

<sup>1</sup> Fourier Neural Operators for Accelerating Earthquake  
<sup>2</sup> Dynamic Rupture Simulations

<sup>3</sup> **Napat Tainpkdipat** <sup>1</sup>, **Mohamed Abdelmeguid** <sup>2</sup>, **Chunhui Zhao** <sup>1</sup>, **Kamyar**  
<sup>4</sup> **Azizzadenesheli** <sup>3</sup>, **Ahmed Elbanna** <sup>1,4</sup>

<sup>5</sup> Department of Civil and Environmental Engineering, University of Illinois Urbana Champaign, Urbana,  
<sup>6</sup> IL 61801, USA.

<sup>7</sup> Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, USA.

<sup>8</sup> <sup>3</sup>Nvidia Corporation, Santa Clara, CA 95051, USA.

<sup>9</sup> Beckman Institute of Advanced Science and Technology, University of Illinois Urbana Champaign,  
<sup>10</sup> Urbana, IL 61801, USA.

<sup>11</sup> **Key Points:**

- <sup>12</sup> We develop a Fourier Neural Operator (FNO) framework to simulate earthquake  
rupture under varying conditions of fractal stress, nucleation sites, and friction.
- <sup>13</sup> The FNO achieves up to  $4 \times 10^5$  times speedup compared to the Spectral Bound-
- <sup>14</sup> ary Integral method, which is currently the fastest available numerical method.
- <sup>15</sup> We demonstrate that the model can generalize to unseen shear stress and frictional  
parameters, including out-of-distribution cases.

19     **Abstract**

20     Dynamic rupture modeling plays a crucial role in unraveling earthquake source processes.  
21     However, the multiscale nature of rupture propagation poses significant challenges, and  
22     classical numerical methods remain computationally expensive. To overcome this hur-  
23     dle, we present a methodology that is both computationally efficient and quantitatively  
24     accurate. Specifically, we introduce a surrogate model, in the form of a Fourier Neural  
25     Operator, for emulating the nonlinear equations governing dynamic rupture propaga-  
26     tion on frictional interfaces. This surrogate is trained on synthetic data generated by mul-  
27     tiple physics-based dynamic rupture simulations and is then applied to unseen problems.  
28     The proposed methodology retains the accuracy of traditional multiscale methods at a  
29     significantly reduced computational cost, achieving a speedup of up to  $4 \times 10^5$  compared  
30     to the state-of-the-art conventional methods. We evaluate this approach using various  
31     examples and demonstrate its efficacy in capturing the spacetime evolution of fault slip  
32     rates for a wide range of stress conditions. This development advances the state of the  
33     art of computational earthquake dynamics and opens new opportunities for accelerat-  
34     ing physics-based rupture forecasts.

35     **Plain Language Summary**

36     Earthquakes begin when stress builds up and causes sudden movement along faults  
37     in the Earth's crust. To simulate how an earthquake spreads, traditional numerical mod-  
38     els are often used. However, these models can be computationally intensive and time-  
39     consuming due to the complexity of earthquake processes. This study introduces a faster  
40     way to model earthquakes using a machine learning tool, specifically a Fourier Neural  
41     Operator framework. The model learns patterns from previous earthquake simulations  
42     and uses that knowledge to predict how a fault slips under different stress and friction  
43     conditions. This approach achieves a speedup of up to  $4 \times 10^5$  times compared to tra-  
44     ditional methods, potentially enabling the rapid exploration of many scenarios and im-  
45     proving the ability to study and forecast earthquakes efficiently.

46     **1 Introduction**

47     Earthquakes are among the most destructive natural hazards, yet available data  
48     remains limited. Large earthquake events are rare, and their recorded histories are short,  
49     posing challenges for purely data-driven analyses. Furthermore, because earthquakes orig-  
50     inate at depths where direct measurements are impossible, observations rely on sparse  
51     surface instrumentation, offering an incomplete view of fault dynamics (Arrowsmith et  
52     al., 2022). Earthquake modeling is therefore essential to interpret limited observations  
53     and infer subsurface processes that cannot be directly measured (Nielsen et al., 2000; Duan  
54     & Oglesby, 2006; Bhat et al., 2007; Ma & Elbanna, 2019; Ripperger et al., 2007; Gabriel  
55     et al., 2012; Xu et al., 2015).

56     Modeling earthquakes is inherently challenging, as it requires integrating data from  
57     multiple sources, including field observations and laboratory experiments, with physics-  
58     based simulations.(Johnson et al., 2006; Gallović et al., 2019a, 2019b; Aochi & Twardzik,  
59     2020; Van Zelst et al., 2019). Additionally, key fault properties remain uncertain since  
60     they cannot be directly measured. The multiscale nature of earthquakes further com-  
61     plicates the problem. The rupture occurs within seconds, but the recurrence intervals  
62     span years to centuries. Physically, earthquakes propagate over kilometers, whereas the  
63     rupture process zone might be confined to a few meters (Lapusta et al., 2000; Ben-Zion,  
64     2008; Chester & Chester, 1998). Accurately capturing rupture dynamics requires extremely  
65     fine discretization, often at the sub-meter scale, leading to immense computational costs  
66     (Ulrich et al., 2019).

Traditional numerical methods such as the finite element method (FEM) (Oglesby et al., 1998, 2000; Aagaard et al., 2001), finite difference method (FDM) (Andrews, 1973; Day, 1982; Madariaga et al., 1998; Andrews, 1999; Dalguer & Day, 2007; Moczo et al., 2007), and boundary integral methods (BIM) (Das, 1980; Andrews, 1985; Cochard & Madariaga, 1994; Geubelle & Rice, 1995) have been widely used for earthquake simulations . Recent advances focus on reducing computational costs while maintaining accuracy, including hybrid approaches like the finite element–spectral boundary integral method (Ma et al., 2019; Abdelmeguid et al., 2019; Abdelmeguid & Elbanna, 2022). However, despite these advances, the primary bottleneck in dynamic rupture modeling remains the inherent computational intensity, especially in the context of dynamic inversion for fault stress and friction. Each inversion requires solving the equations governing fault slip across a wide range of possible initial conditions, material properties, and frictional parameters, often necessitating large number of forward model evaluations. Therefore, developing more efficient approaches remains essential to overcome these computational challenges. Rapid simulations allow the study of fault behavior under varying conditions. They also enable large-scale statistical analyses, helping to identify rare or extreme events, which are essential for seismic hazard assessment. Additionally, accelerated accurate simulations reduce the computational burden of traditional methods, making large-scale inverse analyses more feasible.

In recent years, machine learning methods and reduced-order models have emerged as a powerful tool for addressing these computational challenges in various domains, including seismology (Zhu et al., 2019; L. Yang et al., 2022; Zhu & Beroza, 2019; Mousavi et al., 2020; Ross et al., 2018, 2019; Zhu et al., 2022; Rekoske et al., 2025). Deep learning approaches, specifically neural operators, have shown promise in approximating solutions to partial differential equations (PDEs), which govern many physical processes (Kovachki et al., 2023; Azizzadenesheli et al., 2024; Rahman et al., 2022). Unlike traditional neural networks, which map finite-dimensional input vectors to output vectors, neural operators learn mappings between entire functions. This capability allows them to solve complex problems governed by PDEs without requiring retraining, enabling rapid and efficient solution generation.

In seismology, neural operators have been applied to 2D acoustic wave equation (B. Li et al., 2023; Y. Yang et al., 2021), the elastic wave equation (Lehmann et al., 2023, 2024; Zou et al., 2024; Zhang et al., 2023), viscoelastic wave models (Wei & Fu, 2022), and full-waveform inversion (Y. Yang et al., 2023). While these applications have demonstrated the effectiveness of neural operators in handling wave propagation and seismic inversion tasks, their potential for modeling dynamic rupture propagation remains unexplored.

Dynamic rupture modeling presents unique challenges for machine learning due to the highly nonlinear and multiscale nature of fault slip processes. Capturing the evolution of rupture fronts, where stresses and slip rates exhibit sharp gradients, requires fine spatial and temporal resolution, leading to high computational costs. One potential approach that can address such challenges is Fourier Neural Operators (FNOs) (Z. Li et al., 2020). Through spectral representation in Fourier space, FNOs efficiently capture long-range correlations and high-dimensional dependencies. By adding more Fourier modes, it is also possible to capture increasingly higher frequencies. Moreover, their superior trade-off between computational cost and accuracy makes them well-suited for this problem (de Hoop et al., 2022).

Here, we present a framework for accelerating earthquake dynamic rupture simulations using FNOs. We use FNOs to generate the full spatio-temporal evolution of the fault slip rate. The proposed approach leverages the efficiency and accuracy of FNOs to address the large computational demands required by physics-based numerical methods. Specifically, we train FNOs on synthetic datasets incorporating heterogeneous distributions of initial shear stress, initial slip rate, frictional parameters, and stress perturbation for nucleation. The performance of the FNO framework is evaluated across various

120 datasets, demonstrating its ability to generalize and accurately predict fault slip dynamics  
 121 with significant computational speedup.

122 The remainder of the paper is organized as follows. In Section 2 we give an overview  
 123 of FNOs and describe the dynamic rupture problem setup. We discuss the nature of the  
 124 different datasets used in training the model in Section 3. We summarize our results in  
 125 Section 4. We discuss the implications of our findings and summarize our conclusions  
 126 in Section 5.

## 127 2 Preliminaries and Problem Setup

### 128 2.1 Fault Model Setup

129 We consider a 2D planar fault embedded in a 3D homogeneous, isotropic, and elas-  
 130 tic domain  $\Omega$ . The domain includes a traction boundary  $\Gamma_T$ , a displacement boundary  
 131  $\Gamma_u$ , and a fault surface located along  $\Gamma_f$ . We adopt the TPV101 benchmark problem setup  
 132 from the SCEC/USGS dynamic earthquake rupture code verification exercise (Harris et  
 133 al., 2018, 2009; Kammer et al., 2021). The governing equations are given by:

$$134 \rho \frac{\partial^2 u_i}{\partial t^2} - \frac{\partial \sigma_{ij}}{\partial x_j} - b_i = 0 \quad \text{in } \Omega, \quad (1)$$

$$\sigma_{ij} n_j = T_i \quad \text{on } \Gamma_T, \text{ eq : traction}_b \quad (2)$$

$$u_i = u_i^0 \quad \text{on } \Gamma_u, \quad (3)$$

$$(u_i^+ - u_i^-) = \delta_i, \quad T_{f,i}^+ = -T_{f,i}^- \quad \text{on } \Gamma_f. \quad (4)$$

134 Here,  $\rho$  is the density,  $u_i$  represents the displacement vector,  $b_i$  denotes the body  
 135 force vector, and  $\delta_i$  is the slip. The superscripts + and - refer to the positive and neg-  
 136 ative sides of the fault, respectively. The stress-strain relationship can be defined by lin-  
 137 ear elasticity:

$$\sigma_{ij} = \lambda \delta_{ij} \epsilon_{kk} + 2\mu \epsilon_{ij}, \quad (5)$$

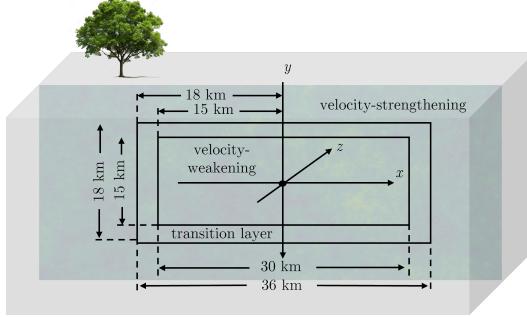
138 where  $\lambda$  and  $\mu$  are the Lamé parameters,  $\delta_{ij}$  is the Kronecker delta, and  $\epsilon_{ij}$  is the  
 139 strain tensor. The domain is characterized by its density ( $\rho$ ), shear wave speed ( $c_s$ ), and  
 140 pressure wave speed ( $c_p$ ), as summarized in Table 1.

141 The shear component of the fault boundary condition is governed by a regularized  
 142 rate-and-state friction law with an aging formulation (Dieterich, 1979; Ruina, 1983; Ben-  
 143 Zion & Rice, 1997; Lapusta et al., 2000; Ampuero & Rubin, 2008). This formulation re-  
 144 lates the fault's strength to the slip rate ( $V$ ), normal stress ( $\sigma$ ), and state variable ( $\theta$ )  
 145 through the following equations.

$$\tau = a\sigma \operatorname{arcsinh} \left[ \frac{V}{2V^*} \exp \left( \frac{f^* + b \ln(V^* \theta / D_{RS})}{a} \right) \right], \quad (6)$$

146 In expression (6), the parameter combination of  $a$ , and  $b$  dictates the stability of  
 147 the frictional interface.  $a-b > 0$  corresponds to a steady state rate-strengthening fric-  
 148 tional response (VS) in which sliding is stable. In contrast,  $a-b < 0$  corresponds to a  
 149 steady state rate-weakening frictional response (VW) which may lead to unstable slip  
 150 and stick-slip sequences.

$$\frac{d\theta}{dt} = 1 - \frac{V\theta}{D_{RS}}. \quad (7)$$



**Figure 1.** Schematic illustration of the 3D fault model, showing the central VW regime surrounded by the VS regime separated by a transition layer. The shear stress follows a fractal distribution, and the hypocenter is varied within the VW region.

A schematic representation of the fault geometry is shown in Figure 1. The fault consists of a central velocity-weakening (VW) regime, defined by  $-W < x < W$  and  $-W/2 < y < W/2$ , with  $W = 15$  km. This region defines the boundaries of the seismogenic zone which is characterized by a negative value of  $a - b$ . Surrounding the VW region is a velocity-strengthening (VS) regime, which is characterized by a positive value of  $a - b$ , with a transition layer of width  $w = 3$  km. To ensure a smooth transition between these regions, the parameter  $a$  varies smoothly based on a mathematically smoothed version of the boxcar function. The variation in  $a$  is governed by:

$$a = a_0 + \Delta a(x, y), \quad (8)$$

$$\Delta a(x, y) = \Delta a_0 \left[ 1 - B(x; W, w) B\left(y; \frac{W}{2}, w\right) \right], \quad (9)$$

where the function  $B(x; W, w)$  is defined as:

$$B(x; W, w) = \begin{cases} 1, & |x| \leq W, \\ \frac{1}{2} \left[ 1 + \tanh \left( \frac{w}{|x| - W - w} + \frac{w}{|x| - W} \right) \right], & W < |x| < W + w, \\ 0, & |x| \geq W + w. \end{cases} \quad (10)$$

We introduce variations into the initial shear stress field ( $\tau_0$ ) to capture spatial stress heterogeneity, which studies suggest follows a fractal-like distribution consistent with the roughness of fault surfaces (Andrews, 1980; Renard & Candela, 2017). Moreover, models incorporating fractal distributions align with key seismological patterns, such as the Gutenberg-Richter law (Hirata, 1989). We generate a fractal shear stress field with a specified fractal dimension  $D$ , mean, and standard deviation. The frequency-domain representation is constructed using a power-law scaling:

$$P(k) \propto \frac{1}{k^{2.5-D}}, \quad (11)$$

where  $k$  is the normalized wavenumber. This scaling yields larger amplitudes at low wavenumbers and smaller amplitudes at high wavenumbers.

169 We assign a random phase  $\phi$  to each frequency component. The phase is drawn from  
 170 a uniform distribution  $\phi \sim \mathcal{U}(0, 2\pi)$ . The complex frequency-domain representation is  
 171 constructed as:

$$\text{Spectrum} = P(k) \cdot [\cos(\phi) + i \sin(\phi)], \quad (12)$$

172 After constructing the spectrum, we apply an inverse FFT to transform the data  
 173 back into the spatial domain. The resulting real-valued shear stress field is normalized  
 174 to ensure it matches the target statistical properties, including the specified mean and  
 175 standard deviation:

$$S'(x) = \frac{S(x) - \mu}{\sigma} \cdot \sigma' + \mu', \quad (13)$$

176 where  $\mu$  and  $\sigma$  are the mean and standard deviation of the raw field  $S(x)$ , and  $\mu'$   
 177 and  $\sigma'$  are the mean and standard deviation of the scaled field  $S'(x)$ , respectively.

178 The dynamic rupture dataset is generated using the Spectral Boundary Integral  
 179 (SBI) method, which is currently the fastest available solver for simulating rupture prop-  
 180 agation along a planar fault in a homogeneous medium (Geubelle & Rice, 1995). This  
 181 method involves solving the coupled equations governing traction and displacement con-  
 182 tinuity along the fault surface. As a result, it eliminates the need to solve the govern-  
 183 ing equations throughout the entire domain. According to (Geubelle & Rice, 1995), the  
 184 response of the governing equations 1–4 is given by

$$\tau_i(x_1, x_3, t) = \tau_i^0 - \eta_{ij}^\pm \dot{u}_j^\pm(x_1, x_3, t) + f_i^\pm(x_1, x_3, t), \quad (14)$$

185 where  $\tau_i$  is the traction at the fault's surface on the half-space lying in the  $x_1$ – $x_3$   
 186 plane.  $\tau_i^0$  denotes the far-field traction.  $\dot{u}_i$  denotes the particle velocity,  $\eta_{ij}$  is the radi-  
 187 ation damping coefficient matrix, and  $f_i$  represents an integral term of the deformation  
 188 history, computed via time convolution in the spectral domain.

189 In SBI simulations, the fault's initial slip rate is prescribed as the constant  $V_{ini} =$   
 190  $10^{-12}$  m/s. To satisfy the friction law, the initial state variable is computed for each spa-  
 191 tial location, incorporating the spatial variability of the parameter  $a$  and the imposed  
 192 fractal shear stress distribution. The rupture is started by artificially overstressing a fault  
 193 segment. The hypocenter location varies across realizations but remains within the VW  
 194 regime. Details of this nucleation procedure are provided in Appendix A. We adopt UGUCA  
 195 code (Kammer et al., 2021) to generate the dynamic rupture dataset.

## 196 2.2 Fourier Neural Operators

197 The Fourier Neural Operator (FNO) is a learning-based operator designed to map  
 198 between infinite-dimensional spaces using input-output pairs,  $\{a_j, u_j\}_{j=1}^N$ . FNO replaces  
 199 the kernel integral operator in traditional neural operators with a convolution operator  
 200 defined in Fourier space. Subsequently, an inverse Fourier transform is applied, along with  
 201 a linear transformation. The resulting values are then passed through an activation func-  
 202 tion, which is applied to the sum of the transformed features. Mathematically, the neu-  
 203 ral operator follows an iterative update process,  $v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_T$ , where  
 204 the update from  $v_0$  to  $v_M$  and the output  $u(x)$  are defined as follows:

**Table 1.** Fault model parameters in the Spectral Boundary Integral (SBI) framework

Medium Parameter	Symbol	Value
Shear wave speed (km/s)	$c_s$	3.464
Pressure wave speed (km/s)	$c_p$	6
Density (kg/m <sup>3</sup> )	$\rho$	2670
<b>Fault Parameters</b>		
Reference coefficient of friction	$f^*$	0.6
Characteristic slip (m)	$D_{RS}$	0.02
Reference slip velocity (m/s)	$V^*$	$10^{-6}$
Length of VW patch in $x$ direction (km)	$2W$	30
Width of VW patch in $y$ direction (km)	$W$	15
Length of transition (km)	$w$	3
Length of the fault in $x$ direction (km)	$L_{fx}$	72
Width of the fault in $y$ direction (km)	$L_{fy}$	36
Evolution effect parameter	$b$	0.012
Steady state velocity dependence in VW patch	$a_{VW} - b$	varies
Steady state velocity dependence in VS patch	$a_{VS} - b$	varies
Initial velocity (m/s)	$V_{ini}$	$10^{-12}$
Initial normal stress (MPa)	$\sigma_{ini}$	120
Target mean of fractal shear stress (MPa)	$\mu'$	75
Target standard deviation of fractal shear stress (MPa)	$\sigma'$	5
<b>Nucleation Parameters</b>		
Nucleation radius (km)	$R$	3
Maximum nucleation amplitude (MPa)	$\Delta\tau_0$	25
Final nucleation time (s)	$T$	1

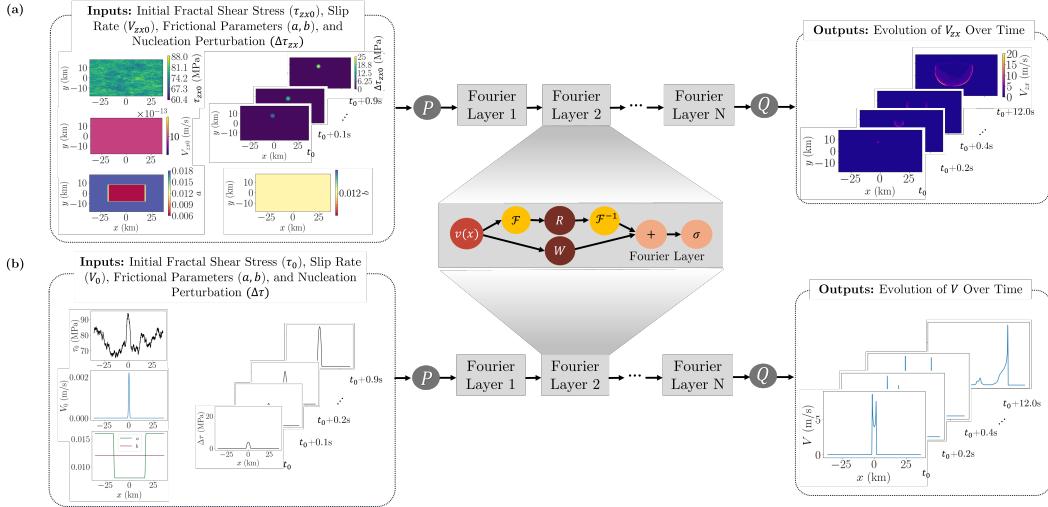
$$\begin{aligned}
 v_0(x) &= P(a(x)) \\
 v_{m+1}(x) &= \sigma \left( W_m v_m(x) + \int k(x, y) v_m(y) dy \right), \quad m = 0, \dots, M-1 \quad (15) \\
 u(x) &= Q(v_M(x))
 \end{aligned}$$

where  $P$  is a pointwise lifting operator parameterized with a neural network that projects the point values of input function  $a(x)$  to a higher dimension.  $W_m$  is a linear transformation applied to  $v_m(x)$  to account for non-periodic boundary conditions.  $\int k(x, y) v_m(y) dy$  represents a kernel integral operator.  $Q$  is a pointwise projection operator parameterized with a neural network that maps back to the target dimension.  $\sigma$  is a nonlinear activation function.

In FNO, we replace the kernel integral operator with a convolution operator using the fast Fourier transform. Thus, we can rewrite the kernel integral operator as follows:

$$\int k(x, y) v_m(y) dy = \mathcal{F}^{-1}(\mathcal{F}(k_m) \cdot \mathcal{F}(v_m)) \quad (16)$$

where  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  are the Fourier transform and its inverse, respectively. Using the Fourier transform enhances computational efficiency, allows the model to handle global dependencies, and enables resolution invariance, meaning the model can generalize across different spatial and temporal resolutions without requiring retraining. Specifically, when



**Figure 2.** Schematic diagram of the proposed FNO framework. (a) FNO framework for a 2D fault plane embedded in a 3D bulk. (c) FNO framework for a 1D fault embedded in a 2D plane. The inputs include the distribution of initial shear stress, slip rate, frictional parameters, and stress perturbation of nucleation. The outputs are snapshots of slip rate over time.

the input function is provided in regular grids, the Fourier transformer is carried using fast Fourier transform, a celebrated and fast algorithm for Fourier analysis. It also leads to better generalization across grids, reduces computational costs for high-dimensional problems, and improves the model's ability to handle complex, nonlinear dynamics. We refer the reader to (Z. Li et al., 2020) for a discussion of this architecture and related features.

Here, the FNO is designed to take as inputs the initial shear stress, the slip rate  $V$  for the 2D case or its component  $V_{zx}$  in 3D, frictional parameters, and the nucleation stress perturbation. The model outputs the evolution of the slip rate  $V$  in 2D or  $V_{zx}$  in 3D, as illustrated in Figure 2. We use this framework to allow for variations in (1) the initial fractal distribution of shear stress, (2) the initial stage of the rupture realization (i.e., the input can be provided at different time point in the simulation), (3) the values of frictional parameters  $a$  and  $b$ , and (4) the location of nucleation sites. This setup captures the variability and uncertainty inherent in natural faults, where initial stress conditions, frictional properties, and nucleation behavior are often poorly constrained or spatially heterogeneous.

### 2.3 Evaluating Model Predictions

The accuracy of predictions is assessed using a measure of the difference between the ground truth and the predicted slip rate over time. We present two error metrics: the relative  $L_2$  error and the normalized root mean squared error (NRMSE). During the training and testing stages, the loss function is calculated using the relative  $L_2$  error:

$$\text{Relative } L_2 \text{ error} = \frac{\|V_{\text{pred}}(\mathbf{x}, t) - V_{\text{true}}(\mathbf{x}, t)\|_2}{\|V_{\text{true}}(\mathbf{x}, t)\|_2} \quad (17)$$

where  $V_{\text{pred}}(\mathbf{x}, t)$  is the predicted slip rate at location  $\mathbf{x}$  and time  $t$ , and  $V_{\text{true}}(\mathbf{x}, t)$  is the corresponding true value. However, this metric can be misleading sometimes be-

cause the slip rate is close to zero at some points. As a result, the metric can be disproportionately influenced.

Moreover, shifts in space and time are inherent in the dynamic rupture problem (Barall & Harris, 2015). These shifts introduce bias into the relative  $L_2$  error calculation, especially when the shifted prediction is compared to a small ground truth value. These biases will be discussed in each of the results sections. To address this issue, we introduce another metric, NRMSE:

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left( V_{\text{pred},j}^{(i)} - V_{\text{true},j}^{(i)} \right)^2}}{V_{\max} - V_{\min}} \quad (18)$$

where  $M$  is the total number of time steps,  $N$  is the total number of spatial points,  $V_{\text{pred},j}^{(i)}$  is the predicted value at spatial point  $j$  and time step  $i$ , and  $V_{\text{true},j}^{(i)}$  is the corresponding true value obtained from the numerical simulations.  $V_{\max}$  and  $V_{\min}$  are the maximum and minimum values of  $V_{\text{true}}$  over the entire space-time domain. NRMSE compares the error with the observed range of the ground truth. As such, the predictions will not be disproportionately influenced by small ground truth values.

### 3 Data Configuration

In the SBI solver, the key parameters used in the model are summarized in Tables 1 and 2. The fractal dimensions are chosen to represent natural faults, with  $D$  varying between 1.2 and 1.6 (Renard & Candela, 2017). The rate-and-state frictional parameters fall within the observed range from laboratory experiments (Ikari et al., 2011; Barbot, 2022). The dataset is obtained by solving the dynamic rupture problem over the time interval  $[0, 15]$  s.

Two datasets are generated, corresponding to 2D and 3D dynamic rupture simulations. The 3D simulation treats the fault as a 2D plane embedded in a 3D bulk, which is trained using FNO-2D, as shown in Figure 2a. It employs a 2D Fourier Transform to capture spatial correlations in both dimensions. The 2D simulation treats the fault as a 1D cross-section along the hypocenter in the  $x$ -axis. This dataset is trained using a separate FNO-1D, as shown in Figure 2b. It applies a 1D Fourier Transform to extract frequency features and efficiently learn spatial dependencies.

For the initial shear stress distribution, expressed in Equation 11, the normalized wave number  $k_i$  for the 1D fault is defined as:

$$k_i = \frac{i}{n_x} \quad (19)$$

where  $n_x$  is the total number of spatial points. We set  $k_0 = 1$  for  $i = 0$  to avoid division by zero. This choice is arbitrary as it only affects the mean value of the stress distribution which gets overridden later by matching the target mean value.

For the 2D fault, each grid point  $(i, j)$  in Fourier space corresponds to a wave number pair  $(k_x, k_y)$ , defined as:

$$k_x = \frac{i}{n_x}, \quad k_y = \frac{j}{n_y}, \quad k = \sqrt{k_x^2 + k_y^2} \quad (20)$$

where  $n_x$  and  $n_y$  are the total number of points in the  $x$  and  $y$  directions, respectively.

277        **3.1 2D Dynamic Rupture Dataset**

278        In the 2D dataset, the fault is represented as a one-dimensional line embedded in  
279        a 2D elastic bulk, assuming plane strain conditions. A total of 8,200 realizations are gen-  
280        erated, with parameters listed in Table 2. Among these, 4,000 realizations have the hypocen-  
281        ter at the center of the fault, 2,100 have hypocenters located at the point of maximum  
282        shear stress in the VW region, and the remaining 2,100 have randomly located hypocen-  
283        ters drawn from a uniform distribution along the fault.

284        Each realization is processed to create four different starting points based on a slip  
285        rate threshold,  $V_{\text{th}}$ . We define this threshold as the time step when the maximum slip  
286        rate in the domain first exceeds  $V_{\text{th}}$ . From that point, we extract the input features: ini-  
287        tial shear stress, slip rate, and nucleation perturbation. The frictional parameters are  
288        kept unchanged across all starting points. We use four threshold values:  $V_{\text{th}} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ .  
289        These thresholds are typically used to define the onset of seismic events in multi cycles  
290        simulations. This approach results in four sets of data for each of the original 8,200 re-  
291        alizations, giving a total of 32,800 data samples. Of these, 28,700 are used for training  
292        and 4,100 for testing.

293        The FNO is trained to approximate the mapping from the initial shear stress  $\tau_0$ ,  
294        slip rate  $V_0$ , frictional parameters  $a$  and  $b$ , and stress perturbation  $\Delta\tau$  to the sequence  
295        of slip rates over the time interval of interest. The input data is structured as a tensor  
296        of dimensions  $(N, X, C_{\text{in}})$ , while the output data has dimensions  $(N, X, C_{\text{out}})$ , where  $N$   
297        denotes the number of realizations,  $X$  represents the number of spatial discretization points,  
298         $C_{\text{in}}$  and  $C_{\text{out}}$  correspond to the number of input and output channels, respectively.

299        The spatial discretization consists of 2,880 points, i.e.,  $X = 2,880$ , consistent with  
300        the resolution used in the SBI solver. The input channels include the spatial distribu-  
301        tions of the initial shear stress  $\tau_0$  and slip rate at  $t_0$ , frictional parameters  $a$  and  $b$ , and  
302        ten time steps of the nucleation stress perturbation  $\Delta\tau$ , sampled at 0.1 s intervals start-  
303        ing from the time  $t_0$ , when the slip rate first exceeds the threshold  $V_{\text{th}}$ . These ten per-  
304        turbation snapshots correspond to  $t = t_0, t_0 + 0.1 \text{ s}, \dots, t_0 + 0.9 \text{ s}$ , contributing ten ad-  
305        ditional input channels. In total, this yields  $C_{\text{in}} = 14$ . This potentially opens up the  
306        possibility of using different nucleation over-stress distribution.

307        We solve the governing equations using the SBI solver with a time step of  $\Delta t =$   
308        0.001 s, which satisfies the Courant–Friedrichs–Lewy (CFL) condition,  $\Delta t \leq f\Delta x/c_p$ ,  
309        where  $f$  is a constant of order one. Here, we choose  $f = 0.25$ . Since FNOs are inde-  
310        pendent of the numerical discretization, we adopt a coarser time step of 0.2 s in the FNO  
311        model. The FNO is trained to perform a single-shot prediction of 60 time steps of slip  
312        rate, starting from time  $t_0$ , resulting in  $C_{\text{out}} = 60$ .

313        Before training, the dataset is normalized, scaling each feature in both the input  
314        and output to the range  $[0, 1]$ . After training, the original physical scale is restored us-  
315        ing an inverse transformation. This normalization enhances numerical stability and fa-  
316        cilitates efficient training of the FNO model (Cuomo et al., 2022).

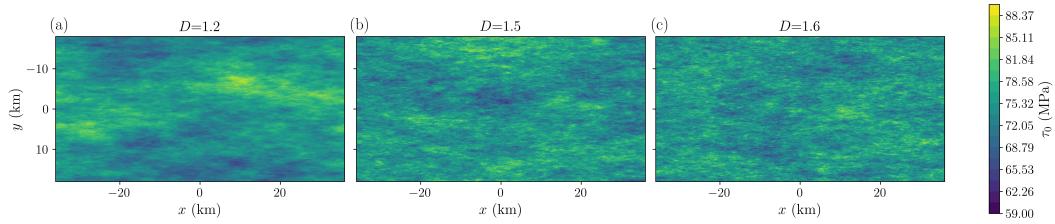
317        **3.2 3D Dynamic Rupture Dataset**

318        In the 3D dataset, the fault is modeled as a 2D plane embedded in a 3D elastic bulk.  
319        The variations in fractal dimension and frictional parameters  $a$  and  $b$  are listed in Ta-  
320        ble 2. Examples of fractal initial shear stress distributions generated using different frac-  
321        tal dimensions are illustrated in Figure 3.

322        Similar to the 2D case, 8,200 realizations are generated: 4,000 with the hypocen-  
323        ter at the fault center, 2,100 at the location of maximum shear stress in the VW region,  
324        and 2,100 with uniformly random hypocenter locations. Each realization is processed in  
325        the same manner as the 2D case to produce four distinct initial conditions based on the

**Table 2.** Fractal dimensions and variation of frictional parameter  $a$  and  $b$  for 2D and 3D dynamic rupture dataset

$D$	$b$	$a$	
		$a_0$	$\Delta a_0$
1.2, 1.5, 1.6	0.012, 0.014	0.009	0.006
		0.008	0.008
		0.007	0.010
		0.006	0.012
		0.0085	0.007
		0.0075	0.009
		0.0065	0.011



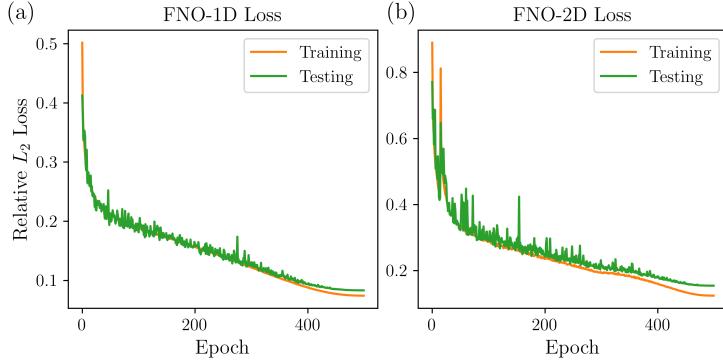
**Figure 3.** Initial fractal shear stress distributions  $\tau_0$  with fractal dimensions of (a) 1.2, (b) 1.5, and (c) 1.6. The target mean is set to 75 MPa, with a target standard deviation of 5 MPa.

slip rate threshold  $V_{\text{th}} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ . This results in 32,800 data samples. From these, 28,700 are used for training and 4,100 for testing.

As in the 2D case, the FNO is trained to approximate the mapping from the initial shear stress  $\Delta\tau_{zx0}$ , slip rate  $V_{zx0}$ , frictional parameters  $a$  and  $b$ , and stress perturbation  $\Delta\tau_{zx}$  to the sequence of component of slip rates  $V_{zx}$  over the time interval of interest. The input data is structured as a tensor of dimensions  $(N, X, Y, C_{\text{in}})$ , while the output data has dimensions  $(N, X, Y, C_{\text{out}})$ , where  $X$  and  $Y$  represent the number of spatial discretization points in the  $x$  and  $y$  directions, respectively.

The 2D fault is discretized into  $720 \times 360$  spatial points, i.e.,  $X = 720$  and  $Y = 360$ . The input channels consist of the spatial distributions of  $\tau_0$ ,  $V_0$ ,  $a$ ,  $b$ , and ten steps of  $\Delta\tau$ , sampled from  $t = t_0$  to  $t = t_0 + 0.9$  s, resulting in  $C_{\text{in}} = 14$ .

The SBI solver uses a time step of  $\Delta t = 0.01$  s to satisfy the CFL condition with  $\Delta x = \Delta y = 100$  m and  $f = 0.6$ . Solutions are output at a frequency of 0.1 s. For FNO training, we generate 60 time steps of  $V_{zx}$  evolution at 0.2 s intervals starting from  $t = t_0$ , resulting in output dimensions  $(N, X, Y, C_{\text{out}}) = (28700, 720, 360, 60)$ , which corresponds to approximately 2 TB of data. Given the available memory and GPU limitations, training with a single batch at this scale is challenging. Thus, we reduce the number of spatial and temporal points by sub-sampling from  $(X, Y, C_{\text{out}}) = (720, 360, 60)$  to  $(360, 180, 60)$  for FNO training, reducing the dataset size to approximately 416 GB. In this case, the output channels correspond to snapshots of  $V_{zx}$  spatially sampled at half the resolution of the original simulations and temporally recorded at 0.2 s intervals. This adjustment makes training feasible on the available NVIDIA A100 GPU with 64 GB of memory.



**Figure 4.** Training and testing losses for (a) 2D dynamic rupture dataset and (b) 3D dynamic rupture dataset.

As with the 2D dataset, we also normalize the 3D dataset, scaling each feature in both the input and output to the range [0, 1]. After training, the original physical scale is restored using an inverse transformation.

## 4 Results

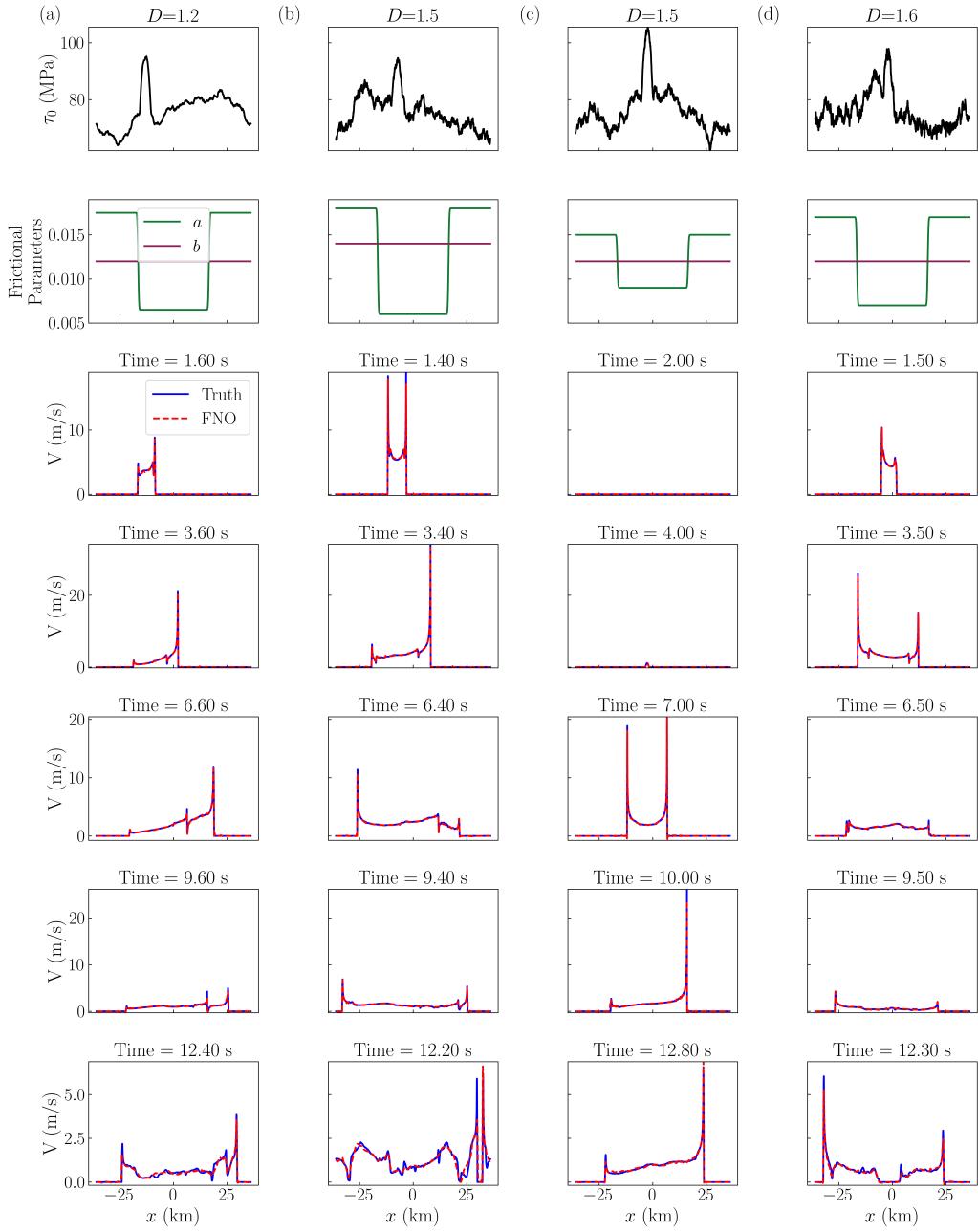
### 4.1 FNO-1D for 2D Dynamic Rupture Dataset

#### 4.1.1 Training and Testing Performance

We optimize the hyperparameters of FNO-1D, including the number of modes, Fourier layers, lifting and projection layers, and the learning rate. Details of hyperparameter tuning and training strategies are listed in Appendix B. The model selection criterion balances accuracy with computational efficiency by maintaining a minimal number of parameters. Based on this tuning process, we configure the model with four Fourier layers ( $m = 4$ ), while the lifting network  $P$  and projection network  $Q$  each consist of 128 neurons. The number of retained modes after applying the Fourier transform and subsequent linear transformation is set to 16. The activation function is the Gaussian Error Linear Unit (GELU) (Hendrycks & Gimpel, 2016). The model is trained using a batch size of 10, with a relative  $L_2$  loss function and the Adam optimizer (Kingma & Ba, 2014), adopting a learning rate of  $10^{-3}$  and a weight decay of  $10^{-4}$  with a cosine annealing schedule. The training process is conducted for 500 epochs. The training and testing losses are shown in Figure 4a. Both training and testing losses consistently decrease without large divergence between them, indicating no signs of overfitting.

The performance of FNO-1D is evaluated on the testing dataset. The model predictions, compared against the ground truth, are presented in Figure 5. FNO-1D effectively captures a wide range of magnitudes arising from different distributions of  $\tau_0$ ,  $V_0$ ,  $\Delta\tau$ ,  $a$ , and  $b$ . The model successfully reproduces the global slip rate, particularly in regions characterized by sharp gradients, such as the rupture front. Discrepancies arise in regions where resolving closely spaced high-frequency components is required. These features reflect a phenomenon known as spectral bias (Rahaman et al., 2019; Cao et al., 2019; Kong et al., 2025), which implies that deep learning models predict high-frequency features less accurately than lower-frequency ones. In Appendix B, we discuss a potential strategy for improving the model's performance by training on more datasets.

We summarize the NRMSE and relative  $L_2$  error on the training and testing datasets in Figure 6. Testing samples show slightly higher errors but still follow the same distri-



**Figure 5.** Testing of the trained FNO-1D model on the 2D dynamic rupture dataset. (a) Inputs include the initial fractal shear stress distribution  $\tau_0$  with  $D = 1.2$ , initial slip rate at the time step where  $\max(V) > V_{\text{th}}$  with  $V_{\text{th}} = 1e-2$ , frictional parameters  $a$  and  $b$  where  $a_0 = 0.0065$ ,  $\Delta a_0 = 0.011$ , and  $b = 0.012$ , and 10 steps of stress perturbation  $\Delta\tau$  starting from the same thresholded step. The outputs are predicted slip rate snapshots at selected time steps following this point (rows 3–7). (b–d) Same as (a), but with varying conditions: (b)  $D = 1.5$ ,  $a_0 = 0.006$ ,  $\Delta a_0 = 0.012$ ,  $b = 0.014$ ,  $V_{\text{th}} = 10^{-4}$ ; (c)  $D = 1.5$ ,  $a_0 = 0.009$ ,  $\Delta a_0 = 0.006$ ,  $b = 0.012$ ,  $V_{\text{th}} = 10^{-2}$ ; (d)  $D = 1.6$ ,  $a_0 = 0.007$ ,  $\Delta a_0 = 0.010$ ,  $b = 0.012$ ,  $V_{\text{th}} = 10^{-3}$ . In all cases, predictions are generated starting from the time step where the slip rate exceeds the specified threshold.

**Table 3.** Median and median absolute deviation (MAD) of NRMSE and relative  $L_2$  error in the bracket shown in Figure 6 for 28,700 training samples and 4,100 testing samples of 2D and 3D dynamic rupture datasets.

	2D Dynamic Rupture	3D Dynamic Rupture
	NRMSE (Relative $L_2$ Error)	NRMSE (Relative $L_2$ Error)
Training	$0.00357 \pm 0.000957$ ( $0.0665 \pm 0.01581$ )	$0.00317 \pm 0.000612$ ( $0.115 \pm 0.0239$ )
Testing	$0.00388 \pm 0.00114$ ( $0.0270 \pm 0.00909$ )	$0.00366 \pm 0.000940$ ( $0.133 \pm 0.0368$ )

bution as the training set, suggesting no major overfitting. The majority of samples have small errors and are highly skewed towards zero, with the median and median absolute deviation (MAD) of the errors presented in Table 3. According to Figure 6b, more than 95% of the training set shows a relative  $L_2$  error of less than 10%, while this fraction drops to 90% for the testing set. This level of the error suggests that the model generalizes well to unseen data. Examples of the predictions corresponding to different selected values of the relative  $L_2$  error are presented in the supplementary information.

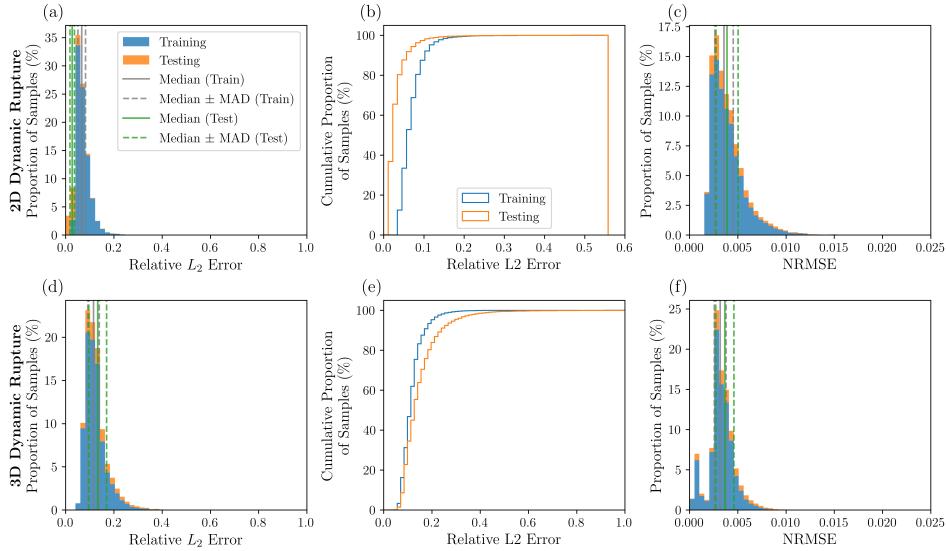
#### 4.1.2 Generalization to Unseen Initial Shear Stress

An advantage of FNOs is that they can target the underlying operator, and learn to handle a family of problems rather than a single instance. Accordingly, in this section we test the generalization of our trained operator to some unseen initial conditions.

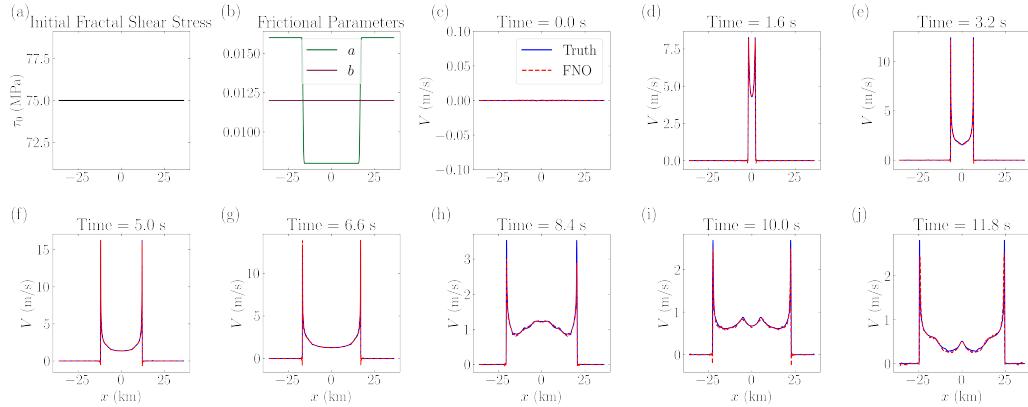
We evaluate the trained FNO-1D model, originally trained on initial fractal shear stress  $\tau_0$ , using a uniform initial shear stress consistent with TPV101 problem description (Harris et al., 2018, 2009; Kammer et al., 2021). The initial shear stress of  $\tau_0 = 75$  MPa is prescribed along the entire fault. This value is within the range of shear stress values in the fractal dataset. However, the network has not seen a spatially uniform case before this test. We compare predictions from the FNO-1D model against the ground truth in Figure 7. The computed NRMSE and relative  $L_2$  error between the predictions and ground truth are 0.00560 and 0.0381, respectively. FNO-1D shows ability to capture slip rate evolution, particularly during rupture propagation in the VW regime and transitioning into the VS region. We observe minor noise near the rupture front in the predicted solution, along with slight mismatches in peak magnitudes at later times. However, the global features of slip rate remains consistent with the ground truth. These results highlight the robustness of FNO-1D, as the model successfully generalizes to this unseen distribution of  $\tau_0$ .

Additionally, we evaluate the performance of FNO-1D on an unseen fractal dimension  $D = 1.3$ , which was not included in the training set. We generate 100 realizations with  $D = 1.3$  and randomly sampled frictional parameters from the ranges listed in Table 2. After preprocessing the data using slip rate thresholds  $V_{\text{th}} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ , we obtain a total of 400 realizations for testing. The distribution of relative  $L_2$  error and NRMSE is shown in Figures 8a and 8c. The distribution shows a high variance with some cases have relative  $L_2$  error close to 1.0. Expanding the training dataset is recommended to improve prediction accuracy.

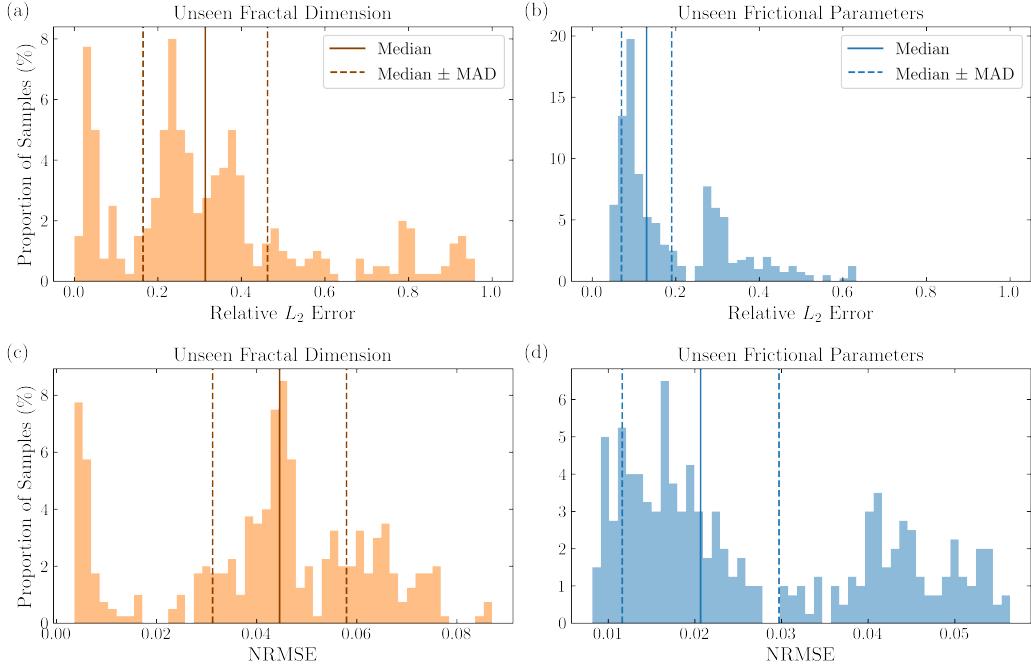
Figure 9 shows an example of results from FNO-1D compared to the ground truth, with NRMSE and relative  $L_2$  error of 0.00254 and 0.0131, respectively. FNO-1D effectively captures both the low-frequency and high-frequency components of the slip rate



**Figure 6.** Error analysis for 2D and 3D dynamic rupture predictions. (a) and (d) show histograms of the relative  $L_2$  error distributions for the training (blue) and testing (orange) datasets in the 2D and 3D cases, respectively. Similarly, (c) and (f) show histograms of the NRMSE distributions for the training (blue) and testing (orange) datasets in the 2D and 3D cases. The vertical lines indicate the median error for the training (gray solid) and testing (green solid) datasets, as well as the median  $\pm$  median absolute deviation (MAD) range for the training (gray dashed) and testing (green dashed) datasets. (b) and (e) show the cumulative histograms of the relative  $L_2$  error distributions for the training (blue) and testing (orange) datasets in the 2D and 3D cases, respectively.



**Figure 7.** Results of FNO-1D testing on an unseen shear stress distribution from the TPV101 SCEC/USGS benchmark. Panels (a) and (b) show selected input features: (a) a uniform initial shear stress  $\tau_0 = 75$  MPa, and (b) the frictional parameters  $a$  and  $b$ , with  $a_0 = 0.008$ ,  $\Delta a_0 = 0.008$ , and  $b = 0.012$ . Other model inputs (e.g., initial slip rate  $V_0$  with a threshold  $V_{th} = 0$  m/s and nucleation stress perturbation  $\Delta\tau$ ) are not shown. Panels (c)-(j) show the outputs consist of predicted slip rate profiles at selected time steps.



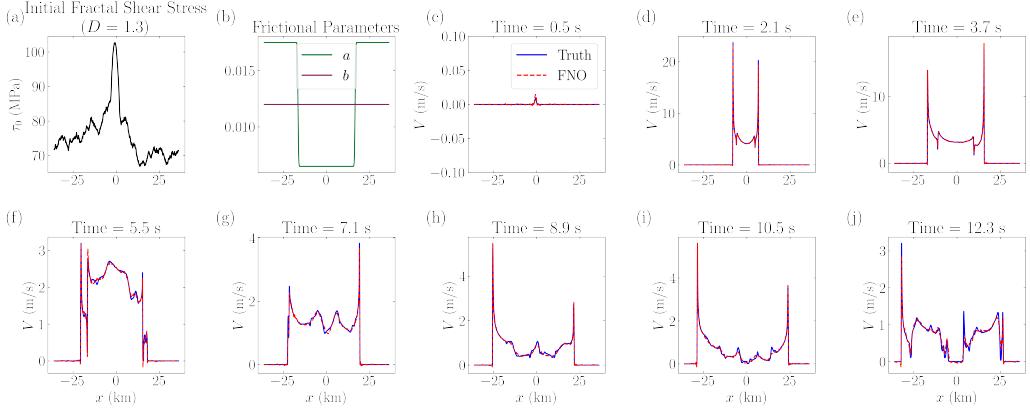
**Figure 8.** The distribution of (a) the relative  $L_2$  error and (c) the NRMSE evaluated on unseen fractal dimension cases, and (b) the relative  $L_2$  error and (d) the NRMSE evaluated on unseen frictional parameters for the 2D model. These results are based on 400 realizations for each case. Vertical lines indicate the median and MAD. For unseen fractal dimension cases, the median  $\pm$  MAD of the relative error is  $0.313 \pm 0.149$  and of the NRMSE is  $0.0446 \pm 0.0134$ . For unseen frictional parameters cases, the median  $\pm$  MAD of the relative error is  $0.130 \pm 0.0600$  and of the NRMSE is  $0.0207 \pm 0.00905$ .

with minor discrepancies at the peak at 12.3 s. The predicted rupture front also closely matches the ground truth during propagation into the VS region. This test further corroborates the potential of FNO-1D in predicting slip rates under unseen initial shear stress conditions.

#### 4.1.3 Generalization to Unseen Frictional Parameters

We test the FNO-1D model on unseen frictional parameters, specifically the values of parameters  $a$  and  $b$  listed in Table 4. We generate 100 realizations by randomly sampling the parameters uniformly from the specified ranges in the table. Then, we apply the same pre-processing using  $V_{\text{th}} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ . As such, we obtain a total of 400 testing sets. The distributions of the relative  $L_2$  error and NRMSE across these 400 testing sets are shown in Figures 8b and 8d. It is worth noting that more than 60% of the testing sets exhibit a relative  $L_2$  error below 0.20, demonstrating the ability of the proposed FNO to generalize across different frictional parameters.

An example comparison between the predicted and ground truth slip rates is shown in Figure 10. In this case, the values of parameter  $a$  are 0.0067 in the VW region and 0.0177 in the VS region, while the parameter  $b$  is set uniformly to 0.012 along the fault. The initial shear stress  $\tau_0$  follows a fractal distribution with a fractal dimension of  $D = 1.6$ . This realization yields an NRMSE and a relative  $L_2$  error of 0.00827 and 0.0464, respectively. The FNO-1D model successfully captures the overall slip rate evolution,



**Figure 9.** Example of results of FNO-1D testing on an unseen fractal dimension  $D = 1.3$ . Panels (a) and (b) show selected input features: (a) initial fractal shear stress distributions ( $\tau_0$ ) with  $D = 1.3$  and (b) frictional parameters  $a$  and  $b$ , with  $\Delta a_0 = 0.0065$ ,  $a_0 = 0.011$ , and  $b = 0.012$ . Other model inputs (i.e., initial slip rate  $V_0$  with a threshold  $V_{th} = 10^{-3}$  m/s and nucleation stress perturbation  $\Delta\tau$ ) are not shown. Panels (c)-(j) show the outputs consist of predicted slip rate profiles at selected time steps.

although some discrepancies are observed at high frequencies. The predicted rupture front closely follows the ground truth, with only slight mismatches at the peaks of rupture front. Minor oscillations occur at the transition from the initial slip rate  $V_0$  to the sharp gradient at the rupture front along the fault.

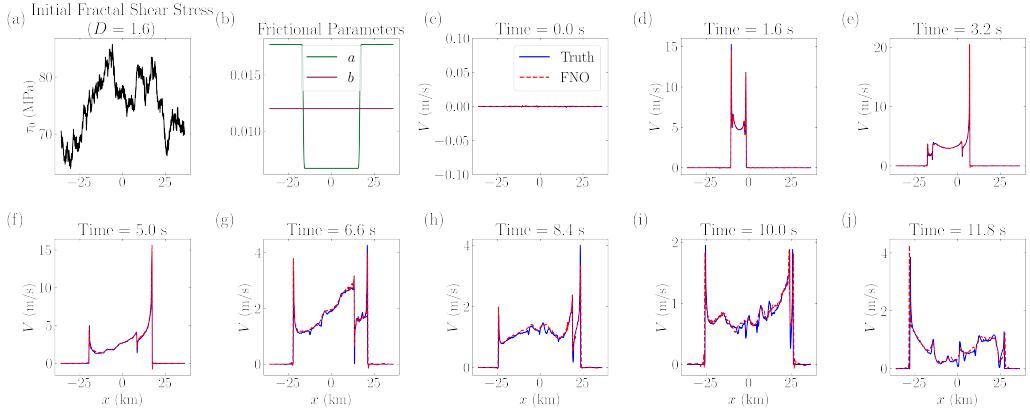
**Table 4.** Fractal dimensions and variation of frictional parameters  $a$  and  $b$  for testing the generalization to unseen frictional parameters

$D$	$b$	$a$	
		$a_0$	$\Delta a_0$
1.5	0.012, 0.014	0.0087	0.007
		0.0077	0.009
		0.0067	0.011

## 4.2 FNO-2D for 3D Dynamic Rupture Dataset

### 4.2.1 Training and Testing Performance

We perform hyperparameter tuning for FNO-2D, resulting in an architecture consisting of four Fourier layers. The lifting and projecting fully connected neural networks each contain 128 neurons. After applying the Fourier transform, 32 modes are retained in the linear transformation, and the GELU activation function is employed. The model is trained using the Adam optimizer with a batch size of 10, a learning rate of  $10^{-3}$ , and a weight decay of  $10^{-4}$ , following a cosine annealing schedule. Training spans 500 epochs until the loss stabilizes. Training and testing losses over optimization epochs are shown in Figure 4b. Both training and testing losses consistently decrease without significant divergence, indicating no signs of overfitting.



**Figure 10.** Example of results of FNO-1D testing on unseen  $a - b$ . Panels (a) and (b) show selected input features: (a) initial fractal shear stress distributions ( $\tau_0$ ) with  $D = 1.6$  and (b) frictional parameters  $a$  and  $b$ , with  $\Delta a_0 = 0.0067$ ,  $a_0 = 0.011$ , and  $b = 0.012$ . Other model inputs (i.e., initial slip rate  $V_0$  with a threshold  $V_{th} = 0$  m/s and nucleation stress perturbation  $\Delta\tau$ ) are not shown. Panels (c)-(j) show the outputs consist of predicted slip rate profiles at selected time steps.

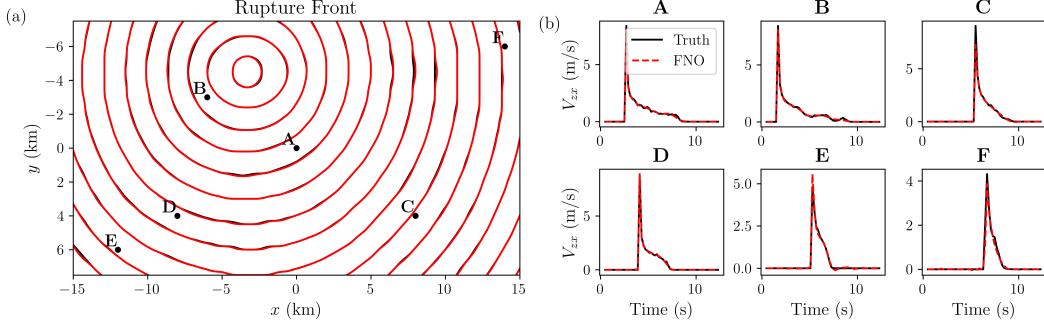
Figure 11 illustrates a successful example of the predictions compared to the ground truth. It shows the rupture front contours in the VW region at 0.5-second intervals, along with time series at six selected points in the VW region. FNO-2D accurately captures the rupture front contours with NRMSE of 0.003035 and relative  $L_2$  error of 0.1002. The predicted rupture arrival times and waveform shapes align closely with the ground truth, although the nucleation site is not centered. Minor discrepancies can be observed in peak amplitudes, particularly at points C, E and F. However, the overall agreement between the model prediction and the ground truth remains high.

The distributions of relative  $L_2$  error and NRMSE during the prediction phase, shown in Figures 6d, 6e, and 6f, further confirm that no overfitting is occurring, as the testing dataset follows the same trend as the training dataset. Moreover, the distribution is skewed toward zero. More than 90% of the training dataset has relative  $L_2$  errors less than 20%, while this proportion drops to 80% for the testing dataset. The modest decrease in this proportion confirms that the model generalizes well and does not overfit. Interestingly, we have found that even in cases where the model predictions show higher relative  $L_2$  errors, the model still successfully captures the major features of rupture dynamics. Examples of predictions with higher relative error are provided in the supplementary information.

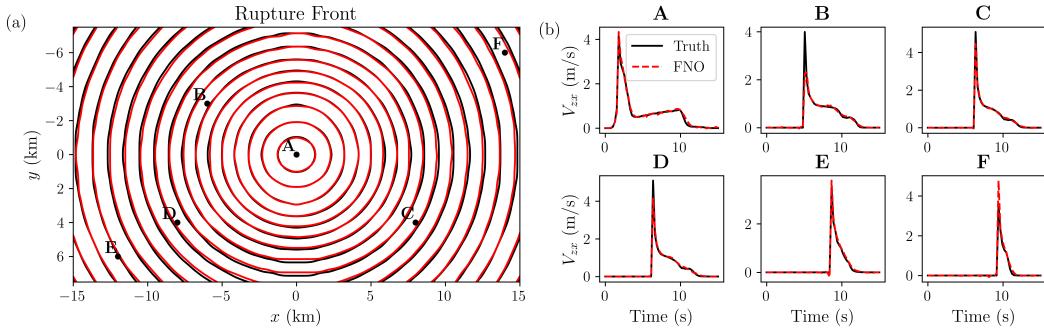
#### 4.2.2 Generalization to Unseen Initial Shear Stress

Similar to the 2D model, we evaluate the trained FNO-2D model using an unseen initial shear stress distribution, specifically the uniform stress distribution defined in the TPV101 benchmark from the SCEC/USGS dynamic earthquake rupture code verification exercise. In this benchmark, the initial shear stress is uniformly set to 75 MPa, while frictional parameters remain within the training range of the 3D dynamic rupture dataset.

Figure 12 shows the rupture contours of the predicted component of slip rate  $V_{zx}$  compared to the ground truth, as well as the time history at six selected points in the VW region. FNO-2D can capture the evolution of  $V_{zx}$ . The predicted rupture front aligns



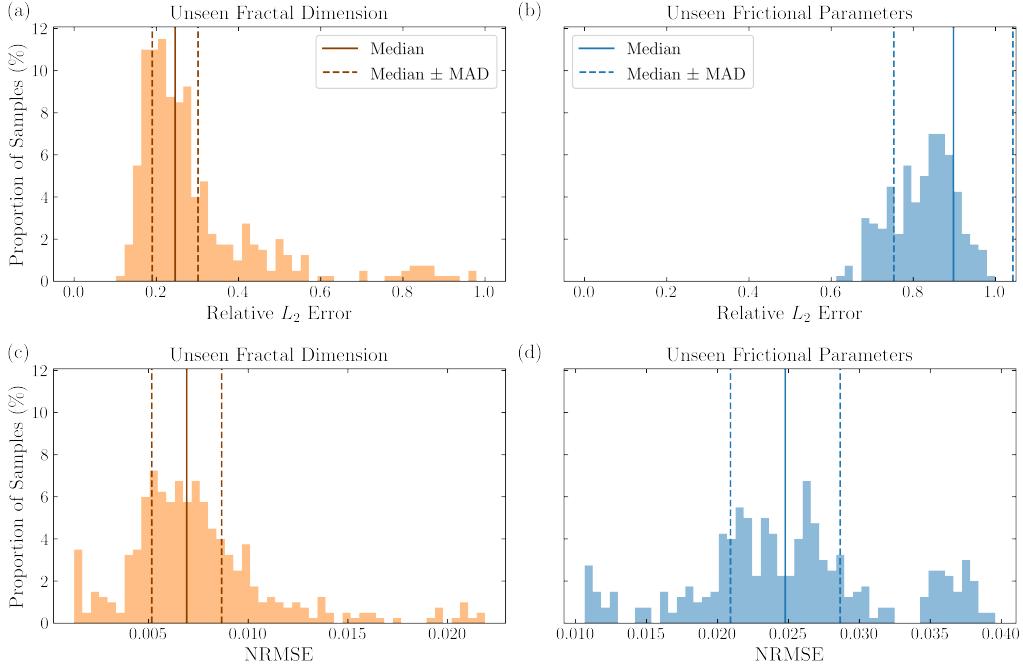
**Figure 11.** Results of FNO-2D on the testing dataset for 3D dynamic rupture. Inputs include a fractal initial shear stress distribution with  $D = 1.6$ , initial slip rate  $V_{zx0}$  extracted at the threshold  $V_{th} = 10^{-3}$  m/s, ten temporal steps of stress perturbation fields, and spatially varying frictional parameters:  $a_0 = 0.007$ ,  $\Delta a_0 = 0.010$ , and  $b = 0.012$ . (a) Predicted rupture front contours shown at 0.5 s intervals. (b) Time series of slip rate at selected locations within the VW region.



**Figure 12.** Results of FNO-2D testing on an unseen shear stress distribution from the TPV101 SCEC/USGS benchmark with a uniform shear stress of 75 MPa,  $\Delta a_0 = 0.008$ ,  $a_0 = 0.008$ ,  $b = 0.012$ , and  $V_{th} = 0$  m/s. (a) Rupture front contour plot of the ground truth (black) and predictions (red), showing progression at 0.5 s intervals. (b) Time histories of slip rate at selected points in the VW region.

closely with the ground truth with NRMSE of 0.01505 and relative  $L_2$  error of 0.3821. The predicted time series also match well with the ground truth, capturing the peaks and fine-scale details of  $V_{zx}$  over time. Minor time shifts appear at later stages as the rupture approaches the VW boundaries. Although there are some discrepancies in peak magnitudes, which contribute to the overall error, the model successfully captures the rupture arrival time and the overall waveform shape. This demonstrates the potential of FNO-2D in capturing the highly nonlinear evolution of slip rate over time, even under an initial shear stress distribution not previously seen during training.

Additionally, we test the model on a fractal initial shear stress distribution with an unseen fractal dimension  $D = 1.3$ , similar to the FNO-1D case. We generate an additional 100 realizations with  $D = 1.3$  and randomly sampled frictional parameters from the seen ranges during training, as listed in Table 2. We preprocess the data such that  $V_{zx0}$  is chosen at the step following thresholds  $V_{th} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$  m/s. This results in 400 realizations. The distributions of relative  $L_2$  error and NRMSE are shown



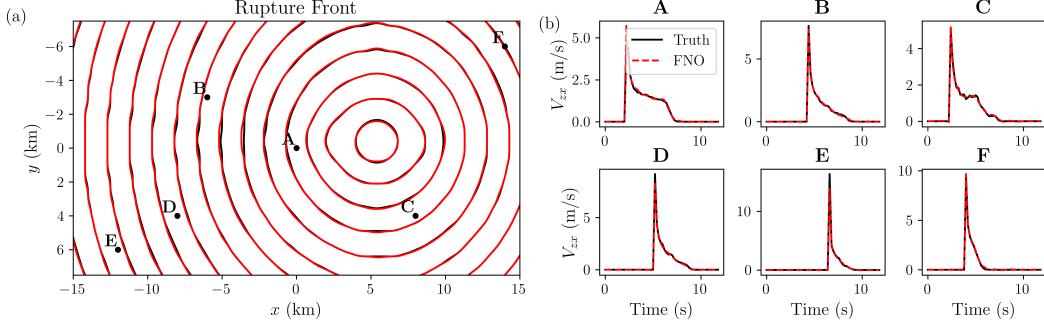
**Figure 13.** The distribution of (a) the relative  $L_2$  error and (c) the NRMSE evaluated on unseen fractal dimension cases, and (b) the relative  $L_2$  error and (d) the NRMSE evaluated on unseen frictional parameters for the 3D model. These results are based on 400 realizations for each case. Vertical lines indicate the median and MAD. For unseen fractal dimension cases, the median  $\pm$  MAD of the relative error is  $0.245 \pm 0.0557$  and of the NRMSE is  $0.00692 \pm 0.00175$ . For unseen frictional parameters cases, the median  $\pm$  MAD of the relative error is  $0.898 \pm 0.145$  and of the NRMSE is  $0.0248 \pm 0.00386$ .

in Figures 13a and 13c. Both distributions are skewed towards zero, showing generalizability to unseen fractal dimension cases.

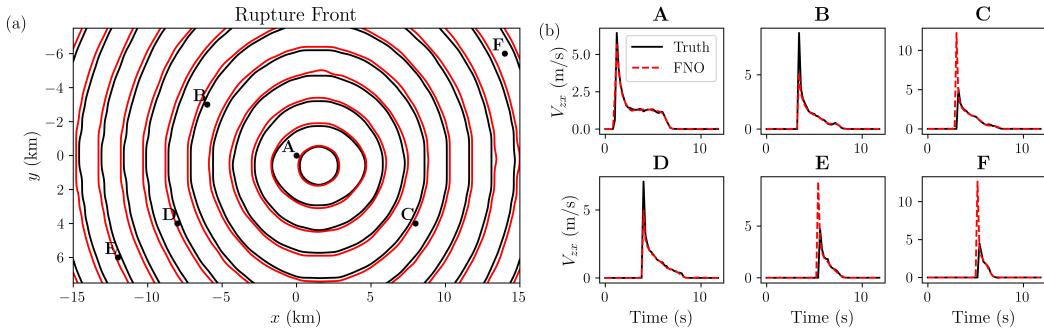
An example of the predictions is presented in terms of the rupture front and time series at selected points in Figure 14. The frictional parameters are  $\Delta a_0 = 0.011$ ,  $a_0 = 0.0065$ , and  $b = 0.012$ , and here,  $V_{\text{th}} = 0$  m/s. The rupture front contour aligns with the ground truth, resulting in an NRMSE of 0.00422 and a relative  $L_2$  error of 0.1284 for this realization. The model accurately captures the rupture arrival time. Peak amplitudes exhibit very small mismatches at points D and E. Furthermore, the predictions of post-peak decay closely match the ground truth. This demonstrates that FNO-2D generalizes effectively to different fractal dimensions of  $\tau_0$ .

#### 4.2.3 Generalization to Unseen Frictional Parameters

We evaluate the generalization performance of the trained FNO-2D model on previously unseen frictional parameter values. Specifically, we generate 100 realizations using frictional parameters sampled from combinations not included in the training dataset, as summarized in Table 4. After preprocessing the data based on velocity thresholds  $V_{\text{th}} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ , we obtain a total of 400 realizations with varying  $V_{zx0}$  fields. Figures 13b and 13d show the distributions of the relative  $L_2$  error and the NRMSE, respectively. Although the relative  $L_2$  error is high, with a median of 0.898, the range of the NRMSE in the 3D model is smaller than that in the 2D model. Despite the lower ac-



**Figure 14.** Results of FNO-2D testing on an unseen fractal dimension  $D = 1.3$  of the initial fractal shear stress. The frictional parameters are  $\Delta a_0 = 0.011$ ,  $a_0 = 0.0065$ , and  $b = 0.012$ . The initial  $V_{zx0}$  is chosen at the time step where  $V_{th} = 0$  m/s: (a) Rupture front contour plot of the ground truth (black) and predictions (red), showing progression at 0.5 s intervals. (b) Time histories of slip rate at selected points in the VW region.



**Figure 15.** Results of FNO-2D testing on an unseen  $a - b$  distribution. Inputs consist of frictional parameters  $a$  and  $b$ , with  $\Delta a_0 = 0.011$ ,  $a_0 = 0.0067$ , and  $b = 0.012$ , along with an initial shear stress distribution  $\tau_0$  with a fractal dimension  $D = 1.5$  and  $V_{th} = 0$  m/s. (a) Rupture front contour plot showing progression at 0.5 s intervals. (b) Time histories of slip rate at selected points in the VW region.

511 accuracy as measured by the relative  $L_2$  error, the 3D model demonstrates greater stability  
512 in terms of normalized error when compared with the observed range in  $V_{zx}$ .

513 Figure 15 illustrates a representative case, including the predicted rupture front and  
514 time histories of the slip rate component  $V_{zx}$  at selected spatial locations. In this exam-  
515 ple, the frictional parameter  $a$  is set to 0.0067 in the VW region and 0.0177 in the VS  
516 region, while parameter  $b$  is uniformly fixed at 0.012 along the fault. The initial shear  
517 stress  $\tau_0$  is constructed using a fractal distribution with a fractal dimension  $D = 1.5$ .

518 We observe spatial and temporal discrepancies in the rupture front contours and  
519 the time histories at points C, E, and F, which contribute to elevated errors, with an NRMSE  
520 of 0.0219 and a relative  $L_2$  error of 0.6767. In particular, mismatches in the peak mag-  
521 nitudes of  $V_{zx}$  at these locations account for a significant portion of the error. However,  
522 the FNO-2D model successfully captures the overall temporal evolution of  $V_{zx}$ , includ-  
523 ing the nucleation, propagation, and arrest phases of rupture. In the following discus-  
524 sion, we address the specific challenges in generalizing to unseen frictional properties and  
525 propose strategies to improve predictive accuracy.

**Table 5.** Comparison of different models in terms of parameters, training time, testing performance, and speed-up. The testing times are calculated by taking the median of 100 random cases. The speed-up is compared with numerical simulation run time of 345 s for 2D and 1182 s for 3D.

# Parameter (-)	Training (s/epoch)	Testing	
		Prediction time (s)	Speed-up vs. numerical simulation (times)
FNO-1D	2,348,492	30	$2 \times 10^5$
FNO-2D	268,739,404	594	$4 \times 10^5$

### 4.3 Computational Efficiency Analysis

We summarize the computational efficiency of FNO-1D and FNO-2D in Table 5. Training and testing times are reported for an NVIDIA A100 GPU, while the 3D dynamic rupture dataset is trained using four parallel NVIDIA A100 GPUs. Once trained on the NCSA Delta system via ACCESS allocation (Boerner et al., 2023), these models can be directly employed as efficient alternatives to numerical simulators.

For the application we are proposing, the trained model is used directly for subsequent predictions. Therefore, we compare the FNO model’s performance to the computational time required by a numerical solver. To evaluate the potential computational speed-up, we compare the numerical simulation run time on an AMD EPYC 7763 “Milan” (PCIe Gen4) CPU to the FNO’s testing time. The run times are 345 s for the 2D simulation and 1182 s for the 3D simulation. The predictions from the 2D and 3D simulations are  $2 \times 10^5$  and  $4 \times 10^5$  times faster than the conventional numerical simulation, respectively.

Notably, the prediction time does not increase significantly, despite the substantial increase in the number of parameters, from 2D to 3D. The training time is slower for the 3D simulation. However, we prioritize prediction accuracy and testing time over the off-line training time.

## 5 Discussion and Conclusion

To overcome the computational bottleneck of classical physics-based earthquake models, we present an FNO-based surrogate model to accelerate dynamic rupture simulations. In the surrogate model, we predict the evolution of the slip rate distribution on the fault plane, given the initial distribution of shear stress, slip rate, stress perturbation, and frictional parameters  $a$  and  $b$  in the rate-and-state friction law. The ground truth is generated using the SBI method with different initial fractal shear stress, slip rate, nucleation site, and frictional parameter distributions in both 2D and 3D domains. Each dataset is trained and tuned separately using FNO-1D and FNO-2D. We select the hyperparameters based on a balance between computational efficiency and accuracy as discussed in Appendix B. As a result, we achieve speedups of  $4 \times 10^5$  and  $2 \times 10^5$  compared to the SBI method in 3D and 2D dynamic rupture problems, respectively. This increase in speed potentially enables more efficient dynamic rupture inversion, statistical analysis, and identification of extreme events in ensembles of earthquake rupture forecasts.

Predicted slip rates are evaluated using pointwise metrics, including the NRMSE, which compares predictions with the observed range of the ground truth, and the relative  $L_2$  error, which assesses the discrepancy at discrete points in the domain. We demonstrate that the training process of FNO-based models does not exhibit overfitting. Test set predictions show strong agreement with the ground truth. Specifically, for the 2D

564 dynamic rupture dataset, the median NRMSE is 0.388% and the median relative  $L_2$  error  
565 is 2.70%. For the 3D dynamic rupture dataset, the median NRMSE is 0.366% and  
566 the median relative  $L_2$  error is 13.3%.

567 Higher errors in the 3D problem are primarily due to down-sampling to a lower res-  
568 olution than the physical problem, resulting in a grid size that may smear some details  
569 in the process zone region. This limitation is due to constrained computational resources.  
570 Access to additional resources would allow training at higher resolution, potentially re-  
571 ducing these errors to a level comparable to what is seen in the 2D problem. The errors  
572 are generally more pronounced in regions with high-frequency features, reflecting the spec-  
573 tral bias, which states that deep learning models tend to favor low-frequency components.  
574 To further improve model performance, we suggest training on a larger dataset.

575 We also evaluate the models under unseen initial stress distributions and frictional  
576 parameters. The FNO-based models show robustness in out-of-distribution cases, such  
577 as a uniform initial stress distribution corresponding to the SCEC/USGS verification ex-  
578 ercise. They also perform well on a shear stress distribution with an unseen fractal di-  
579 mension of  $D = 1.3$ , achieving median relative  $L_2$  errors of 31.3% in 2D and 24.5% in  
580 3D. By contrast, when tested on unseen frictional parameters, the model yields a me-  
581 dian relative  $L_2$  error of 13.0% in 2D but 89.8% in 3D. One important source of error  
582 in FNO predictions is spatial and temporal shifts. When the slip rate is small, the rel-  
583 ative  $L_2$  error can be amplified, particularly where sharp, high-value rupture fronts shift  
584 in space and time relative to small slip rate values. This reduced robustness may be at-  
585 tributed to the fact that variations in frictional parameters modify the underlying ma-  
586 terial model and consequently change the properties of the governing operator. In con-  
587 trast, changing the shear stress distribution affects only the initial conditions without  
588 modifying the governing equations. This limitation could be mitigated by training with  
589 a larger datasets. As demonstrated in Section Appendix B, increasing the size of the train-  
590 ing dataset improves performance. A shared community database of large-scale dynamic  
591 rupture simulations would be highly beneficial.

592 We highlight a couple of key aspects of FNO as applied to our problem. The FNO  
593 neural approximation effectively captures the evolution of slip rate over extended peri-  
594 ods. The FNO formulation is time-continuous and can be discretized as needed for train-  
595 ing and application. Furthermore, the FNO enables the use of data generated at vary-  
596 ing spatial and temporal discretizations. Traditional solvers must adhere to stability con-  
597 dition to maintain stability, whereas this constraint does not apply to FNOs, allowing  
598 for greater flexibility in discretization and time stepping. Third, the FNO shows poten-  
599 tial for successful generalization to unseen stress and frictional condition although our  
600 findings suggest that the performance will further improve by training on larger data sets.  
601 In addition, incorporating stress perturbations for rupture nucleation enables the model  
602 to learn the operator more flexibly, as one can vary the nucleation site location, nucle-  
603 ation radius, or even the spatial distribution of the perturbation. Lastly, this model serves  
604 as a proof-of-concept for simulating rupture scenarios given the initial conditions at any  
605 point in time.

606 We conclude with a discussion of limitations and potential future directions. First,  
607 this work assumes a specific mean and standard deviation for the input distributions.  
608 Future work should explore training the FNO on datasets with more diverse parameter  
609 distributions and non-dimensional quantities. Second, the current FNO is trained on datasets  
610 generated using a constant normal stress. To be more realistic, training the model on  
611 a heterogeneous normal stress is desirable. Third, the FNO is trained for predictions within  
612 a fixed time interval. Extending the prediction horizon beyond a specific interval could  
613 involve recursive training, enabling the model to predict slip evolution iteratively. Fur-  
614 thermore, a hybrid approach combining traditional numerical solvers with FNOs could  
615 enable long-range predictions. A key aspect of this approach would be defining a robust  
616 criterion, such as a physics-guided error threshold, for switching between FNO and tra-

ditional solvers. In addition, investigating more advanced surrogate model architectures may further enhance performance. Finally, access to larger training datasets and more GPUs will further improve the performance of FNOs and their ability to generalize. A community effort for creating a database for dynamic rupture simulations would be beneficial in that respect.

## Appendix A Details on Model Setup

Rupture is nucleated by applying a time- and space-dependent perturbation to the horizontal shear traction. The perturbation grows smoothly from zero to its maximum amplitude  $\Delta\tau_0$  over a finite time interval ( $T$ ) and is confined to a circular region of radius ( $R$ ) centered on the hypocenter. The nucleation perturbation is expressed as:

$$\Delta\tau(x, t) = \Delta\tau_0 F(x - x_0) G(t), \quad (\text{A1})$$

where the spatial function  $F(r)$  and temporal function  $G(t)$  are given by:

$$F(r) = \begin{cases} \exp\left(\frac{r^2}{r^2 - R^2}\right), & r < R, \\ 0, & r \geq R, \end{cases} \quad (\text{A2})$$

$$G(t) = \begin{cases} \exp\left[\frac{(t-T)^2}{t(t-2T)}\right], & 0 < t < T, \\ 1, & t \geq T. \end{cases} \quad (\text{A3})$$

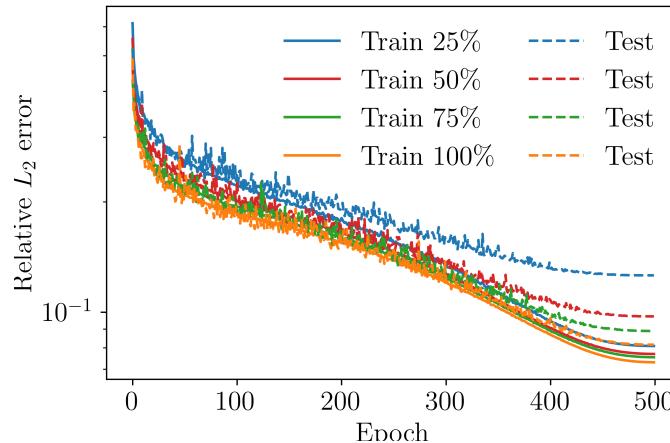
## Appendix B FNO Hyperparameters Tuning and Training Strategies

This section presents the training strategies and the influence of hyperparameters on prediction accuracy. We reuse the hyperparameters tuned on a three-feature model (initial shear stress and frictional parameters). The only change is preprocessing: feature-wise normalization is recomputed to accommodate 14 input features. While this choice reduces compute, it may be slightly suboptimal. For the 2D dynamic rupture dataset, we tune the model by varying the number of Fourier modes, the number of Fourier layers, and the learning rate over 50 epochs. The number of retained Fourier modes is initially selected as 16 and 32, while the number of Fourier layers is set to 4, 5, and 6. The depth of the fully connected layers  $P$  and  $Q$  is chosen from 32, 64, 128, and 256. We consider two learning rates, 0.001 and 0.0005, while fixing the batch size at 10. Based on Table B1, we find that models with four Fourier layers generally yield the lowest error, with the best-performing candidates highlighted in bold. These candidates are selected based on a balance between the total number of parameters in the FNO model and prediction accuracy. We further train these models for 500 epochs, and the differences in relative  $L_2$  error among the top three models remain insignificant, specifically within a 1% difference. Consequently, we select the final model as the one with the lowest number of parameters, consisting of four Fourier layers, 32 Fourier modes, a width of 128 for the lifting and projection layers, and a learning rate of 0.001. This model is then further trained for up to 10,000 epochs until the training and testing losses saturate. This approach helps save computational resources and allows for more efficient hyperparameter selection, especially for large datasets such as the 3D dynamic rupture dataset.

To improve generalizability, we propose training on a larger number of realizations while keeping the architecture and hyperparameters fixed. We assess the effect of the training set size by using 25%, 50%, 75%, and 100% of the available 28,700 realizations. The relative  $L_2$  error is then evaluated on a testing set of 3,280 realizations. The training losses

**Table B1.** Relative  $L_2$  errors for different hyperparameter configurations, including the number of Fourier layers, modes, learning rate, and the widths of  $P$  and  $Q$ , on the 2D dynamic rupture testing dataset.

Fourier Layers	Learning Rate	Modes	Width of $P$ and $Q$			
			32	64	128	256
4	0.001	16	0.327423	0.279825	<b>0.199039</b>	0.225360
		32	0.319402	0.263659	0.205039	0.206671
	0.0005	16	0.376180	0.306192	0.241678	0.214775
		32	0.361853	0.271715	0.238622	<b>0.198188</b>
5	0.001	16	0.324249	0.338498	0.328546	0.323920
		32	0.309520	0.340786	0.324978	0.321918
	0.0005	16	0.379091	0.340786	0.316518	0.312066
		32	0.352530	0.329741	0.322074	0.305138
6	0.001	16	0.330633	0.334641	0.324744	0.328914
		32	0.297148	0.343834	0.320379	0.321232
	0.0005	16	0.359399	0.275518	0.237807	0.425277
		32	0.328937	0.250830	0.273538	0.212822



**Figure B1.** Training and testing losses for different numbers of training realizations: 25%, 50%, 75%, and 100% of 7,380 sets. Solid lines show training losses, and dashed lines show testing losses.

for the different training set sizes are shown in Figure B1. At epoch 500, the generalization gap between the training and testing losses decreases as the number of training realizations increases. Likewise, the relative  $L_2$  error on the testing set decreases with larger training sets. The relative  $L_2$  errors for training on 25%, 50%, 75%, and 100% of the data are 0.0377, 0.0304, 0.0281, and 0.0266, respectively. This improvement demonstrates the benefit of training on a larger dataset.

For the 3D dynamic rupture dataset, we primarily tune the model while keeping the number of Fourier layers fixed at four, based on insights from the 2D dynamic rupture dataset. The number of retained Fourier modes is initially selected as 16, 32, and 64, while the depth of  $P$  and  $Q$  is chosen from 32, 64, and 128. We consider learning rates of 0.001 and 0.0005. The results of the tuning process are presented in Table B2. The

665 optimal model is selected with four Fourier layers, 32 Fourier modes, 128 neurons in  $P$   
 666 and  $Q$ , and a learning rate of 0.001.

**Table B2.** Relative  $L_2$  errors for different hyperparameter configurations, including the number of Fourier layers, modes, learning rate, and the widths of  $P$  and  $Q$ , tested on data from a 3D simulation of dynamic rupture.

Learning Rate	Modes	Width of $P$ and $Q$		
		32	64	128
0.0005	16	0.127383	0.115418	0.113272
	32	0.108283	0.105937	<b>0.080965</b>
	64	0.102793	0.126559	0.114296
	16	0.123509	0.093694	0.107529
	32	0.101131	0.096700	0.083106
	64	0.096357	0.106043	0.091511

## Appendix C Quantitative Analysis of Prediction Error

As an additional validation metric, we compare the FNO predictions against the ground truth using the  $Q$  metric, which is specifically designed for time-series comparisons in dynamic rupture simulations (Barall & Harris, 2015). The  $Q$  metric is defined as

$$Q(t_s) = \frac{\|f(t) - g(t - t_s)\|_2}{\|f(t)\|_2 + \|g(t - t_s)\|_2}, \quad (\text{C1})$$

where  $f(t)$  denotes the vector of slip rate predictions from the FNO model,  $g(t)$  represents the corresponding ground truth values, and  $t_s$  is a temporal shift applied to align the signals. The optimal time shift is obtained as

$$t_s^* \in \arg \min_{t_s} Q(t_s). \quad (\text{C2})$$

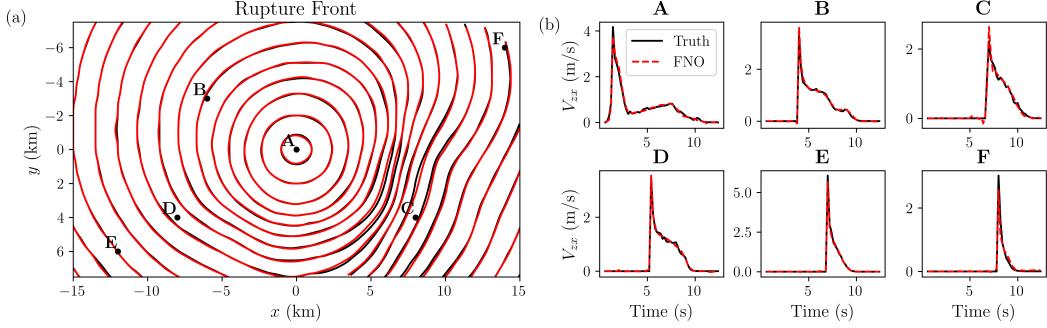
A smaller  $Q$  value indicates closer agreement between prediction and ground truth.

For illustration, we select three representative cases from the 3D dynamic rupture dataset, corresponding to relative  $L_2$  errors equal to the median, median + 2MAD, and median + 4MAD. For each case, we compute  $Q$  and report the associated optimal shift  $t_s^*$ .

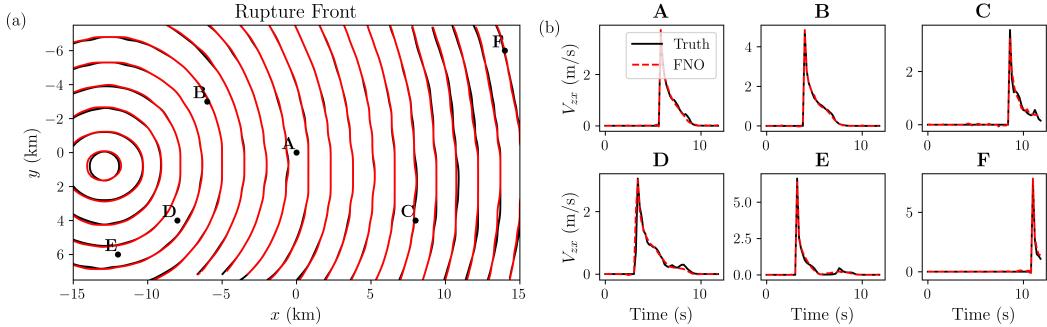
Figures C1–C3 present the rupture front evolution and slip rate histories for these cases. The computed  $Q$  values are 12.94%, 20.6%, and 39.6%, with corresponding optimal time shifts  $t_s^*$  of 0.005 s, 0.004 s, and 0.025 s, respectively. Figure C4 shows the relative error of the peak  $V_{zx}$  as a function of time step for the case with relative  $L_2$  error closest to the median. The error increases after time step 40, which marks the stage when the rupture arrives at the VS boundaries. This introduces higher-frequency components into the slip rate field that are more challenging to capture. Additional contributing factors include the limited availability of training samples, particularly for cases where the rupture reaches the boundaries, the restriction imposed by the chosen number of modes in the FNO, and the inherent spectral bias of neural operators.

## Open Research

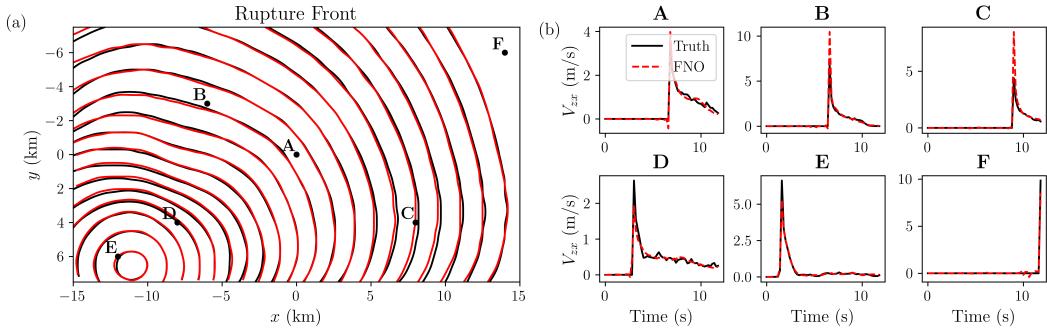
All data used in the analysis are available online at Tainpakdipat et al. (2025). The FNO software used to conduct the numerical experiments will be made available upon acceptance.



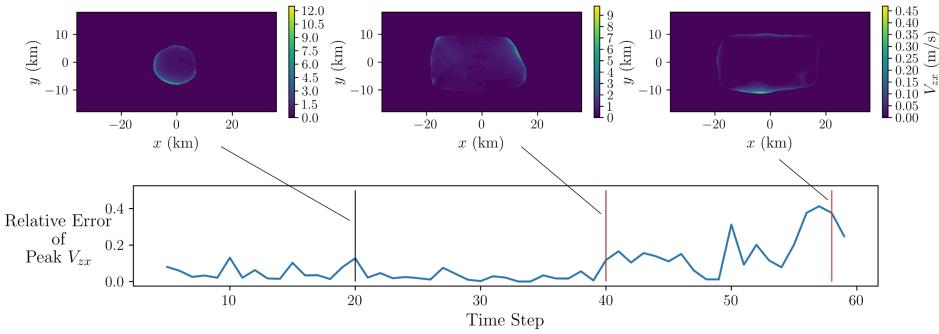
**Figure C1.** Results of FNO-2D on the testing dataset for a 3D dynamic rupture case with a relative  $L_2$  error of 0.130 (median) and an NRMSE of 0.00354. The inputs include the initial fractal shear stress ( $D = 1.5$ ), the initial  $V_{zx}$  field ( $V_{th} = 10^{-3}$  m/s), a nucleation perturbation, and frictional parameters  $a$  ( $a_0 = 0.008$ ,  $\Delta a_0 = 0.008$ ) and  $b = 0.012$ . (a) Rupture front contours at 0.5 s intervals. (b) Time histories of slip rate at selected spatial locations.



**Figure C2.** Results of FNO-2D on the testing dataset for a 3D dynamic rupture case with a relative  $L_2$  error of 0.207 (median + 2MAD) and an NRMSE of 0.00531. The inputs include the initial fractal shear stress ( $D = 1.2$ ), the initial  $V_{zx}$  field ( $V_{th} = 10^{-3}$  m/s), a nucleation perturbation, and frictional parameters  $a$  ( $a_0 = 0.008$ ,  $\Delta a_0 = 0.008$ ) and  $b = 0.012$ . (a) Rupture front contours at 0.5 s intervals. (b) Time histories of slip rate at selected spatial locations.



**Figure C3.** Results of FNO-2D on the testing dataset for a 3D dynamic rupture case with a relative  $L_2$  error of 0.408 (median + 4MAD) and an NRMSE of 0.0108. The inputs include the initial fractal shear stress ( $D = 1.6$ ), the initial  $V_{zx}$  field ( $V_{th} = 0$  m/s), a nucleation perturbation, and frictional parameters  $a$  ( $a_0 = 0.008$ ,  $\Delta a_0 = 0.008$ ) and  $b = 0.012$ . (a) Rupture front contours at 0.5 s intervals. (b) Time histories of slip rate at selected spatial locations.



**Figure C4.** Top: Contours of the ground-truth slip rate  $V_{zx}$  at time steps 20, 40, and 58, illustrating rupture propagation and the stages when rupture reaches the fault boundaries. Bottom: Temporal evolution of the relative error in the peak slip rate, with vertical lines marking the reference time steps (20, 40, and 58). The last two vertical lines (in red) indicate when the rupture front reaches the boundaries, leading to higher error due to the emergence of high-frequency components.

#### Acknowledgments

The study was supported by the National Science Foundation (CAREER award No. 1753249 and OAC-2311207) and the Southern/Statewide California Earthquake Center (based on NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047). We also acknowledge funding provided by DOE EERE Geothermal Technologies Office to Utah FORGE and the University of Utah under Project DE-EE0007080 Enhanced Geothermal System Concept Testing and Development at the Milford City, Utah Frontier Observatory for Research in Geothermal Energy (Utah FORGE) site. This work used Delta at the National Center for Supercomputing Applications (NCSA) through allocation EES230038 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by U.S. National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

#### References

- Aagaard, B. T., Heaton, T. H., & Hall, J. F. (2001). Dynamic earthquake ruptures in the presence of lithostatic normal stresses: Implications for friction models and heat production. *Bulletin of the Seismological Society of America*, 91(6), 1765–1796.
- Abdelmeguid, M., & Elbanna, A. (2022). Modeling sequences of earthquakes and aseismic slip (seas) in elasto-plastic fault zones with a hybrid finite element spectral boundary integral scheme. *Journal of Geophysical Research: Solid Earth*, 127(12), e2022JB024548.
- Abdelmeguid, M., Ma, X., & Elbanna, A. (2019). A novel hybrid finite element-spectral boundary integral scheme for modeling earthquake cycles: Application to rate and state faults with low-velocity zones. *Journal of Geophysical Research: Solid Earth*, 124(12), 12854–12881.
- Ampuero, J.-P., & Rubin, A. M. (2008). Earthquake nucleation on rate and state faults—aging and slip laws. *Journal of Geophysical Research: Solid Earth*, 113(B1).
- Andrews, D. (1973). A numerical study of tectonic stress release by underground explosions. *Bulletin of the Seismological Society of America*, 63(4), 1375–1391.
- Andrews, D. (1980). A stochastic fault model: 1. static case. *Journal of Geophysical*



- 780 Dieterich, J. H. (1979). Modeling of rock friction: 1. experimental results and consti-  
781 tutive equations. *Journal of Geophysical Research: Solid Earth*, 84(B5), 2161–  
782 2168.
- 783 Duan, B., & Oglesby, D. D. (2006). Heterogeneous fault stresses from previous  
784 earthquakes and the effect on dynamics of parallel strike-slip faults. *Journal of*  
785 *Geophysical Research: Solid Earth*, 111(B5).
- 786 Gabriel, A.-A., Ampuero, J.-P., Dalguer, L., & Mai, P. M. (2012). The transition  
787 of dynamic rupture styles in elastic media under velocity-weakening friction.  
788 *Journal of Geophysical Research: Solid Earth*, 117(B9).
- 789 Gallovič, F., Valentová, Ľ., Ampuero, J.-P., & Gabriel, A.-A. (2019a). Bayesian  
790 Dynamic Finite-Fault Inversion: 1. Method and Synthetic Test. *Journal*  
791 *of Geophysical Research: Solid Earth*, 124(7), 6949–6969. doi: 10.1029/  
792 2019JB017510
- 793 Gallovič, F., Valentová, L., Ampuero, J.-P., & Gabriel, A.-A. (2019b). Bayesian dy-  
794 namic finite-fault inversion: 2. application to the 2016 mw 6.2 amatrice, italy,  
795 earthquake. *Journal of Geophysical Research: Solid Earth*, 124(7), 6970–6988.
- 796 Geubelle, P. H., & Rice, J. R. (1995). A spectral method for three-dimensional elas-  
797 todynamic fracture problems. *Journal of the Mechanics and Physics of Solids*,  
798 43(11), 1791–1824.
- 799 Harris, R. A., Barall, M., Aagaard, B., Ma, S., Roten, D., Olsen, K., ... others  
800 (2018). A suite of exercises for verifying dynamic earthquake rupture codes.  
801 *Seismological Research Letters*, 89(3), 1146–1162.
- 802 Harris, R. A., Barall, M., Archuleta, R., Dunham, E., Aagaard, B., Ampuero, J. P.,  
803 ... others (2009). The scec/usgs dynamic earthquake rupture code verification  
804 exercise. *Seismological Research Letters*, 80(1), 119–126.
- 805 Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv*  
806 preprint arXiv:1606.08415.
- 807 Hirata, T. (1989). A correlation between the b value and the fractal dimension  
808 of earthquakes. *Journal of Geophysical Research: Solid Earth*, 94(B6), 7507–  
809 7514.
- 810 Ikari, M. J., Marone, C., & Saffer, D. M. (2011). On the relation between fault  
811 strength and frictional stability. *Geology*, 39(1), 83–86.
- 812 Johnson, K. M., Burgmann, R., & Larson, K. (2006). Frictional properties on the  
813 san andreas fault near parkfield, california, inferred from models of afterslip  
814 following the 2004 earthquake. *Bulletin of the Seismological Society of Amer-*  
815 *ica*, 96(4B), S321–S338.
- 816 Kammer, D. S., Albertini, G., & Ke, C.-Y. (2021). Uguca: A spectral-boundary-  
817 integral method for modeling fracture and friction. *SoftwareX*, 15, 100785.
- 818 Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*  
819 preprint arXiv:1412.6980.
- 820 Kong, Q., Zou, C., Choi, Y., Matzel, E. M., Azizzadenesheli, K., Ross, Z. E., ...  
821 Clayton, R. W. (2025). Reducing frequency bias of fourier neural operators in  
822 3d seismic wavefield simulations through multi-stage training. *arXiv preprint*  
823 arXiv:2503.02023.
- 824 Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., &  
825 Anandkumar, A. (2023). Neural operator: Learning maps between function  
826 spaces with applications to pdes. *Journal of Machine Learning Research*,  
827 24(89), 1–97.
- 828 Lapusta, N., Rice, J. R., Ben-Zion, Y., & Zheng, G. (2000). Elastodynamic analysis  
829 for slow tectonic loading with spontaneous rupture episodes on faults with  
830 rate-and state-dependent friction. *Journal of Geophysical Research: Solid*  
831 *Earth*, 105(B10), 23765–23789.
- 832 Lehmann, F., Gatti, F., Bertin, M., & Clouteau, D. (2023). Fourier neural oper-  
833 ator surrogate model to predict 3d seismic waves propagation. *arXiv preprint*  
834 arXiv:2304.10242.

- Lehmann, F., Gatti, F., Bertin, M., & Cloueau, D. (2024). 3d elastic wave propagation with a factorized fourier neural operator (f-fno). *Computer Methods in Applied Mechanics and Engineering*, 420, 116718.
- Li, B., Wang, H., Feng, S., Yang, X., & Lin, Y. (2023). Solving seismic wave equations on variable velocity models with fourier neural operator. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–18.
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.
- Ma, X., & Elbanna, A. (2019). Dynamic rupture propagation on fault planes with explicit representation of short branches. *Earth and Planetary Science Letters*, 523, 115702.
- Ma, X., Hajarolasvadi, S., Albertini, G., Kammer, D. S., & Elbanna, A. E. (2019). A hybrid finite element-spectral boundary integral approach: Applications to dynamic rupture modeling in unbounded domains. *International Journal for Numerical and Analytical Methods in Geomechanics*, 43(1), 317–338.
- Madariaga, R., Olsen, K., & Archuleta, R. (1998). Modeling dynamic rupture in a 3d earthquake fault model. *Bulletin of the Seismological Society of America*, 88(5), 1182–1197.
- Moczo, P., Kristek, J., Galis, M., Pazak, P., & Balazovjech, M. (2007). The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion. *Acta physica slovaca*, 57(2).
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1), 3952.
- Nielsen, S. B., Carlson, J., & Olsen, K. B. (2000). Influence of friction and fault geometry on earthquake rupture. *Journal of Geophysical Research: Solid Earth*, 105(B3), 6069–6088.
- Oglesby, D. D., Archuleta, R. J., & Nielsen, S. B. (1998). Earthquakes on dipping faults: The effects of broken symmetry. *Science*, 280(5366), 1055–1059.
- Oglesby, D. D., Archuleta, R. J., & Nielsen, S. B. (2000). The three-dimensional dynamics of dipping faults. *Bulletin of the Seismological Society of America*, 90(3), 616–628.
- Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., ... Courville, A. (2019). On the spectral bias of neural networks. In *International conference on machine learning* (pp. 5301–5310).
- Rahman, M. A., Ross, Z. E., & Azizzadenesheli, K. (2022). U-no: U-shaped neural operators. *arXiv preprint arXiv:2204.11127*.
- Rekoske, J. M., May, D. A., & Gabriel, A.-A. (2025). Reduced-order modelling for complex three-dimensional seismic wave propagation. *Geophysical Journal International*, 241(1), 526–548.
- Renard, F., & Candela, T. (2017). Scaling of fault roughness and implications for earthquake mechanics. *Fault zone dynamic processes: Evolution of fault properties during seismic rupture*, 195–215.
- Ripperger, J., Ampuero, J.-P., Mai, P., & Giardini, D. (2007). Earthquake source characteristics from dynamic rupture with constrained stochastic fault stress. *Journal of Geophysical Research: Solid Earth*, 112(B4).
- Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, 123(6), 5120–5129.
- Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2019). Phaselink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, 124(1), 856–869.
- Ruina, A. (1983). Slip instability and state variable friction laws. *Journal of Geophysical Research: Solid Earth*, 88(B12), 10359–10370.

- 890 Tainpakdipat, N., Abdelmeguid, M., Zhao, C., & Elbanna, A. (2025). *Fourier neural*  
891 *operators for accelerating earthquake dynamic rupture simulations* [data set].  
892 Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.15085622> doi:  
893 10.5281/zenodo.15085622
- 894 Ulrich, T., Gabriel, A.-A., Ampuero, J.-P., & Xu, W. (2019). Dynamic viability  
895 of the 2016 mw 7.8 kaikōura earthquake cascade on weak crustal faults. *Nature*  
896 *Communications*, 10, 1213. doi: 10.1038/s41467-019-09125-w
- 897 Van Zelst, I., Wollherr, S., Gabriel, A.-A., Madden, E. H., & van Dinther, Y. (2019).  
898 Modeling megathrust earthquakes across scales: One-way coupling from geo-  
899 dynamics and seismic cycles to dynamic rupture. *Journal of Geophysical*  
900 *Research: Solid Earth*, 124(11), 11414–11446.
- 901 Wei, W., & Fu, L.-Y. (2022). Small-data-driven fast seismic simulations for com-  
902 plex media using physics-informed fourier neural operators. *Geophysics*, 87(6),  
903 T435–T446.
- 904 Xu, J., Zhang, H., & Chen, X. (2015). Rupture phase diagrams for a planar fault in  
905 3-d full-space and half-space. *Geophysical Journal International*, 202(3), 2194–  
906 2206.
- 907 Yang, L., Liu, X., Zhu, W., Zhao, L., & Beroza, G. C. (2022). Toward improved ur-  
908 ban earthquake monitoring through deep-learning-based noise suppression. *Sci-  
909 ence advances*, 8(15), eabl3564.
- 910 Yang, Y., Gao, A. F., Azizzadenesheli, K., Clayton, R. W., & Ross, Z. E. (2023).  
911 Rapid seismic waveform modeling and inversion with neural operators. *IEEE*  
912 *Transactions on Geoscience and Remote Sensing*, 61, 1–12.
- 913 Yang, Y., Gao, A. F., Castellanos, J. C., Ross, Z. E., Azizzadenesheli, K., & Clay-  
914 ton, R. W. (2021). Seismic wave propagation and inversion with neural  
915 operators. *The Seismic Record*, 1(3), 126–134.
- 916 Zhang, T., Trad, D., & Innanen, K. (2023). Learning to solve the elastic wave equa-  
917 tion with fourier neural operators. *Geophysics*, 88(3), T101–T119.
- 918 Zhu, W., & Beroza, G. C. (2019). Phasenet: a deep-neural-network-based seismic  
919 arrival-time picking method. *Geophysical Journal International*, 216(1), 261–  
920 273.
- 921 Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and  
922 decomposition using deep neural networks. *IEEE Transactions on Geoscience*  
923 *and Remote Sensing*, 57(11), 9476–9488.
- 924 Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P., & Beroza, G. C. (2022). An end-  
925 to-end earthquake detection method for joint phase picking and association  
926 using deep learning. *Journal of Geophysical Research: Solid Earth*, 127(3),  
927 e2021JB023283.
- 928 Zou, C., Azizzadenesheli, K., Ross, Z. E., & Clayton, R. W. (2024). Deep neural  
929 helmholtz operators for 3d elastic wave propagation and inversion. *Geophysical*  
930 *Journal International*, ggae342.