject to uncertainty, and the observed differences do not necessarily imply error from a single source.



Effect of waveform preprocessing settings

We investigate the impact of waveform preprocessing settings on the performance of LocNO. As a full-waveform location method, LocNO leverages all information contained in the input waveforms to infer earthquake hypocenters. Therefore, the preprocessing configuration may significantly influence the results. Under various waveform filtering bands and window lengths, Table 2 compares the location performance in terms of MAE between LocNO predictions and the catalog locations in the easting, northing, and depth directions for all test samples.

TABLE 2

Effect of Waveform Preprocessing Settings (Filtering and Waveform Length) on Location Performance, Measured by Mean Absolute Error (MAE)

Waveform Setting		MAE (Easting)	MAE (Northing)	MAE (Depth)
1–10 Hz	30 s	10.71 km	9.75 km	4.32 km
	20 s	12.61 km	9.02 km	4.02 km
	10 s	16.63 km	8.28 km	4.33 km
	5 s	23.18 km	9.87 km	4.77 km
30 s	1–20 Hz	10.98 km	11.40 km	4.18 km
	1–10 Hz	10.71 km	9.75 km	4.32 km
	1–5 Hz	12.47 km	10.20 km	4.50 km

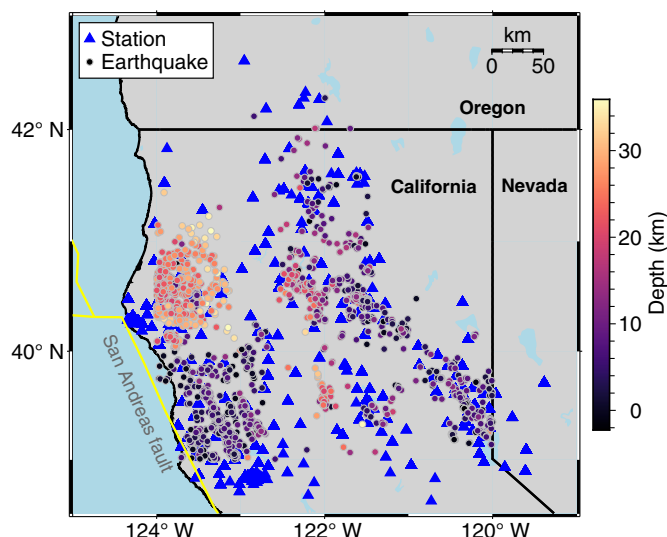


Figure 10. Dataset of in-network earthquakes near the MTJ region (1987–2021). This dataset serves as a benchmark for comparing LocNO with a baseline deep learning model for earthquake location using full-waveform data. The locations of these events are well determined by the onshore seismic network. The color version of this figure is available only in the electronic edition.

First, using waveforms filtered in the 1–10 Hz band, we evaluate the effect of different window lengths: 5, 10, 20, and 30 s. A clear trend is observed in the easting direction: location errors increase because the window length shortens. This may be attributed to the uneven station distribution because all stations are located east of the offshore earthquakes in the MTJ region. The full-waveform information preserved in longer time windows helps mitigate errors caused by this nonideal station geometry. In the northing direction, where stations are distributed more symmetrically, the effect of window length is less pronounced, with an average location error of around 10 km. Depth errors are relatively stable across all settings, likely due to the intrinsic difficulty of resolving depth using surface stations at large epicentral distances and limited azimuthal coverage.

Overall, waveform length shows a stronger influence on location accuracy, particularly in the easting direction due to the asymmetric station geometry. Notably, with a 5 s window (~ 4 s after *P*-wave arrival), the location error increases to 16.53 km (averaged over easting and northing), which is comparable to the performance of the bEPIC algorithm (14 km) in the MTJ region (Williamson *et al.*, 2023). This highlights the potential of LocNO for earthquake early warning applications in the MTJ area using short-window waveforms and only a few stations.

Next, fixing the waveform length at 30 s, we investigate the effect of applying band-pass filters at 1–5 Hz, 1–10 Hz, and 1–20 Hz to the observed waveform data before feeding it into LocNO. Among these, the 1–10 Hz band yields the best performance in the easting and northing directions, with an average MAE of 10.23 km, indicating that overly narrow filtering

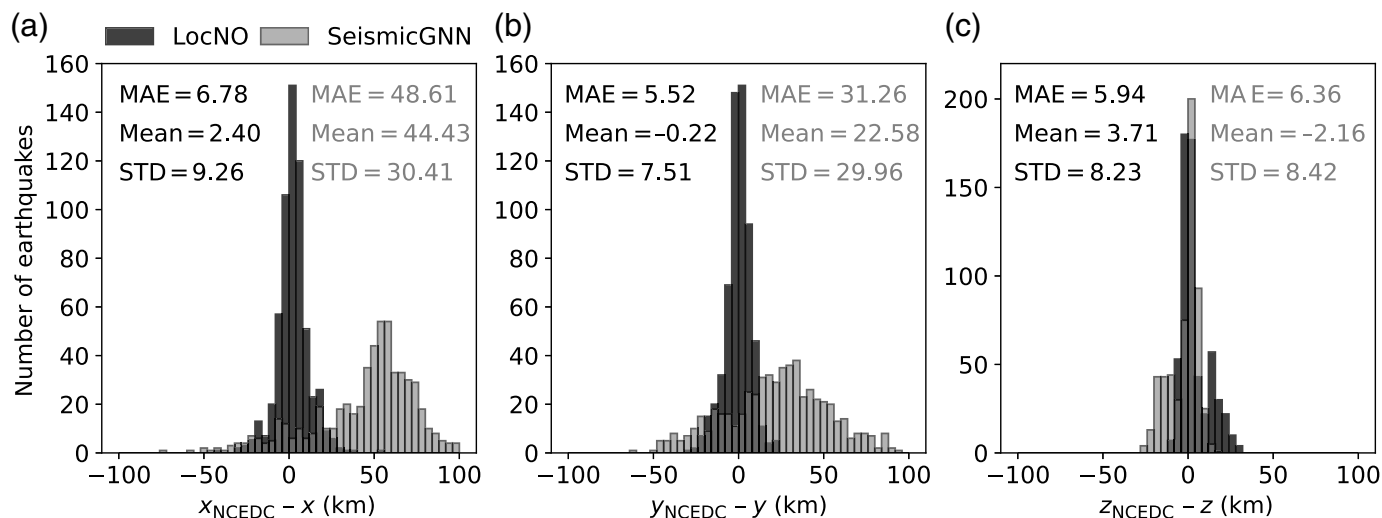
may remove useful high-frequency information needed for accurate location, whereas wider bands may retain noise. These trends are consistent with expectations: raising the upper frequency bound to 20 Hz may introduce noise, while lowering it to 5 Hz may filter out informative signal components. In terms of depth, the error increases gradually from 4.18 km (1–20 Hz) to 4.50 km (1–5 Hz). Still, the degradation from using the 1–5 Hz band is not substantial. Given that most regional velocity models are better suited for modeling low-frequency waveforms, the 1–5 Hz band may provide a reasonable trade-off between accuracy and generalizability, especially for training LocNO with synthetic waveform datasets simulated on a known velocity model.

Benchmark with SeismicGNN

In addition to evaluating LocNO's performance on out-of-network earthquake location, we benchmark its capabilities against other state-of-the-art deep learning models using well-monitored earthquake datasets, which have been extensively studied. Specifically, we compile a dataset of onshore earthquakes near the MTJ region (Fig. 10), covering the period from 1987 to 2021. This dataset provides an ideal testing ground due to the high-quality monitoring afforded by the dense onshore seismic network, which ensures relatively more reliable location data for comparison. Similar to the out-of-network dataset, each event in this onshore dataset is associated with 4–15 stations. In line with the previous approach, the dataset was split chronologically, with 1344 events recorded before 2016 used for training and 605 events from 2017 to 2021 reserved for testing.

Figure 11 presents a comparative analysis of the location performance between LocNO and SeismicGNN (Van Den Ende and Ampuero, 2020) on the in-network test dataset. SeismicGNN, a graph neural network framework, incorporates spatial information for seismic source characterization by facilitating information exchange between seismic stations. Node features at individual stations are processed using a CNN and an MLP to estimate earthquake locations. The baseline model was retrained on the same dataset as LocNO to ensure a fair comparison. However, the SeismicGNN model retrained with the NCEDC dataset compiled in this study performs worse than the original model reported in Van Den Ende and Ampuero (2020). The original model was trained on events with a minimum magnitude of 3 and at least 21 associated stations. In contrast, the dataset used here includes events with magnitudes as low as 1 and only 4–15 stations per event. This suggests that SeismicGNN may be less effective on more challenging datasets compared to high-quality training data.

The results demonstrate that LocNO outperforms SeismicGNN in geographic location accuracy, highlighting its ability to effectively leverage operator learning for earthquake location with full-waveform data (Fig. 11). LocNO achieves MAEs of 6.78 km (easting) and 5.52 km (northing),



compared with 48.61 and 31.26 km for SeismicGNN, respectively, clearly showing LocNO's improvement in both directions. In addition, with a depth MAE of 5.94 km for LocNO compared to 6.36 km for SeismicGNN, LocNO shows comparable depth prediction, which is traditionally challenging when relying solely on surface observations.

Furthermore, a comparison of the histograms in Figures 6 and 11 reveals that LocNO achieves significantly lower location errors for in-network earthquakes compared to out-of-network earthquakes across all evaluation metrics. This finding aligns with the intuitive understanding that locating out-of-network earthquakes is inherently a more challenging problem due to limited station coverage. The complexity of this task underscores the necessity of training specialized deep-learning models tailored to out-of-network earthquake location.

Application to an offshore earthquake sequence: The 22 September 2017 M_w 5.7 event, 218 km west of Ferndale, California

A typical seismic monitoring workflow includes earthquake detection, seismic phase picking, initial location, and relocation. Traditional initial location methods rely on travel-time algorithms using picked phase arrivals, and they generally perform well when the seismic network has favorable geometry (see Yu *et al.*, 2025, for a benchmark study using synthetic controlled experiments). However, in regions with sparse or uneven station coverage, such as offshore settings, solutions to travel-time-based inverse problems are often unstable and can result in large location errors that are difficult to correct during relocation. In contrast, LocNO directly utilizes the spatiotemporal information contained in the multistation waveforms to infer earthquake hypocenters. As shown in Table 2, longer waveform segments better correct artifacts caused by uneven station distribution, confirming that waveform data provide additional constraints for mitigating such errors.

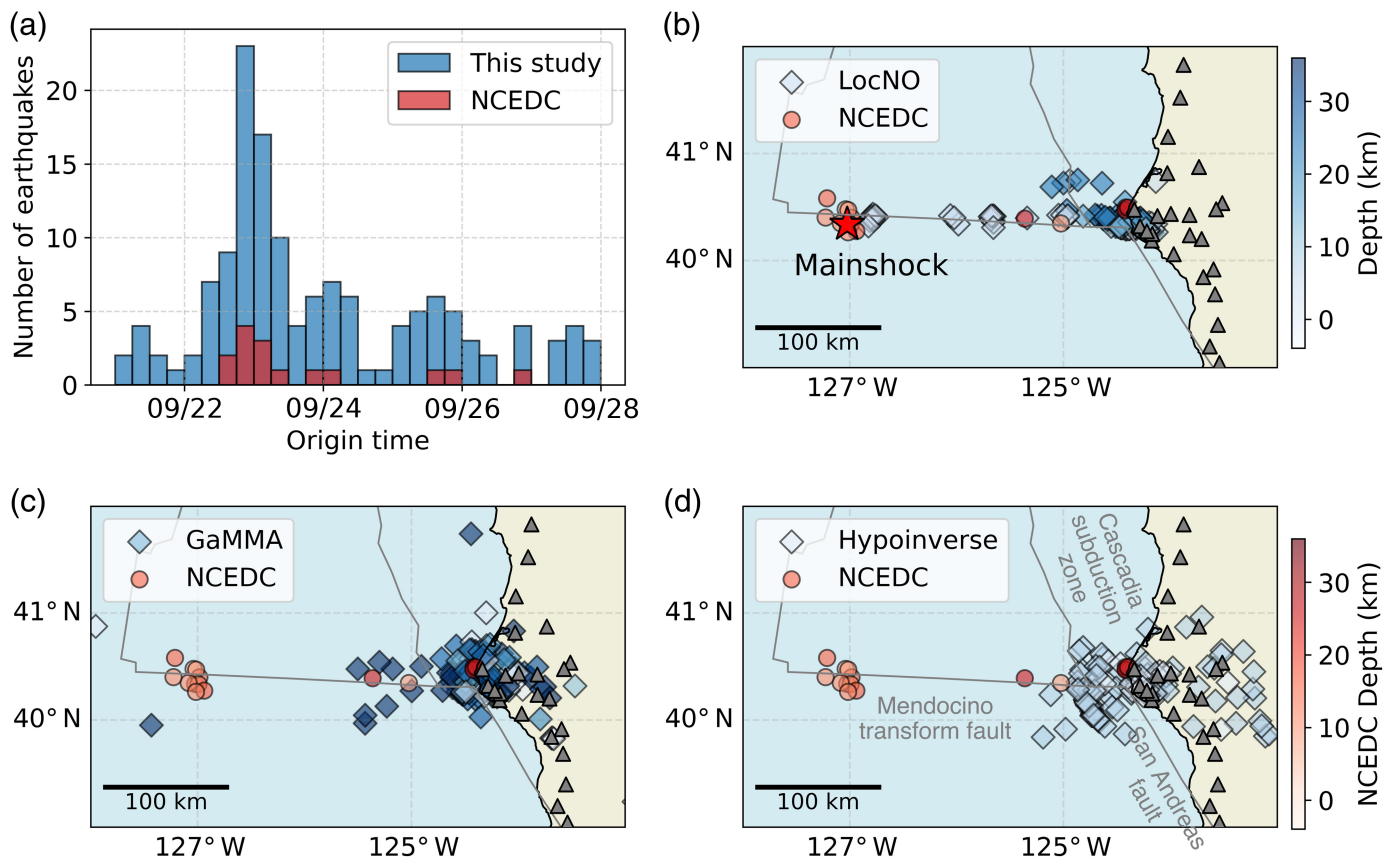
Since events before 2016 were used to construct the training dataset, we select an earthquake sequence that occurred after

Figure 11. (a–c) Comparison of the locations determined by LocNO and SeismicGNN on the in-network test dataset in the MTJ. LocNO yielded mean absolute errors in horizontal locations that are less than 18% of the errors yielded by SeismicGNN. The errors in the determined depths were comparable between the two methods.

2016 for this case study. According to the NCEDC catalog, an M_w 5.7 event occurred on 22 September 2017, ~218 km west of Ferndale, California. To analyze this earthquake sequence, we downloaded continuous waveform data from 25 onshore stations in Northern California (NC) and Plate Boundary Observatory Borehole (PB) networks via NCEDC, covering 21–28 September 2017. The selected seismic network consists of a combination of the three-component EH (high-gain short-period) channels, the three-component broadband HH (high-gain broadband) channels, and the one-component EH channels, as listed in [Data and Resources](#).

LocNO can be applied after any detection algorithm, even when the method does not involve phase picking. In this study, however, we adopt a workflow that begins with phase picking using PhaseNO (Sun *et al.*, 2023), which performs both detection and picking directly on continuous seismic data. The resulting picks are then associated into events using GAMMA (Zhu, McBrearty, *et al.*, 2022), and LocNO is subsequently used to estimate the event locations based on the full-waveform information. For comparison, we also determine the locations of the earthquakes in the new catalog using HYPOINVERSE, a travel-time-based location algorithm that has been widely used for decades (Klein, 2002).

As a result, we construct an earthquake catalog for this offshore sequence containing 141 events over seven days. Each event is required to have at least eight associated phases, including both the *P*-wave and the *S*-wave arrivals. The number of earthquakes in this new catalog is nearly one order of magnitude greater than the 15 events reported by NCEDC for the same period (Fig. 12a). Figure 12b,d compares the locations



of these newly detected events with those in the NCEDC catalog. Figure 12c shows the locations estimated by GaMMA during phase association. For travel-time-based location, we use the 1D velocity model from Morton *et al.* (2023). Because of poor network geometry and a large primary azimuthal gap, events in the far western portion of the sequence, near the mainshock, are difficult to locate accurately using travel-time-based methods. In contrast, locations estimated by LocNO are closer to the Mendocino transform fault and to the NCEDC catalog events. In addition, several earthquakes near the mainshock hypocenter are incorrectly located by GaMMA and HYPOINVERSE, whereas LocNO provides locations that are more consistent with the mainshock's hypocenter.

DISCUSSION

This study introduces LocNO and an effective training strategy to address the challenges of earthquake location in regions with sparse station coverage and large azimuthal gaps. Advanced operator-learning techniques have proven highly adaptable to complex network geometries, establishing themselves as powerful tools to uncover the spatiotemporal patterns in seismic data for earthquake location. The architecture of LocNO is designed to leverage the waveform information directly, allowing it to mitigate artifacts even when the station coverage is sparse or the azimuthal gaps are large. This makes it highly applicable to offshore regions such as the MTJ, where traditional travel-time-based methods often struggle due to the lack

Figure 12. 22 September 2017 M_w 5.7 offshore earthquake sequence and their locations. (a) Comparison of earthquake counts between the newly detected catalog and the NCEDC catalog. Comparison of NCEDC catalog locations with earthquake locations determined by (b) the LocNO model, (c) GaMMA during phase association, and (d) HYPOINVERSE. The color version of this figure is available only in the electronic edition.

of permanent stations in the marine environment. Quantitatively, LocNO achieves mean absolute errors of ~ 10 km in horizontal location and 4 km in depth when compared to the NCEDC catalog. Furthermore, LocNO can be integrated into deep-learning-based seismic monitoring workflows, where many of the detected small-magnitude events are recorded by only a few stations. In such cases, including additional distant stations may not meaningfully improve the azimuthal coverage because these stations often fail to record high-quality signals of small events. LocNO's ability to provide stable locations under these conditions demonstrates its potential as a standard component for modern catalog-building pipelines.

A training dataset designed for locating out-of-network earthquakes is particularly valuable for monitoring seismic activity in offshore regions. Using the MTJ region as an example, frequent offshore earthquakes near the Mendocino transform fault pose significant risks to surrounding communities. However, the inherent challenge of locating these out-of-

network events stems from the limited number of stations in marine environments. Rather than incorporating historical seismicity as a prior in a Bayesian framework (Williamson *et al.*, 2023), we leverage past seismic data to construct a comprehensive training dataset and directly train a deep learning model to locate these offshore events using full-waveform information. Our results indicate that the challenge of locating out-of-network earthquakes can be effectively mitigated through advanced deep learning methods, provided that a sufficiently representative training dataset is available.

As a data-driven approach, LocNO relies on a well-curated training dataset with diverse earthquake locations across the monitoring region, ensuring comprehensive wavepath coverage in the MTJ area and enhancing the model's generalization ability to new earthquakes in this region. By leveraging this dataset, the model directly determines earthquake source locations from waveform data without an explicit velocity model. Although LocNO does not require an explicit velocity model during inference, the training labels come from catalog locations that were determined using a velocity model. As a result, the model is still influenced by the assumptions of that velocity model. For example, the deep learning model that directly maps waveforms to source parameters may encounter challenges when applied to regions with velocity distributions that significantly differ from those represented in the training dataset. This limitation highlights the critical need to construct training datasets that capture the geological and tectonic diversity of the target region. Therefore, leveraging historical seismicity data remains an effective strategy for accurately and efficiently locating newly detected earthquakes within the studied area.

Traditional template matching methods locate earthquakes by cross-correlating new waveforms against a library of known events, effectively detecting and refining the locations of repeating earthquakes with similar waveforms (Waldhauser and Ellsworth, 2000; Shelly *et al.*, 2006, 2007). In contrast, LocNO does not rely on waveform similarity between training and testing events. Although trained on catalog events, its goal is to learn a functional mapping from waveforms to source locations that generalizes to new events with previously unseen waveform characteristics. This generalization capability is analogous to deep learning-based phase pickers such as PhaseNet (Zhu and Beroza, 2019) and PhaseNO (Sun *et al.*, 2023), which can pick phases for earthquakes beyond the training sets. Moreover, while template matching is mainly used for event detection and relocation through cross correlation, LocNO directly infers absolute source locations from multistation waveforms. Unlike double-difference (DD) methods that require high waveform similarity, LocNO is not limited to repeated events and is thus applicable to general earthquake location tasks.

Despite the encouraging performance of LocNO in out-of-network earthquake location, significant errors are observed

for events with poor station geometry, such as clusters of stations. Although the location errors for out-of-network events can be substantial, the results may still be useful for distinguishing between the in-network and the out-of-network events, as demonstrated in Annunziata *et al.* (2025). Furthermore, it is important to emphasize that simply increasing the number of stations does not necessarily improve location accuracy if the added stations do not enhance the overall station coverage by reducing the primary azimuthal gap (Fig. 7), particularly for out-of-network earthquakes.

To overcome the challenge of training the LocNO model in low seismicity areas, we may generate a synthetic training dataset to supplement the real-world dataset (Leong and Zhu, 2024). By sampling a known 3D velocity model with various earthquake locations, a comprehensive training dataset could be synthesized through seismic forward modeling to compensate for the scarcity of observed data. Moreover, the synthetic data may be further augmented using a generative model trained to map event locations to seismograms that exhibit the characteristics of the field data for a specific area (Spurio Mancini *et al.*, 2021).

The key question when training with synthetic data is how accurate and realistic the simulations must be to ensure that LocNO generalizes well to real-world observations. This strategy may be particularly effective in regions where a smooth velocity model sufficiently captures long-distance wave propagation. However, in local areas with significant scattering and small-scale heterogeneity, a high-resolution velocity model may be necessary to simulate seismic wavefields more accurately and ensure that the synthetic data can effectively enhance the model's performance. Because of the lack of high-resolution velocity models, synthetic data are typically inaccurate at higher frequencies. In this study, LocNO demonstrates the ability to infer earthquake source locations with an uncertainty of ~ 10 km in the MTJ region using data filtered in the 1–10 Hz frequency band. Although limiting the input to the 1–5 Hz band may exclude some informative high-frequency components, the resulting degradation in performance is relatively minor (Table 2). Most regional velocity models are better suited for simulating low-frequency waveforms, so simulating data in the 1–5 Hz band offers a practical compromise between accuracy and generalizability when training LocNO with synthetic waveform datasets.

Low SNR remains a significant challenge for accurate earthquake location. Traditional travel-time-based methods can sometimes provide reliable solutions for low SNR events when combined with robust phase pickers. In particular, deep-learning-based pickers such as PhaseNO have demonstrated their ability to extract *P*- and *S*-wave arrivals even from noisy waveforms, enabling subsequent location using conventional travel-time-based methods (Sun *et al.*, 2023). However, when the seismic network has poor geometry, such as sparse station coverage or large azimuthal gaps, these travel-time-based location

methods may still produce unstable or biased estimates despite having accurate picks. This limitation is particularly evident for the out-of-network events, as demonstrated in the 22 September 2017 M_w 5.7 offshore earthquake sequence. In contrast, LocNO is trained on out-of-network examples that reflect these geometric challenges, allowing it to learn and correct for spatial biases using full-waveform information. To further improve LocNO's robustness for low SNR events, waveform denoising techniques could be incorporated before model inference. This approach could enhance the clarity of input signals, which is commonly employed in seismic full-waveform inversion. Overall, while travel-time-based methods remain effective in well-instrumented regions, waveform-based approaches like LocNO offer distinct advantages for events in challenging network settings.

The DD earthquake location is well known for achieving high relative accuracy and reduced sensitivity to the velocity model errors because it minimizes travel time differences between pairs of events recorded at common stations (Waldhauser and Ellsworth, 2000). However, its effectiveness is inherently constrained by the geometry of the seismic network (Waldhauser, 2001). In sparse or uneven station distributions, or when the azimuthal coverage is poor, DD methods cannot effectively constrain the absolute locations of event clusters, often resulting in systematic spatial drift. This limitation arises because DD techniques only update relative positions between events and cannot correct the initial absolute location of event clusters, which is typically determined before starting the DD relocation. Furthermore, constructing a sufficient number of strongly linked event pairs, typically defined by eight or more high-quality observations, is difficult in networks with limited station coverage. Consequently, the benefits of DD approaches diminish significantly for isolated earthquakes or in regions where stations are few and unevenly distributed.

In addition to algorithmic advancements such as the development of LocNO, a complementary strategy to improve the earthquake location accuracy in offshore or out-of-network regions is to enhance the seismic network by adding more stations and increasing the azimuthal coverage. The use of distributed acoustic sensing (DAS) on submarine fiber-optic cables (Gou *et al.*, 2025) and deployments of ocean-bottom seismometers (OBSs; Toomey *et al.*, 2014; Alongi *et al.*, 2021; Morton *et al.*, 2023) have shown promising results in extending the seismic monitoring capabilities into oceanic regions. These offshore instruments significantly improve the azimuthal coverage and reduce the path-length uncertainties, which are critical for accurate hypocenter determination. However, DAS deployment in the MTJ region is currently limited by the availability of dark fiber infrastructure, and OBS deployments are typically temporary, providing improved hypocenter constraints only during active recording periods. Integrating offshore deployments with LocNO may further improve location accuracy, but this requires careful evaluation in future studies.

CONCLUSION

This study proposes the LocNO, a deep learning framework for earthquake location using full waveforms recorded by multiple stations with arbitrary network geometry. Its application to offshore earthquakes in the MTJ region highlights the effectiveness of LocNO in addressing the persistent challenges of locating offshore events using onshore seismic stations. By utilizing operator learning frameworks and full-waveform seismic data, LocNO effectively captures spatiotemporal relationships in seismic signals, enabling robust performance even in regions with sparse monitoring networks. Quantitatively, LocNO achieves mean absolute errors of ~ 10 km in horizontal location and 4 km in depth when compared to the NCEDC catalog. The combined use of PhaseNO for earthquake detection and LocNO for location provides a practical path to study MTJ seismotectonics with deep-learning-enhanced catalogs, permitting reliable locations of newly detected small offshore events even when only a few onshore stations record them. Future research could enhance the training dataset with synthetic data to address low-seismicity regions, extend LocNO's application to other tectonic settings, and integrate it into real-time monitoring systems. These efforts would further increase LocNO's potential for mitigating seismic risks and improving regional earthquake monitoring in poorly instrumented areas.

DATA AND RESOURCES

The Python package location neural operator (LocNO) is available at <https://github.com/sun-hongyu/LocNO>. Python code SeismicGNN is available at doi: [10.6084/m9.figshare.12231077](https://doi.org/10.6084/m9.figshare.12231077). Detection of events for 22 September 2017 M_w 5.7 earthquake sequence, located 218 km west of Ferndale, California, was performed using the Python package phase neural operator (PhaseNO) (Sun *et al.*, 2023), available at <https://github.com/sun-hongyu/PhaseNO>. The case study used stations from the BK, NC, and PB networks. Station codes include BK.JCC.HH, NC.GHO.EH, NC.GNA.EH, NC.GTC.EH, NC.KBS.EH, NC.KCPB.HH, NC.KCRE.HH, NC.KCS.EH, NC.KCT.HH, NC.KHMB.HH, NC.KMPB.HH, NC.KMR.HH, NC.KRMB.HH, NC.KRP.HH, NC.KSM.EH, NC.KSXB.HH, PB.B045.EH, PB.B046.EH, PB.B047.EH, PB.B049.EH, PB.B932.EH, PB.B933.EH, PB.B934.EH, and PB.B935.EH. Waveform data, metadata, and data products for this study were accessed through the Northern California Earthquake Data Center, doi: [10.7932/NCEDC](https://doi.org/10.7932/NCEDC). All websites were last accessed in September 2025.

DECLARATION OF COMPETING INTERESTS

The author acknowledges that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

The author thanks Associate Editor Thomas M. Brocher and the two anonymous reviewers for their thoughtful reviews, which greatly improved the article. This work was supported in part by the UTEP University Research Institute Grant, the UTEP Institute for Strategic and Sustainable Resources, and the UT System STARS Program. H.S. also acknowledges support during the revision of this article from the U.S. Department of Energy, Office of Science, Energy

Earthshots Initiative (Award Number DE-SC0024703). This work used Bridges-2 (Brown *et al.*, 2021) at Pittsburgh Supercomputing Center [EES240076] from the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) program, which is supported by National Science Foundation Grant Numbers 2138259, 2138286, 2138307, 2137603, and 2138296.

REFERENCES

- Alongi, T., S. Y. Schwartz, H. R. Shaddox, and D. T. Small (2021). Probing the southern Cascadia plate interface with the dense amphibious Cascadia initiative seismic array, *J. Geophys. Res.* **126**, e2021JB022180, doi: [10.1029/2021JB022180](https://doi.org/10.1029/2021JB022180).
- Annunziata, D., M. Savoia, C. Martino, F. Giampaolo, V. Convertito, F. Piccialli, and G. C. Beroza (2025). A robust and rapid grid-based machine learning approach for inside and off-network earthquakes classification in dynamically changing seismic networks, *Seismol. Res. Lett.* **96**, 933–948.
- Aquib, T. A., and P. M. Mai (2024). Broadband ground-motion simulations with machine-learning-based high-frequency waves from Fourier neural operators, *Bull. Seismol. Soc. Am.* **114**, 2846–2868.
- Bird, P. (2003). An updated digital model of plate boundaries, *Geochem. Geophys. Geosys.* **4**, 1–52.
- Brown, S. T., P. Buitrago, E. Hanna, S. Sanielevici, R. Scibek, and N. A. Nystrom (2021). Bridges-2: A platform for rapidly-evolving and data intensive research, *Presented at the Practice and Experience in Advanced Research Computing 2021: Evolution Across All Dimensions*, Association for Computing Machinery.
- Burkett, E. R., D. D. Given, and L. M. Jones (2014). ShakeAlert—An earthquake early warning system for the United States west coast, *Tech. Rept. US Geol. Surv.*, doi: [10.3133/fs20143083](https://doi.org/10.3133/fs20143083)
- Cabieces, R., E. Bufo, S. Cesca, and A. Pazos (2020). Focal parameters of earthquakes offshore Cape St. Vincent using an amphibious network, *Pure Appl. Geophys.* **177**, 1761–1780.
- Castro, J. D., C. Meneses-Ponce, R. Ortega, J. Beltran-Gracia, and H. Gonzalez-Huizar (2024). Single station seismic event location with a neural-guided mixture model: Model-based weighting for improved accuracy, *J. Geophys. Res.* **1**, e2024JH000346, doi: [10.1029/2024JH000346](https://doi.org/10.1029/2024JH000346).
- Chen, Y., O. M. Saad, A. Savvaidis, Y. Chen, and S. Fomel (2022). 3D microseismic monitoring using machine learning, *J. Geophys. Res.* **127**, e2021JB023842, doi: [10.1029/2021JB023842](https://doi.org/10.1029/2021JB023842).
- Eisermann, A. S., A. Ziv, and H. G. Wust-Bloch (2018). Array-based earthquake location for regional earthquake early warning: Case studies from the Dead Sea transform, *Bull. Seismol. Soc. Am.* **108**, 2046–2053.
- Ellsworth, W. L., and S. W. Roecker (1981). Sensitivity of the earthquake location problem to network geometry, in *Seismicity and Tectonics of the Pamir-Hindu Kush Region of Central Asia*, Ph.D. Thesis, Massachusetts Institute of Technology, 57–74 (Chapter in Ph.D. thesis by Steven W. Roecker).
- Elsayed, H. S., O. M. Saad, M. S. Soliman, Y. Chen, and H. A. Youness (2023). EQConvMixer: A deep learning approach for earthquake location from single-station waveforms, *IEEE Geosci. Remote Sens. Lett.* **20**, doi: [10.1109/LGRS.2023.3312324](https://doi.org/10.1109/LGRS.2023.3312324).
- Franceschi, L., M. Niepert, M. Pontil, and X. He (2019). Learning discrete structures for graph neural networks, *International Conf. on Machine Learning*, PMLR, 1972–1982.
- Gilmer, J., S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl (2017). Neural message passing for quantum chemistry, *International Conf. on Machine Learning (ICML)*, PMLR, 1263–1272.
- Given, D. D., R. M. Allen, A. S. Baltay, P. Bodin, E. S. Cochran, K. Creager, R. M. de Groot, L. S. Gee, E. Hauksson, T. H. Heaton, *et al.* (2018). Revised technical implementation plan for the ShakeAlert system—An earthquake early warning system for the west coast of the United States, *U.S. Geol. Surv. Open-File Rept. 2018-1155*. (Supersedes USGS Open-File Report 2014-1097).
- Gou, Y., R. M. Allen, W. Zhu, T. Taira, and L.-W. Chen (2025). Leveraging submarine DAS arrays for offshore earthquake early warning: A case study in Monterey Bay, California, *Bull. Seismol. Soc. Am.* **115**, 516–532.
- He, K., X. Zhang, S. Ren, and J. Sun (2016). Deep residual learning for image recognition, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 770–778.
- Hendrycks, D., and K. Gimpel (2016). Gaussian error linear units (GELUs), arXiv preprint arXiv:1606.08415.
- Jung, S., J.-H. Park, Y. J. Seong, and D.-H. Sheen (2023). Automatic determination of back azimuth based on a single mini array for an earthquake early warning system, *Bull. Seismol. Soc. Am.* **113**, 1311–1323.
- Klein, F. W. (2002). User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes, *Tech. Rept. U.S. Geol. Surv. Open-File Rept. 2002-171*.
- Kohler, M. D., D. E. Smith, J. Andrews, A. I. Chung, R. Hartog, I. Henson, D. D. Given, R. de Groot, and S. Guivits (2020). Earthquake early warning ShakeAlert 2.0: Public rollout, *Seismol. Res. Lett.* **91**, 1763–1775.
- Kovachki, N., Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar (2023). Neural operator: Learning maps between function spaces with applications to PDEs, *J. Machine Learn. Res.* **24**, 1–97.
- Kuang, W., C. Yuan, Z. Zou, J. Zhang, and W. Zhang (2024). Autonomous earthquake location via deep reinforcement learning, *Seismol. Res. Lett.* **95**, 367–377.
- Leong, Z. X., and T. Zhu (2024). Machine learning-assisted micro-earthquake location workflow for monitoring the Newberry enhanced geothermal system, *J. Geophys. Res.* **1**, e2024JH000159, doi: [10.1029/2024JH000159](https://doi.org/10.1029/2024JH000159).
- Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar (2020). Neural operator: Graph kernel network for partial differential equations, arXiv preprint arXiv:2003.03485.
- Li, Z., N. B. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar (2021). Fourier neural operator for parametric partial differential equations, arXiv preprint, doi: [10.48550/arXiv.2010.08895](https://doi.org/10.48550/arXiv.2010.08895).
- McBrearty, I. W., and G. C. Beroza (2022). Earthquake location and magnitude estimation with graph neural networks, *2022 IEEE International Conf. on Image Processing (ICIP)*, IEEE, 3858–3862.
- McBrearty, I. W., and G. C. Beroza (2023). Earthquake phase association with graph neural networks, *Bull. Seismol. Soc. Am.* **113**, 524–547.
- McBrearty, I. W., and G. C. Beroza (2025). Double difference earthquake location with graph neural networks, *Earth Planets Space* **77**, 127.

- Morton, E. A., S. L. Bilek, and C. A. Rowe (2023). Cascadia subduction zone fault heterogeneities from newly detected small magnitude earthquakes, *J. Geophys. Res.* **128**, e2023JB026607, doi: [10.1029/2023JB026607](https://doi.org/10.1029/2023JB026607).
- Mousavi, S. M., and G. C. Beroza (2020). Bayesian-deep-learning estimation of earthquake location from single-station observations, *IEEE Trans. Geosci. Remote Sens.* **58**, 8211–8224.
- Mousavi, S. M., Y. Sheng, W. Zhu, and G. C. Beroza (2019). STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI, *IEEE Access* **7**, 179464–179476.
- Netanel, M., A. S. Eisermann, and A. Ziv (2021). Off-network earthquake location by earthquake early warning systems: Methodology and validation, *Bull. Seismol. Soc. Am.* **111**, 3090–3102.
- Nof, R. N., A. I. Chung, H. Rademacher, L. Dengler, and R. M. Allen (2019). MEMS accelerometer mini-array (MAMA): A low-cost implementation for earthquake early warning enhancement, *Earthq. Spectra* **35**, 21–38.
- Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4**, e1700578, doi: [10.1126/sciadv.1700578](https://doi.org/10.1126/sciadv.1700578).
- Rubinstein, J. L., and G. C. Beroza (2007). Full waveform earthquake location: Application to seismic streaks on the Calaveras fault, California, *J. Geophys. Res.* **112**, doi: [10.1029/2006JB004463](https://doi.org/10.1029/2006JB004463).
- Shelly, D. R., G. C. Beroza, and S. Ide (2007). Non-volcanic tremor and low-frequency earthquake swarms, *Nature* **446**, 305–307.
- Shelly, D. R., G. C. Beroza, S. Ide, and S. Nakamura (2006). Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip, *Nature* **442**, 188–191.
- Si, X., X. Wu, Z. Li, S. Wang, and J. Zhu (2024). An all-in-one seismic phase picking, location, and association network for multi-task multi-station earthquake monitoring, *Commun. Earth Environ.* **5**, 22.
- Spurio Mancini, A., D. Piras, A. M. G. Ferreira, M. P. Hobson, and B. Joachimi (2021). Accelerating Bayesian microseismic event location with deep learning, *Solid Earth* **12**, 1683–1705.
- Sun, H. (2025). Location neural operator (LocNO): Source code, *GitHub*, available at <https://github.com/sun-hongyu/LocNO> (last accessed September 2025).
- Sun, H., Z. E. Ross, W. Zhu, and K. Azizzadenesheli (2023). Phase neural operator for multi-station picking of seismic arrivals, *Geophys. Res. Lett.* **50**, e2023GL106434, doi: [10.1029/2023GL106434](https://doi.org/10.1029/2023GL106434).
- Sun, H., Y. Yang, K. Azizzadenesheli, R. W. Clayton, and Z. E. Ross (2022). Accelerating time-reversal imaging with neural operators for real-time earthquake locations, arXiv preprint arXiv:2210.06636.
- Toomey, D. R., R. M. Allen, A. H. Barclay, S. W. Bell, P. D. Bromirski, R. L. Carlson, X. Chen, J. A. Collins, R. P. Dziak, B. Evers, et al. (2014). The Cascadia initiative: A sea change in seismological studies of subduction zones, *Oceanography* **27**, 138–150.
- Van Den Ende, M. P. A., and J. Ampuero (2020). Automated seismic source characterization using deep graph neural networks, *Geophys. Res. Lett.* **47**, e2020GL088690, doi: [10.1029/2020GL088690](https://doi.org/10.1029/2020GL088690).
- Waldhauser, F. (2001). hypoDD-A program to compute double-difference hypocenter locations, *Tech. Rept. U.S. Geol. Surv. Open-File Rept. 2001-113*.
- Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California, *Bull. Seismol. Soc. Am.* **90**, 1353–1368.
- Williamson, A., A. Lux, and R. Allen (2023). Improving out of network earthquake locations using prior seismicity for use in earthquake early warning, *Bull. Seismol. Soc. Am.* **113**, 664–675.
- Yu, Y., W. L. Ellsworth, and G. C. Beroza (2025). Accuracy and precision of earthquake location programs: Insights from a synthetic controlled experiment, *Seismol. Res. Lett.* **96**, 1860–1874.
- Zhang, X., W. Reichard-Flynn, M. Zhang, M. Hirn, and Y. Lin (2022). Spatiotemporal graph convolutional networks for earthquake source characterization, *J. Geophys. Res.* **127**, e2022JB024401, doi: [10.1029/2022JB024401](https://doi.org/10.1029/2022JB024401).
- Zhang, X., J. Zhang, C. Yuan, S. Liu, Z. Chen, and W. Li (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method, *Sci. Rep.* **10**, 1941.
- Zhu, W., and G. C. Beroza (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.* **216**, 261–273.
- Zhu, W., I. W. McBrearty, S. M. Mousavi, W. L. Ellsworth, and G. C. Beroza (2022). Earthquake phase association using a Bayesian Gaussian mixture model, *J. Geophys. Res.* **127**, e2021JB023249, doi: [10.1029/2021JB023249](https://doi.org/10.1029/2021JB023249).
- Zhu, W., K. S. Tai, S. M. Mousavi, P. Bailis, and G. C. Beroza (2022). An end-to-end earthquake detection method for joint phase picking and association using deep learning, *J. Geophys. Res.* **127**, e2021JB023283, doi: [10.1029/2021JB023283](https://doi.org/10.1029/2021JB023283).
- Zhu, W., H. Wang, B. Rong, E. Yu, S. Zuzlewski, G. Tepp, T. Taira, J. Marty, A. Husker, and R. M. Allen (2025). California earthquake dataset for machine learning and cloud computing, arXiv preprint arXiv:2502.11500.
- Ziv, A., A. S. Eisermann, O. Volk, and I. Zbada (2024). Introducing ViDA3, an earthquake early warning algorithm for offshore hypocenter determination using onshore seismic networks, *Bull. Seismol. Soc. Am.* **114**, 2446–2456.
- Zollo, A., A. Caruso, G. De Landro, S. Colombelli, and L. Elia (2021). A Bayesian method for real-time earthquake location using multi-parameter data, *J. Geophys. Res.* **126**, e2020JB020359, doi: [10.1029/2020JB020359](https://doi.org/10.1029/2020JB020359).

Manuscript received 24 January 2025

Published online 15 October 2025