

1      **Fourier Neural Operators for Accelerating Earthquake**  
2      **Dynamic Rupture Simulations**

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

5      <sup>1</sup>Department of Civil and Environmental Engineering, University of Illinois Urbana Champaign, Urbana,  
6      IL 61801, USA.

7      <sup>2</sup>Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, USA.

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

9      <sup>4</sup>Beckman Institute of Advanced Science and Technology, University of Illinois Urbana Champaign,  
10     Urbana, IL 61801, USA.

11     **Key Points:**

- 12     • We develop a Fourier Neural Operator (FNO) framework to simulate earthquake  
13     rupture under varying fractal stress under varying conditions of fractal stress, nucleation  
sites, and friction.
- 14     • The FNO achieves up to  $4 \times 10^5$  times speedup compared to the Spectral Bound-  
15     ary Integral method, which is currently the fastest available numerical method.
- 16     • We demonstrate that the model can generalize to unseen shear stress and frictional  
17     parameters, including out-of-distribution cases.
- 18

19      **Abstract**

20      Dynamic rupture modeling plays a crucial role in unraveling earthquake source processes.  
21      However, the multiscale nature of rupture propagation ~~pose~~poses significant challenges,  
22      and classical numerical methods remain computationally expensive. To overcome this  
23      hurdle, we present a methodology that is both computationally efficient and quantita-  
24      tively accurate. Specifically, we introduce a surrogate model, in the form of a Fourier Neu-  
25      ral Operator, for emulating the nonlinear equations governing dynamic rupture prop-  
26      agation on frictional interfaces. This surrogate is trained on synthetic data generated  
27      by multiple physics-based dynamic rupture simulations and is then applied to unseen  
28      problems. The proposed methodology retains the accuracy of traditional multiscale meth-  
29      ods at a significantly reduced computational cost, achieving a speedup of up to  $4 \times 10^5$   
30      compared to the ~~state-of-the-art~~state-of-the-art conventional methods. We evaluate  
31      this approach using various examples and demonstrate its efficacy in capturing the space-  
32      time evolution of fault slip rates for a wide range of stress conditions. This development  
33      advances the state of the art of computational earthquake dynamics and opens new op-  
34      portunities for accelerating 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; Zhu et al., 2019; L. Yang et al., 2022; Zhu & Beroza, 2019; Mousavi et al., 2020; Ross et al., 2018, 2019; Zhu et al., 2018). 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 stressand frictional parameters, initial slip rate, frictional parameters, and stress perturbation for nucleation. The performance of the FNO framework is eval-

120 uated across various datasets, demonstrating its ability to generalize and accurately pre-  
 121 dict fault slip dynamics 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 Fourier Neural Operators

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

$$v_0(x) = P(a(x))$$

$$v_{m+1}(x) = \sigma \left( W_m v_m(x) + \int k(x, y) v_m(y) dy \right), \quad m = 0, \dots, M-1$$

$$u(x) = Q(v_M(x))$$

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

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

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

146 where  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  are the Fourier transform and its inverse, respectively. Using the  
 147 Fourier transform enhances computational efficiency, allows the model to handle global  
 148 dependencies, and enables resolution invariance, meaning the model can generalize across  
 149 different spatial and temporal resolutions without requiring retraining. Specifically, when  
 150 the input function is provided in regular grids, the Fourier transformer is carried using  
 151 fast Fourier transform, a celebrated and fast algorithm for Fourier analysis. It also leads  
 152 to better generalization across grids, reduces computational costs for high-dimensional  
 153 problems, and improves the model's ability to handle complex, nonlinear dynamics. We  
 154 refer the reader to (Z. Li et al., 2020) for a discussion of this architecture and related features.

### 156 2.1 Fault Model Setup

157 We consider a two-dimensional (2D) planar fault embedded in a three-dimensional  
 158 (3D) homogeneous, isotropic, and elastic domain  $\Omega$ . The domain includes a traction bound-

159       ary  $\Gamma_T$ , a displacement boundary  $\Gamma_u$ , and a fault surface located along  $\Gamma_f$ . We adopt  
 160       the TPV101 benchmark problem setup from the SCEC/USGS dynamic earthquake rup-  
 161       ture code verification exercise (Harris et al., 2018, 2009; Kammer et al., 2021). The gov-  
 162       erning equations are given by:

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

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

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

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

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

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

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

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

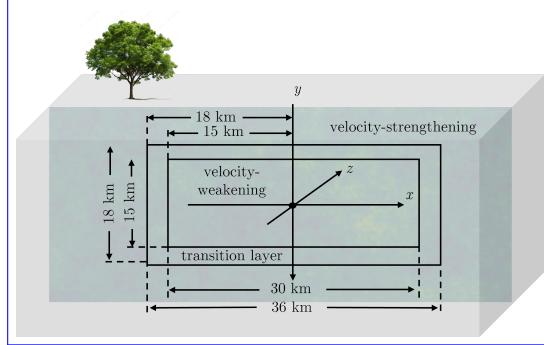
$$\tau = a\sigma \operatorname{arcsinh} \left[ \frac{V}{2V^*} \exp \left( \frac{f^* + b \ln(V^* \theta / D_{RS})}{a} \right) \right], \quad (6)$$

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

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

180       A schematic representation of the fault geometry is shown in Figure 2a. The hypocenter  
 181       is located at  $(x_0, y_0) = (0, 0)$  km. The fault consists of a central velocity-weakening  
 182       (VW) regime, defined by  $-W < x < W$  and  $-W/2 < y < W/2$ , with  $W = 15$  km.  
 183       This region defines the boundaries of the seismogenic zone which is characterized by a  
 184       negative value of  $a-b$ . Surrounding the VW region is a velocity-strengthening (VS) regime,  
 185       which is characterized by a positive value of  $a-b$ , with a transition layer of width  $w =$   
 186       3 km. To ensure a smooth transition between these regions, the parameter  $a$  varies smoothly  
 187       based on a mathematically smoothed version of the boxcar function. The variation in  
 188        $a$  is governed by:

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



**Figure 1.** Schematic diagram of the proposed FNO framework. (a) Schematic illustration of the 3D fault model, showing the central VW regime surrounded by the VS regime, with the hypocenter at the center. (b) FNO framework for a 2D fault plane embedded in a 3D bulk transition layer. (c) FNO framework for a 1D fault embedded in a 2D plane. The inputs include the distribution of initial shear stress follows a fractal distribution, and frictional parameters the hypocenter is varied within the VW region. The outputs are snapshots of slip rate over time.

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

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.

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

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

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

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

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

The fault's dynamic rupture dataset is generated using the Spectral Boundary Integral (SBI) method, which is currently the fastest available solver for simulating rupture propagation along a planar fault in a homogeneous medium (Geubelle & Rice, 1995). This method involves solving the coupled equations governing traction and displacement continuity along the fault surface. As a result, it eliminates the need to solve the governing equations throughout the entire domain. According to (Geubelle & Rice, 1995), the response of the governing equations 1–4 is given by

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

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

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

## 2.2 Fourier Neural Operators

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

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

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

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

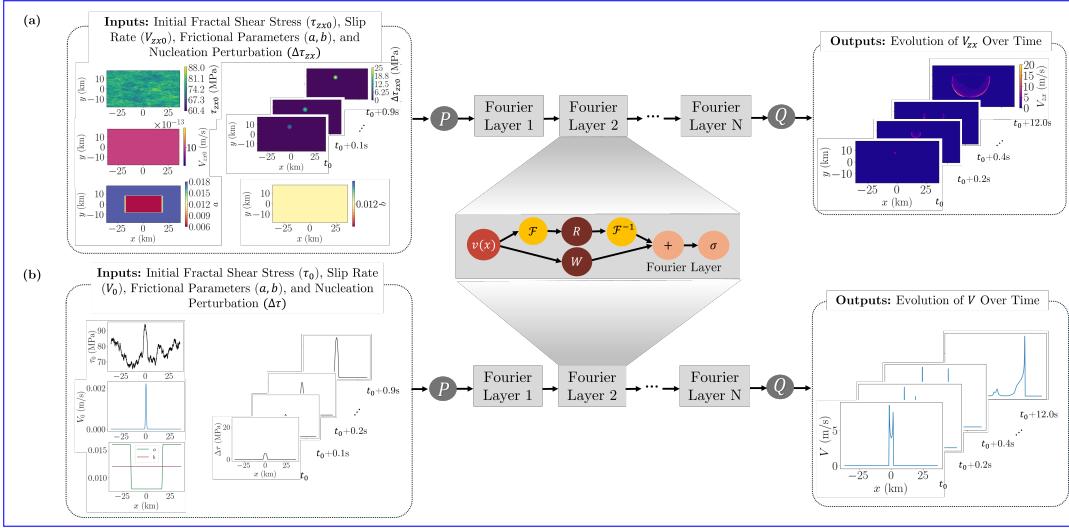
| Medium Parameter  | Symbol         | Value      |
|---|----------------|------------|
| Shear wave speed (km/s)                                 | $c_s$          | 3.464      |
| Pressure wave speed (km/s)                              | $c_p$          | 6          |
| Density (kg/m <sup>3</sup> )                            | $\rho$         | 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          |

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:

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

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 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.



**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.

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 because 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:

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

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.  $V_{\text{max}}$  and  $V_{\text{min}}$  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

The dynamic rupture dataset is generated using the Spectral Boundary Integral (SBI) method, which is currently the fastest available solver for rupture propagation on a planar fault in a homogeneous medium (Geubelle & Rice, 1995; Kammer et al., 2021). The 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) (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 2ba. 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 2eb. 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.

#### 3.1 2D Dynamic Rupture Dataset

In the 2D dataset, the fault is represented as a one-dimensional line embedded in a two-dimensional elastic bulk 2D elastic bulk, assuming plane strain conditions. A total of 4,000 8,200 realizations are generated, with 1,000 realizations for each parameter

315 set parameters listed in Table ?? . Of these, 3,600 realizations are allocated for training,  
 316 while the remaining 400 realizations are reserved for testing. Among these, 4,000 realizations  
 317 have the hypocenter at the center of the fault, 2,100 have hypocenters located at the point  
 318 of maximum shear stress in the VW region, and the remaining 2,100 have randomly located  
 319 hypocenters drawn from a uniform distribution along the fault.

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

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

335 The spatial discretization consists of 2,880 points, i.e.,  $X = 2,880$ , which corresponds  
 336 to the discretization of  $X = 2,880$ , consistent with the resolution used in the SBI solver.  
 337 The input channels comprise include the spatial distributions of the initial shear stress  
 338 , parameter  $a$ , and parameter  $b$ , resulting in  $C_{in} = 3$ .  $\tau_0$  and slip rate at  $t_0$ , frictional parameters  
 339  $a$  and  $b$ , and ten time steps of the nucleation stress perturbation  $\Delta\tau$ , sampled at 0.1 s  
 340 intervals starting from the time  $t_0$ , when the slip rate first exceeds the threshold  $V_{th}$ . These  
 341 ten perturbation snapshots correspond to  $t = t_0, t_0 + 0.1 \text{ s}, \dots, t_0 + 0.9 \text{ s}$ , contributing  
 342 ten additional input channels. In total, this yields  $C_{in} = 14$ . This potentially opens up  
 343 the possibility of using different nucleation over-stress distribution.

344 We solve the system of equations using SBI governing equations using the SBI solver  
 345 with a time step  $\Delta t$  of 0.001 s to satisfy of  $\Delta t = 0.001$  s, which satisfies the Courant–Friedrichs–Lowy  
 346 (CFL) condition,  $\Delta t \leq f \Delta x / c_p$  with  $\Delta t \leq f \Delta x / c_p$ , where  $f$  is of order 1. Here we take  
 347  $f$  equal to 0.25 a constant of order one. Here, we choose  $f = 0.25$ . Since FNOs are in-  
 348 dependent of the numerical discretization, we can use adopt a coarser time step in FNO  
 349 because we aim of 0.2 s in the FNO model. The FNO is trained to perform a single-shot  
 350 prediction up to 15 s, and a time step of 0.001 s poses significant memory challenges. As  
 351 such, the output channels represent snapshots of the slip rate at discrete of 60 time steps  
 352 of 0.2 s, spanning the duration from  $t = 0$  s to  $t = 15$  s, yielding 76 time steps and thus  
 353  $C_{out} = 76$  slip rate, starting from time  $t_0$ , resulting in  $C_{out} = 60$ .

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

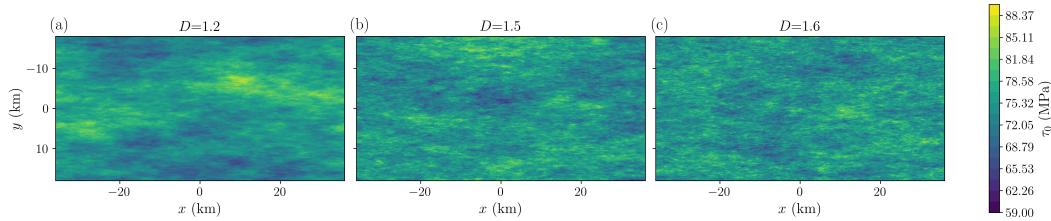
358 Fractal dimensions and variation of frictional parameter  $a$  and  $b$  for 2D and 3D dynamic  
 359 rupture dataset

### 360 3.2 3D Dynamic Rupture Dataset

361 In the 3D dataset, the fault is modeled as a two-dimensional 2D plane embedded  
 362 in a three-dimensional 3D elastic bulk. A total of 4,000 realizations are generated, with  
 363 3,600 realizations used for training and 400 realizations for testing. The variations in frac-

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

| $D$           | $b$          | $a_0$  | $\Delta a_0$ |
|---------------|--------------|--------|--------------|
| 1.2, 1.5, 1.6 | 0.012, 0.014 | 0.009  | 0.006        |
| 1.2, 1.5, 1.6 | 0.012, 0.014 | 0.008  | 0.008        |
| 1.2, 1.5, 1.6 | 0.012, 0.014 | 0.007  | 0.010        |
| 1.2, 1.5, 1.6 | 0.012, 0.014 | 0.006  | 0.012        |
| 1.2, 1.5, 1.6 | 0.012, 0.014 | 0.0085 | 0.007        |
| 1.2, 1.5, 1.6 | 0.012, 0.014 | 0.0075 | 0.009        |
| 1.2, 1.5, 1.6 | 0.012, 0.014 | 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.

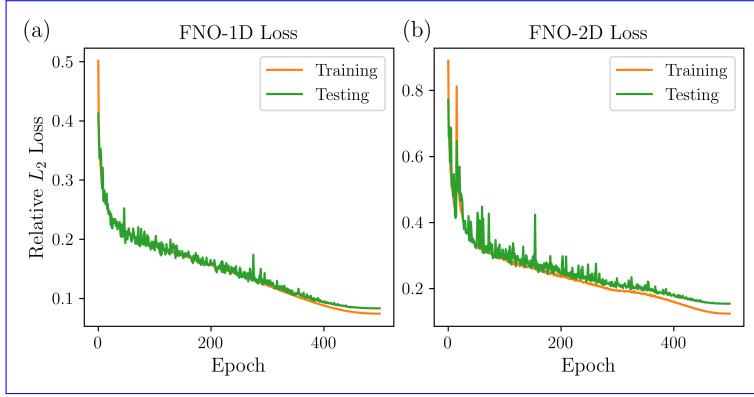
tal dimension and frictional parameters  $a$  and  $b$  are listed in Table ??, with each parameter set containing 1,000 realizations. Examples of fractal initial shear stress distributions generated using different fractal dimensions are illustrated in Figure 3.

Similar to the 2D case, 8,200 realizations are generated: 4,000 with the hypocenter at the fault center, 2,100 at the location of maximum shear stress in the VW region, and 2,100 with uniformly random hypocenter locations. Each realization is processed in the same manner as the 2D case to produce four distinct initial conditions based on the slip rate threshold  $V_{th} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ . 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 and  $\Delta\tau_{zx0}$ , slip rate  $V_{zx0}$ , frictional parameters  $a$  and  $b$ , and stress perturbation  $\Delta\tau_{zx}$  to the sequence of slip rates component of slip rates  $V_{zx}$  over the time interval of interest. The input data is structured as a tensor of dimensions  $(N, X, Y, C_{in})$ , while the output data has dimensions  $(N, X, Y, C_{out})$ , where  $X$  and  $Y$  represent the number of spatial discretization points in the  $x$  and  $y$  directions, respectively.

The two-dimensional-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 initial shear stress, parameter  $a$ , and parameter  $b$ , resulting in  $C_{in} = 3\tau_0, V_0, a, b$ , and ten steps of  $\Delta\tau$ , sampled from  $t = t_0$  to  $t = t_0 + 0.9$  s, resulting in  $C_{in} = 14$ .

It is worth noting that the time step  $\Delta t$  in the SBI solver is 0.01. 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  $f = 0.6$ . We output the SBI solution  $m$  and  $f = 0.6$ . Solutions are output at a frequency of 0.1 s. The discretization consists of  $720 \times 360$  spatial points. This results in  $(N, X, Y, C_{out}) = (3600, 720, 360, 150)$ , producing approximately 500 GB/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  $(N, X, Y, C_{out}) = (28700, 720, 360, 60)$ , which corresponds to approximately



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

2 TB of data. Given the available memory and GPU limitations, training with a single batch at this scale is challenging. To address this, we reduce the number of spatial and temporal points by sub-sampling from  $(X, Y, C_{\text{out}}) = (720, 360, 150)$  to  $(360, 180, 76)$ .  $(X, Y, C_{\text{out}}) = (720, 360, 60)$  to  $(360, 180, 60)$  for FNO training, reducing the dataset size to approximately 66–416 GB. In this case, the output channels correspond to snapshots of slip rate  $V_{zx}$  spatially sampled at half the resolution of the original simulations and temporally recorded at 0.2 s intervals, covering the time range  $t = 0$  s to  $t = 15$  s, thus  $C_{\text{out}} = 76$ . This adjustment makes training feasible on the available NVIDIA A100 GPU with 64 GB of memory.

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 employed 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 10,000–500 epochs. The training and testing losses are shown in Figure 4a. Both training and testing losses consistently decrease without significant large divergence between them, indicating no signs of overfitting. Some fluctuations in the testing loss at the beginning occur due to the selected batch size.

**Table 3.** Median and median absolute deviation (MAD) of NRMSE and relative  $L_2$  error in the bracket shown in Figure 6 for 3600–28,700 training samples and 400–4,100 testing samples of 2D and 3D dynamic rupture datasets.

|          | 2D Dynamic Rupture                               | 3D Dynamic Rupture                              |
|----------|--|---|
|          | NRMSE<br>(Relative $L_2$ Error)                  | NRMSE<br>(Relative $L_2$ Error)                 |
| Training | $0.00261 \pm 0.00357$<br>( $0.0550 \pm 0.0665$ ) | $0.00231 \pm 0.00317$<br>( $0.0854 \pm 0.115$ ) |
| Testing  | $0.00329 \pm 0.00388$<br>( $0.0685 \pm 0.0270$ ) | $0.00387 \pm 0.00366$<br>( $0.153 \pm 0.133$ )  |

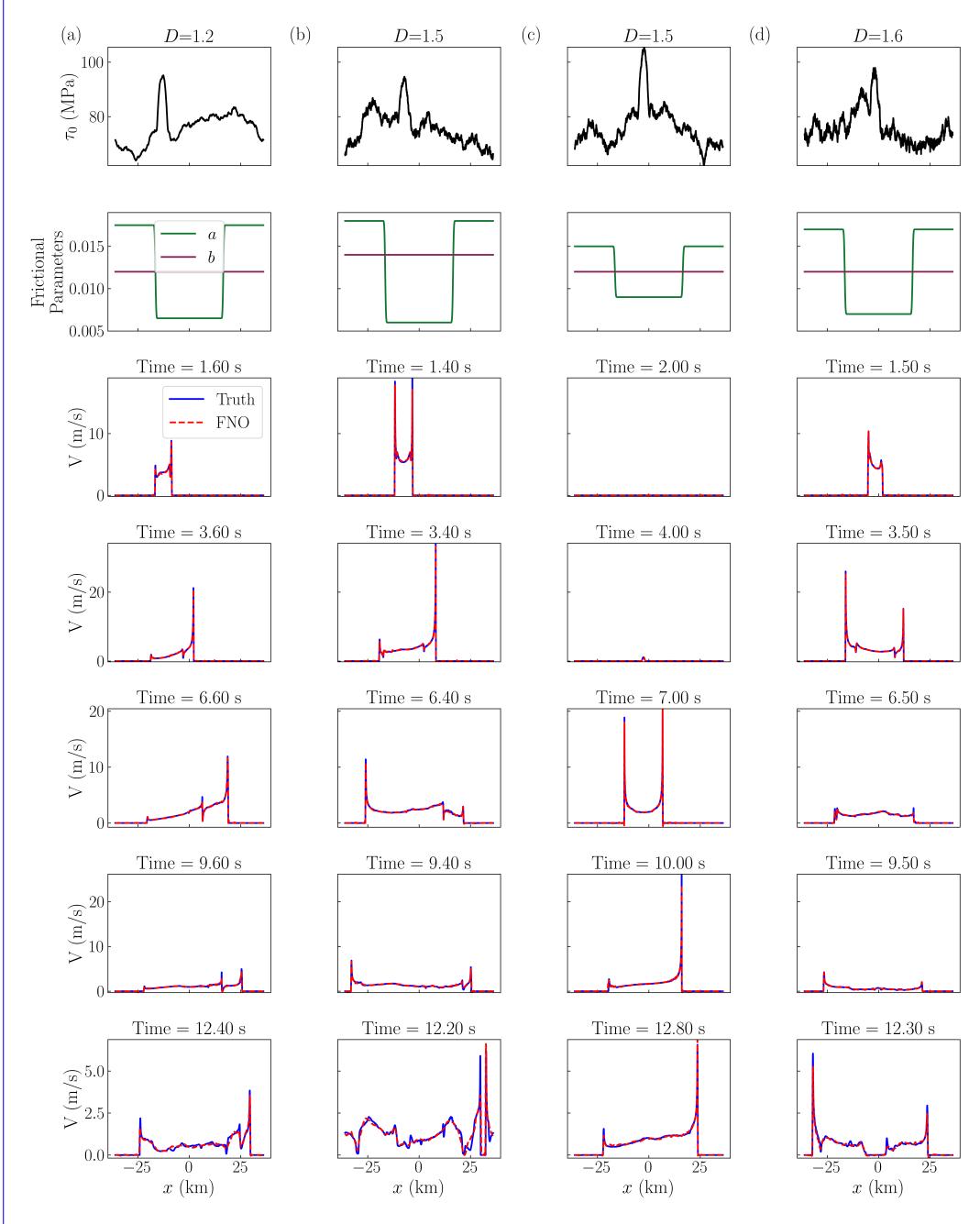
The performance of FNO-1D is evaluated on the testing dataset. The model predictions, compared against the ground truth, are presented in Figure 5. FNO-1D effectively captures a wide range of magnitudes arising from different distributions of  $\tau_0$  and the frictional parameters  $V_0$ ,  $\Delta\tau$ ,  $a$ , 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 at higher frequencies by implementing a specialized training protocol 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 distribution 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 90.95% of the training set shows a relative  $L_2$  error of less than 10%, while this fraction drops to 80.90% 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

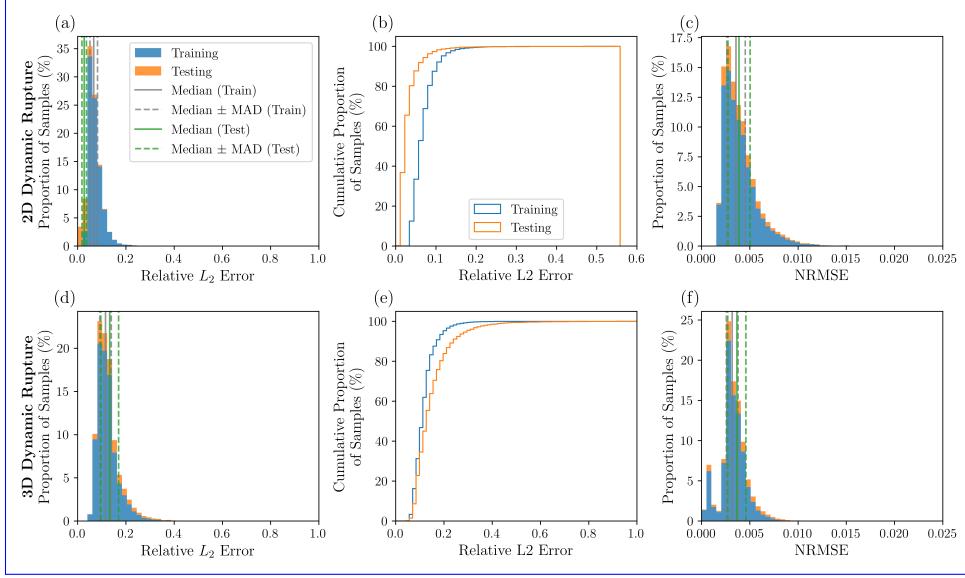
An advantage of FNOs is that they can target the underlying operator, and learn to handle a family of problems rather than a single instance. Accordingly, in this section we test the generalization of our trained operator to some unseen initial conditions.

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 0.00963 and 0.1822, 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 noise, spatial shifts in the predictions, and minor noise near the rupture front in the predicted solution, along with slight mismatches in peak magnitudes at later times. Such discrepancies contribute to the high

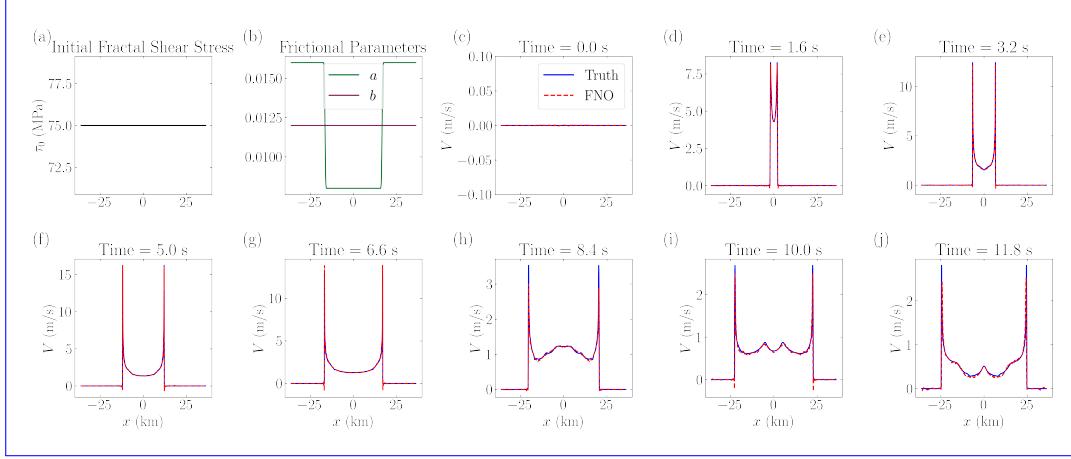


**Figure 5.** Testing of the trained FNO-1D model on the 2D dynamic rupture `test`-dataset.

(a) Inputs include the initial fractal shear stress distributions  $\tau_0$  and (b) distribution  $\tau_0$  with  $D = 1.2$ , initial slip rate at the time step where  $\max(V) > V_{th}$  with  $V_{th} = 1e-2$ , frictional parameters  $a$  and  $b$ ; (c) Outputs consist 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 discrete time steps following this point (rows 3–7). (b-d) Same as (a), but with varying conditions: (b)  $D = 1.5$ ,  $a_0 = 0.006$ ,  $\Delta a_0 = 0.012$ ,  $b = 0.014$ ,  $V_{th} = 10^{-4}$ ; (c)  $D = 1.5$ ,  $a_0 = 0.009$ ,  $\Delta a_0 = 0.006$ ,  $b = 0.012$ ,  $V_{th} = 10^{-2}$ ; (d)  $D = 1.6$ ,  $a_0 = 0.007$ ,  $\Delta a_0 = 0.010$ ,  $b = 0.012$ ,  $V_{th} = 10^{-3}$ . In all cases, predictions are generated starting from the time step where the slip rate exceeds the specified threshold.



**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) Inputs include a uniform initial shear stress ( $\tau_0$ ) of 75  $\tau_0 = 75$  MPa, and (b) the frictional parameters  $a$  and  $b$ , with  $\Delta a_0 = -0.008$ ;  $a_0 = 0.008$ ,  $\Delta a_0 = 0.008$ , and  $a_0 = 0.008$ ;  $b = 0.012$ . Other model inputs (e.g., initial slip rate  $V_0$  with a threshold  $V_{th} = 0$  m/s and nucleation stress perturbation  $\Delta\tau$ ) Outputs are not shown. Panels (c)-(j) show the outputs consist of predicted slip rate profiles at selected time steps.

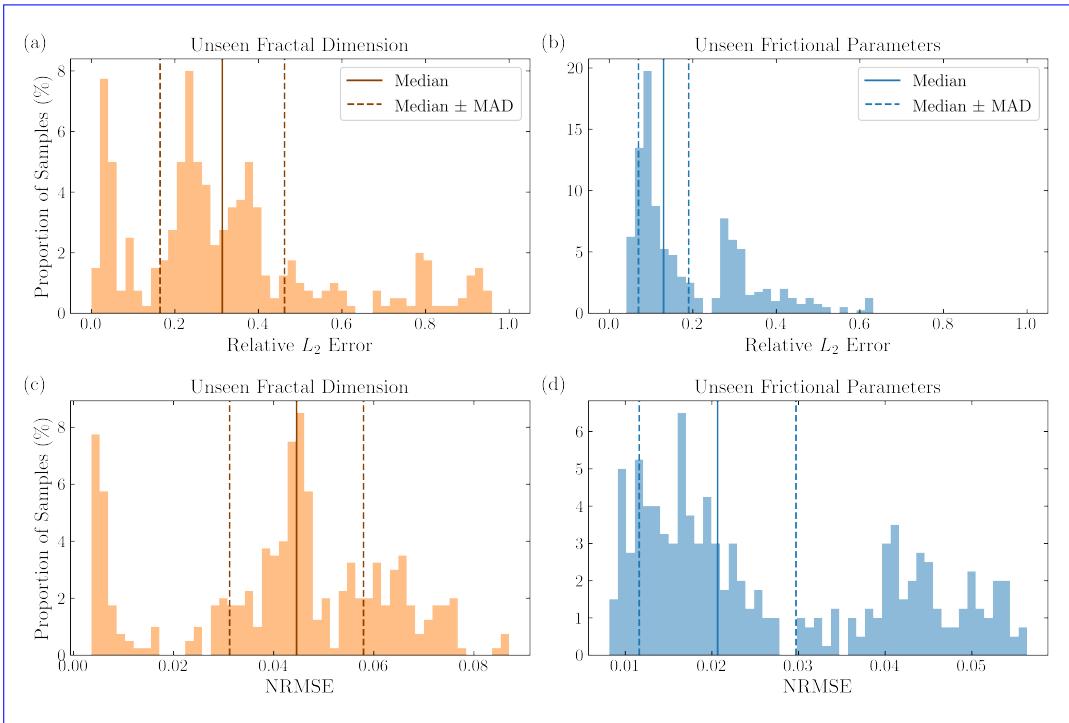
relative  $L_2$ -error. However, the global features of slip rate remains consistent with the ground truth. These results highlight the robustness of FNO-1D, as the model successfully generalizes to this unseen distribution of  $\tau_0$ .

Additionally, we test evaluate the performance of FNO-1D on an unseen fractal dimension  $D$ . Specifically, we evaluate the trained model using a dataset generated with  $D = 1.3$ , which was not part of included in the training set. In this case, the frictional parameter  $a$  follows a boxcar distribution, with  $a = 0.008$  in the VW region and  $a = 0.016$  in the VS region, while the frictional parameter  $b$  is fixed at 0.012. 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  $V_{th} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ , 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.

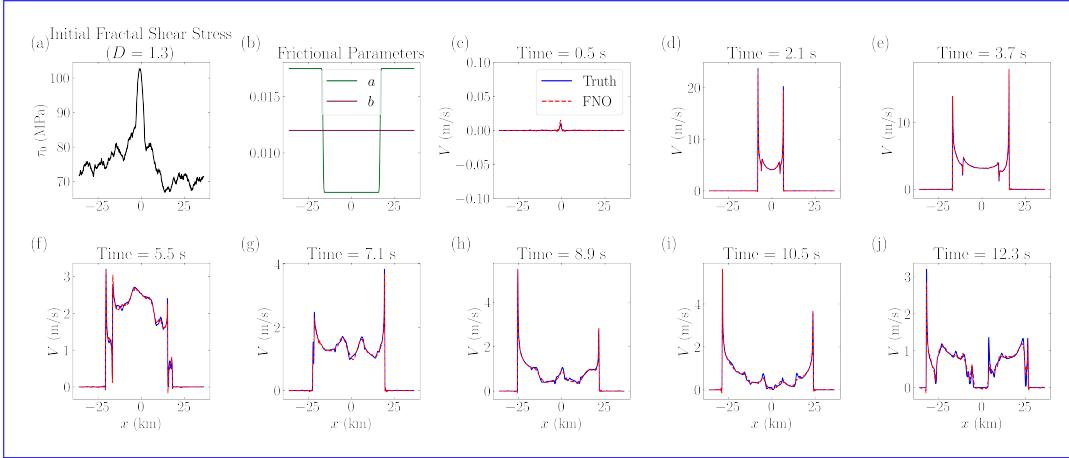
Figure 9 shows the an example of results from FNO-1D compared to the ground truth, with NRMSE and relative  $L_2$  error of 0.00353 and 0.0566, 0.00254 and 0.0131, respectively. FNO-1D effectively captures both the low-frequency and high-frequency components of the slip rate up to approximately 12 with minor discrepancies at the peak at 12.3 s. The magnitude of the rupture front in the predictions also aligns with predicted rupture front also closely matches the ground truth. Some discrepancies occur when predicting the high-frequency component at later times 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



**Figure 8.** Results—The distribution of FNO-1D testing on an unseen fractal dimension  $D = 1.3$ . Inputs consist of (a) initial fractal shear stress distributions—the relative  $L_2$  error and (b) the relative  $L_2$  error and (c) the NRMSE evaluated on unseen fractal dimension cases, and (d) 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, with  $\Delta a_0 = 0.008$  the median ± MAD of the relative error is  $0.313 \pm 0.149$  and  $a_0 = 0.008$ ; (e)–(h) Outputs consist of predicted slip rate profiles at selected time steps; the NRMSE is  $0.0446 \pm 0.0134$ . For unseen frictional parameters cases, the median ± MAD of the relative error is  $0.130 \pm 0.0600$  and of the NRMSE is  $0.0207 \pm 0.00905$ .



**Figure 9.** Example of results of FNO-1D testing on an unseen fractal dimension  $D = 1.3$ .

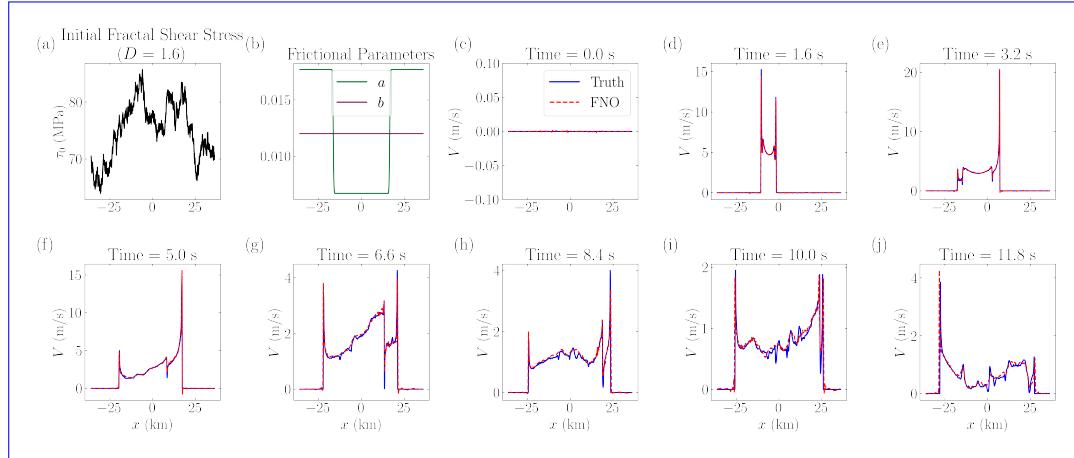
Panels (a) and (b) show selected input features: (a) initial fractal shear stress distributions ( $\tau_0$ ) with  $D = 1.3$  and (b) frictional parameters  $a$  and  $b$ , with  $\Delta a_0 = 0.0065$ ,  $a_0 = 0.011$ , and  $b = 0.012$ . Other model inputs (i.e., initial slip rate  $V_0$  with a threshold  $V_{th} = 10^{-3}$  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.

the parameters uniformly from the specified ranges in the table. Then, we apply the same pre-processing using  $V_{th} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ . 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  $a$ , with values of 0.0075  $a$  are 0.0067 in the VW region and 0.0165 0.0177 in the VS region, while parameter  $b$  the parameter  $b$  is set uniformly to 0.012 along the fault. The initial shear stress  $\tau_0 - \tau_0$  follows a fractal distribution with fractal dimension  $D = 1.5$ . Figure 10 compares the predicted slip rates with the ground truth, showing NRMSE and a fractal dimension of  $D = 1.6$ . This realization yields an NRMSE and a relative  $L_2$  error of 0.0213 and 0.4290 0.00827 and 0.0464, respectively. The FNO-1D model successfully captures the overall evolution trend slip rate evolution, although some discrepancies occur at the peaks of the rupture front in the beginning. The high relative  $L_2$  error results from mismatches in the magnitude between predictions and ground truth that 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 start, as well as spatial shifts occurring around 9 s. Additionally, the slip rates at the initial time step have a very small magnitude close to zero, which amplifies the relative  $L_2$  error value without indicating significant mismatches. Despite these issues, FNO-1D predicts, reasonably well, the major features of the slip rate evolution under this unseen frictional conditions sharp gradient at the rupture front along the fault.

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

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



**Figure 10.** Results Example of results of FNO-1D testing on unseen  $a - b$ . Panels (a) and (b) show selected input features: Inputs consist of (a) initial fractal shear stress distributions ( $\tau_0$ ) with  $D = 1.5$ , along with  $D = 1.6$  and (b) frictional parameters  $a$  and  $b$ , with  $\Delta a_0 = 0.009$ ,  $\Delta a_0 = 0.0067$ ,  $a_0 = 0.011$ , and  $a_0 = 0.0075$ ;  $b = 0.012$ . 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)-(h) Outputs show the outputs consist of predicted slip rate profiles at selected discrete-time steps.

508            **4.2 FNO-2D for 3D Dynamic Rupture Dataset**

509            ***4.2.1 Training and Testing Performance***

510            We perform hyperparameter tuning for FNO-2D, resulting in an architecture  
511            consisting of four Fourier layers. The lifting and projecting fully connected neural networks  
512            each contain 128 neurons. After applying the Fourier transform, 32 modes are retained  
513            in the linear transformation, and the GELU activation function is employed. The model  
514            is trained using the Adam optimizer with a batch size of 10, a learning rate of  $10^{-3}$ , and  
515            a weight decay of  $10^{-4}$ , following a cosine annealing schedule. Training spans 1,000–500  
516            epochs until the loss stabilizes. Training and testing losses over optimization iterations  
517            epochs are shown in Figure 4b. Both training and testing losses consistently decrease  
518            without significant divergence, indicating no signs of overfitting.

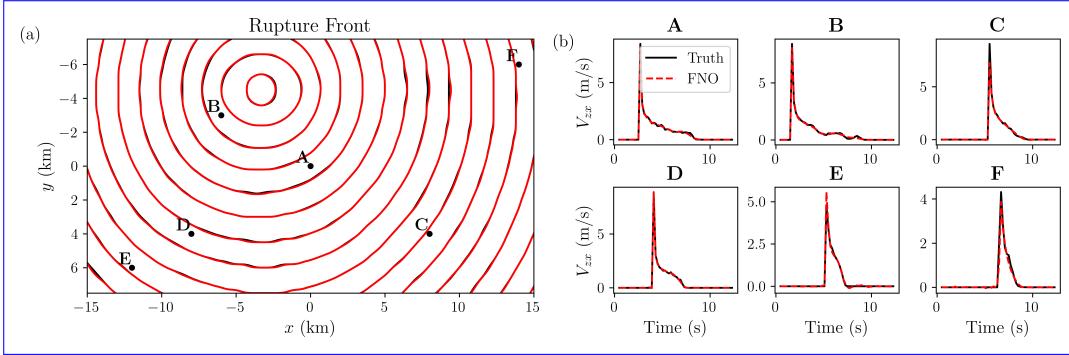
519            Figure 11 illustrates a successful example of the predictions compared to the ground  
520            truth. It shows the rupture front contours in the VW region at 5-second–0.5-second  
521            intervals, along with time series at six selected points in the VW region. FNO-2D accu-  
522            rately captures the spatially irregular shape of the rupture front contours with NRMSE  
523            of 0.00636–0.003035 and relative  $L_2$  error of 0.2448. The model's accuracy is high during  
524            the initial nucleation phase but decreases over time due to minor shifts in the rupture  
525            front arrival times. This loss in accuracy over time is also pointed out in time-series prediction  
526            (Zhang et al., 2023). Nonetheless, the time histories at randomly selected points still exhibit  
527            good overall agreement. The minor time shifts observed during the later part of the rupture  
528            propagation amplifies the relative  $L_2$  error. 0.1002. The predicted rupture arrival times  
529            and waveform shapes align closely with the ground truth, although the nucleation site  
530            is not centered. Minor discrepancies can be observed in peak amplitudes, particularly  
531            at points C, E and F. However, the overall agreement between the model prediction and  
532            the ground truth remains high.

533            The distribution distributions of relative  $L_2$  error and NRMSE during the predic-  
534            tion phase, shown in Figure 6, further confirms Figures 6d, 6e, and 6f, further confirm  
535            that no overfitting is occurring, as the testing dataset follows the same trend as the train-  
536            ing dataset. Moreover, the distribution is skewed toward zero. More than 80%–90% of  
537            the training dataset has relative  $L_2$  errors less than 20%, while this proportion drops to  
538            60%–80% for the testing dataset. The modest decrease in this proportion confirms that  
539            the model generalizes well and does not overfit. Interestingly, we have found that even  
540            in cases where the model predictions show higher relative  $L_2$  errors, the model still suc-  
541            cessfully captures the major features of rupture dynamics. Examples of predictions with  
542            higher relative error are provided in the supplementary information.

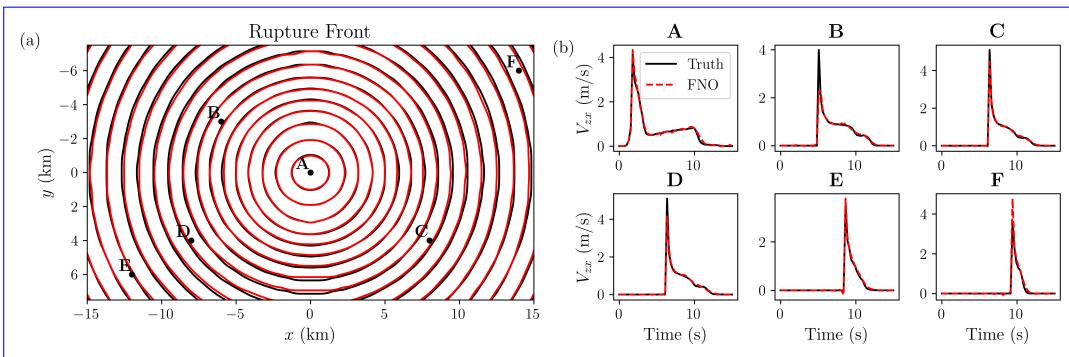
543            ***4.2.2 Generalization to Unseen Initial Shear Stress***

544            We Similar to the 2D model, we evaluate the trained FNO-2D model using an un-  
545            seen initial shear stress distribution, specifically, the uniform stress distribution defined  
546            in the TPV101 benchmark from the SCEC/USGS dynamic earthquake rupture code ver-  
547            ification exercise. In this benchmark, the initial shear stress is uniformly set to 75 MPa,  
548            while frictional parameters remain within the training range of the 3D dynamic rupture  
549            dataset.

550            Figure 12 shows the rupture contours of the predicted component of slip rate  $V_{zx}$   
551            compared to the ground truth, as well as the time history at six selected points in the  
552            VW region. FNO-2D can capture the evolution of  $V_{zx}$ . The predicted rupture front aligns  
553            closely with the ground truth with NRMSE of 0.0100–0.01505 and relative  $L_2$  error of  
554            0.2867–0.3821. The predicted time series also match well with the ground truth, captur-  
555            ing the peaks and fine-scale details of  $V_{zx}$  over time. The errors in the magnitude of peaks  
556            are less than 1%, with some noise observed during the period between the first and second  
557            drops at point A. The accuracy drops for points located closer to the VW domain boundaries,



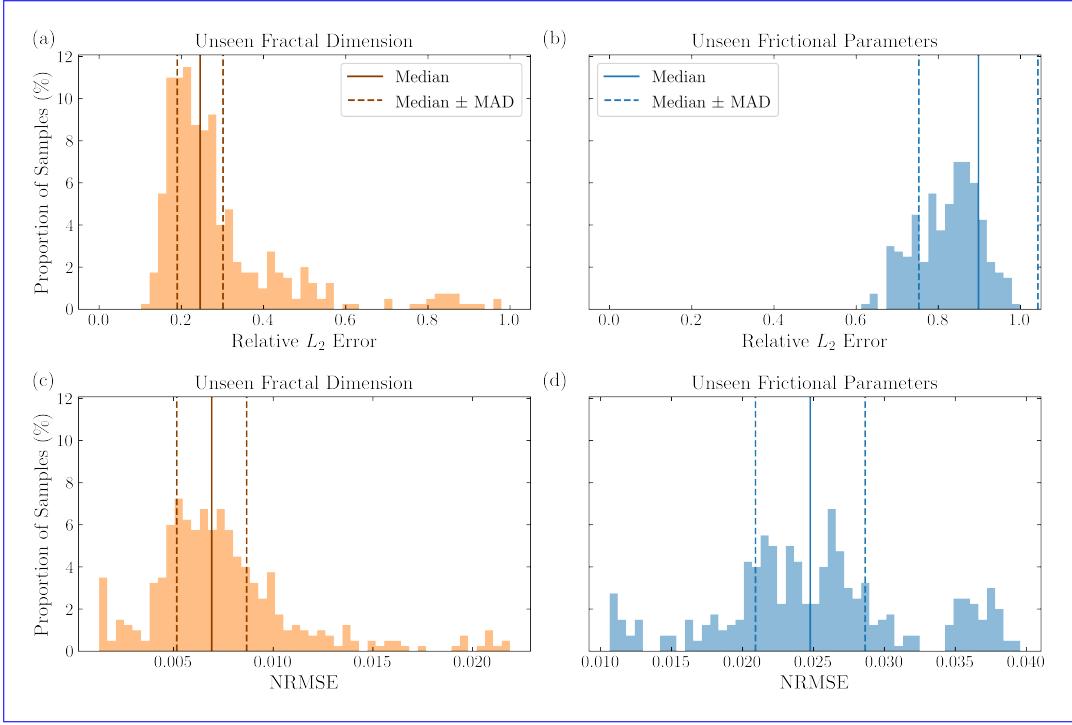
**Figure 11.** Results of FNO-2D on the testing dataset for 3D dynamic rupture. The inputs include the initial fractal shear stress distribution with  $D = 1.5$ ,  $D = 1.6$ , initial slip rate  $V_{zr0}$  extracted at the threshold  $V_{th} = 10^{-3}$  m/s, ten temporal steps of stress perturbation fields, and spatially varying frictional parameters:  $a$  with  $\Delta a_0 = 0.008$  and  $a_0 = 0.008$ ,  $a_0 = 0.007$ ,  $\Delta a_0 = 0.010$ , and  $b = 0.012$ , over the spatial domain. (a) Rupture Predicted rupture front contour plot, showing progression contours shown at 0.5 s intervals. (b) Time histories series of slip rate at selected points in 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,  $\Delta a_0 = 0.008$ ,  $a_0 = 0.008$ , and  $b = 0.012$ , 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.

with some minor shifts in arrival times and slightly larger mismatch in the peaks (e.g., points E and F). However, overall, there is good agreement between the model predictions and the ground truth. 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.

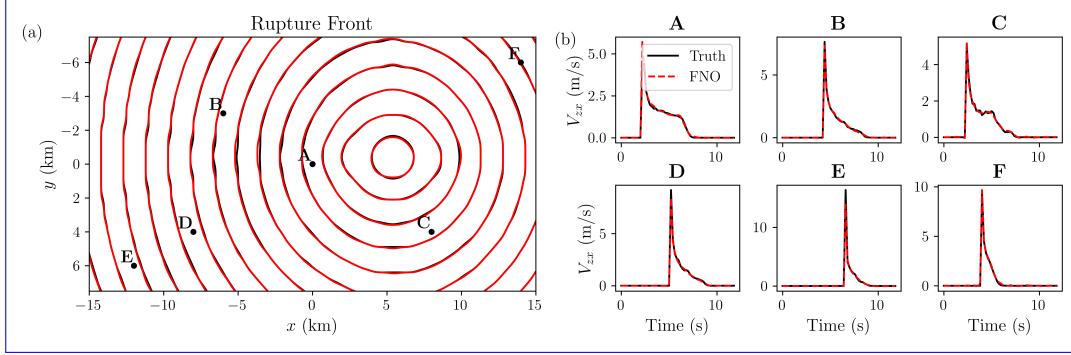
Additionally, we test the model on a fractal initial shear stress distribution with an unseen fractal dimension  $D = 1.3$ , similar to the FNO-1D case. The frictional parameters are identical to those used previously with FNO-1D. The predictions, We generate an additional 100 realizations with  $D = 1.3$  and randomly sampled frictional parameters from the seen ranges during training, as listed in Table 2. We preprocess the data such that



**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  $0.0248 \pm 0.00386$ .

571  $V_{zr0}$  is chosen at the step following thresholds  $V_{th} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$  m/s. This results  
 572 in 400 realizations. The distributions of relative  $L_2$  error and NRMSE are shown in Figures  
 573 13a and 13c. Both distributions are skewed towards zero, showing generalizability to unseen  
 574 fractal dimension cases.

575 An example of the predictions is presented in terms of the rupture front and time  
 576 series at selected points in Figure 14, show agreement with the ground truth. The shape  
 577 of the predicted rupture front 14. The frictional parameters are  $\Delta a_0 = 0.011$ ,  $a_0 = 0.0065$ ,  
 578 and  $b = 0.012$ , and here,  $V_{th} = 0$  m/s. The rupture front contour aligns with the ground  
 579 truth initially, though the match slightly decreases over time, resulting in a NRMSE of  
 580 0.00906, resulting in an NRMSE of 0.00422 and a relative  $L_2$  error of 0.3883–0.1284 for  
 581 this realization. We note that shifts in the rupture front can cause a high relative  $L_2$  error  
 582 despite the visual agreement. This is due to the small ground truth values, close to zero,  
 583 being compared to the peak slip rates in the shifted rupture front. The time series at  
 584 point E shows slight differences in the peak slip rate. However, it still captures the overall  
 585 trend of The model accurately captures the rupture arrival time. Peak amplitudes exhibit  
 586 very small mismatches at points D and E. Furthermore, the evolution predictions of post-peak  
 587 decay closely match the ground truth. This demonstrates that FNO-2D generalizes ef-  
 588 fectively to different fractal dimensions of  $\tau_0$ .



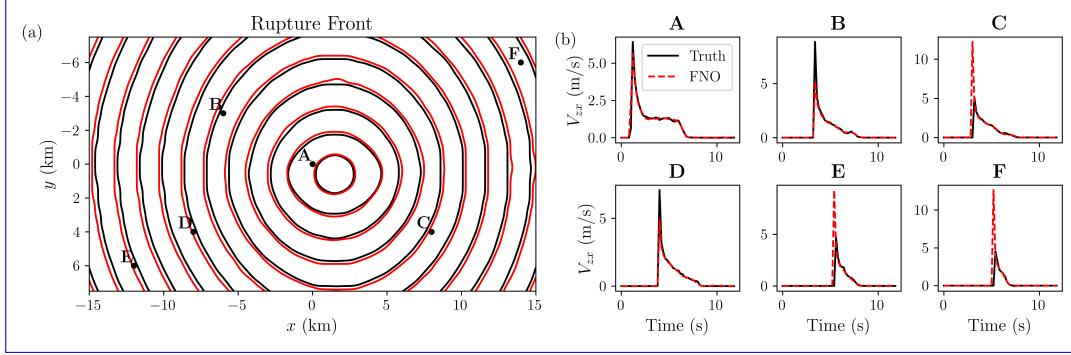
**Figure 14.** Results of FNO-2D testing on an unseen fractal dimension  $D = 1.3$  of the initial fractal shear stress. The frictional parameters are  $\Delta a_0 = 0.008$ ,  $\Delta a_0 = 0.011$ ,  $a_0 = 0.008$ ,  $a_0 = 0.0065$ , and  $b = 0.012$ . The initial  $V_{z0}$  is chosen at the time step 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.

#### 589 4.2.3 Generalization to Unseen Frictional Parameters

590 We evaluate the generalization performance of the trained FNO-2D model on pre-  
 591 viously unseen frictional parameter values. Specifically, ~~parameter  $a$~~  we generate 100 realizations  
 592 using frictional parameters sampled from combinations not included in the training dataset,  
 593 as summarized in Table 4. After preprocessing the data based on velocity thresholds  $V_{th} \in \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$ ,  
 594 we obtain a total of 400 realizations with varying  $V_{z0}$  fields. Figures 13b and 13d show  
 595 the distributions of the relative  $L_2$  error and the NRMSE, respectively. Although the  
 596 relative  $L_2$  error is high, with a median of 0.898, the range of the NRMSE in the 3D model  
 597 is smaller than that in the 2D model. Despite the lower accuracy as measured by the relative  
 598  $L_2$  error, the 3D model demonstrates greater stability in terms of normalized error when  
 599 compared with the observed range in  $V_{z0}$ .

600 Figure 15 illustrates a representative case, including the predicted rupture front and  
 601 time histories of the slip rate component  $V_{zx}$  at selected spatial locations. In this example,  
 602 the frictional parameter  $a$  is set to  $0.0075$ – $0.0067$  in the VW region and  $0.0165$ – $0.0177$   
 603 in the VS region, while parameter  $b$  is uniformly fixed at  $0.012$ – $0.012$  along the fault.  
 604 The initial shear stress  $\tau_0$  is modeled  $\tau_0$  is constructed using a fractal distribution with  
 605 fractal dimension  $D = 1.5$ . Figure 15 presents a successful case of the predicted rupture  
 606 front and corresponding time series of  $V_{zx}$  at selected spatial locations. ~~a fractal dimension~~  
 607  $D = 1.5$ .

608 We observe spatial ~~shifts and temporal discrepancies~~ in the rupture front contours  
 609 ~~, which amplify the NMRSE and relative  $L_2$  error to 0.0171 and 0.6051, respectively. Some~~  
 610 ~~mismatches at the peaks of  $V_{zx}$  over time at selected points also contribute to this error.~~  
 611 ~~While~~ and the time histories at points C, E, and F, which contribute to elevated errors,  
 612 with an NRMSE of 0.0219 and a relative  $L_2$  error of 0.6767. In particular, mismatches  
 613 in the peak magnitudes of  $V_{zx}$  at these locations account for a significant portion of the  
 614 error. However, the FNO-2D ~~still captures the global features of the model successfully~~  
 615 ~~captures the overall~~ temporal evolution of  $V_{zx}$  across  $V_{zx}$ , including the nucleation, prop-  
 616 agation, and arrest phases ~~, we later discuss specific challenges related to of rupture. In~~  
 617 ~~the following discussion, we address the specific challenges in generalizing to unseen fric-~~  
 618 ~~tional properties and how to possibly improve the relative  $L_2$  error~~ propose strategies to  
 619 ~~improve predictive accuracy.~~



**Figure 15.** Results of FNO-2D testing on an unseen  $a - b$  distribution. Inputs consist of frictional parameters  $a$  and  $b$ , with  $\Delta a_0 = 0.009 \Delta a_0 = 0.011$ ,  $a_0 = 0.0075 a_0 = 0.0067$ , and  $b = 0.012$ , along with an initial shear stress distribution  $\tau_0$  with a fractal dimension  $D = 1.5$  and  $V_{tb} = 0$  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.

**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<br>(-) | 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                 | $4 \times 10^5$                           |

620

### 4.3 Computational Efficiency Analysis

621

We summarize the computational efficiency of FNO-1D and FNO-2D in Table 5.

622

The training and testing times are both evaluated on an Nvidia reported for an NVIDIA A100 GPU. Once the models are trained, they, 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 used as an alternative employed as efficient alternatives to numerical simulators.

623

624

625

626

627

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.

628

629

630

631

632

633

634

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.

635

636

637

638

639 **5 Discussion and Conclusion**

640 To overcome the computational bottleneck of classical physics-based earthquake  
641 models, we present an FNO-based surrogate model to accelerate dynamic rupture sim-  
642 ulations. In the surrogate model, we predict the evolution of the slip rate distribution  
643 on the fault plane, given the initial distribution of shear stress, slip rate, stress perturbation,  
644 and frictional parameters  $a$  and  $b$  in the rate-and-state friction law, a commonly used  
645 friction law in earthquake modeling. The ground truth is generated using the SBI method  
646 with different initial fractal shear stress, slip rate, nucleation site, and frictional param-  
647 eter distributions in both 2D and 3D homogeneous domains. Each dataset is trained and  
648 tuned separately using FNO-1D and FNO-2D. We select the hyperparameters based on  
649 a balance between computational efficiency and accuracy as discussed in Appendix-Ap-  
650 pendix B. As a result, we achieve speedups of  $4 \times 10^5$  and  $2 \times 10^5$  compared to the SBI  
651 method in 3D and 2D dynamic rupture problems, respectively. This increase in speed  
652 potentially enables more efficient dynamic rupture inversion, statistical analysis, and iden-  
653 tification of extreme events in ensembles of earthquake rupture forecasts.

654 Predicted slip rates are evaluated using pointwise metrics, including the normalized  
655 root mean squared error (NRMSE) NRMSE, which compares predictions with the ob-  
656 served range of the ground truth, and the relative  $L_2$  error, which assesses the discrep-  
657 ency at discrete points in the domain. We demonstrate that the training process of FNO-  
658 based models does not exhibit overfitting. Test set predictions show strong agreement  
659 with the ground truth. Specifically, for the 2D dynamic rupture dataset, the median NRMSE  
660 is 0.329% 0.388% and the median relative  $L_2$  error is 6.85% 2.70%. For the 3D dynamic  
661 rupture dataset, the median NRMSE is 0.387% 0.366% and the median relative  $L_2$  er-  
662 ror is 15.3% 13.3%.

663 Higher errors in the 3D problem are primarily due to down-sampling to a lower res-  
664 olution than the physical problem, resulting in a grid size that may smear some details  
665 in the process zone region. This limitation is due to constrained computational resources.  
666 Access to additional resources would allow training at higher resolution, potentially re-  
667 ducing these errors to a level comparable to what is seen in the 2D problem. The errors  
668 are generally more pronounced in regions with high-frequency features, reflecting the spec-  
669 tral bias, which states that deep learning models tend to favor low-frequency components.  
670 To further improve model performance, we suggest training on a specialized dataset, with  
671 the most critical hyperparameter being the width of  $P$  and  $Q$ . This opens opportunities  
672 for further improving the model performance by increasing the size of the training dataset  
673 as well as increasing the model complexity. larger dataset.

674 Furthermore, we test the model. We also evaluate the models under unseen initial  
675 stress distributions and frictional parameters. The FNO-based models demonstrate robustness  
676 when applied to show robustness in out-of-distribution cases, such as a uniform initial  
677 stress distribution. Under this distribution, the problem setup corresponds corresponding  
678 to the SCEC/USGS verification exercise. Additionally, we evaluate the model on an initial  
679 They also perform well on a shear stress distribution with an unseen fractal dimension  
680 of  $D = 1.3$ . The FNO-based models maintain good accuracy under these conditions, with  
681 NRMSE values lower than 1%. We also test FNO-1D and FNO-2D  $D = 1.3$ , achieving  
682 median relative  $L_2$  errors of 31.3% in 2D and 24.5% in 3D. By contrast, when tested on  
683 unseen frictional parameters. While we have demonstrated successful predictions for unseen  
684 frictional parameters, the model's performance is less robust compared to tests involving  
685 unseen shear stress. yields a median relative  $L_2$  error of 13.0% in 2D but 89.8% in 3D. One  
686 important source of error in FNO predictions is spatial and temporal shifts. When the  
687 slip rate is small, the relative  $L_2$  error can be amplified, particularly where sharp, high-value  
688 rupture fronts shift in space and time relative to small slip rate values. This reduced ro-  
689 bustness may be attributed to the fact that variations in frictional parameters modify  
690 the underlying material model and consequently change the properties of the governing  
691 operator. In contrast, changing the shear stress distribution affects only the initial con-

692 ditions without modifying the governing equations. This limitation could ~~potentially be~~  
 693 ~~addressed by incorporating a greater variety of frictional parameters during training~~ be  
 694 mitigated by training with a larger datasets. As demonstrated in Section Appendix B,  
 695 increasing the size of the training dataset improves performance. A shared community  
 696 database of large-scale dynamic rupture simulations would be highly beneficial. ~~One important~~  
 697 ~~source of error in FNO predictions is spatial and temporal shifts. When the slip rate is~~  
 698 ~~small, the relative  $L_2$  error can be amplified, particularly where sharp, high-value rupture~~  
 699 ~~fronts shift in space and time relative to small slip rate values.~~

700 We highlight a couple of key aspects of FNO as applied to our problem. The FNO  
 701 neural approximation effectively captures the evolution of slip rate over extended peri-  
 702 ods. The FNO formulation is time-continuous and can be discretized as needed for train-  
 703 ing and application. ~~Furthermore~~ Furthermore, the FNO enables the use of data gener-  
 704 ated at varying spatial and temporal discretizations. Traditional solvers must adhere to  
 705 stability condition to maintain stability, whereas this constraint does not apply to FNOs,  
 706 allowing for greater flexibility in discretization and time stepping. Third, the FNO shows  
 707 potential for successful generalization to unseen stress and frictional condition although  
 708 our findings suggest that the performance will further improve by training on larger data  
 709 sets. ~~In addition, incorporating stress perturbations for rupture nucleation enables the~~  
 710 ~~model to learn the operator more flexibly, as one can vary the nucleation site location,~~  
 711 ~~nucleation radius, or even the spatial distribution of the perturbation. Lastly, this model~~  
 712 ~~serves as a proof-of-concept for simulating rupture scenarios given the initial conditions~~  
 713 ~~at any point in time.~~

714 We conclude with a discussion of limitations and potential future directions. First,  
 715 this work assumes a specific mean and standard deviation for the input distributions.  
 716 Future work should explore training the FNO on datasets with more diverse parameter  
 717 distributions ~~and non-dimensional quantities~~. Second, the current FNO is trained on datasets  
 718 generated using a ~~specific nucleation protocol. To improve generalization~~ constant normal  
 719 ~~stress. To be more realistic~~, training the model on a ~~broad range of nucleation parameters,~~  
 720 ~~including variation in the nucleation site, heterogeneous normal stress~~ is desirable. Third,  
 721 the FNO is trained for predictions within a fixed time interval. Extending the predic-  
 722 tion horizon beyond a specific interval could involve recursive training, enabling the model  
 723 to predict slip evolution iteratively. Furthermore, a hybrid approach combining tradi-  
 724 tional numerical solvers with FNOs could enable long-range predictions. A key aspect  
 725 of this approach would be defining a robust criterion, such as a physics-guided error thresh-  
 726 old, for switching between FNO and traditional solvers. ~~In addition, investigating more~~  
 727 ~~advanced surrogate model architectures may further enhance performance.~~ Finally, ac-  
 728 ~~cess to larger training datasets and more GPUs will further~~ ~~improve~~ improve the per-  
 729 ~~formance of FNOs and their ability to generalize. A community effort for creating a database~~  
 730 ~~for dynamic rupture simulations would be beneficial in that respect.~~

## 731 Appendix A Details on Model Setup

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

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

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

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

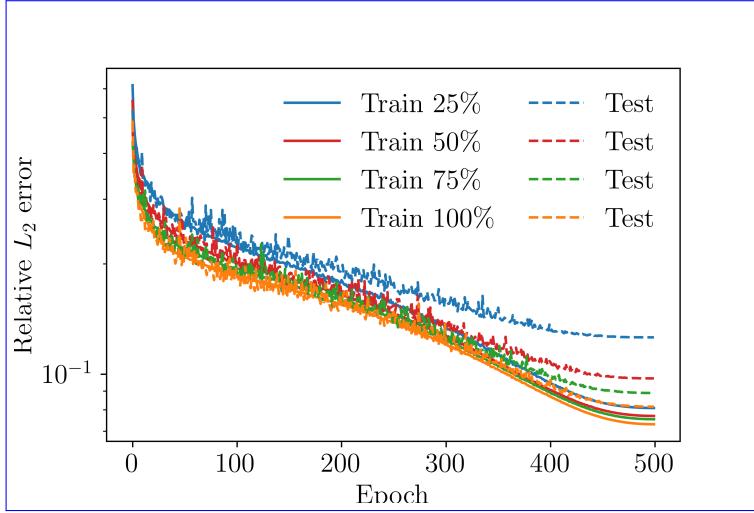
$$G(t) = \begin{cases} \exp\left[\frac{(t-T)^2}{t(t-2T)}\right], & 0 < t < T, \\ 1, & t \geq T. \end{cases} \quad (\text{A3})$$

## 737 Appendix B FNO Hyperparameters Tuning and Training Strategies

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

**Table B1.** Relative  $L_2$  errors for different hyperparameter configurations, including the num-  
 ber 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.** Heat map showing the relative  $L_2$  error of 100 Training and testing sets with inputs losses for different numbers of fractal stress distribution at  $D = 1.6$  training realizations: 25%, 50%, 75%, and frictional parameters  $\Delta a_0 = 0.012$  and  $a_0 = 0.006$ . The number 100% of modes and the widths of  $P$  and  $Q$  are shown on the vertical and horizontal axes, respectively. Solid lines show training losses, with (a) a learning rate of 0.0005 and (b) a learning rate of 0.001. Dashed lines show testing losses.

To improve accuracy and generalizability, we propose training exclusively on a specific dataset with a chosen fractal dimension and frictional parameters. Here, we select a dataset with  $D = 1.6$ ,  $\Delta a_0 = 0.012$ , and  $a_0 = 0.006$  resulting in a training set of 900 realizations and a testing set of 100 realizations. We primarily tune the number of Fourier modes, the width of the lifting and projection layers, and on a larger number of realizations while keeping the architecture and hyperparameters fixed. We assess the effect of the learning rate and training set size by using 25%, 50%, 75%, and 100% of the available 28,700 realizations. The relative  $L_2$  error for the testing set is then evaluated on a testing set of 3,280 realizations. The training losses for the different training set sizes are shown in Figure ???. The number of modes is increased to 16, 32, 64, 128. We observe a decrease in error when increasing the depth of the fully connected layers  $P$  and  $Q$ . The most influential hyperparameter is the width when using a learning rate of 0.001. It is evident that 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 decreases as the widths of  $P$  and  $Q$  increase, as indicated by the blue gradient on the right side of the heat map. Other hyperparameters, including the number of Fourier layers, the number of retained Fourier modes, and on the learning rate, do not exhibit a clear pattern, and the error fluctuates.

After hyperparameter tuning, the revised model demonstrates enhanced predictive accuracy, with the predicted slip rates illustrated in Figure ???. In the zoomed inset of panel (f), the specialized FNO captures some high-frequency components better than the original FNO at  $x \approx -10$  and  $x \approx 10$  km. This improvement paves the way for further increasing the complexity of the FNO in future work. 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.

785 Results from FNO-1D (dashed red) are compared with those from the specialized  
 786 FNO (dashed orange), which is trained on 900 separate sets of fractal stress distributions  
 787 with  $D = 1.6$  and frictional parameters  $a$  and  $b$  (with  $\Delta a_0 = 0.012$  and  $a_0 = 0.006$ ; dashed  
 788 orange) and with the ground truth (solid blue). (a) and (b) show the inputs, including  
 789 the initial shear stress and frictional parameters, while (c) – (f) are the outputs, consisting  
 790 of predicted slip rate snapshots at selected discrete time steps.

791 For the 3D dynamic rupture dataset, we primarily tune the model while keeping  
 792 the number of Fourier layers fixed at four, based on insights from the 2D dynamic rup-  
 793 ture dataset. The number of retained Fourier modes is initially selected as 16, 32, and  
 794 64, while the depth of  $P$  and  $Q$  is chosen from 32, 64, and 128. We consider learning rates  
 795 of 0.001 and 0.0005. The results of the tuning process are presented in Table B2. The  
 796 optimal model is selected with four Fourier layers, 32 Fourier modes, 128 neurons in  $P$   
 797 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             |
| 0.001         | 16    | 0.127383             | 0.115418 | 0.113272        |
|               | 32    | 0.108283             | 0.105937 | <b>0.080965</b> |
|               | 64    | 0.102793             | 0.126559 | 0.114296        |
| 0.0005        | 16    | 0.123509             | 0.093694 | 0.107529        |
|               | 32    | 0.101131             | 0.096700 | 0.083106        |
|               | 64    | 0.096357             | 0.106043 | 0.091511        |

## 798 Appendix C Quantitative Analysis of Prediction Error

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

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

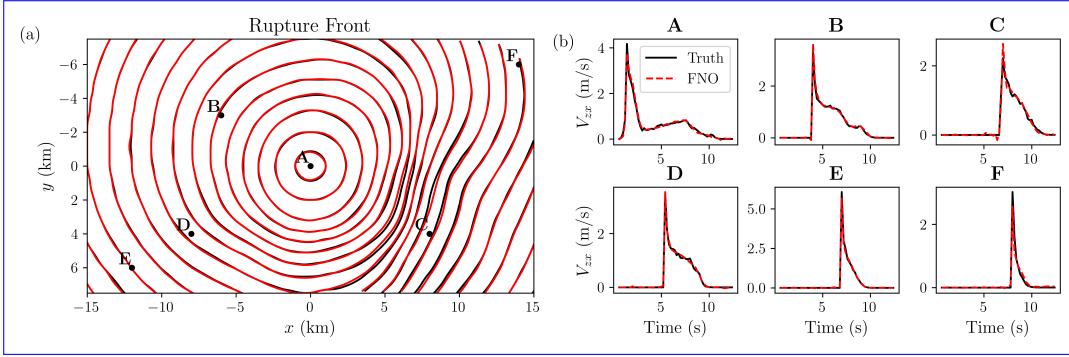
802 where  $f(t)$  denotes the vector of slip rate predictions from the FNO model,  $g(t)$  represents  
 803 the corresponding ground truth values, and  $t_s$  is a temporal shift applied to align the  
 804 signals. The optimal time shift is obtained as

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

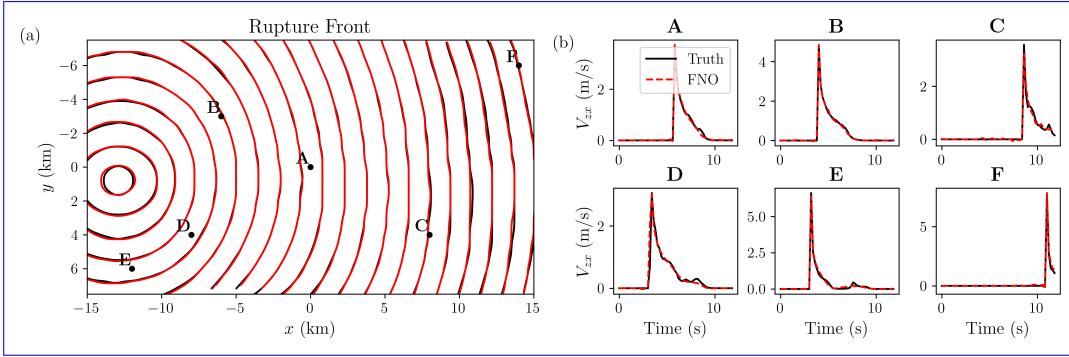
805 A smaller  $Q$  value indicates closer agreement between prediction and ground truth.

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

810 Figures C1–C3 present the rupture front evolution and slip rate histories for these  
 811 cases. The computed  $Q$  values are 12.94%, 20.6%, and 39.6%, with corresponding optimal  
 812 time shifts  $t_s^*$  of 0.005 s, 0.004 s, and 0.025 s, respectively. Figure C4 shows the relative



**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.

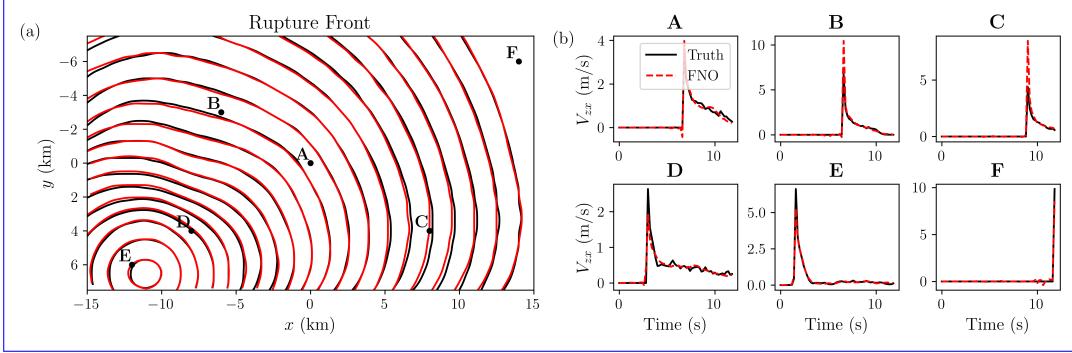


**Figure C2.** Results of FNO-2D on the testing dataset for a 3D dynamic rupture case with a relative  $L_2$  error of 0.207 (median + 2MAD) and an NRMSE of 0.00531. The inputs include the initial fractal shear stress ( $D = 1.2$ ), the initial  $V_{zx}$  field ( $V_{th} = 10^{-3}$  m/s), a nucleation 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.

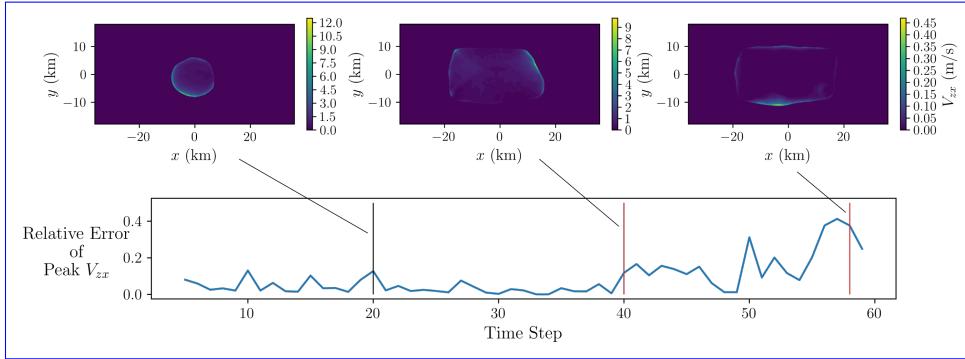
813 error of the peak  $V_{zx}$  as a function of time step for the case with relative  $L_2$  error closes  
 814 to the median. The error increases after time step 40, which marks the stage when the  
 815 rupture arrives at the VS boundaries. This introduces higher-frequency components into  
 816 the slip rate field that are more challenging to capture. Additional contributing factors  
 817 include the limited availability of training samples, particularly for cases where the rupture  
 818 reaches the boundaries, the restriction imposed by the chosen number of modes in the  
 819 FNO, and the inherent spectral bias of neural operators.

## 820 Open Research

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



**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.

824 **Acknowledgments**

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

836 **References**

- 837 Aagaard, B. T., Heaton, T. H., & Hall, J. F. (2001). Dynamic earthquake ruptures  
838 in the presence of lithostatic normal stresses: Implications for friction models  
839 and heat production. *Bulletin of the Seismological Society of America*, 91(6),  
840 1765–1796.
- 841 Abdelmeguid, M., & Elbanna, A. (2022). Modeling sequences of earthquakes and  
842 aseismic slip (seas) in elasto-plastic fault zones with a hybrid finite element  
843 spectral boundary integral scheme. *Journal of Geophysical Research: Solid*  
844 *Earth*, 127(12), e2022JB024548.
- 845 Abdelmeguid, M., Ma, X., & Elbanna, A. (2019). A novel hybrid finite element-  
846 spectral boundary integral scheme for modeling earthquake cycles: Application  
847 to rate and state faults with low-velocity zones. *Journal of Geophysical Re-*  
848 *search: Solid Earth*, 124(12), 12854–12881.
- 849 Ampuero, J.-P., & Rubin, A. M. (2008). Earthquake nucleation on rate and state  
850 faults–aging and slip laws. *Journal of Geophysical Research: Solid Earth*,  
851 113(B1).
- 852 Andrews, D. (1973). A numerical study of tectonic stress release by underground ex-  
853 plosions. *Bulletin of the Seismological Society of America*, 63(4), 1375–1391.
- 854 Andrews, D. (1980). A stochastic fault model: 1. static case. *Journal of Geophysical*  
855 *Research: Solid Earth*, 85(B7), 3867–3877.
- 856 Andrews, D. (1985). Dynamic plane-strain shear rupture with a slip-weakening fric-  
857 tion law calculated by a boundary integral method. *Bulletin of the Seismologi-*  
858 *cal Society of America*, 75(1), 1–21.
- 859 Andrews, D. (1999). Test of two methods for faulting in finite-difference calculations.  
860 *Bulletin of the Seismological Society of America*, 89(4), 931–937.
- 861 Aochi, H., & Twardzik, C. (2020). Imaging of seismogenic asperities of the 2016 ml  
862 6.0 amatrice, central italy, earthquake through dynamic rupture simulations.  
863 *Pure and Applied Geophysics*, 177(5), 1931–1946.
- 864 Arrowsmith, S. J., Trugman, D. T., MacCarthy, J., Bergen, K. J., Lumley, D., &  
865 Magnani, M. B. (2022). Big data seismology. *Reviews of Geophysics*, 60(2),  
866 e2021RG000769.
- 867 Azizzadenesheli, K., Kovachki, N., Li, Z., Liu-Schiavini, M., Kossaifi, J., & Anand-  
868 kumar, A. (2024). Neural operators for accelerating scientific simulations and  
869 design. *Nature Reviews Physics*, 6(5), 320–328.
- 870 Barall, M., & Harris, R. A. (2015). Metrics for comparing dynamic earthquake rup-  
871 ture simulations. *Seismological Research Letters*, 86(1), 223–235.
- 872 Barbot, S. (2022). A rate-, state-, and temperature-dependent friction law with  
873 competing healing mechanisms. *Journal of Geophysical Research: Solid Earth*,  
874 127(11), e2022JB025106.
- 875 Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-  
876 discrete transitions, progressive evolutionary changes, and different dynamic

- 877 regimes. *Reviews of Geophysics*, 46(4).
- 878 Ben-Zion, Y., & Rice, J. R. (1997). Dynamic simulations of slip on a smooth fault  
879 in an elastic solid. *Journal of Geophysical Research: Solid Earth*, 102(B8),  
880 17771–17784.
- 881 Bhat, H. S., Olives, M., Dmowska, R., & Rice, J. R. (2007). Role of fault branches  
882 in earthquake rupture dynamics. *Journal of Geophysical Research: Solid Earth*,  
883 112(B11).
- 884 Boerner, T. J., Deems, S., Furlani, T. R., Knuth, S. L., & Towns, J. (2023). Access:  
885 Advancing innovation: Nsf's advanced cyberinfrastructure coordination ecosystem:  
886 Services & support. In *Practice and experience in advanced research  
887 computing 2023: Computing for the common good* (pp. 173–176).
- 888 Cao, Y., Fang, Z., Wu, Y., Zhou, D.-X., & Gu, Q. (2019). Towards understanding  
889 the spectral bias of deep learning. *arXiv preprint arXiv:1912.01198*.
- 890 Chester, F. M., & Chester, J. S. (1998). Ultracataclasite structure and friction pro-  
891 cesses of the punchbowl fault, san andreas system, california. *Tectonophysics*,  
892 295(1-2), 199–221.
- 893 Cochard, A., & Madariaga, R. (1994). Dynamic faulting under rate-dependent fric-  
894 tion. *pure and applied geophysics*, 142(3), 419–445.
- 895 Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F.  
896 (2022). Scientific machine learning through physics-informed neural networks:  
897 Where we are and what's next. *Journal of Scientific Computing*, 92(3), 88.
- 898 Dalguer, L. A., & Day, S. M. (2007). Staggered-grid split-node method for spon-  
899 taneous rupture simulation. *Journal of Geophysical Research: Solid Earth*,  
900 112(B2).
- 901 Das, S. (1980). A numerical method for determination of source time functions for  
902 general three-dimensional rupture propagation. *Geophysical Journal Interna-  
903 tional*, 62(3), 591–604.
- 904 Day, S. M. (1982). Three-dimensional finite difference simulation of fault dynamics:  
905 rectangular faults with fixed rupture velocity. *Bulletin of the Seismological So-  
906 ciety of America*, 72(3), 705–727.
- 907 de Hoop, M. V., Huang, D. Z., Qian, E., & Stuart, A. M. (2022). The cost-  
908 accuracy trade-off in operator learning with neural networks. *arXiv preprint  
909 arXiv:2203.13181*.
- 910 Dieterich, J. H. (1979). Modeling of rock friction: 1. experimental results and consti-  
911 tutive equations. *Journal of Geophysical Research: Solid Earth*, 84(B5), 2161–  
912 2168.
- 913 Duan, B., & Oglesby, D. D. (2006). Heterogeneous fault stresses from previous  
914 earthquakes and the effect on dynamics of parallel strike-slip faults. *Journal of  
915 Geophysical Research: Solid Earth*, 111(B5).
- 916 Gabriel, A.-A., Ampuero, J.-P., Dalguer, L., & Mai, P. M. (2012). The transition  
917 of dynamic rupture styles in elastic media under velocity-weakening friction.  
918 *Journal of Geophysical Research: Solid Earth*, 117(B9).
- 919 Gallovič, F., Valentová, Ľ., Ampuero, J.-P., & Gabriel, A.-A. (2019a). Bayesian  
920 Dynamic Finite-Fault Inversion: 1. Method and Synthetic Test. *Journal  
921 of Geophysical Research: Solid Earth*, 124(7), 6949–6969. doi: 10.1029/  
922 2019JB017510
- 923 Gallovič, F., Valentová, Ľ., Ampuero, J.-P., & Gabriel, A.-A. (2019b). Bayesian dy-  
924 namic finite-fault inversion: 2. application to the 2016 mw 6.2 amatrice, italy,  
925 earthquake. *Journal of Geophysical Research: Solid Earth*, 124(7), 6970–6988.
- 926 Geubelle, P. H., & Rice, J. R. (1995). A spectral method for three-dimensional elas-  
927 todynamic fracture problems. *Journal of the Mechanics and Physics of Solids*,  
928 43(11), 1791–1824.
- 929 Harris, R. A., Barall, M., Aagaard, B., Ma, S., Roten, D., Olsen, K., ... others  
930 (2018). A suite of exercises for verifying dynamic earthquake rupture codes.  
931 *Seismological Research Letters*, 89(3), 1146–1162.

- Harris, R. A., Barall, M., Archuleta, R., Dunham, E., Aagaard, B., Ampuero, J. P., ... others (2009). The scec/usgs dynamic earthquake rupture code verification exercise. *Seismological Research Letters*, 80(1), 119–126.
- Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.
- Hirata, T. (1989). A correlation between the b value and the fractal dimension of earthquakes. *Journal of Geophysical Research: Solid Earth*, 94(B6), 7507–7514.
- Ikari, M. J., Marone, C., & Saffer, D. M. (2011). On the relation between fault strength and frictional stability. *Geology*, 39(1), 83–86.
- Johnson, K. M., Burgmann, R., & Larson, K. (2006). Frictional properties on the san andreas fault near parkfield, california, inferred from models of afterslip following the 2004 earthquake. *Bulletin of the Seismological Society of America*, 96(4B), S321–S338.
- Kammer, D. S., Albertini, G., & Ke, C.-Y. (2021). Uguca: A spectral-boundary-integral method for modeling fracture and friction. *SoftwareX*, 15, 100785.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kong, Q., Zou, C., Choi, Y., Matzel, E. M., Azizzadenesheli, K., Ross, Z. E., ... Clayton, R. W. (2025). Reducing frequency bias of fourier neural operators in 3d seismic wavefield simulations through multi-stage training. *arXiv preprint arXiv:2503.02023*.
- Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89), 1–97.
- Lapusta, N., Rice, J. R., Ben-Zion, Y., & Zheng, G. (2000). Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate-and state-dependent friction. *Journal of Geophysical Research: Solid Earth*, 105(B10), 23765–23789.
- Lehmann, F., Gatti, F., Bertin, M., & Clouteau, D. (2023). Fourier neural operator surrogate model to predict 3d seismic waves propagation. *arXiv preprint arXiv:2304.10242*.
- Lehmann, F., Gatti, F., Bertin, M., & Clouteau, 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).

- 987 Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020).  
988     Earthquake transformer—an attentive deep-learning model for simultaneous  
989     earthquake detection and phase picking. *Nature communications*, 11(1), 3952.
- 990 Nielsen, S. B., Carlson, J., & Olsen, K. B. (2000). Influence of friction and fault ge-  
991     ometry on earthquake rupture. *Journal of Geophysical Research: Solid Earth*,  
992     105(B3), 6069–6088.
- 993 Oglesby, D. D., Archuleta, R. J., & Nielsen, S. B. (1998). Earthquakes on dipping  
994     faults: The effects of broken symmetry. *Science*, 280(5366), 1055–1059.
- 995 Oglesby, D. D., Archuleta, R. J., & Nielsen, S. B. (2000). The three-dimensional  
996     dynamics of dipping faults. *Bulletin of the Seismological Society of America*,  
997     90(3), 616–628.
- 998 Rahman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., ...  
999     Courville, A. (2019). On the spectral bias of neural networks. In *Inter-  
1000     national conference on machine learning* (pp. 5301–5310).
- 1001 Rahman, M. A., Ross, Z. E., & Azizzadenesheli, K. (2022). U-no: U-shaped neural  
1002     operators. *arXiv preprint arXiv:2204.11127*.
- 1003 Rekoske, J. M., May, D. A., & Gabriel, A.-A. (2025). Reduced-order modelling for  
1004     complex three-dimensional seismic wave propagation. *Geophysical Journal In-  
1005     ternational*, 241(1), 526–548.
- 1006 Renard, F., & Candela, T. (2017). Scaling of fault roughness and implications for  
1007     earthquake mechanics. *Fault zone dynamic processes: Evolution of fault proper-  
1008     ties during seismic rupture*, 195–215.
- 1009 Ripperger, J., Ampuero, J.-P., Mai, P., & Giardini, D. (2007). Earthquake source  
1010     characteristics from dynamic rupture with constrained stochastic fault stress.  
1011     *Journal of Geophysical Research: Solid Earth*, 112(B4).
- 1012 Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P wave arrival picking and first-  
1013     motion polarity determination with deep learning. *Journal of Geophysical Re-  
1014     search: Solid Earth*, 123(6), 5120–5129.
- 1015 Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2019).  
1016     Phaselink: A deep learning approach to seismic phase association. *Journal  
1017     of Geophysical Research: Solid Earth*, 124(1), 856–869.
- 1018 Ruina, A. (1983). Slip instability and state variable friction laws. *Journal of Geo-  
1019     physical Research: Solid Earth*, 88(B12), 10359–10370.
- 1020 Tainpkdipat, N., Abdelmeguid, M., Zhao, C., & Elbanna, A. (2025). Fourier neural  
1021     operators for accelerating earthquake dynamic rupture simulations [data set].  
1022     Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.15085622> doi:  
1023     10.5281/zenodo.15085622
- 1024 Ulrich, T., Gabriel, A.-A., Ampuero, J.-P., & Xu, W. (2019). Dynamic viability  
1025     of the 2016 mw 7.8 kaikōura earthquake cascade on weak crustal faults. *Nature  
1026     Communications*, 10, 1213. doi: 10.1038/s41467-019-09125-w
- 1027 Van Zelst, I., Wollherr, S., Gabriel, A.-A., Madden, E. H., & van Dinther, Y. (2019).  
1028     Modeling megathrust earthquakes across scales: One-way coupling from geo-  
1029     dynamics and seismic cycles to dynamic rupture. *Journal of Geophysical  
1030     Research: Solid Earth*, 124(11), 11414–11446.
- 1031 Wei, W., & Fu, L.-Y. (2022). Small-data-driven fast seismic simulations for com-  
1032     plex media using physics-informed fourier neural operators. *Geophysics*, 87(6),  
1033     T435–T446.
- 1034 Xu, J., Zhang, H., & Chen, X. (2015). Rupture phase diagrams for a planar fault in  
1035     3-d full-space and half-space. *Geophysical Journal International*, 202(3), 2194–  
1036     2206.
- 1037 Yang, L., Liu, X., Zhu, W., Zhao, L., & Beroza, G. C. (2022). Toward improved ur-  
1038     ban earthquake monitoring through deep-learning-based noise suppression. *Sci-  
1039     ence advances*, 8(15), eabl3564.
- 1040 Yang, Y., Gao, A. F., Azizzadenesheli, K., Clayton, R. W., & Ross, Z. E. (2023).  
1041     Rapid seismic waveform modeling and inversion with neural operators. *IEEE*

- 1042                    *Transactions on Geoscience and Remote Sensing*, 61, 1–12.
- 1043       Yang, Y., Gao, A. F., Castellanos, J. C., Ross, Z. E., Azizzadenesheli, K., & Clay-  
1044                    ton, R. W. (2021). Seismic wave propagation and inversion with neural  
1045                    operators. *The Seismic Record*, 1(3), 126–134.
- 1046       Zhang, T., Trad, D., & Innanen, K. (2023). Learning to solve the elastic wave equa-  
1047                    tion with fourier neural operators. *Geophysics*, 88(3), T101–T119.
- 1048       Zhu, W., & Beroza, G. C. (2019). Phasenet: a deep-neural-network-based seismic  
1049                    arrival-time picking method. *Geophysical Journal International*, 216(1), 261–  
1050                    273.
- 1051       Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and  
1052                    decomposition using deep neural networks. *IEEE Transactions on Geoscience  
1053                    and Remote Sensing*, 57(11), 9476–9488.
- 1054       Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P., & Beroza, G. C. (2022). An end-  
1055                    to-end earthquake detection method for joint phase picking and association  
1056                    using deep learning. *Journal of Geophysical Research: Solid Earth*, 127(3),  
1057                    e2021JB023283.
- 1058       Zou, C., Azizzadenesheli, K., Ross, Z. E., & Clayton, R. W. (2024). Deep neural  
1059                    helmholtz operators for 3d elastic wave propagation and inversion. *Geophysical  
1060                    Journal International*, ggae342.