Detecting and predicting earthquake ground motion directionality patterns using machine learning tools

Alan Poulos
Blume Earthquake Engineering Center
Stanford University, USA
apoulos@stanford.edu

Rodrigo Silva-Lopez
Blume Earthquake Engineering Center
Stanford University, USA
rsilval@stanford.edu

Abstract

Earthquakes generate ground shaking that varies with orientation. At any given site, there exist an orientation of maximum ground shaking. This project developed models to predict the absolute difference between the orientations of maximum ground shaking between two sites. A database of worldwide earthquake ground motions was used to compute these orientations and obtain relevant seismological, geological, and geographical attributes. The data was then used to calibrate a linear model using ridge regression and a neural network. These types of model can then be generalized for more sites and used to improve the accuracy of city-scale seismic risk assessments.

1 Introduction

The horizontal intensity of ground shaking during earthquakes depends on orientation, a phenomenon commonly known as ground motion directionality. The orientation of maximum ground shaking is relatively well modeled by a uniform distribution, in other words, all orientations are equally likely. However, observations of several earthquakes show that these orientations are correlated between sites that are close to each other, as shown in Figure 1 for an example earthquake near Anchorage, Alaska.

This project aims to develop a model f receives seismological, geographical, and geological characteristics of two given recording sites, s_i and s_j, and properties of the earthquake E_k and predicts the mean absolute difference between the angles of maximum ground motion intensity of these sites, $\Delta\theta_{s_i,s_j}$:

$$\Delta\theta^k_{i,j} = f(s_i, s_j, E_k) \tag{1}$$

Taking this into account we will: (1) analyze and clean a seismic database to select parameters that are reliable and are physically related to the phenomenon; (2) construct the attributes used for regression by pairing all sites that recorded the same earthquake, computing the distances between them, and constructing the outputs of the model by computing $\Delta\theta$; (3) calibrate a ridge regression model with the data obtained in Step (2) to predict $\Delta\theta$; and (4) calibrate a neural network for the same task. If successful, this result could be used to improve our current earthquake loss estimations and help to improve the resilience of cities to future earthquake events.

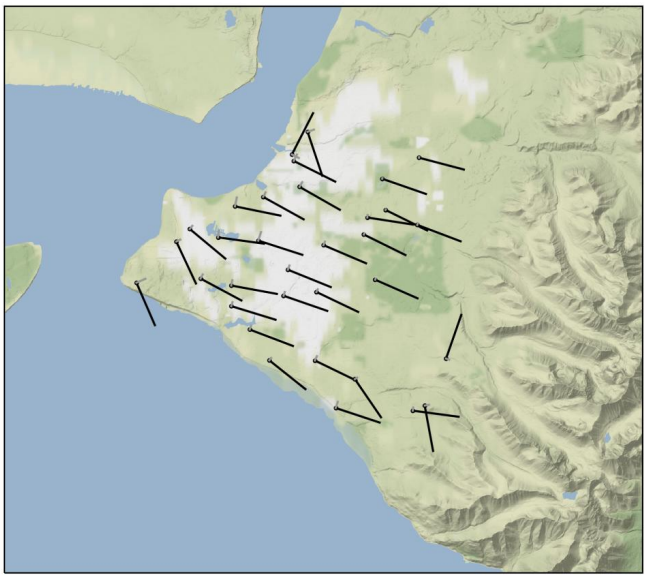

Figure 1: Directions of maximum ground motions intensity at different sites within the city of Anchorage during a M_W 5.7 earthquake in 2018.

2 Related work

The most widely used ground motion intensity measure in earthquake engineering is the pseudo-spectral accelerations, which varies with orientation. The maximum intensity from all orientations in the horizontal plane is known as RotD100 (Boore, 2010), and is usually used by building codes to design infrastructure (e.g., American Society of Civil Engineers, 2016). Hong and Goda (2007) studied the orientation dependence of ground motions recorded in several earthquakes and reported that, on average, the RotD100 intensity is on average between 39% to 68% higher than the intensity in the perpendicular direction. Several reasons have been given in the literature for this orientation dependence, such as topographic irregular-

ities (Spudich et al., 1996), local geologic heterogeneities (Bonamassa and Vidale, 1991), and basin edge effects (Heresi et al., 2020). Somerville et al. (1997) noted that for records that are close to the fault, the orientation of RotD100 is more likely to occur in an orientation that is close to being normal to the orientation of the fault. However, for record that are more than ∼5 km from the fault, which is usually the case, all orientations appear to be equally likely (Shahi and Baker, 2014). Thus, using a uniform distribution is a reasonable assumption to sample the orientation of RotD100. However, as seen in Figure 1, for a given earthquake, the orientation of RotD100 for sites that are close to each other are correlated. No model exists in the literature to sample the the orientations of RotD100 at multiple sites simultaneously, and hence this effect is not considered when estimating the effect of earthquakes on infrastructure.

3 Dataset and Features

3.1 Data cleaning

For this project, we use the NGA-West2 database, which contains approximately 20 thousand ground motion records from worldwide shallow crustal earthquakes (Ancheta et al., 2014). To only use records with engineering significance, we first filtered out all earthquakes that have a magnitude smaller than 5.0, yielding a database of around 5 thousand records. Each record in the database has 274 attributes, which include information about the earthquake source, the geology of the site, and the instrument used to measure the ground shaking. These attributes can be either categorical or numerical, and there are some parameters that do not have information for all seismic records. To clean the database, we selected meaningful attributes one by one, checking that the information was reliable, that the attribute could have a physical relation with directionality, and that the value of the parameter was not the same for all records. After this selection, we obtained 71 attributes to be used to predict $\Delta\theta$. Some examples of these attributes are:

- Earthquake source data: earthquake magnitude; type of focal mechanism; rake and dip angles; fault rupture width, length, and area; and depth to the hypocenter and top of the rupture surface.
- Different distance measures between the earthquake and each of the two sites: epicentral, hypocentral, rupture and Joyner-Boore distances.
- Soil properties of the two sites: average shear wave velocity of the top 30 m of soil (Vs30) and NEHRP site class.
- Ground motion intensities at the two sites: peak ground acceleration, velocity, and displacements.
- Distance between the two sites.

Finally, we normalized numerical attributes to have mean zero and unit variance, and encoded categorical attributes as a one-hot numeric array.

3.2 Improving seismic database

Given that the objective of this study is to estimate the difference in directionality angles between two given sites, we created the examples by pairing all possible sites that recorded the same earthquake and combined their attributes. We also computed the distance between the sites as an additional attribute. Moreover, we had to compute $\Delta\theta$ for all of the examples because they are not provided in the database, which required us to process the ground motion acceleration waveforms. As a result of this process, we obtained approximately 500 thousand individual examples. Taking into account that the results of this study are meant to be used for city-scale analysis of earthquakes, we limited the distance between sites to 5 km. This threshold reduced the number of examples to approximately 5.3 thousand.

3.3 Feature selection using backward and forward analysis

Besides using expert criteria to filter the features that should be included in the regression model, we further refined feature selection by using a backward and forward analysis based on a ridge regression model. First, we ranked each feature according to the decrease on R^2 on test data if the feature was removed from the model. A bigger decrease would indicate that the feature has a more important role at predicting difference in directionality. This process is named backwards analysis. Then, once all features have been ranked, we add one by one the features in the order provided by the ranking, in a process called forward analysis. We stopped including features once the inclusion of an additional variable into the model would decrease the value of R^2 over the test set due to overfitting. We perform the previous analysis iteratively with the hyperparameter selection. Specifically, we select a value of α as the regularization parameter, then we select variables, and explore the optimal α given the selected variables. Out of this analysis, 21 features are selected, with the distance d between sites and the peak ground acceleration (PGA) as the variables with a greater importance. Figures 2 and 3 show scatter plots of $\Delta\theta$ vs these two selected features. Although the variability of $\Delta\theta$ is significant for all attribute values, the mean value tends to increase with separation distance and to decrease with peak ground acceleration.

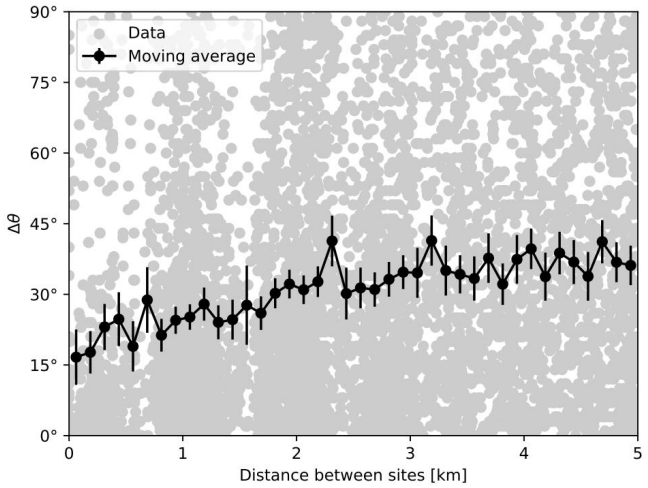

Figure 2: $\Delta\theta$ as a function of the distance between sites. Error bars represent 95% confidence intervals of the mean.

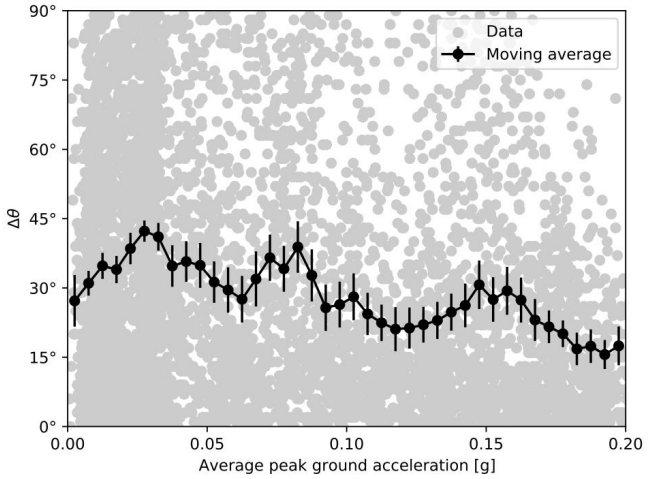

Figure 3: $\Delta\theta$ as a function of the average peak ground acceleration measured at the sites. Error bars represent 95% confidence intervals of the mean.

4 Methods and Results

To compute the angular difference in directionality for two given sites we present the results of a ridge regression and a neural network. In this section we also define the benchmark for comparison as a simple linear regression that only uses the distance between sites. All learning algorithms were implemented using scikit-learn (Pedregosa et al., 2011).

4.1 Benchmark

To compare the results of the regression models proposed in this study, we defined a benchmark by performing a simple linear regression using the distance between the stations d to predict the difference in the directionality between sites $\Delta\theta$. Given the trends observed in the data, we explored different functional forms on the distance for the simple regression, concluding that a linear function of the form $\Delta\theta = c_1 + c_2 d$, instead of using any nonlinear parameter over d. Using this regression, we got the results shown in Figure 4. We observe that this simple benchmark performs poorly at predicting the difference in directionality angle both in the training and the test data.

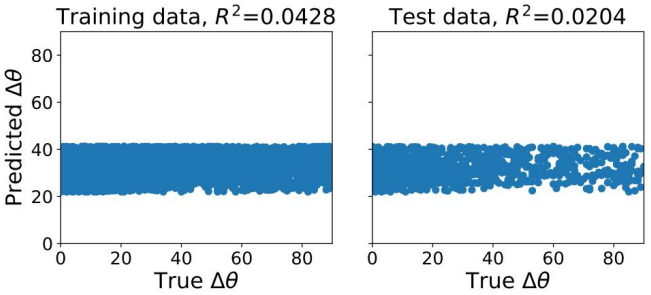

Figure 4: True vs predicted values by the linear regression model over the training and test sets.

4.2 Ridge Regression

The first model that we calibrated was a linear least squares model with L2 regularization. A Ridge regression aims to minimize the prediction error, defined as the squared error between the real value, $y^{(i)}$ and the predicted value $\beta^T x^{(i)}$, while penalizing the norm of the coefficient of the prediction β, as shown in Equation (2).

$$\min_{\beta} \sum_{i=1}^{n} \left(y^{(i)} - \beta^T x^{(i)} \right)^2 + \alpha ||\beta||_2^2 \qquad (2)$$

To implement the ridge regression we used 80% of the examples for training, 10% to look for the α parameter of the Ridge regression in Equation (2), and 10% for testing. In terms of the variables, we used the variables selected in the feature selection, which maximize accuracy through a backward-forward analysis. From the validation set, the value of α is 1.5. Using these parameters we got the results shown in Figures 5, 6 and 7. Figure 5 shows a scatter plot comparing predicted values and results from the model. Figures 6 and 7 show the ability of the model to predict the mean value of $\Delta\theta$ as a function of some parameters of the regression.

4.3 Neural Network

The final regression method that we used to predict $\Delta\theta$ is a neural network. In the same way as a the Ridge regression, a neural network aims to minimize the error between observed and predicted data. However, instead of predicting values using a linear function, a neural network transforms the inputs of the model using activation functions over several layers, generating a complex nonlinear model. Regarding the neural network implemented in this study, its inputs are all features in the dataset, since

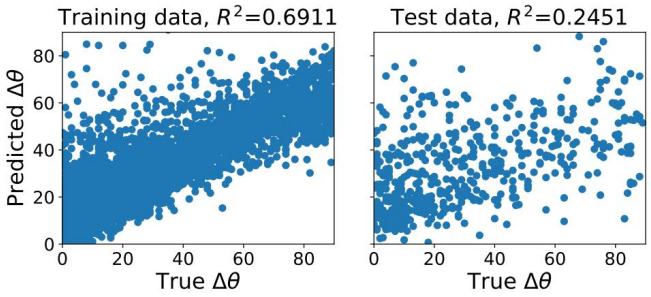

Figure 5: True vs predicted values by the ridge linear regression model over the training and test sets

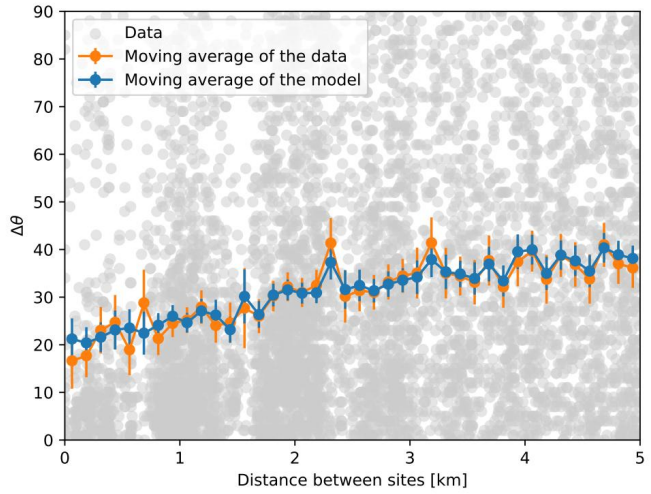

Figure 6: Comparison of true and predicted values for windowed average over distance for ridge regression.

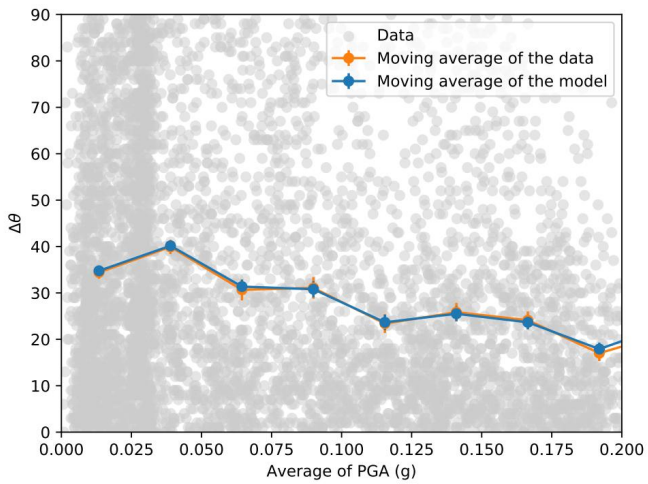

Figure 7: Comparison of true and predicted values for windowed average over PGA for ridge regression.

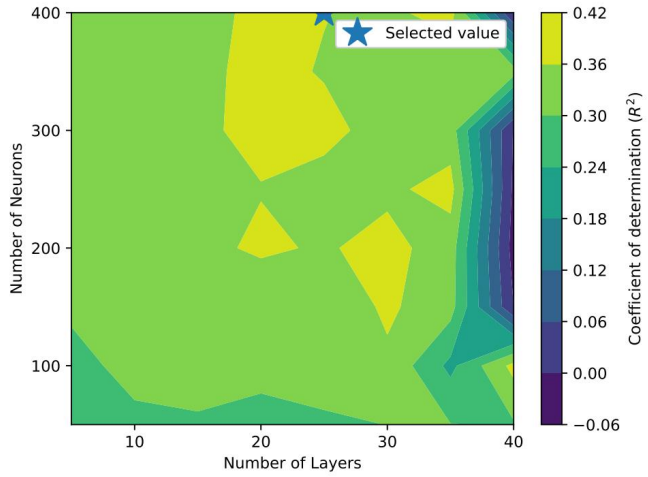

Figure 8: Gridsearch over hyperparameters of the neural network.

it proved to perform better in terms of R^2 for the test data than the features selected in the forward-backward analysis. We divided the examples into 80% training set, 10% validation set, and 10% as the test set. As an optimization algorithm we used the ADAM algorithm and the activation function used was ReLU. To select the hyperparameters of the neural network we developed a gridsearch on the validation set. The hyperparameters that we tuned were the number of layers in the model, the number of neurons on each layer, and the α parameters used for the optimization. The optimality criteria used to select the hyperparameters was the value of R^2 on the validation set. The results of the gridsearch are shown in Figure 8 for the number of layers in the model and the number of neurons per layer. Using a 25 layers with 400 neurons each of them, with a learning parameters of $\alpha = 0.1$, we got the results shown in Figures 9,10 and 11.

The neural network is able to capture the mean behaviour of the data, as shown in Figures 10 and 10 that compare the predicted and observed mean trends on the dataset. Regarding the training data, we observe that the neural network has a high R^2, which could suggest overfitting. We explored different values of hyperparameters and all of them yielded high values of R^2 for the training data; hence we concluded that this overfitting could not be avoided unless significant modifications were performed in the model, such as changing inputs and outputs. Despite overfitting on the training data, the neural network has the highest R^2 value on the test set among the models we implemented in this study.

5 Discussion

We measured the accuracy of the three developed models using two metric: the coefficient of determination and the mean absolute error, which are presented in Table 1 for the test set. The benchmark regression model that only uses the distance between sites performs very poorly, having the lowest R^2 and highest mean absolute error. The ridge regression with all selected features improves both metrics significantly. As expected, the neural network model has the highest R^2 and lowest mean absolute error because its

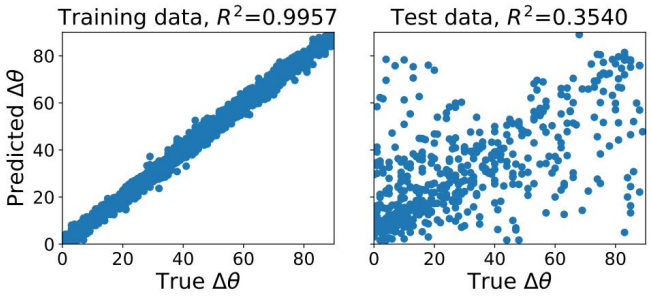

Figure 9: True vs predicted values by the neural network over the training and test sets.

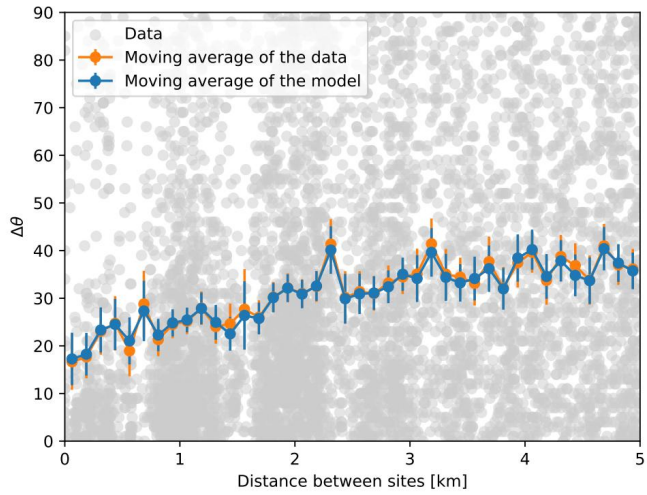

Figure 10: Comparison of true and predicted values for windowed average over distance for the neural network.

use of multiple layers with several neurons can produce very expressive functions that can better fit the data.

Table 1: Performance metrics of each model.

Model	R^2	Mean absolute error
Benchmark	0.025	20.2
Ridge regression	0.245	16.9
Neural network	0.354	13.2

Note that the values of R^2 for all models are not high. As we observed on Section 3 there is a high inherit uncertainty in the data, with same values of $\Delta\theta$ for different feature combinations. Given this uncertainty, the maximum value achievable of accuracy of a model on a test set tends to be low. Despite not showing a good value of R^2, we observed that the proposed models are able to capture the mean trend of the data, as showed in Figures 6, 6,7,10 and 11. This combination of uncertainty, but a good predicting capability of the mean trend suggest that in a future work these models could be reformulated to directly quantify the dispersion of the data and the mean,

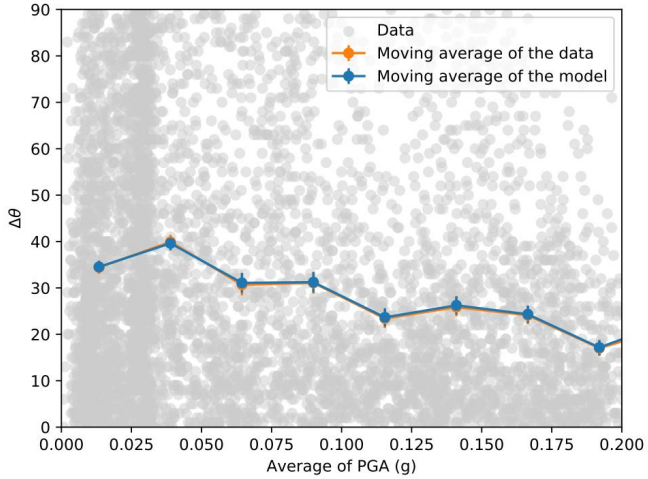

Figure 11: Comparison of true and predicted values for windowed average over PGA for the neural network.

given some parameters of the dataset.

6 Conclusion and Future Work

In this work we constructed different models to predicts the absolute angular difference between the orientations of maximum ground motion intensity at two sites for a given earthquake. We started by a very simple benchmark regression model that only uses the distance between the two sites. Then we trained a ridge regression and a neural network using all features. The highest accuracy on the test set was achieved by the neural network model, which we expected due to its higher expressive power and flexibility.

The results show that the developed models are good to predict the average values of $\Delta\theta$; however, there is a significant amount of uncertainty that the models are not capturing and that is leading to relatively low values of accuracy. A better model could be developed in the future that accounts for this uncertainty by predicting, for example, the mean and the variance. Moreover, this work could be extended by considering an arbitrary number of sites instead of only two, which would enable the simulation of ground motion directionality at multiple sites. These types of models could then be coupled with seismic hazard and risk assessment tools to estimate the effects of future earthquakes on the built environment.

7 Contributions

We both contributed equally to each aspect of the project.

References

American Society of Civil Engineers (2016). *Minimum design loads and associated criteria for buildings and other struc-*

tures. American Society of Civil Engineers, Reston, VA, asce/sei 7-16 edition.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., et al. (2014). NGA-West2 database. *Earthquake Spectra*, 30(3):989–1005.

Bonamassa, O. and Vidale, J. E. (1991). Directional site resonances observed from aftershocks of the 18 October 1989 Loma Prieta earthquake. *Bull. Seismol. Soc. Am.*, 81(5):1945–1957.

Boore, D. M. (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bulletin of the Seismological Society of America*, 100(4):1830–1835.

Heresi, P., Ruiz-García, J., Payán-Serrano, O., and Miranda, E. (2020). Observations of Rayleigh waves in Mexico City Valley during the 19 september 2017 Puebla–Morelos, Mexico earthquake. *Earthq. Spectra*, 36(2S):62–82.

Hong, H. P. and Goda, K. (2007). Orientation-dependent ground-motion measure for seismic-hazard assessment. *Bulletin of the Seismological Society of America*, 97(5):1525–1538.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Shahi, S. K. and Baker, J. W. (2014). NGA-West2 models for ground motion directionality. *Earthquake Spectra*, 30(3):1285–1300.

Somerville, P. G., Smith, N. F., Graves, R. W., and Abrahamson, N. A. (1997). Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismological research letters*, 68(1):199–222.

Spudich, P., Hellweg, M., and Lee, W. H. K. (1996). Directional topographic site response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake: implications for mainshock motions. *Bull. Seismol. Soc. Am.*, 86(1B):S193–S208.