SCEC Report: Proposal 23186

A Data-Driven Site Response Module for the Broadband Platform: Development and Prototype Implementation

PI: Domniki Asimaki
Graduate Students: Flora Xia; Yaozhong Shi

July 31, 2025

Abstract

The proliferation of seismic record repositories through the expansion of seismic networks and high-performance computer simulations of earthquake scenarios, have made available ground-motion databases rich enough to make possible data-driven models of ground-motion synthesis that account for complex site effects. Data-driven methods offer a novel approach to describing these processes by directly learning the governing laws from sufficiently rich training data, while avoiding the use of simplified assumptions that limit the realism of models developed with traditional statistical tools. In this study, we demonstrate this new paradigm of learning the underlying physics in a data-driven framework –and quantifying as such, the sources of ground motion epistemic uncertainty, by developing a Fourier Neural Operator (FNO) which modifies the outcrop ground motions from the SCEC BBP to account for the full nonlinear response of the near-surface soil layers. FNO was trained on non-linear one-dimensional wave propagation through smooth Bay Area velocity profiles using the site-response software, PySeismoSoil. A key advantage of the neural operator architecture in FNO compared to traditional neural networks, is its ability to learn the mapping between continuous function spaces as opposed to finite-dimensional sets, rendering the training and application of the model resolutions invariant (i.e., training can include input signals of different sampling frequencies without loss of information or generation of artifacts, while prediction can be performed on sampling frequencies independent of training). Verification analyses through residual and goodness of fit evaluations demonstrate that FNO can correctly estimate the nonlinear amplification for ground motions and profiles not included in the training dataset in the 0.1 to 30z frequency range. By appropriately conditioning data-driven algorithms, our work demonstrates the potential of using these methods to learn increasingly complex physics and their uncertainty over the entire frequency range of engineering interest, and to modulate on demand time-histories appropriate for engineering design with high degree of realism.

1 Introduction

Modern advancements in computer processing power have enabled significant improvements in physics-based, broadband ground motion simulations for seismic hazard mapping and engineering applications. Our goal for developing an easy-to-use, data-driven nonlinear site response module is to further improve the ability for ground motion models to accurately represent the site-specific nonlinear response of materials in the shallow crust. We take advantage of cutting edge machine learning architecture to develop a data-driven model that can accurately capture the nonlinear response of weathered rock and sedimentary soils, through which high frequencies propagate in the near surface.

Historically, researchers developed site amplification factors (also known as ground motion attenuation relations) by fitting a functional form equation to data through regression analysis. In this project, we introduce a site response module that departs from the historically used functional form representation in order to accurately and efficiently perform nonlinear site response analysis. Like many site modules of ground motion models, our model is designed to apply on a rock outcrop ground motion the appropriate amplification according to site-specific shear wave velocity properties, as to produce the corresponding ground surface site response. However, by taking advantage of the strengths of data-driven models, our site response module is able to directly learn the nonlinear relationship between rock outcrop and surface response, parameterized by an entire input acceleration time series and a multi-variable representation of shear wave velocity profile. Our model is therefore able to avoid the biases associated with predefining a functional form, and it is also able to reduce the loss of information associated with a limited representation of input motion intensity and velocity profile.

In the following sections, we give a primer on neural operator machine learning architectures and our FNO achitecture in particular, followed by a description of the input ground motions and velocity profiles used in the training, the nonlinear model used in the analyses, the training and testing datasets, and a description of the model validation, focusing specifically on the goodness of fit performance as formulated by (1) across multiple metrics.

2 Neural Operators

The deep learning architecture used for site-FNO is the neural operator, which is a novel paradigm well-suited for solving differential equations such as those governing 1D wave propagation. This is because neural operators are designed to learn how to map between two infinite dimensional function spaces, in comparison with traditional neural networks that are limited to mapping between finite dimensional spaces.

2.1 site-FNO: A Data-Driven Method

With adequate time and computational resources, it is sufficient to provide a reference ground motion and site characteristics to a numerical code, and directly compute the nonlinear site response to get the surface ground motion. However, there are many situations where a quick and accurate estimate of the site response is desired – such as uncertainty quantification, etc. In such situations, we want to be able to provide similar inputs (ground motion, site characteristics such as Vs) and get out a faster and less computationally expensive estimate of the site response.

For example, the Statewide California Earthquake Center (SCEC) Broadband Platform (BBP) has two implemented modules for computing the site response on generated ground motions: the GP method and PySeismoSoil. These modules are designed to be quick and easy to use.

In this work we present site-FNO, a data-driven Fourier neural operator model, which has been trained to produce a site-specific surface ground motion time series given a reference site time series and easily acquirable information on the local soil profile of interest. It is a simplified tool for accurate approximation of nonlinear site response which can produce outputs at speeds of over three orders of magnitude faster than traditional solvers over a broad frequency range (0.1 - 20 Hz). We call site-FNO a data-driven method, because only the data controls how the neural operator model relates input variables to the output. This can be contrasted with traditional regression methods that are more prone to bias, where a functional form is predetermined and regression methods are used to fit it to data characteristics.

3 Training Datasets

3.1 Nonlinear Site Response

We generated the dataset of nonlinear simulations that would be used to train and test the neural operator model using the open-source, 1D seismic site response analysis code, PySeismoSoil (**?**). For nonlinear analysis, PySeismoSoil computes the solution in the time domain using the finite difference method (FDM). There are three major modeling aspects for nonlinear simulation that PySeismoSoil improves on. First, PySeismoSoil models small-strain damping using the memory-variable technique proposed by Liu and Archuleta (2006), which more accurately simulates frequency-independent small-strain damping than Rayleigh or Caughey damping. Second, Pyseismosoil models hysteresis behavior using the model proposed by Li and Assimaki (2010), which produces narrower and more realistic hysteresis loops than those produced using Masing rules. Finally, PySeismoSoil models the stress-strain and damping behaviors of soil using the hybrid hyperbolic (HH) model (2), which can capture both small-strain and large-strain soil behaviors. PySeismoSoil was developed over the past 15 years with partial support from SCEC.

3.2 SCEC BBP Ground Motions

The simulated ground motion time series used to generate the training and testing datasets were generated using the Southern California Earthquake Center's Broadband Platform (SCEC BBP v17.3.0) using the Graves and Pitarka (2015) simulation module (BBP-GP). The database was developed by (1) in collaboration with BBP-GP developer Robert Graves, and contains 113 events ranging from M 5.5 to 7.2 on strike-slip and from M 7.0 to 7.2 on reverse faults, which were recorded at stations on reference site conditions (Vs30 = 760 m/s) at rupture distances between 0 to 200 km. The marginal distribution of the ground motion characteristics of the data is shown in Figure 1.

In this study, we take the N-S and E-W components of each record to be separate time series, for a total of 11,000 ground motion time series that were extracted from the (**?**) database for use in training and testing the neural operator model. We computed in PySeismoSoil the nonlinear site response for the 11,000 ground motions on each of the 33 velocity profiles produced in this study. The surface response was recorded for each nonlinear simulation for a grand total of 363,000 input-output ground motion site response samples.

3.3 Shear Wave Velocity Model

The velocity profiles were generated using the parameterized, region-specific near-surface velocity model algorithm developed by (**?** 3), which takes as input the V_{S30} and a depth vector, and produces a generic 1D velocity profile for use in wave propagation simulations. The algorithm is based on the one originally proposed by (**?**) for SCEC, but includes new scaling relationships to augment the model's predictive and extrapolation capabilities. For the training dataset, we focused on eleven different V_{S30} values in this study, evenly spaced 50 m/s apart between 250 m/s and 750 m/s to capture a wide range of shear wave velocity profiles. Each V_{S30} value was used to generate three shear wave velocity profiles: an upper and a lower bound, as well as median estimate. In total, 33 velocity profiles were produced at a resolution of 1 meter up to a depth of 275 meters. The maximum shear wave velocity for each profile was capped to 1000 m/s.

3.4 Nonlinear constitutive model: Hybrid-Hyperbolic

Within PySeismoSoil, we captured the nonlinear behavior of the soil using the Hybrid Hyperbolic (HH) 1D soil constitutive model developed by (2). The HH model is formulated as composite of two hyperbolic models (aka KZ models): the modified hyperbolic model (MKZ) proposed by (4) for the low-to-medium strain range, and a flexible KZ model (FKZ) in the high strain range. Because the HH model is designed to represent dynamic soil behavior over a wide range of strains, it is able to accurately capture the shear strength of soils in near-surface layers that have lower velocities without under-predicting the soil stress. In addition, it only requires a 1D V_S profile to calibrate its parameters. The HH

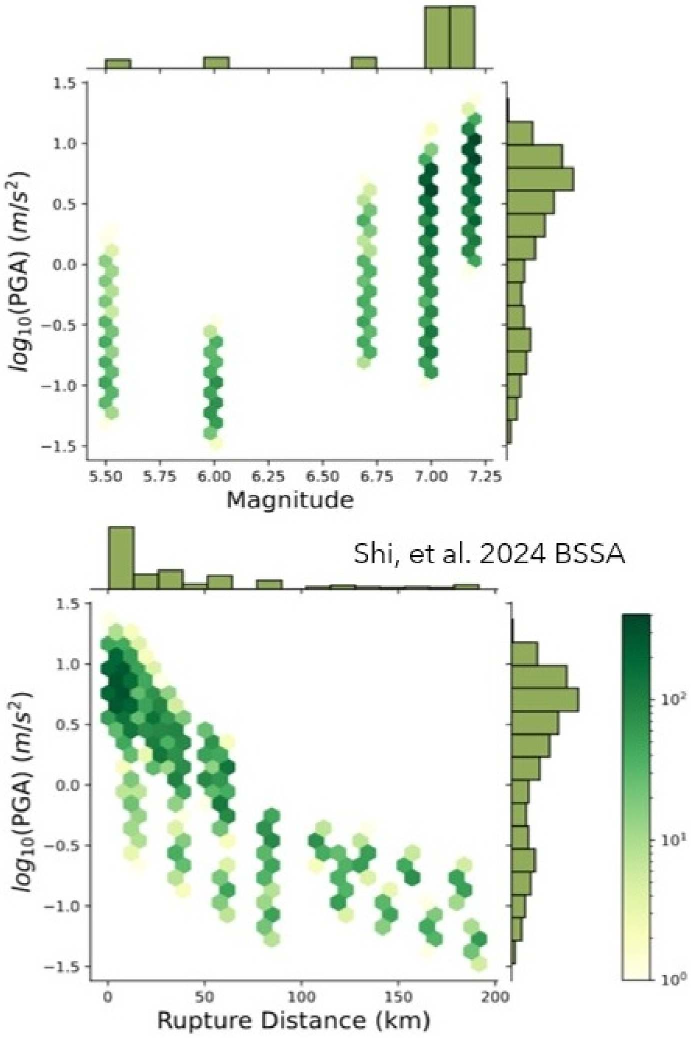


Figure 1: Caption

model was developed by the PI with SCEC funding over the last 15 years, and more information can be found in (2) and in papers and SCEC reports therein.

A schematic that depicts how the elements of soil profiles, rock outcrop ground motions and nonlinear site response were integrated to formulate the training and testing datasets is shown in Figure 3.

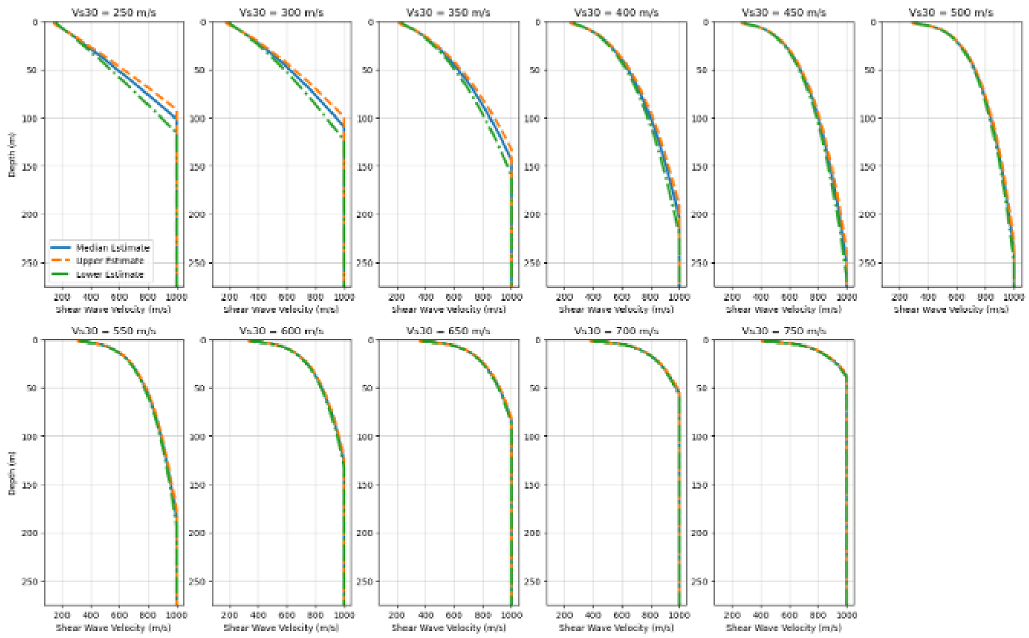


Figure 2: Velocity profiles in training and test datasets.

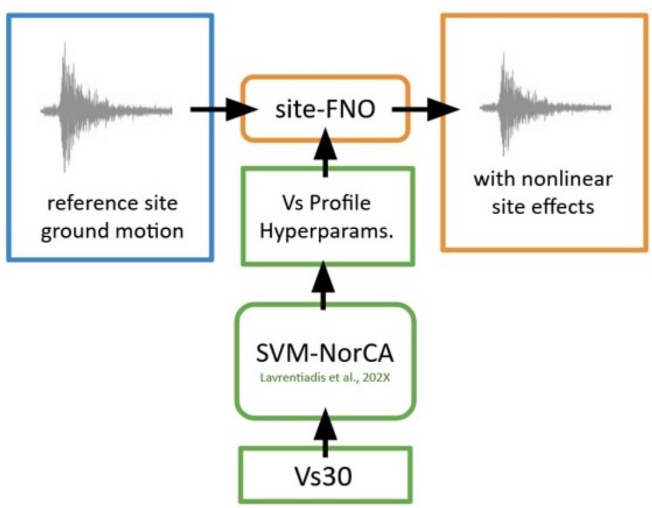


Figure 3: Schematic representation of modular set of elements that we synthesized into the training and testing datasets.

4 Model Training

The U-NO model trained in this study takes as input seven parameters, each represented as a vector of length 10,000. This arises due to a restriction of Fourier neural operators which requires all inputs to be of the same size for the Fourier decomposition. In addition, we chose to normalize the time series by the peak ground acceleration in both the input and the output for greater computational stability in light of the wide range of ground motion amplitudes in the dataset, which span two orders of magnitude. The log of input time series PGA and output time series PGA were thus included as parameters to be able to recover the input-output amplification relation of the original site response analysis.

The first parameter contains the ground motion information as an acceleration time series sampled at 100hz over a time domain of 100 seconds then normalized by peak ground acceleration (PGA). The second parameter is a constant vector of the log of the peak ground acceleration. The remaining five parameters contain a parameterized representation of the velocity profile, the values of which come from the velocity model algorithm developed by (3).

The output of the model is two vectors, the first of which represents the surface acceleration time series normalized by output PGA, and the second of which is a constant vector containing the log of the output PGA.

We trained the model using mini batch gradient descent, and the model parameters were updated after each mini batch to minimize the L2 norm error between the U-NO model output and the "true" solution generated by nonlinear site response analysis in PySeismoSoil. We split the dataset of 363,000 samples randomly, so that 80 percent (290,400 samples) were used for training and 20 percent (72,600 samples) for testing. The PGA distributions of the training and testing datasets were checked to ensure that they were reflective of each other and the original total dataset . The model was trained for 300 epochs on the Caltech Resnick High Performance Computing Cluster, which took about 25 hours to complete using one Nvidia P100 GPU. The histogram of the training and testing dataset is shown in Figure 4 while the algorithm UNO schematic is shown in Figure 5.

5 Model Evaluation

We devised a series of tests to assess the performance of the trained FNO model using the goodness-of-fit (GoF) scoring scheme from (2), which scores the fit between a measured waveform and a simulated waveform. Generally, the goal is to evaluate how well a model can capture site effects to produce simulated waveforms which closely match measured waveforms; in our case, the measured waveforms are the 1d nonlinear site response results from PySeismoSoil, and the simulated waveforms are the outputs from the trained U-NO model.

The GoF scheme rates a pair of acceleration time series according to nine different

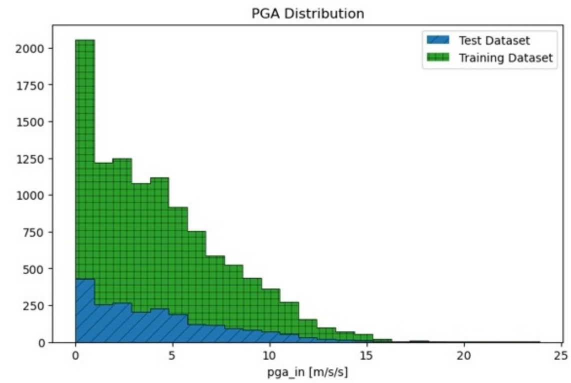
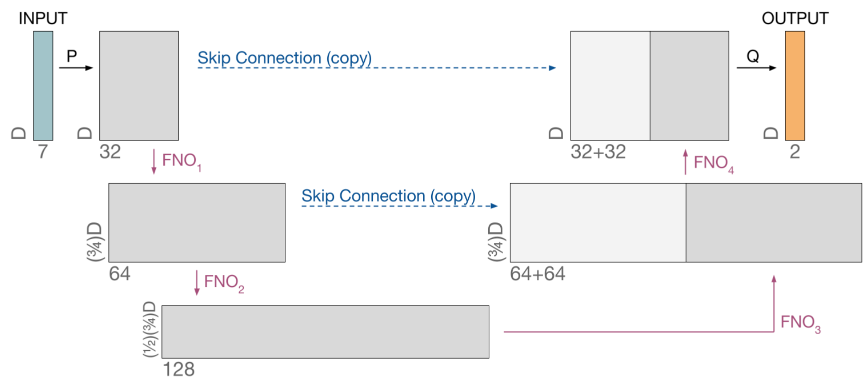


Figure 4: Histogram depicting the training and hold-off dataset PGA distributions.



Operation	Input Shape*	Output Shape*	FNO Modes	FNO Scalings
Lifting (P)	(7, 10000)	(32, 10000)		
FNO$_1$	(32, 10000)	(64, 7500)	3072	3/4
FNO$_2$	(64, 7500)	(128, 5625)	1600	1/2
FNO$_3$	(128, 5625)	(64, 7500)	1600	2/1
FNO$_4$	(64+64, 7500)	(32, 10000)	3072	4/3
Projecting (Q)	(32+32, 10000)	(2, 10000)		

* for a training dataset sampled at 100Hz (dt = 0.01s), such that D = 10,000 steps to represent the time domain spanning 100 seconds

INPUT Channels
Input time series normalized by PGA_in
ln(PGA_in)
ln(vs30)
n (velocity profile hyperbolic fit param)
a (velocity profile hyperbolic fit param)
k (velocity profile hyperbolic fit param)
ln(vs0)

OUTPUT Channels
Output time series normalized by PGA_out
ln(PGA_out)

Figure 5: Schematic of the UNO FNO architecture and details about the dataset size and model parameters.

metrics. Each scoring metric has a range from -10 to 10, where 0 represents a perfect match, negative values indicate underprediction by the model, and positive values indicate overprediction by the model. We evaluated the model at fs=100hz on the 2,200 test samples that the trained model had not seen; in addition, we augmented the testing dataset by scaling the pga of the testing dataset to evaluate the model performance at higher strains, considering the scarcity of high amplitude ground motions in the dataset compared to low-middle amplitude ground motions. We finally evaluated the model on a completely different ground motion dataset of recorded (NGA-W2) rather than simulated (BBP) time series; where we assessed both the goodness of fit of the results for 100 Hz, the resolution that the module was trained on, and for 200 Hz, a resolution higher than the module had been trained on.

Results shown in Figure 7 demonstrate, via the histogram of gof density distributions, that in all cases the module's performance is optimal (median gof is equal to zero, regardless of the strain intensity of the ground motion dataset) and the spread is never wider than ± 1. Focusing on the testing dataset and NGA datasets in the strain range where the module was tested (medium to high strain range) we see that the goodness of fit is invariant with the strain amplitude (here depicted by means of its proxy, $pgv/V_{s,30}$. The capacbility of the module to extrapolate in higher strains and higher frequencies is purely the result of the use of NO, that assumes a continuous function of the underlying PDE solution, and as such it's performance is invariant to the temporal and strain range extrapolation testing data.

The ML module will be made available to the SCEC IT team for integration in the BBP platform as an optional site module.

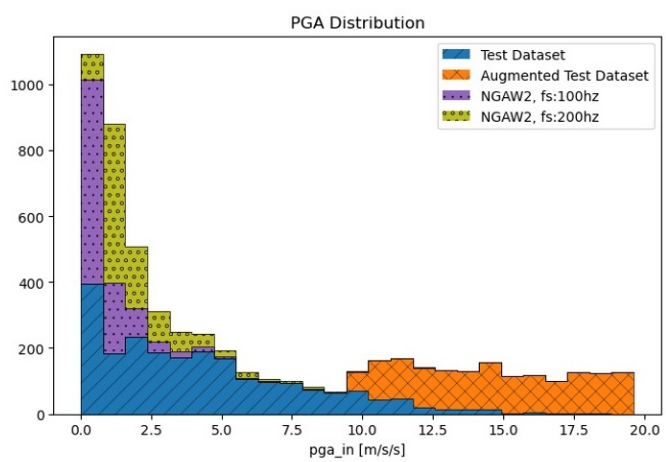


Figure 6: Histogram depicting the testing dataset PGA distributions of the BBP dataset and the NGA W2 dataset.

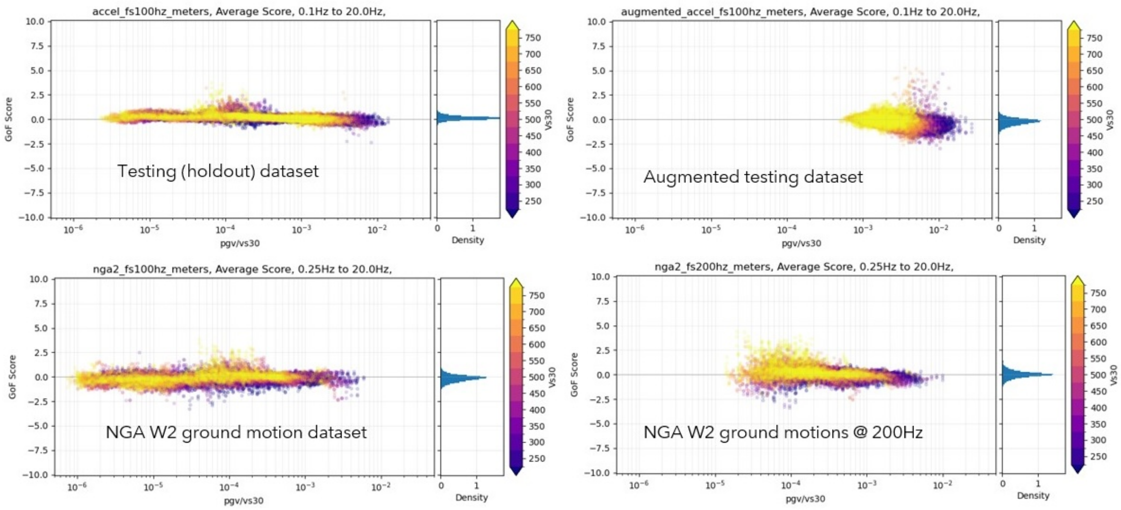


Figure 7: GOF scores of the performance of the module comparing it to the amplification computed by nonlinear site response using SeismoSoil.

References

[1] J. Shi, D. Asimaki, R. W. Graves, Sensitivity to nonlinear site amplification factors to the models used to simulate and constrain rare events, Report to USGS, Grant (G17AP00042) (2018).

[2] J. Shi, D. Asimaki, From stiffness to strength: Formulation and validation of a hybrid hyperbolic nonlinear soil model for site-response analyses, Bulletin of the Seismological Society of America 107 (3) (2017) 1336–1355.

[3] G. Lavrentiadis, E. Seylabi, F. Xia, H. Tehrani, D. Asimaki, D. McCallen, Data-driven characterization of near-surface velocity in the san francisco bay area: A stationary and spatially varying approach, Earthquake Spectra (2024) 87552930251320666.

[4] N. Matasović, M. Vucetic, Cyclic characterization of liquefiable sands, Journal of Geotechnical Engineering 119 (11) (1993) 1805–1822.