

COMPARING ARTIFICIAL NEURAL NETWORKS WITH
TRADITIONAL GROUND-MOTION MODELS FOR SMALL
MAGNITUDE EARTHQUAKES IN SOUTHERN CALIFORNIA

by

ALEXIS KLIMASEWSKI

A THESIS

Presented to the Department of Earth Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

December 2020

THESIS APPROVAL PAGE

Student: Alexis Klimasewski

Title: Comparing Artificial Neural Networks with Traditional Ground-Motion Models for Small Magnitude Earthquakes in Southern California

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Department of Earth Sciences by:

Valerie Sahakian	Advisor
Amanda Thomas	Chairperson
Brittany Erickson	Member

and

Kate Mondloch	Interim Vice Provost and Dean of the Graduate School
---------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2020

© 2020 Alexis Klimasewski

THESIS ABSTRACT

Alexis Klimasewski

Master of Science

Department of Earth Sciences

December 2020

Title: Comparing Artificial Neural Networks with Traditional Ground-Motion Models for Small Magnitude Earthquakes in Southern California

Traditional, empirical ground-motion models (GMMs) are developed by prescribing a functional form between predictive parameters and ground-motion intensity measures. Machine learning techniques may serve as a fully data-driven alternative to regression techniques as they do not require explicitly defining these relationships; however, there are few studies that assess performance of the methods side-by-side. We compare these two approaches: a mixed-effects maximum-likelihood (MEML) model, and a feed-forward artificial neural network (ANN). We develop both models on the same dataset from Southern California and test on the 2019 Ridgecrest sequence to examine model portability. We find that with our small set of input parameters, the ANN shows more site-specific predictions than the MEML model and performs better than their corresponding MEML model when applied “blind” to our testing dataset.

This thesis is co-authored by Valerie Sahakian and Amanda Thomas and is accepted at the *Bulletin of The Seismological Society of America*.

CURRICULUM VITAE

NAME OF AUTHOR: Alexis Klimasewski

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
University of Rochester, Rochester, NY

DEGREES AWARDED:

Master of Sciences, Earth Sciences, 2020, University of Oregon
Bachelor of Sciences, Physics and Astronomy, 2017, University of Rochester

AREAS OF SPECIAL INTEREST:

Seismology
Ground-Motion Modeling

PROFESSIONAL EXPERIENCE:

Research Assistant, US Geological Survey Golden, CO, June – September 2020

Graduate Employee, Department of Earth Sciences, University of Oregon,
September 2018 – December 2020

Student Researcher, US Geological Survey Menlo Park, CA, March-July 2018

GRANTS, AWARDS, AND HONORS:

Department of Earth Sciences Research Grant, University of Oregon, 2020

Raymund Graduate Fellowship, University of Oregon, 2018

Promising Scholar Award, University of Oregon, 2018

PUBLICATIONS:

Klimasewski, A., Sahakian, V., Thomas, A., Comparing artificial neural networks with traditional ground-motion models for small magnitude earthquakes in Southern

California. (under second review at the *Bulletin of the Seismological Society of America*).

Klimasewski, A., Sahakian, V., Baltay, A., Boatwright, J., Fletcher, J. B., Baker, L. M., Broadband site spectra in Southern California from source model-constrained inversion. *Bulletin of the Seismological Society of America*, (2019).109, no. 5 1878– 1889, doi: 10.1785/0120190037.

ACKNOWLEDGMENTS

I would like to thank my advisor Valerie Sahakian and my committee members Amanda Thomas and Brittany Erickson. I also want to thank my partner, Gabriel Ferragut, and my family for their support

Funding for this work was supported in part by a University of Oregon Raymond Graduate Fellowship, the USGS Earthquake Hazards Program award #G19AP00071, and SCEC award #18119.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. DATA AND METHODS.....	3
Data.....	3
Mixed-Effects Maximum Likelihood Method.....	6
ANN Method	8
III. RESULTS	12
Performance/fit of MEML models.....	12
Performance/fit of ANN models.....	12
Performance/fit between methods.....	13
Site Residuals vs. κ_0 and V_{S30}	16
Performance/fit of the models on Ridgecrest testing data	17
IV. DISCUSSION.....	20
MEML vs. ANN Methods	20
Performance with κ_0 and V_{S30}	21
Capturing region- and path-specific effects.....	22
ANN and MEML on unseen Ridgecrest data	24
V. CONCLUSION.....	26

Chapter	Page
APPENDIX: SUPPLEMENTAL TABLES AND FIGURES	28
REFERENCES CITED.....	40

LIST OF FIGURES

Figure	Page
1. Study regions with event locations and stations labeled with station name	4
2. a) Magnitude vs. distance, b) PGA vs. distance, and c) magnitude vs. PGA for the entire dataset	4
3. Details of the no site term models. AIC and standard deviation of residuals, residuals vs. distance, residuals vs. magnitude	11
4. PGA vs distance and PGA vs magnitude for station WMC	14
5. PGA vs distance and PGA vs magnitude for station SWS	15
6. Comparing average residuals per site from ANN models and mixed-effects maximum-likelihood models	17
7. Performance of the no site term models with the Ridgecrest test data and comparison with study testing data	19
S1. Contour plots of predicted PGA as a function of magnitude and distance (testing data).....	29
S2. Median observed PGA values gridded in distance and magnitude and colored by value of PGA for testing data.....	30
S3. Average residuals between observed and predicted ground motions for each site plotted vs. site.	31
S4. Average residuals between observed and predicted ground motions for each site plotted vs. site V_{S30}	32
S5. Histogram of residuals between observed and predicted PGA for training, validation, and testing data	33
S6. Details of the V_{S30} models.....	34
S7. Details of the κ_0 models	35
S8. Model predictions vs. distance and magnitude for the top 10 performing models of each configuration	36
S9. Performance of the V_{S30} models with the Ridgecrest test data and comparison with study testing data.	37

Figure	Page
S10. Performance of the κ_0 models with the Ridgecrest test data and comparison with study testing data.	38
S11. Map of events recorded on SWS with $\log_{10}(\text{PGA}) < -5$ at distances between 16 and 100 kilometers.	39

LIST OF TABLES

Table	Page
1. Mixed-Effects Maximum-Likelihood model performance, standard deviation of residuals between observed and predicted PGA for testing data	12
S1. Mixed-Effects Maximum-Likelihood model performance, standard deviation of residuals between observed and predicted PGA for testing data	28

CHAPTER I

INTRODUCTION

Empirical ground-motion models (GMMs) are one of the key components of seismic hazard assessment. Traditional models are developed by regressing existing seismic observations to obtain coefficients for a prescribed functional form, describing the relationship between ground-motion intensity measures, and other earthquake parameters. Regression models have traditionally relied on an “average” regional or global physical description for the functional form, along with observational amplification terms (i.e., for site effects). These models can include a large number of coefficients, and as the number of dependent variables increases, the regression process becomes more complicated and the risk of overfitting becomes greater. Recently, the availability of larger datasets allows for data-driven, region-specific, fully non-ergodic, and non-parametric models (Landwehr et al., 2016; Kuehn, Abrahamson, Walling, 2019; Douglas and Edwards, 2016).

In recent years, machine learning methods have become more common in seismology for a variety of applications from earthquake phase picking to seismic tomography (see Kong et al., 2018 for a recent review). Artificial neural networks (ANNs) have been used in developing a non-parametric, data-driven alternative to regression GMMs. Unlike traditional regression methods, machine learning allows for fully non-parameterized models without having to specify complex physical relationships, or fix parameters. While machine learning methods are often considered to be “black box” algorithms, they are helpful in informing human understanding of relationships between input parameters and ground motions. Machine learning techniques

like ANNs have been used to predict peak ground motions with data from Western North America (Emami, Iwao, Harada 1996; Trugman and Shearer, 2018), the Next Generation Attenuation of Ground Motion (NGA) database (Alavi and Gandomi 2011; Aagaard, 2017; Dhanya and Rachukanth, 2018), Europe (Derras, Bard, Cotton, 2014), Japan (Derras, Bard, Cotton, 2012), Central and Eastern North America (Khosravikia et al., 2018), and Northwest Turkey (Günaydın and Günaydın, 2008). While many of these papers compare their ANN GMMs to existing GMMs, they do not develop and compare their ANN model to a regression GMM developed with the exact same dataset.

The main aim of this work is to compare the performance and behavior of two methods of creating GMMs: a more traditional, regression based mixed-effects maximum likelihood (MEML) method, and a feed-forward neural network (ANN) method. Both models are simple, and created with the same input parameters and developed and tested with the same datasets. With both methods, we create models with three sets of input parameters. All sets include moment magnitude (M) and hypocentral distance (R_{hyp}). In some cases, we include a site parameter, (V_{S30} or κ_0) to test the efficacy of these parameters and to compare our ANN models to MEML models formulated both with and without a site parameter. The differences and similarities between models and methods can elucidate regional differences in observed and predicted ground motions, and inform future region-specific models.

We also test our models with an independent dataset of main events and aftershocks from the Ridgecrest sequence. We do so to evaluate the reliability of our models with unseen events and stations and in a new region of Southern California, as well as data leakage that may be present in our original dataset and models.

CHAPTER II

DATA AND METHODS

Data

We create our models using the same dataset from Southern California in order to directly compare the two methods, and subsequently evaluate the models on unseen events and stations with a dataset from the 2019 Ridgecrest sequence. Our overall Southern California dataset is from a previous paper (Klimasewski et al., 2019) where we calculated κ_0 for the 16 stations. This first dataset consists of 3,357 crustal earthquakes **M** 2.8-5.7 recorded on 16 stations in Southern California for a total of 52,297 records (Fig. 1, 2). Our stations include 13 ANZA network stations: BZN, CPE, CRY, FRD, KNW, LVA2, PFO, RDM, SMER, SND, SOL, TRO, WMC and three Southern California Seismic Network (CI) network stations: ERR, PMD, SWS (California Institute of Technology [Caltech], 1926; Berger et al., 1984; Vernon, 1989; Southern California Earthquake Data Center [SCEDC], 2013). Due to the time period of our catalog, many of our events are aftershocks of the 2010 **M**7.2 El Mayor-Cucapah earthquake (Wei et al., 2011). While Although our data consists of small magnitude events, they can help us understand region-specific seismology (Baltay et al., 2017; Sahakian et al., 2019). This is because there are no effects from a globally determined set of coefficients, which may not represent physical properties in this region, or complicated source effects from large ruptures.

We use the horizontal components of broadband velocity seismograms and cut each record to start 2 s before and 60 s after the theoretical shear-wave arrival to capture the shear-wave signal, calculated using event time, propagation distance, and a regional

average crustal velocity of 3.5 km/s. We correct for instrument response and apply an antialiasing band-pass, cosine-tapered filter from 0 to 0.001 Hz and 35 to 50 Hz. We calculate PGA by taking the gradient of the velocity time series and finding the largest absolute value in the acceleration time series.

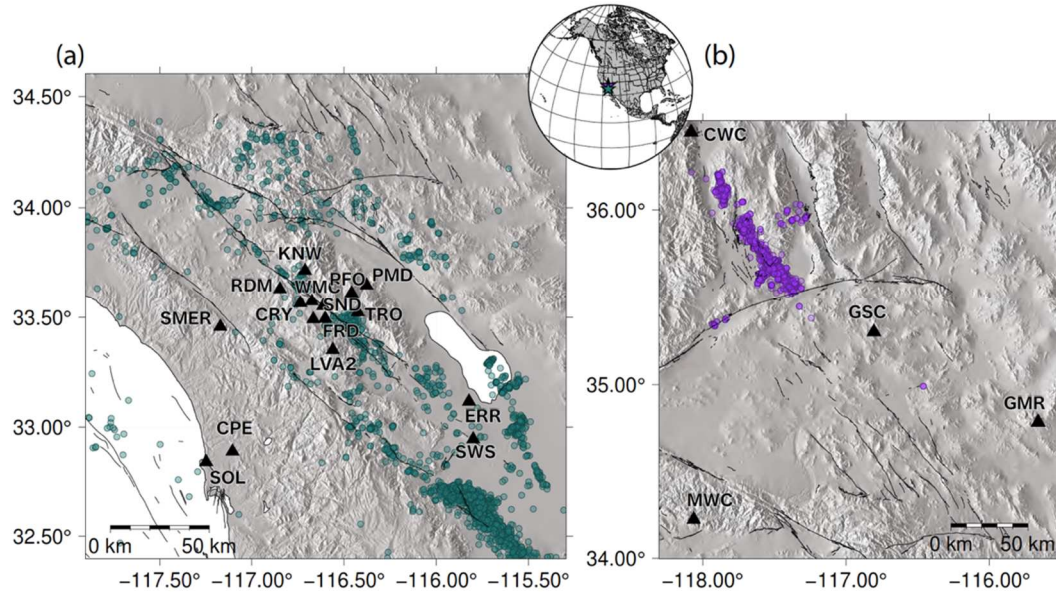


Figure 1: Study regions with event locations (dots), stations (triangles) labeled with station name, and U.S. Geological Survey mapped Holocene to Latest Pleistocene faults (lines) (USGS, 2020) a) main study region b) Ridgecrest region. The stars on the inset globe show the two regions.

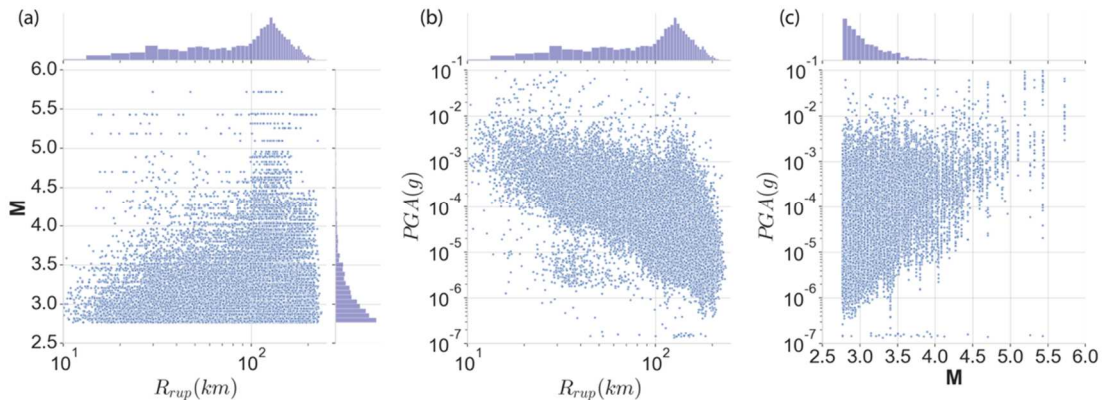


Fig 2: a) Magnitude vs. distance, b) PGA vs. distance, and c) magnitude vs. PGA for the entire dataset.

The events and stations in our Ridgecrest dataset are also in Southern California, but outside our study area (Figure 1b). We select earthquakes with $M > 2.8$ occurring in

July 2019 with hypocenter latitude between 35°N and 36.25°N and -118.9°E and -116.5°E (~2400 earthquakes), including the July 2019 **M**6.4 and **M**7.1 events (Goldberg et al., 2020). Similar to the Southern California dataset, we convert reported M_L to **M** for events smaller than **M**3.5 using the relationship from Ross et al. (2016). We chose four Southern California Seismic Network (CI) network stations MWC, CWC, GMR, and GSC with measured V_{S30} . The records are preprocessed in the same manner of our main study data. We calculate a simple signal-to-noise (SNR) ratio and select all records with $\text{SNR} > 3$. We set a maximum R_{hyp} of 235 km (maximum of study data). Site V_{S30} values are from Yong et al., (2013) and κ_0 is computed from V_{S30} with the relationship from Van Houtte et al. (2011), equation 6. Because we do not have independently calculated κ_0 for these four stations, the Ridgecrest dataset is not ideal for testing our κ_0 models; however, it should be representative to other two models since we do have measured values of V_{S30} . Our final Ridgecrest dataset consists of 1,335 events and a total of 1,894 records. Our models predict peak ground acceleration (PGA) as our dependent variable. We consider the input parameters: moment magnitude (**M**), hypocentral distance (R_{hyp}), and either V_{S30} , defined as the time-averaged shear-wave velocity in the top 30 meters of the crust, or κ_0 , the near site attenuation of high frequency energy (Anderson and Hough, 1984). We convert events with local magnitude to moment magnitude using Ross et al., (2016). Site effects on PGA are often parameterized by V_{S30} ; however, some studies have found that the inclusion of V_{S30} does not always help predict ground motions (Gallipoli and Mucciarelli, 2009; Yong et al., 2012; Derras et al., 2016, 2017; Thompson and Wald, 2016; Sahakian et al., 2018; Klimasewski et al., 2019). κ_0 , the near site attenuation of high frequency energy (Anderson and Hough, 1984), has been suggested as

a predictor of site effects on ground motions (van Houtte et al, 2014; Laurendeau et al., 2016). For this reason, we test sets of input parameters with no site term, a V_{S30} site term, and a κ_0 site term. V_{S30} for each site is reported by (Sahakian et al., 2018) from the MASW and terrain-based proxy method from Yong et al. (2012). Only four of our 16 stations have a measured V_{S30} : ERR, PFO, PMD, and SWS. κ_0 is calculated for each site with a modified version of the Andrews (1986) spectral decomposition method from Klimasewski et al. (2019).

ANNs are created by training the model on a subset of the dataset, the training data. The model hyperparameters and architecture are tuned using a separate subset of the data, the validation data. The final, tuned models are evaluated with the test data. Unlike machine learning models, regression models are typically created and evaluated with one dataset, leading to possible bias in the model. In this study, we use the same training, validation, and testing data with both methods. We randomly split all of our events into 60% training, 20% validation, and 20% testing data using the same random split for each method and set of input parameters. Splitting our data by event prevents leakage of individual event information between the three data splits. With only 16 stations in our dataset, preventing station leakage is difficult so we include records from all 16 stations in each split. Each model, regardless of method, is created using the training data, tuned with the validation data, and evaluated with the testing data.

Mixed-Effects Maximum-Likelihood method

Our regression method is a mixed-effects maximum-likelihood (MEML) technique similar to Sahakian et al. (2018), which we justify as a simple but good approximation because of similarity between datasets, and the small magnitude nature of

our events. The functional form has either six coefficients ($a_1 - a_6$), or five coefficients excluding the V_{S30} term ($a_1 - a_5$).

$$f(M, R_{hyp})_{ij} = a_1 + a_2 M + a_3 (8.5 - M)^2 + a_4 \ln(R_{hyp}) + a_5 + a_6 \left(\frac{V_{S30}}{V_{ref}} \right) \quad (1)$$

For our set of inputs with κ_0 , we chose the functional form of the κ_0 term after Laurendeau et al., 2013; Van Houtte et al., 2011. We also tested both a κ_0 scaling linearly with $\ln(\text{PGA})$ and $\ln(\kappa_0)$ scaling with $\ln(\text{PGA})$ by comparing the standard deviation of residuals on the validation data without the random site term added for various reference κ_0 values (0.02, 0.03, 0.04, 0.05, 0.06). We found that all models had similar fits, but the best model had a sixth term similar to a reference V_{S30} term,

$$a_6 \ln \left(\frac{\kappa_0}{\kappa_{ref}} \right), \quad \kappa_{ref} = 0.06 \quad (2)$$

The MEML model is a combination of the fixed effects ($f(M, R_{rup})_{ij}$) (effects and relationships that exist regardless of the dataset selection), random effects ($\delta E_i + \delta S_j$) (effects that exist due to bias from dataset selection), and aleatory residuals (δW_{ij}). Here we include the event and site as random effects, for any event i and any site j :

$$y_{ij} = f(M, R_{rup})_{ij} + \delta E_i + \delta S_j + \delta W_{ij} \quad (3)$$

In order to sufficiently compare methods, we fit the random effects (event and site terms) to the training data and add the event and site terms to the function. When comparing the validation and testing data, we report our results both without including any random effects, as well as by adding the site term. This assumes that in an unseen dataset, event terms would be unknown but site terms might be known a priori. However, if we were to

apply the model to different sites or a different geographic region, we would not add either event terms or site terms.

Sahakian et al. (2018) created a five-coefficient and six-coefficient models for small magnitude earthquakes in Southern California (different than our dataset, but with some overlap). In some of their models, coefficients are explored and prescribed to prevent unrealistic values that come from correlated terms, such as a_4 and a_5 . They found an absence of correlation between V_{S30} and site terms for all of their models, and therefore chose a five-coefficient model with no V_{S30} term and a_4 set to -1.2 as their preferred model.

We create both five and six-coefficient models with no prescribed coefficients as our initial coefficient values were realistic (Table S1). To verify, however, we created models prescribing a_4 to both -1 and -1.2 and found model fit and behavior almost identical between models with prescribed a_4 and those without. Our six-coefficient models include either V_{S30} or κ_0 as the a_6 parameter. We label our five-coefficient as $MEML_{ns}$. We label our six-coefficient models with either $MEML_{V_{S30}}$ and $MEML_{\kappa_0}$.

ANN method

Over the past decade, artificial neural networks have gained popularity as an alternative to regression methods of creating ground motion models (Alavi & Gandomi, 2011; Derras et al., 2012; Derras et al., 2014; Aagard, 2017; Dhanya and Rachukanth, 2018; Khosravikia et al., 2018). An artificial neural network is a collection of weights and biases that represent the connections between neurons, which connect an input layer of dependent variables and an output layer of independent variables. The weights and biases are initialized as small values around zero (e.g. Glorot initialization using a

gaussian distribution centered around zero), and then refined during training using optimizers (such as gradient descent). The contribution from each node is found with iterative forward and back propagation of error (Geron, 2017). The process continues for a number of epochs until the optimal configuration of weights and biases is found. We use keras with the tensorflow backend to build our models (Adabi et al., 2015). Features are often normalized or standardized so that the scale and distribution of each feature is similar. We use the standard scaler method from Scikit-learn to standardize our input features (Pedregosa et al., 2011). The mean of each feature column is removed and is scaled by unit variance as fit to the training data. Our standardized input features are in linear and the model predicts $PGA(g)$ in natural log space.

After an initial grid search with a subset of model layer and unit architectures, we determined $\sigma(x) = \tanh(x)$ was the best activation function for the input layer as well as the hidden layers. We use a linear activation function for the output layer as is common for regression models. We use the gradient descent optimizer with a learning rate of 0.01. We use a batch size of 32 to speed up training and stabilize the model. The model is compiled with a mean squared error loss function, but evaluated with a mean absolute error loss metric because we look at residuals (not residuals squared). We choose this in an attempt to have the best model fit possible. We test models with one, two, and three hidden layers in order to capture complex source, path, and site phenomena. We prevent overtraining by comparing training error to validation error and ensuring that as the training error decreases the validation error does as well.

We choose our best models using the Akaike information criterion (AIC) computed on the validation data (equation 3) where n = number of data points, m =

number of model hyperparameters (weights and biases), mse = mean squared error of validation data (Derras et al., 2012):

$$AIC = n \times (mse) + 2 \times m \quad (4)$$

The AIC represents the tradeoff between model fit and simplicity. Using the AIC ensures that between two models with similar fits, the simpler model is preferred over the more complex model.

We choose our number of hidden layers and hidden units per layer with a hyperparameter grid search. We search models with one, two, and three hidden layers with between 1 and 14 units per layer for a total of 326 models for each set of input parameters. After the initial model runs, we choose the number of epochs by finding the point when validation error reaches an asymptote. We choose 200 epochs for our models with no site and V_{S30} , and 400 epochs for our model with κ_0 . Compared to other GMMs, this is a large number of epochs, but we found that our deeper models required more training and we ensured that validation error was not increasing.

Figure 3 shows that for our model with no site term, the standard deviation of residuals plateaus with more than ~ 100 hyperparameters, while the AIC plateaus around ~ 140 hyperparameters. The lowest AIC model has 3 layers of 8, 6, and 8 hidden units, and we label it ANN_{ns} . The model with V_{S30} shows standard deviations of residuals that plateau around 0.87 and an AIC that also reaches a minimum value around 100 hyperparameters (figure S6 a,d). The final V_{S30} model, $ANN_{V_{S30}}$, has 3 layers with 10, 4, and 3 hidden units. The κ_0 model has a standard deviation of residuals that decreases with the number of hyperparameters to ~ 0.82 and an AIC that reaches a minimum around 100

hyperparameters and then slightly increases past ~ 300 hyperparameters (figure S7a,d).

The final κ_0 model, ANN_{κ_0} , has 2 layers of size 12 and 10 units.

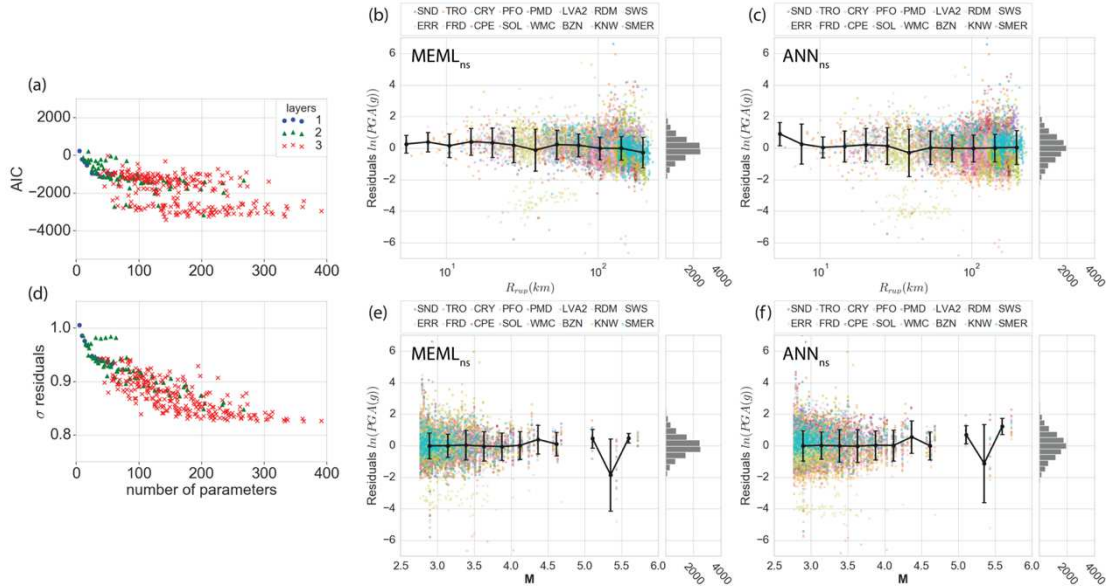


Fig 3: Details of the no site term models. a,d) AIC and standard deviation of residuals for the validation data in the hyperparameter grid search for ANN model, b) residuals vs. distance of testing data for mixed-effects maximum-likelihood model, c) residuals vs. distance of testing data for ANN model, e) residuals vs. magnitude of testing data for mixed-effects maximum-likelihood model, f) residuals vs. magnitude of testing data for ANN model.

All ANN models implement 5-fold cross-validation. The training data is split into 5 subsets by event and one instance of the model (sub-model) is trained for each subset. For each of the five sub-models, four of the data splits are used for training and the last split is used for validation, with the validation set changing with each run. We find that the sub-models perform similarly on each data split showing that our data folds and sub-models are relatively consistent. The final model is a generalized model created by averaging the five sub-model predictions to minimize overfitting (Diamantidis et al., 2000; Baykan and Yılmaz et al., 2011).

CHAPTER III

RESULTS

We compare the standard deviation of residuals between observed and predicted ground motions of the test data for all models (Table 1). Residuals are in natural log space. The distribution of residuals shows a normal shape and are centered around zero for all models (Fig. S5). The residuals show no trend with magnitude or distance (Fig. 3, S6, S7).

Table 1: Mixed-Effects Maximum-Likelihood model performance, standard deviation of residuals between observed and predicted PGA for testing data.

Model	Training	Training+ $\delta E_i + \delta S_j$	Validation	Validation+ δS_j	Testing	Testing+ δS_j
5 coeff	0.95327	0.6010	0.9925	0.8595	0.9680	0.8340
6 coeff V_{s30}	0.9472	0.5905	0.9846	0.8501	0.9615	0.8261
6 coeff κ_0	0.9523	0.5905	0.9922	0.8501	0.9673	0.8261
ANN no site		0.9374		0.9810		0.9547
ANN Vs30		0.8251		0.8715		0.8455
ANN kappa		0.7876		0.8347		0.8137

Performance/fit of the MEML models

For the Southern California dataset, we find that the three MEML models have very similar standard deviations of residuals for all data splits for fixed effects only without the random effect site term added (MEML_{ns}: $\sigma = 0.9680$, MEML_{Vs30}: $\sigma = 0.9615$, MEML _{κ_0} : $\sigma = 0.9673$). With the random effect site term, the vs30 and kappa models are nearly identical with the only difference after 4 significant digits (MEML_{ns}: $\sigma = 0.8340$, MEML_{Vs30}: $\sigma = 0.8261$, MEML _{κ_0} : $\sigma = 0.8261$).

Performance/fit of ANN models

The uncertainty of the ANN model developed and tested on the Southern California dataset shows more variation between sets of input parameters than the nearly identical MEML models. We find that ANN_{ns} has a larger standard deviation of residuals

than the two ANNs including a site parameter, and ANN_{k0} has slightly better fit than ANN_{VS30}. (ANN_{ns}: $\sigma = 0.9547$, ANN_{VS30}: $\sigma = 0.8455$, ANN_{k0}: $\sigma = 0.8137$).

Performance/fit between methods

ANN_{ns} has similar performance to the MEML models without the random effect site term added to the residuals, but with the random effect site term considered in the residuals, the MEML model has a better fit (ANN_{ns}: $\sigma = 0.9547$ vs. MEML_{ns}: $\sigma = 0.8340$). ANN_{VS30} and ANN_{k0} have significantly better performance than the MEML_{VS30} and MEML_{k0} models without the random effect site term. ANN_{VS30} and ANN_{k0} have similar uncertainty to the MEML models including the random effect site terms (ANN_{VS30}: $\sigma = 0.8455$, ANN_{k0}: $\sigma = 0.8137$, MEML_{VS30}: $\sigma = 0.8261$, MEML_{k0}: $\sigma = 0.8261$).

While the MEML method results in a coefficient for each variable parameterization, the ANN model is non-parametric and allows for much more freedom in the underlying relationships between parameters, but is also more complex and difficult for human interpretation. We examine the behavior of the ANN and MEML methods by plotting both GMM curves against distance and magnitude for each site and set of input parameters. Here we look in detail at two sites; WMC located in the center of the Anza network which represents a site in our dataset with a relatively smooth model in distance and magnitude, and SWS, a Southern California Seismic Network (CI) station located south of the Salton Sea on Superstition Mountain which shows anomalous site behavior compared with other sites (Fig. 4,5).

In general, the ANN_{ns} curves show larger deviation from the MEML models because, without a term to differentiate between sites, they are the same for every site

while the MEML_{ns} curves are linear with a constant shift up or down from the random effects site residual (Fig. 4b, 5b). The ANN_{ns} curves show changes in slope at various distances, and to some degree, magnitudes (Fig 4,5). From 40-60 km and 110 km – 140 km, the smaller magnitude ANN_{ns} curves show a shallower slope than the MEML curves. At other distances, 79 – 100 km and 140-200 km, the ANN_{ns} curves show a steeper slope than the MEML_{ns} curves.

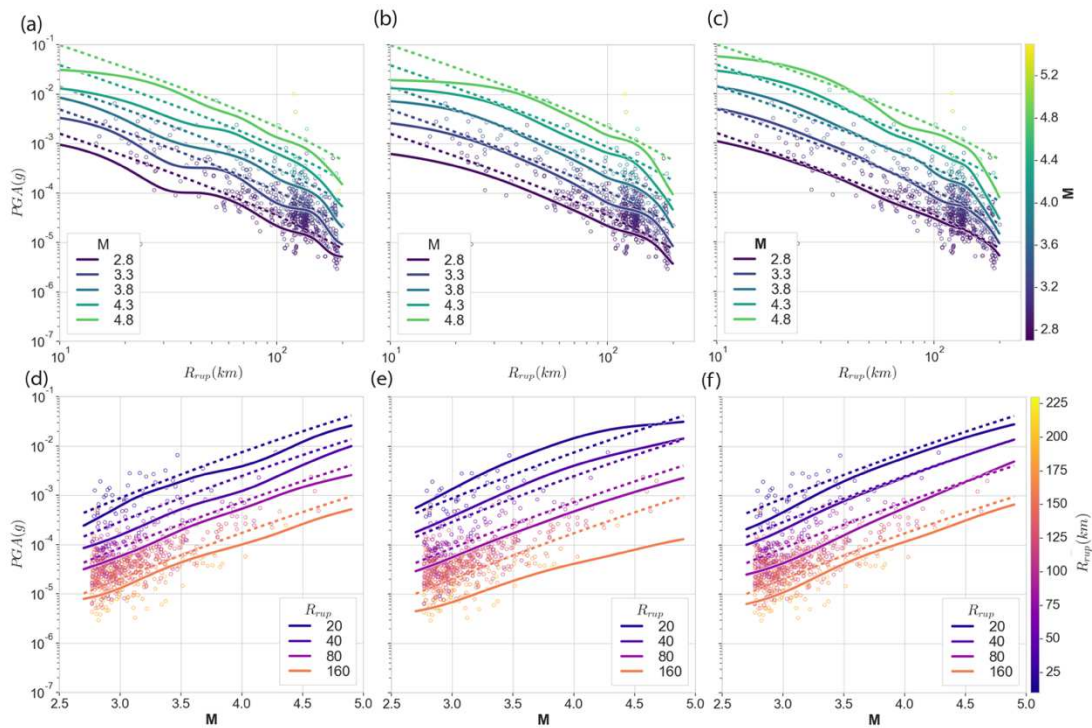


Figure 4: PGA vs distance and PGA vs magnitude for station WMC. Observations of test data (points) and model predictions for various magnitudes and distances with the mixed-effects maximum-likelihood models (dashed lines) and artificial neural networks (solid lines). a,d) with no site term, b,e) with V_{S30} site term, c,f) and with κ_0 site term.

The ANN models with κ_0 or V_{S30} show variations with each site. Since they contain a parameter to differentiate between sites, the neural network is essentially able to create a different model with varying relationships between parameters for each site. For

WMC, both ANN_{VS30} and ANN_{κ_0} (Fig. 4c,d) follow the MEML curves fairly well. The larger deviations are at the smallest and largest distances where there are limited data.

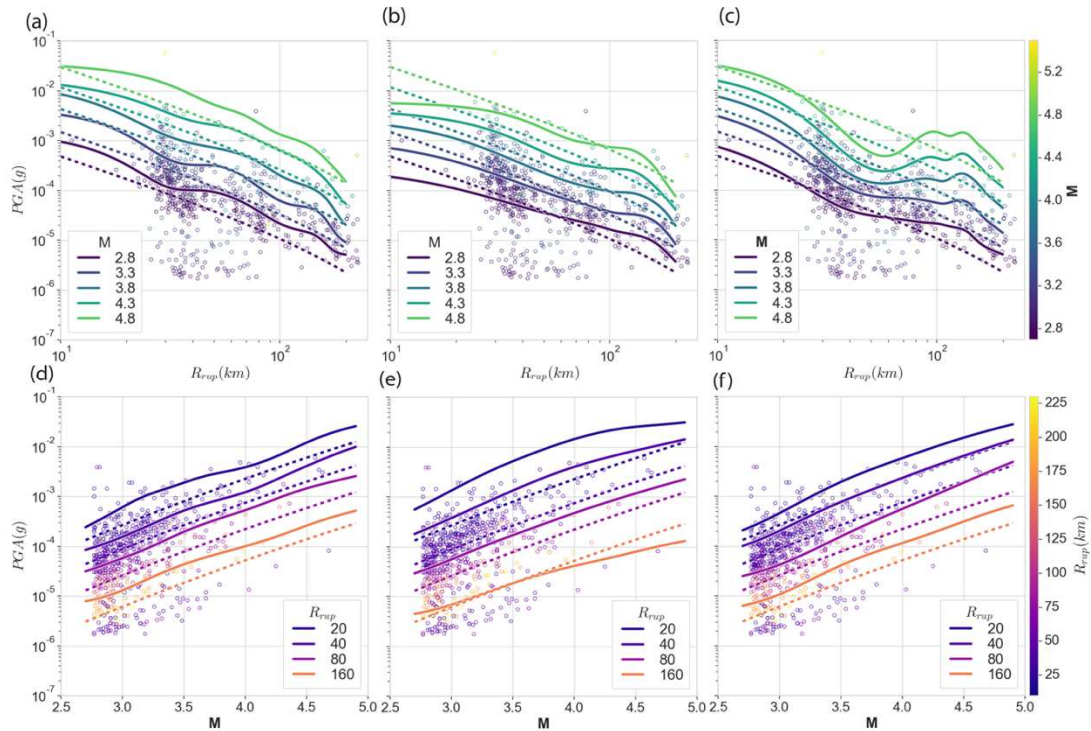


Figure 5: PGA vs distance and PGA vs magnitude for station SWS. Observations of test data (points) and model predictions for various magnitudes and distances with the mixed-effects maximum-likelihood models (dashed lines) and artificial neural networks (solid lines). a,d) with no site term, b,e) with V_{S30} site term, c,f) and with κ_0 site term.

SWS has a group of anomalously low ground motions from 25-100 km (101.4-102.0) from events from a range of azimuths, but particularly a group of El Mayor-Cucapah aftershocks. These low ground motions are not captured in the MEML models or ANN_{ns} , but they are captured to some extent in ANN_{VS30} and ANN_{κ_0} . The ANN_{κ_0} curves show strong deviation from $MEML_{\kappa_0}$ with the distance curves dipping steeply from about 20-50 km and then either sloping upward ($M=3.8, 4.3, 3.8$) or flattening from about 50-160 km ($M=2.8, 3.3$) with the exact distances varying slightly with magnitude. At distances past 160 km, the curves slope downward with similar slopes to $MEML_{\kappa_0}$.

Site residuals vs. κ_0 and V_{S30}

To understand the role of site parameters in our models, we compare our three MEML site residuals to both site parameters (V_{S30} , and κ_0). We define site residuals for MEML models as the random effect site term, δS_j , for any site j . Sahakian et al. (2018) found none of their 5 or 6-coefficient MEML models show a correlation between random effect site residuals and V_{S30} . Similarly, we find that none of our three MEML models show a correlation with V_{S30} (MEML_{ns}: $R = -0.2018$, 6-coefficient MEML _{V_{S30}} : $R = 0.0180$, MEML _{κ_0} : $R = -0.2405$) (Fig. S3). While a few studies have implemented random effects into their ANN method in order to capture event and site residuals (Derras et al., 2014; Derras et al., 2016), we simply compute site residuals for the ANN models as the average of the residuals for all records at a given station, to represent its un-modeled contributions to ground-motions. The MEML random effect site terms are inverted for simultaneously with the model coefficients (fixed effects) and are technically included in model predictions. The ANN site residuals are not included in the predictions because they are calculated after the models are finalized and not during the model development step like the random effect residuals. We also find no evidence of a correlation between the average site residuals from our three ANN models and V_{S30} (ANN_{ns}: $R = -0.1959$, ANN _{V_{S30}} : $R = -0.0192$, ANN _{κ_0} : $R = -0.3815$) (Fig. S3).

Klimasewski et al. (2019) found a correlation between 5-coefficient site residuals of Sahakian et al. (2018) and κ_0 . The correlation suggested that including κ_0 could help improve the model. However, for the site residuals computed with our data, we see no correlation between κ_0 and the MEML site residuals (MEML_{ns}: $R = -0.0866$, MEML _{V_{S30}} : $R = -0.1978$, MEML _{κ_0} : $R = -0.0070$), which implies no correlation between κ_0 and each

sites' contributions to the observed ground-motion. We also see no correlation between κ_0 and any of the ANN models site residuals (ANN_{ns}: R= -0.0857, ANN_{VS30}: R= -0.0293, ANN κ_0 : R= -0.0371) (Fig. S4).

Between ANN vs. MEML models with the same input parameters, we compare residuals per site (Fig. 6). We find strong correlation between residuals from MEML_{ns} and ANN_{ns}. Residuals are nearly one-to-one (Fig. 6a: R=0.9978, p=0.0000, power = 0.9612). Site residuals are quite consistent for all three MEML models, but including a site input parameter to the ANN models results in a narrower distribution of site residuals. The models with site parameters show site residuals that are also correlated, but less strongly (Fig. 6b, ANN_{VS30}: R=0.6292, p=0.0090, power = 0.6531; Fig. 6c, ANN κ_0 : R=0.8216, p=0.0001, power = 0.8665).

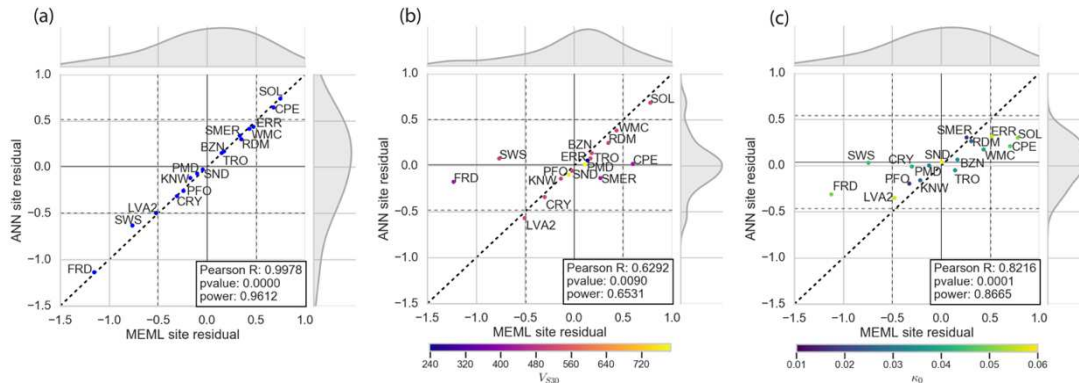


Figure 6: Comparing average residuals per site from ANN models and mixed-effects maximum-likelihood models for a) models with no site term b) models with V_{S30} c) models with κ_0 .

Performance/fit of the models on Ridgecrest testing data

We find that the mixed-effects maximum-likelihood models perform very well with the Ridgecrest earthquakes and aftershock dataset. All three MEML models have better fit on the Ridgecrest dataset than the study testing dataset ($\sigma_{\text{Ridge}} \approx 0.71$). Figures 7, S6, and S7, show histograms of both ANN and MEML models on the testing data and

Ridgecrest data. ANN_{ns} demonstrates a better fit on Ridgecrest data than the testing dataset ($\sigma_{\text{Ridge}} = 0.8195$ vs. $\sigma_{\text{Test}} = 9619$). ANN_{VS30} exhibits a worse fit on Ridgecrest data than testing data ($\sigma_{\text{Ridge}} = 1.1543$ vs. $\sigma_{\text{Test}} = 8556$). ANN_{k0} has a very similar fit on Ridgecrest data and testing data ($\sigma_{\text{Ridge}} = 8195$ vs. $\sigma_{\text{Test}} = 8112$). The residuals vs. magnitude and distance plots show some trends with distance, but it may be related to only 4 stations and each station has records from a narrow range of distances. The MEML residuals from the **M**6.4 and **M**7.1 earthquakes are centered around zero while the ANN models have positive residuals except for CWC (Fig. 7g,h, S6g,h, S7g,h). Because the ANN is trained on smaller magnitude earthquakes, the model appears to be underpredicting PGA for the larger earthquakes outside the training domain. Finally, we find an average of zero for bins of residuals for earthquakes of **M**<3.5, indicating that the local to moment magnitude conversion of Ross et al. (2016) is sufficiently applicable in this region.

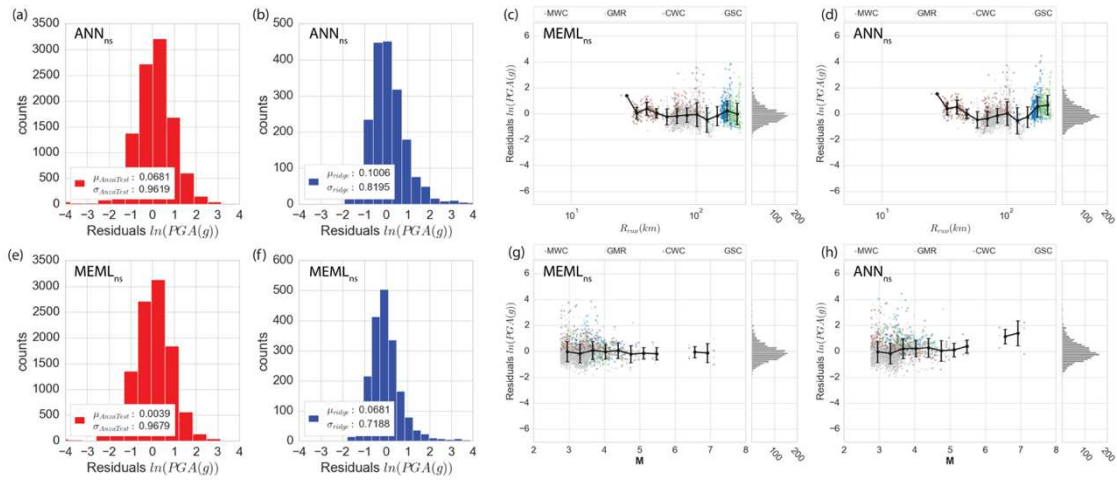


Fig 7: Performance of the no site term models with the Ridgecrest test data and comparison with study testing data. Histogram of residuals between observed and predicted PGA for a) study testing data with ANN no site model, e) study testing data with mixed-effects maximum-likelihood with no site term model, b) Ridgecrest data with ANN no site model, f) Ridgecrest data with mixed-effects maximum-likelihood with no site term model. c) residuals vs. distance and g) residuals vs. magnitude of Ridgecrest data for mixed-effects maximum-likelihood no site term model, d) residuals vs. distance and h) residuals vs. magnitude of Ridgecrest data for ANN no site term model.

CHAPTER IV

DISCUSSION

MEML vs. ANN Methods

We first compare, with regards to performance (standard deviation), the MEML models to the ANN models with the same input parameters. Between models with no site input parameter, the MEML method performs much better than the ANN. This is likely because the MEML includes the random effects site residual, while the ANN has no way to differentiate between sites. Interestingly, MEML_{ns} without the site residuals added (without considering the effects of the random effect site term when computing model residual standard deviations) has very similar performance to the ANN_{ns} model ($\sigma = 0.9680$ and $\sigma = 0.9547$; Table 1). ANN_{VS30} and ANN _{κ 0} have significantly smaller uncertainties than MEML_{VS30} and MEML _{κ 0} without the random effect site terms included, but including the random effect site terms performance is comparable. Next, we compare the resulting distributions of site residuals, with respect to how well the ANN vs. MEML methods integrate site properties into their ground-motion estimations. Site residuals between the two methods correlate well, between models created with the same parameters (i.e., between MEML_{ns} and ANN_{ns}), but to varying degrees (Fig. 6a). The random effect site terms from MEML_{ns} are very close in value to the ANN_{ns} site residuals, despite being calculated differently. While we add the random site term to the MEML predictions, we do not add it to the ANN predictions. Between our three MEML models, the distribution of site residuals is consistent (see x-axis histograms on Fig. 6a-c). The site residuals from MEML_{VS30} have a narrower distribution than the MEML_{ns} and MEML _{κ 0} residuals (Fig. 6). The site residuals resulting from

MEML κ_0 are less correlated with the site residuals from ANN κ_0 ; site residuals between MEML V_{S30} and ANN V_{S30} are even less correlated. This indicates that including V_{S30} captures some of the site effects on PGA so that the site uncertainty is decreased for the ANN model, but the MEML model stays the same. The MEML site residuals are systematically slightly higher than the ANN site residuals for the models with V_{S30} . MEML κ_0 site residuals correlate well with the ANN κ_0 site residuals, but follow less of a one-to-one trend compared to the models with no site term.

Together, these observations indicate that without explicit identification of the site within the ANN model (i.e., one-hot encoding, Potdar et al., 2017), the ANN models learn relationships between these site parameters, and ground-motion effects. This suggests that the non-parameterization in the ANN could be promising for its predictive power, and for better understanding the relationship between site parameters and ground motions when compared to the current prescribed form between PGA, and κ_0 or V_{S30} (equations 1, 2). However, with our relatively small number of stations, the κ_0 model may effectively be differentiating between sites by unique κ_0 values instead of finding a physical relationship between κ_0 and site effects. The ANN models are able to capture linear and nonlinear site effects while the MEML models must follow the prescribed relationship $\ln(\text{PGA}) \propto \ln(V_{S30})$ or $\ln(\kappa_0)$ (equation 1).

Performance with κ_0 and V_{S30}

Including an input site parameter greatly improves the ANN models, but does not appreciably affect the fit of the MEML model. However, there is no evidence of a correlation between either input site terms and the site residuals for any of our models (Fig. S3, S4). This means that differentiating between sites in the ANN is important, but

that the physical relationship given by a particular site parameter (V_{S30} , κ_0) in the functional form is not well-constrained or defined.

A previous study found that κ_0 correlated well ($R = -0.6128$, $p\text{-value} = 0.0116$, $\text{power} = 0.6302$) with MEML site residuals (Klimasewski et al., 2019). Sites with larger κ_0 values generally had negative site residuals indicating that the more attenuating sites tended to have lower ground motions than predicted, while sites with smaller κ_0 values tended to have positive site residuals indicating that the less attenuating sites had larger observed ground motions than predicted. The same site residuals showed no evidence of a correlation with V_{S30} values, although since most V_{S30} values for these sites are proxy values, they may not represent the actual V_{S30} values at the sites.

The lack of a correlation between site residuals and both V_{S30} and κ_0 suggests that we may not have the correct parameterization in the MEML model or that κ_0 and V_{S30} are not important in predicting for PGA in our data set, but would be important to Fourier amplitude spectra (FAS, Abrahamson and Bayless, 2018) or other independent ground-motion parameters.

Capturing region- and path-specific effects

In GMMs, the path is often represented only by geometric spreading term and anelastic attenuation terms. Epistemic path uncertainty is a significant contributor to overall uncertainty (Lin et al., 2010; Kuehn, Abrahamson, Walling, 2019). Fortunately, path effects can be studied with data from small magnitude events and applied in the prediction of ground motions for larger magnitude events (Baltay et al., 2017; Abrahamson et al., 2018; Sahakian et al., 2019). Capturing spatial variability in path effects is important for non-ergodic models. Deviations in the ANN models from the

MEML predictions communicate information about potential path and site effects that are not captured in the functional form of the MEML model. As seen in figures 3 and 5, our ANN models are very dataset specific; however, they are not over-trained since the training, validation, and testing data all show consistent fits. Our ANN GMMs are not models that would be used in practice for hazard applications, because they are specific to our narrow distribution of events and stations both in space and time. Figures 4 and 5 show that in domains where we do not have a lot of data (large magnitudes and close distances), model curves are not consistent. In domains where we have many data points, such as El Mayor-Cucupah aftershocks recorded on Anza stations (distances ~112-140km) we see shallower slopes between model predicted PGA and distance in many of the ANN models. This could be caused by a common source, path, or site effect.

We seek to use the ANN models to 1) compare fit to the more traditional regression based MEML method 2) harness the complexities found in the ANN models to learn about path- and site-specific effects in our study region, and their potential contributions to developing non-ergodic models.

The dip seen in the SWS curves (Fig. 5b,c) for the ANN models with a site term is not seen at any other stations. These low ground motions are seen in the training, validation, and testing data. It is not obvious what causes these low ground motions. SWS has an average V_{S30} (measured) and κ_0 . Figure S11 shows all earthquakes recorded on SWS with PGA less than 10^{-5} m/s² with event to station distances between 16 and 100 km. Most of the lower than expected ground motions recorded on SWS are small magnitude El Mayor-Cucupah aftershocks. They are relatively shallow, and most events are located at a less than 10 km depth. (Fig. S11). They could be due to raypaths traveling

in shallow metamorphosed ancient Lake Cahuilla and Gulf of California rift sediments, with lower velocity, and/or more attenuating (Hauksson et al., 2006; Han et al., 2016; Sahakian et al., 2016). While we do not further explore these observations, we expect that they may be useful for informing studies that seek to incorporate physical properties into fully non-ergodic ground-motion models (Baltay et al., 2017; Sahakian et al., 2019), for both seismic hazard and earthquake early warning applications.

ANN and MEML on unseen Ridgecrest data

While our results have demonstrated that the ANN appears to be a better model for our dataset with respect to model fit, as well as representation of non-ergodic behavior not represented in the MEML model, several questions remain regarding the model applicability to new regions with different crustal properties, new stations (for which there were no random effect site term to include in estimating ground-motion), and earthquake magnitudes outside the original models' range. To test this, we evaluate our models on an independent dataset of Ridgecrest earthquakes. For this dataset, we have no random effects terms in the MEML because these are new events and stations. We find that the MEML models have a better fit than the ANNs, but ANN_{ns} and ANN_{k0} perform well on the new data. ANN_{VS30} has a worse fit than $MEML_{VS30}$. V_{S30} may not accurately represent site effects for these four sites. Interestingly, ANN_{k0} has much better performance on the new data than ANN_{VS30} despite the kappa values being a function of V_{S30} . We note, however, that we apply hypocentral distance (R_{hyp}) instead of closest distance to rupture (R_{rup}) in our original models. This was a valid assumption for the original, small-magnitude Southern California dataset, however the larger events in the Ridgecrest sequence do not behave as point sources. The difference between R_{hyp} and

R_{rup} is Between 0.8% and 11.7% difference for the **M**6.4 earthquake and between 11.4% and 44.8% difference for the **M**7.1 earthquake (Goldberg et al., 2020). This distance assumption may alter and bias our results in applying these models, and future studies should incorporate more representative distance metrics such as R_{rup} .

These results indicate that while the ANN models are generally similar or outperform the MEML models on the original training dataset, they may not be applicable in a new region without additional constraints. For example, our study has demonstrated that the ANN seems to learn more complex, non-ergodic path and site effects not represented in the MEML functional form. However, when applied to a new region, these original effects learned by the ANN are no longer applicable. ANN models developed with a more balanced dataset of stations and events may be more portable to new regions. Although advantageous for region-specific studies in areas with many seismic observations, an ANN approach would likely be deficient for regions with a dearth of seismic data to use in a training dataset. In these cases, a MEML model may be preferred, or numerical simulations of earthquakes could be used to fill this “data gap” in an ANN.

CHAPTER V

CONCLUSION

This study compares two methods of creating GMMs for small magnitude earthquakes in Southern California: (1) a more traditional, statistically-based, maximum likelihood mixed effects regression, and (2) a machine learning, nonparametric artificial neural net. This work shows that the methods perform similarly on identical testing data, but that the ANN may learn more detailed behavior without the need to predefine relationships between parameters. When applied to an unseen dataset, the MEML model generally outperforms the ANN model, indicating that the detailed regional behavior learned by the ANN model is not applicable to new regions without additional constraints.

Our models only include two or three input parameters, but our results indicate that machine learning methods will be more effective for datasets with many more dependent parameters – particularly those that are less physics-based in their prescribed functional form (i.e., faulting type terms, hanging wall terms, etc.). Studies such as Derras et al., (2014) and Aagaard et al., (2017) show that including more input variables generally increases model performance. Unlike a regression model where the functional form and potential tradeoffs must be determined before include a new input parameter, it is easy to add more parameters to an ANN model compared to a regression method.

In the future, using one-hot encoding to differentiate between sites may help illuminate and quantify the physical basis that other site parameters contribute to predicted ground motions, as well as expanding to larger databases with a wider variety of magnitudes. Including further dependent parameters and intensity measures beyond

PGA, such as Fourier amplitude spectrum, will also be important for evaluating the effectiveness of machine learning models (Abrahamson & Bayless, 2018).

APPENDIX

SUPPLEMENTAL TABLES AND FIGURES

This appendix contains a table of the mixed-effects maximum-likelihood model coefficients. Figures S1 and S2 show contour and gridded plots of PGA for our test dataset. Figures S3-S5 show the distribution of model residuals and with κ_0 , and V_{S30} . Figures S6 and S7 give details of the V_{S30} and κ_0 models hyperparameter grid search and both MEML and ANN residuals vs. magnitude and distance (similar figure 3 for the no site term models). Figure 8 is a plot of the top 10 best performing models as determined by the AIC for the three neural network models. Figures S9 and S10 are similar to figure 7, but they show the Ridgecrest test for the V_{S30} and κ_0 models.

Table S1: Table with mixed-effects maximum-likelihood model coefficient values.

coefficient	Estimate 5-coefficient	Estimate 6-coefficient V_{S30}	Estimate 6-coefficient κ_0
Intercept	-1.9649	-2.1551	-2.0166
m	0.8926	0.8926	0.8926
m^2	-0.1247	-0.1247	-0.1247
$\ln(R)$	-1.4353	-1.4353	-1.4353
R_{rup}	-0.0056	-0.0056	-0.0056
κ_0 or V_{s30}	-	-0.4520	-0.1063

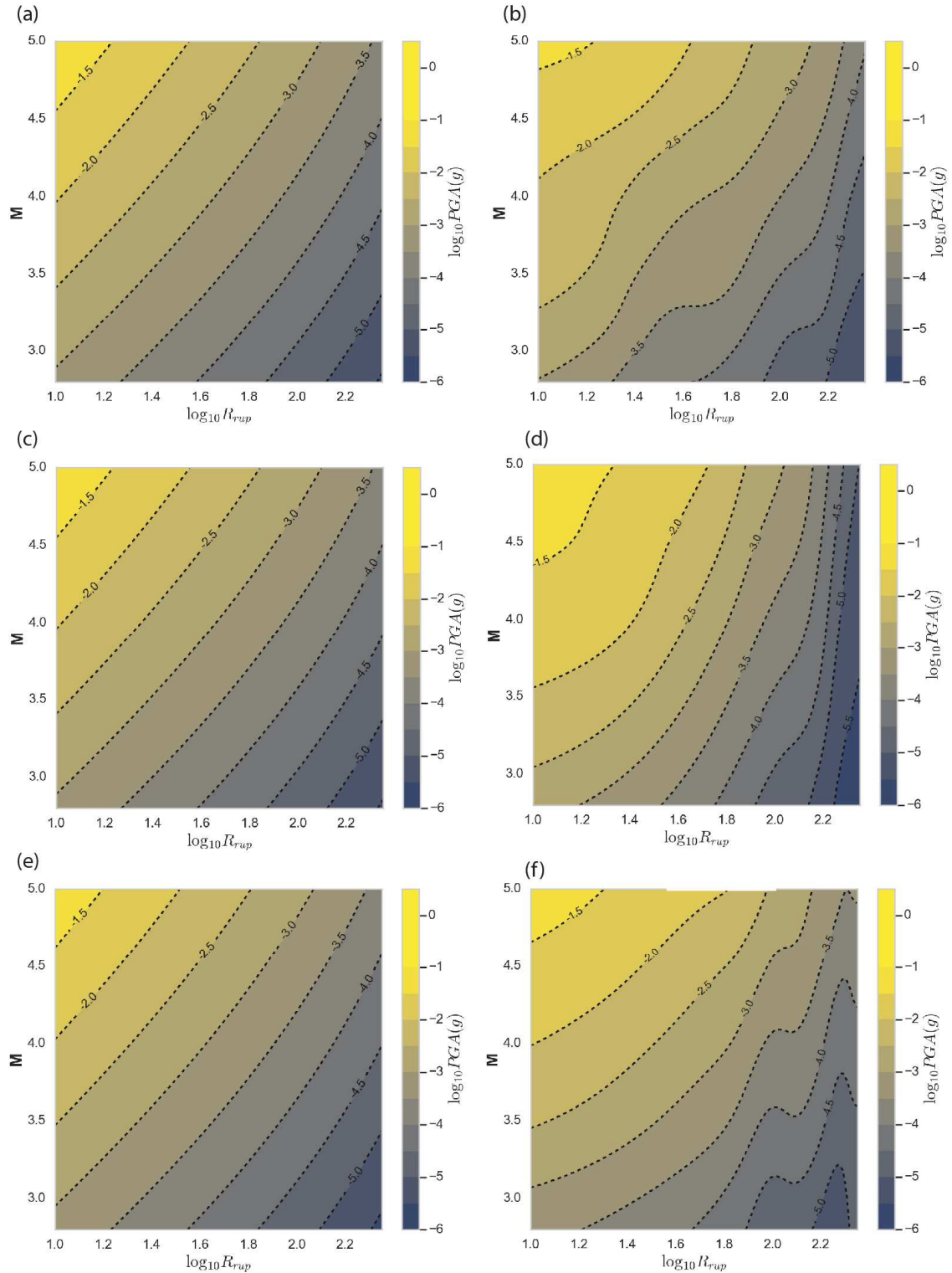


Figure S1: Contour plots of predicted PGA as a function of magnitude and distance (testing data) for a) 5-coefficient mixed-effects maximum-likelihood model b) artificial neural network with no site term c) 6 coefficient V_{S30} mixed-effects maximum-likelihood model d) artificial neural network with no site term with V_{S30} e) 6-coefficient κ_0 mixed-effects maximum-likelihood model f) artificial neural network with no site term with κ_0 .

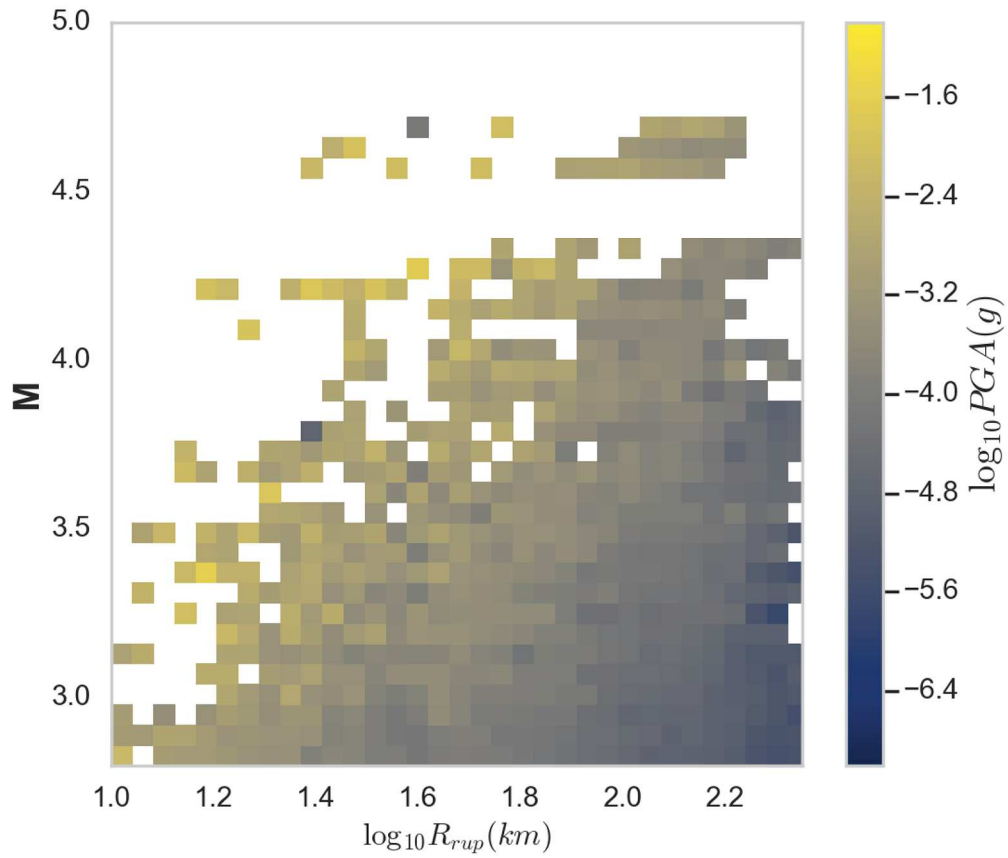


Figure S2: Median observed PGA values gridded in distance and magnitude and colored by value of PGA for testing data.

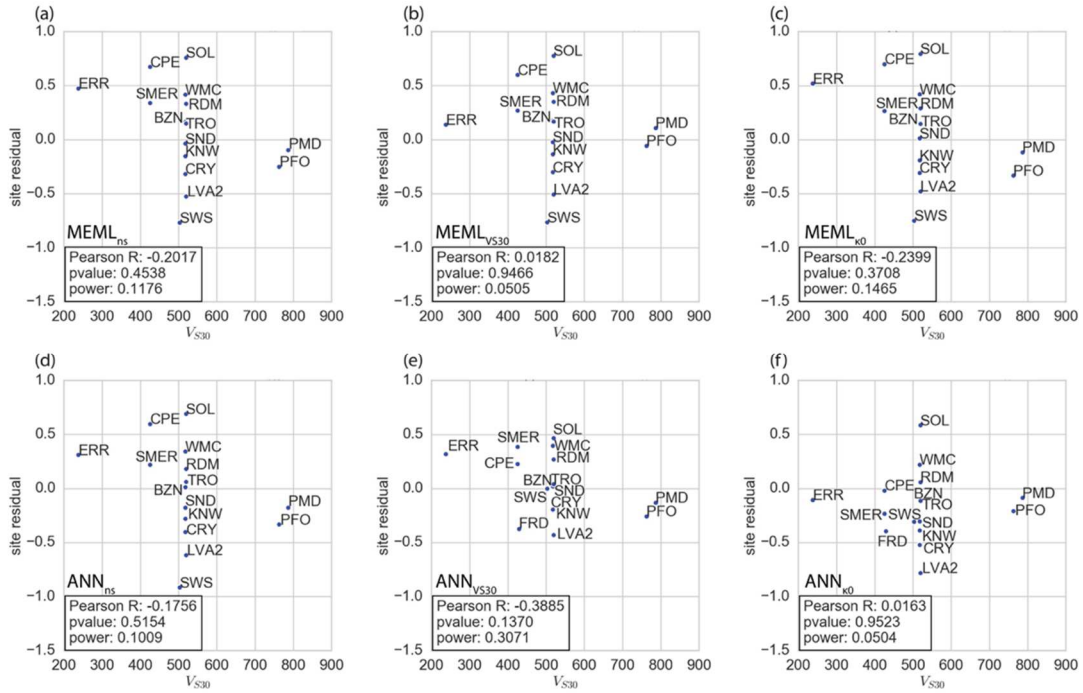


Figure S3: Average residuals between observed and predicted ground motions for each site plotted vs. site V_{S30} for mixed-effects maximum-likelihood models with a) 5-coefficient model with no site term b) 6-coefficient model with V_{S30} site term c) 6-coefficient model with κ_0 site term d) ANN model with no site term e) ANN model with V_{S30} site term f) ANN model with κ_0 site term.

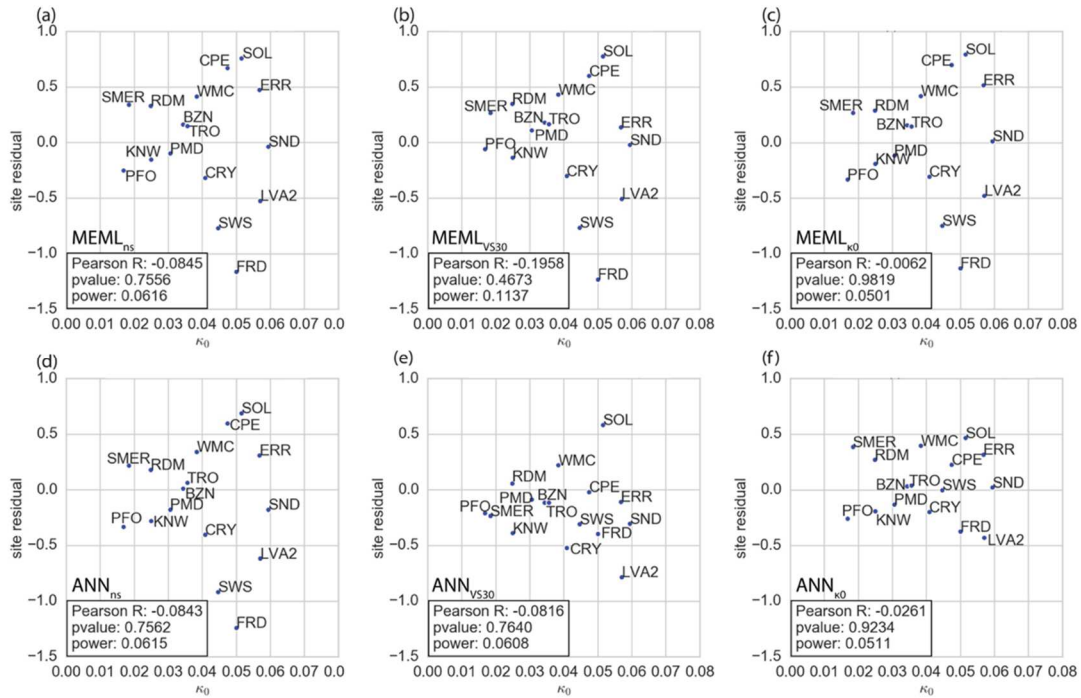


Figure S4: Average residuals between observed and predicted ground motions for each site plotted vs. site κ_0 for mixed-effects maximum-likelihood models with a) 5-coefficient model with no site term b) 6-coefficient model with V_{S30} site term c) 6-coefficient model with κ_0 site term d) ANN model with no site term e) ANN model with V_{S30} site term f) ANN model with κ_0 site term.

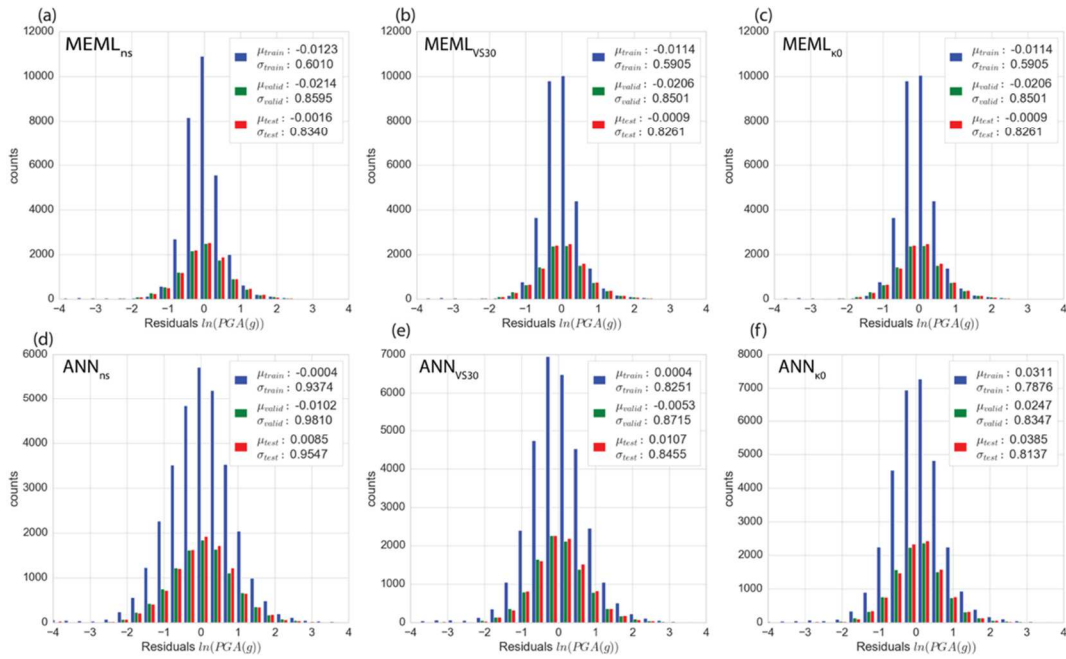


Figure S5: Histogram of residuals between observed and predicted PGA for training, validation, and testing data for a) 5-coefficient mixed-effects maximum-likelihood model b) 6-coefficient V_{S30} mixed-effects maximum-likelihood model c) 6-coefficient κ_0 mixed-effects maximum-likelihood model d) Artificial Neural Network with no site parameter e) Artificial Neural Network with V_{S30} site parameter f) Artificial Neural Network with κ_0 site parameter.

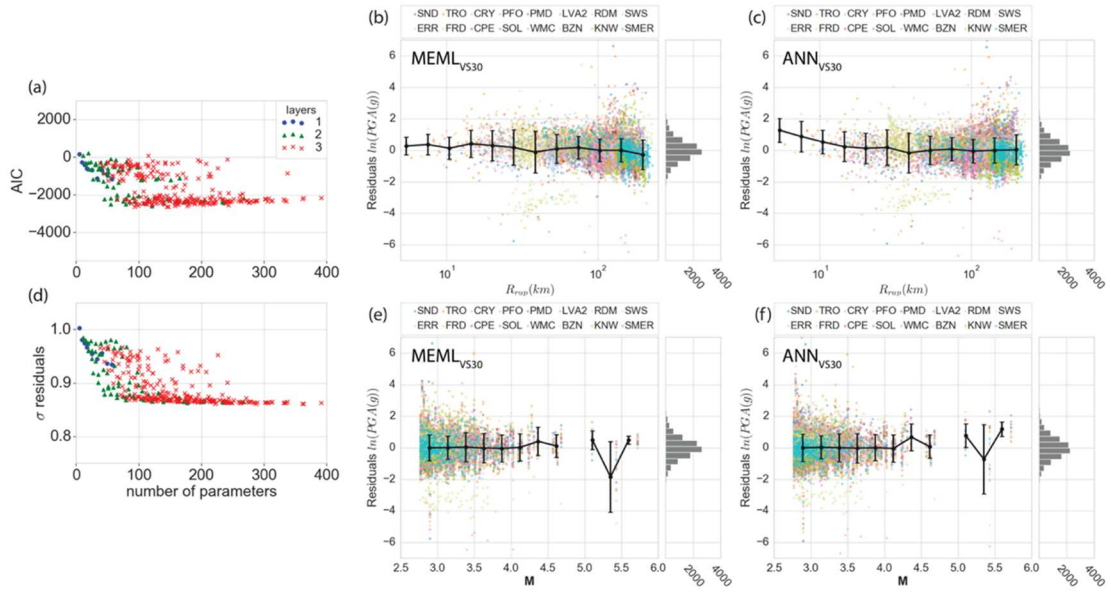


Figure S6: Details of the VS₃₀ models. a,d) AIC and standard deviation of residuals for the validation data in the hyperparameter grid search for ANN model, b) residuals vs. distance of testing data for mixed-effects maximum-likelihood model, c) residuals vs. distance of testing data for ANN model, e) residuals vs. magnitude of testing data for mixed-effects maximum-likelihood model, f) residuals vs. magnitude of testing data for ANN model.

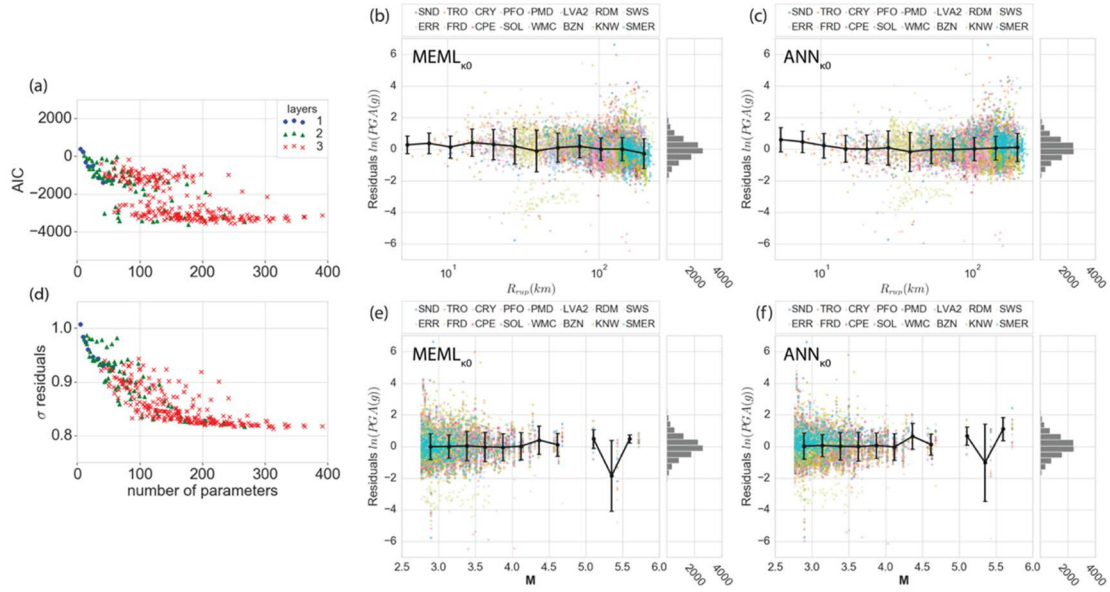


Figure S7: Details of the κ_0 models. a,d) AIC and standard deviation of residuals for the validation data in the hyperparameter grid search for ANN model, b) residuals vs. distance of testing data for mixed-effects maximum-likelihood model, c) residuals vs. distance of testing data for ANN model, e) residuals vs. magnitude of testing data for mixed-effects maximum-likelihood model, f) residuals vs. magnitude of testing data for ANN model.

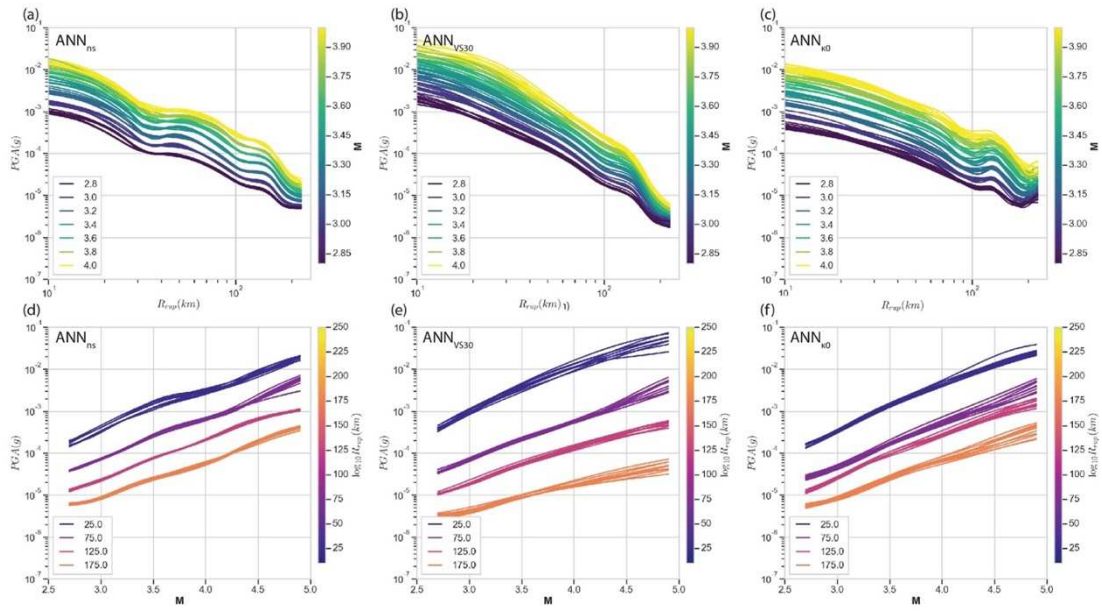


Figure S8: Model predictions vs. distance and magnitude for the top 10 performing models of each configuration a,d) no site parameter b,e) V_{S30} site parameter c,f) κ₀ *note consistency (small variance) between models and narrow spread.

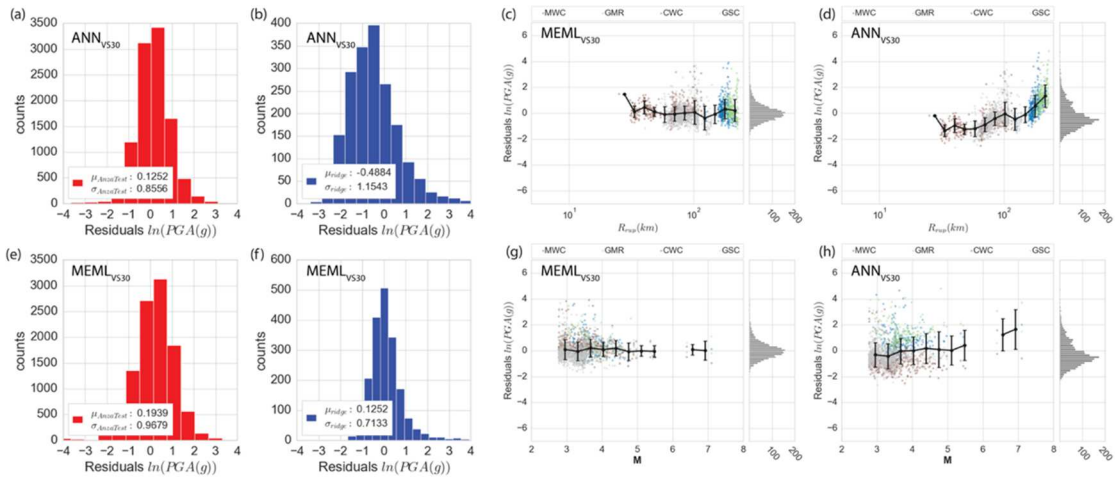


Fig S9: Performance of the V_{S30} models with the Ridgecrest test data and comparison with study testing data. Histogram of residuals between observed and predicted PGA for a) study testing data with ANN V_{S30} model, e) study testing data with mixed-effects maximum-likelihood V_{S30} model, b) Ridgecrest data with ANN V_{S30} model, f) Ridgecrest data with mixed-effects maximum-likelihood V_{S30} model. c) residuals vs. distance and g) residuals vs. magnitude of Ridgecrest data for mixed-effects maximum-likelihood V_{S30} model, d) residuals vs. distance and h) residuals vs. magnitude of Ridgecrest data for ANN V_{S30} model.

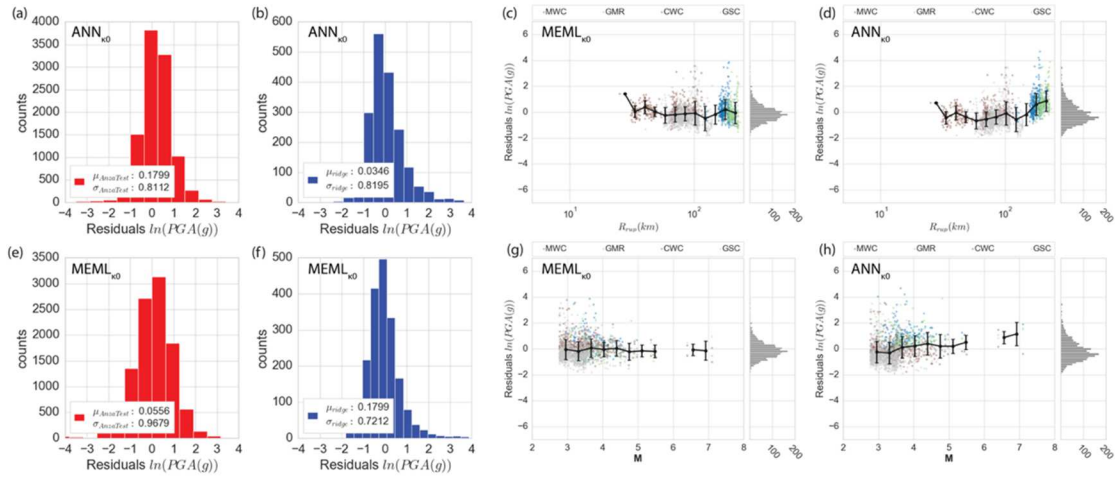


Fig S10: Performance of the κ_0 models with the Ridgecrest test data and comparison with study testing data. Histogram of residuals between observed and predicted PGA for a) study testing data with ANN κ_0 model, e) study testing data with mixed-effects maximum-likelihood κ_0 model, b) Ridgecrest data with ANN κ_0 model, f) Ridgecrest data with mixed-effects maximum-likelihood κ_0 model. c) residuals vs. distance and g) residuals vs. magnitude of Ridgecrest data for mixed-effects maximum-likelihood κ_0 model, d) residuals vs. distance and h) residuals vs. magnitude of Ridgecrest data for ANN κ_0 model.

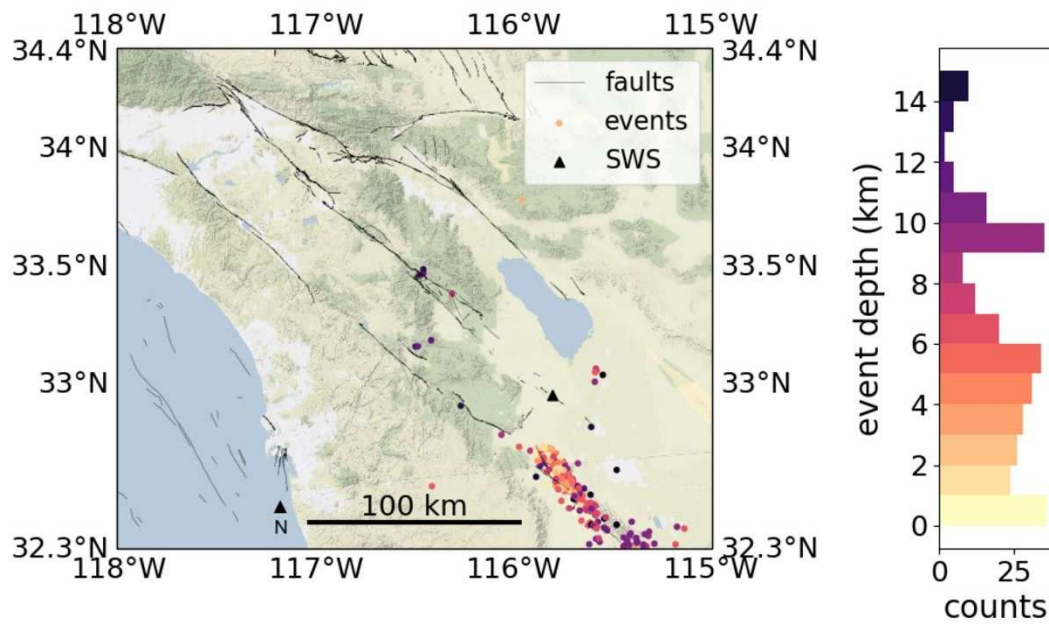


Figure S11: Map of events recorded on SWS with $\log_{10}(\text{PGA}) < -5$ at distances between 16 and 100 kilometers.

REFERENCES CITED

- Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous distributed systems, <https://arxiv.org/pdf/1603.04467>
- Aagaard, B. T. (2017). A Ground-Motion Prediction Equation for California Constructed Using an Artificial Neural Network. Poster presented at the Seismological Society of America, 18–20 April 2017 Denver, Colorado.
- Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous distributed systems, <https://arxiv.org/pdf/1603.04467>
- Abrahamson, N. A., W. J. Silva, and R. Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions, *Earthquake Spectra*, **30**, no. 3, 1025–1055, doi: 10.1193/070913EQS198M.
- Abrahamson, N.M., M.W. Kuehn, and N. Landwehr (2019). Probabilistic Seismic Hazard Analysis in California Using Nonergodic Ground Motion Models. *Bull. Seismol. Soc. Am.* **109**, 1235-1249.
- Alavi A. H., and A. H. Gandomi (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. *Comput. Struct.* **89**, 2176–2194.
- Anderson, J. G., and S. E. Hough (1984). A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies, *Bull. Seismol. Soc. Am.* **74**, 1969–1993.
- Andrews, D. J. (1986). Objective determination of source parameters and similarity of earthquakes of different size, in *Earthquake Source Mechanics*, S. Das, J. Boatwright, and C. H. Scholz (Editors), Am. Geophys. Union, Washington, D.C.
- Baykan, N. A., and N. Yılmaz (2011). A mineral classification system with multiple artificial neural network using k-fold cross validation. *Math. Comput. Appl.* **16(1)**, 22-30.
- Bayless J., and N. A. Abrahamson (2018b). An empirical model for Fourier amplitude spectra using the NGA-West2 database, PEER Report No. 2018/07, Pacific Earthquake Engineering Research Center, University of California, Berkeley, California.
- Berger, J., L. M. Baker, J. N. Brune, J. B. Fletcher, T. C. Hanks, and F. L. Vernon (1984). The Anza array: A high-dynamic, range, broadband, digitally radio telemetered seismic array, *Bull. Seismol. Soc. Am.* **74**, no. 4, 1469–1481.

- Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.* **75**, no. 26, 4997–5009, doi: 10.1029/JB075i026p04997.
- California Institute of Technology (Caltech) (1926). Southern California Seismic Network, International Federation of Digital Seismograph Networks, Other/Seismic/Network, doi: 10.7914/SN/CI.
- Derras B., P.Y. Bard, and F. Cotton (2014). Towards fully data driven ground-motion prediction models for Europe. *Bull. Earthquake. Eng.*, **12 (1)**, 495-516. doi: 10.1007/s1051801394810.
- Derras, B., P. Y. Bard, and F. Cotton (2016). Site-condition proxies, ground motion variability, and data-drive GMPEs: Insights from the NGA- West2 and RESORCE datasets, *Earthquake Spectra* **32**, no. 4, 2027–2056.
- Derras, B., P. Y. Bard, and F. Cotton (2017). VS30, slope, H800, and f0: Performance of various site-condition proxies in reducing ground- motion aleatory variability and predicting nonlinear site response, *Earth Planets Space* **69**, no. 1, 133.
- Derras, B., P.Y. Bard, F. Cotton, and A. Bekkouche (2012). Adapting the neural network approach to PGA prediction: an example based on the KiK-net data. *Bull. Seismol. Soc. Am.* **102**, 1446–1461 <http://www.bssaonline.org/content/102/4/1446.abstract>
- Dhanya, J. and S. T. G. Raghukanth (2018). Ground motion prediction model using artificial neural network. *Pure Appl. Geophys.*, **175 (3)**, 10351064, doi: 10.1007/s0002401717513.
- Diamantidis, N. A., D. Karlis, and E.A. Giakoumakis (2000). Unsupervised stratification of cross-validation for accuracy estimation, *Artificial Intelligence* **116 (1-2)**, 1-16.
- Douglas, J. and B. Edwards (2016). Recent and future developments in earthquake ground motion estimation. *Earth Sci. Rev.*, **160**, 203-219, doi: 10.1016/j.earsirev.2016.07.005.
- Emami, S. M. R., Y. Iwao, and T. Harada (1996). A method for prediction of peak horizontal acceleration by artificial neural networks. In Proceedings of *Eleventh World Conference on Earthquake Engineering*. Paper no. 1238.
- Gallipoli, M. R., and M. Mucciarelli (2009). Comparison of site classification from VS30, VS10, and HVSR in Italy, *Bull. Seismol. Soc. Am.* **99**, no. 1, 340–351.
- Geron, A. (2017). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, Inc.

- Goldberg, D. E., Melgar, D., Sahakian, V. J., Thomas, A. M., Xu, X., Crowell, B. W., & Geng, J. (2020). Complex rupture of an immature fault zone: A simultaneous kinematic model of the 2019 Ridgecrest, CA earthquakes. *Geophysical Research Letters*, 47(3), e2019GL086382.
- Güllü, H. and E. Erçelebi (2007). A neural network approach for attenuation relationships: An application using strong ground motion data from Turkey. *Eng. Geol.*, **93(3-4)**, 65:81.
- Günaydın, K. and A. Günaydın (2008). Peak ground acceleration prediction by artificial neural networks for northwestern Turkey. *Math Probl. Eng.* doi:10.1155/2008/919420.
- Han, L., J. A. Hole, J. M. Stock, G. S. Fuis, A. Kell, N. W Driscoll, G.M. Kent, A. J. Harding, M. J. Rymer, A. González-Fernández, et al. (2016). Continental rupture and the creation of new crust in the Salton Trough rift, Southern California and northern Mexico: results from the Salton Seismic Imaging Project. *J. Geophys. Res. Solid Earth*, **121(10)**, 7469-7489.
- Hauksson, E., and P. M. Shearer (2006). Attenuation models (QP and QS) in three dimensions of the southern California crust: Inferred fluid saturation at seismogenic depths. *J. Geophys. Res. Solid Earth*, **111**, no. B05302.
- Khosravikia, F., Y. Zeinali, Z. Nagy, P. Clayton, and E. M. Rathje (2018). Neural network-based equations for predicting PGA and PGV in Texas, Oklahoma, and Kansas. In *Geotechnical Earthquake Engineering and Soil Dynamics GSP 291*, pages 538-549.
- Klimasewski, A., V. J. Sahakian, A. S. Baltay, J. Boatwright, J. B. Fletcher, and L. M. Baker (2019). κ_0 and Broadband Site Spectra in Southern California from Source Model-Constrained Inversion. *Bull. Seismol. Soc. Am.* **109**, no. 5 1878–1889.
- Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, P. Gerstoft (2018). Machine Learning in Seismology: Turning Data into Insights. *Seismol. Res. Lett.* **90** (1): 3–14.
- Kuehn N. M., N. A. Abrahamson, and M. A. Walling (2019). Incorporating non-ergodic path effects into the NGA-West2 ground-motion prediction equations, *Bull. Seismol. Soc. Am.* **109**, no. 2, 575–585, doi: <https://doi.org/10.1785/0120180260>.
- Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016). A nonergodic ground-motion model for California with spatially varying coefficients, *Bull. Seismol. Soc. Am.* **106**, no. 6, 2574–2583, doi: 10.1785/0120160118.
- Laurendeau, A., F. Cotton, O. J. Ktenidou, L. F. Bonilla, and F. Hollender (2013). Rock and stiff-soil amplification: Dependency on VS30 and kappa (κ_0), *Bull. Seismol. Soc. Am.* **103**, no. 6, 3131–3148.

- Lin, P. S., B. Chiou, N. Abrahamson, M. Walling, C. T. Lee, and C. T. Cheng (2011). Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models, *Bull. Seismol. Soc. Am.* **101**, no. 5, 2281–2295, doi: 10.1785/0120090312.
- Pedregosa, F. G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and, É. Duchesnay (2011). Scikit-learn: Machine Learning in Python **12**(85): 2825–2830,.
- Potdar, K., T. S. Pardawala, and C. D. Pai (2017). A comparative study of categorical variable encoding techniques for neural network classifiers, *Int. J. Comput. Appl.*, **175**, no. 4, 7-9.
- Ross, Z. E., Y. Ben-Zion, M. C. White, and F. L. Vernon (2016). Analysis of earthquake body wave spectra for potency and magnitude values: implications for magnitude scaling relations. *Geophys. J. Int.* 207, 1158–1164.
- Sahakian, V. J., A. S. Baltay, T. C. Hanks, J. Buehler, F. L. Vernon, D. Kilb, and N. Abrahamson (2018). Decomposing leftovers: Event, path, and site residuals for a small magnitude ANZA region GMPE, *Bull. Seismol. Soc. Am.* **108**, 2478–2492, doi: 10.1785/0120170376.
- Sahakian, V., A. Kell, A. Harding, N. Driscoll, and G. Kent (2016). Geophysical evidence for a San Andreas subparallel transtensional fault along the northeastern shore of the Salton Sea. *Bull. Seismol. Soc. Am.* **106**(5), 1963-1978.
- Sahakian, V.J., A.S. Baltay, T.C. Hanks, J. Buehler, F.L. Vernon, D. Kilb, N. Abrahamson (2019). Ground motion residuals, path effects, and crustal properties: A pilot study in Southern California, *J. Geophys. Res.*, **124**(6), 5738-5753, doi: 10.1029/2018JB016796.
- Southern California Earthquake Data Center (SCEDC) (2013). Southern California Earthquake Data Center, Caltech. Dataset, doi: 10.7909/C3WD3xH1.
- Thompson, E., and D. Wald (2016). Uncertainty in VS30-based site response, *Bull. Seismol. Soc. Am.* **106**, no. 2, 453–463.
- Trugman, D. T., and P. M. Shearer (2018). Strong correlation between stress drop and peak ground acceleration for recent M 1–4 earthquakes in the San Francisco Bay area. *Bull. Seismol. Soc. Am.* **108**(2), 929-945.
- U.S. Geological Survey, Quaternary fault and fold database for the United States, accessed May 1, 2017, at: <https://www.usgs.gov/natural-hazards/earthquake-hazards/faults>.
- Van Houtte, C. S. Drouet, and F. Cotton (2011). Analysis of the origins of κ (kappa) to compute hard rock to rock adjustment factors for GMPEs, *Bull. Seismol. Soc. Am.* **101**, no. 6, 2926–2941.

- Van Houtte, C., O. J. Ktenidou, T. Larkin, and C. Francois-Holden (2014). Hard-site κ_0 (Kappa) calculations for Christchurch, New Zealand, and comparison with local ground-motion prediction models, *Bull. Seismol. Soc. Am.* **104**, 1899–1913, doi: 10.1785/0120130271.
- Van Houtte, C., S. Drouet, and F. Cotton (2011). Analysis of the Origins of κ (Kappa) to Compute Hard Rock to Rock Adjustment Factors for GMPEs. *Bull. Seismol. Soc. Am.* **101(6)**, 2926-2941.
- Vernon, F. L. (1989). Analysis of data recorded on the ANZA seismic network, Ph.D. Thesis, University of California, San Diego.
- Wei, S., E. Fielding, E., S. Leprince, S., A. Sladen, A., J. P. Avouac, J. P., D. Helmberger, D., E. Hauksson, R. Chu, M. Simons, K. Hudnut, et al... & Herring, T. (2011). Superficial simplicity of the 2010 El Mayor–Cucapah earthquake of Baja California in Mexico. *Nature Geoscience*, *4(9)*, 615-618.
- Wessel, P., and W. H. Smith (1998). New, improved version of Generic Mapping Tools released. *Eos Trans. AGU* **79**, no. 47, 579. <https://doi.org/10.1029/98EO00426>.
- Yong, A., S. E. Hough, J. Iwahashi, and A. Braverman (2012). A terrain- based site-conditions map of California with implications for the contiguous United States, *Bull. Seismol. Soc. Am.* **102**, no. 1, 114–128.