Regional-Scale 3D Ground-Motion Simulations of Moderate Earthquakes on the Hayward Fault

By *Taka'aki Taira, Jeffrey Sung, Julie Hoang*

Abstract

SW4 earthquake simulations can predict the behavior of real earthquakes in the Bay Area. We used different python libraries to clean and visualize SW4 and observed data from different geographical stations near the Hayward fault. We found that the SW4 peak amplitude and duration had a correlation with the distance from the earthquake source while the observed data did not. Further research is needed on the peak amplitude calculations to strengthen the accuracy of the SW4 simulation.

Introduction

The question our project is attempting to answer is how accurate are our SW4 earthquake simulations in different areas. Does the accuracy of the simulation vary based on distance from the source or which side of the Hayward fault it resides? SW4 is an anelastic finite difference code that is produced with the aid of HPC (High Performance Computing). The data was previously obtained by the research team through [UC Berkeley's Savio computing cluster]. Quantifying the accuracy of the SW4 model compared to observed data is instrumental to structural decisions made in the future relying on the SW4 model.

Data and Methods

Our data is a time series of ground movements, and our goal of this project was to visualize the frequency, duration, peak amplitude, and differences between the simulated SW4 program and the observed data. A local earthquake occurred at Berkeley was selected for our project. The magnitude of this earthquake is 4.4, which produces strong ground shaking over the Bay Area. We used a combination of mathematical models and existing Python libraries to clean and understand the data obtained. Our main applications included Jupyter Notebook in addition to GitHub to collaborate remotely.

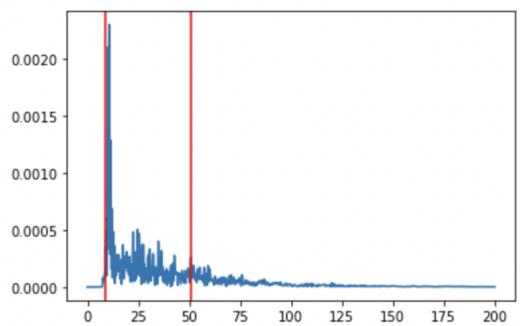
Python Libraries: NumPy, pandas, obspy, Matplotlib, Plotly, GeoPy, SciPy

NumPy, pandas, and Matplotlib are all general-purpose data science libraries for doing computational work, organizing data, and plotting basic graphs respectively. ObsPy is an open source framework for processing seismic data, and it was what we used to actually parse the seismograms and access the waveform data. GeoPy was only used for its *distance* function, so we could easily calculate how far away each station was from the epicenter using the coordinates. SciPy was specifically used to apply the Fourier transform on the seismic traces, so we could visualize the frequencies. Plotly was used for more sophisticated visualizations, mainly geographic plots.

Duration:

To calculate the duration of seismic data at each station, we used a cutoff of 10% of the maximum recorded amplitude as the end time of the duration at that location. In addition, we filtered the data with a bandpass filter in the range of [0.03, 0.25] Hz. Increasing the frequency bandfilter may give us more data points but may include more outliers. The bandpass filter can be modified depending on the level of noise in the data.

```
BK.BKS.HN
duration:  [8.8922384634329852, 50.753581399339957]
```



This is an example of one station (BK.BKS.HN). The red line indicates the cutoff point.

Bearing Angle Calculation:

As our research progressed, we were interested in visualizing systematic spatial variations of our model parameters (duration and peak amplitude) relative to the Hayward fault (i.e. East or West). Using our latitude and longitude coordinates in the station file, we used a simple mathematical model to calculate the bearing angle. The bearing angle can be thought of as the angle going clockwise from a vertical line (N-S) at your reference point to the target point. This tutorial was provided by IgisMap:

Bearing from point A to B, can be calculated as,

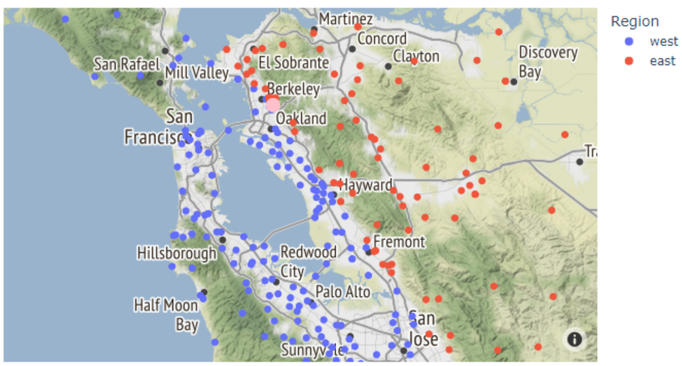
$$\beta = \text{atan2}(X,Y),$$

where, **X** and **Y** are two quantities and can be calculated as:

$$X = \cos\theta b * \sin\Delta L$$

$$Y = \cos\theta a * \sin\theta b - \sin\theta a * \cos\theta b * \cos\Delta L$$

Though the Hayward fault is only approximately [74 miles](), we differentiated stations to be either the East or West side of the fault. Our rough estimate distinguishing the E/W sides of the fault can be shown above.



This is not a precise regional label since we cannot get an accurate angular relationship between the source and fault. The "East" is estimated to be greater than zero degrees and less than 146 degrees or between negative forty five degrees and zero degrees. The "West" is defined as any angle greater than negative 180 while being less than negative forty five degrees, in addition to any angle greater than 146 degrees.

Our final dataframe for both the simulated and observed data looks like the following:

	Point	Station	Magnitudes	Peak	Duration	lat	lon	Bearing Angle	Region	Distance (miles)	Peak 0-1	Peak z-score
0	BK.BDM.HN	BK.BDM	[1.07668e-06, 1.07772e-06, 1.08061e-06, 1.0815...	0.001627	35.750	37.953970	-121.865540	72.551819	East	21.561090	0.005120	-0.457625
1	BK.BKS.HN	BK.BKS	[1.60992e-06, 1.6149e-06, 1.62033e-06, 1.62513...	0.004559	15.600	37.876220	-122.235580	19.518857	East	1.100331	0.015257	-0.300966

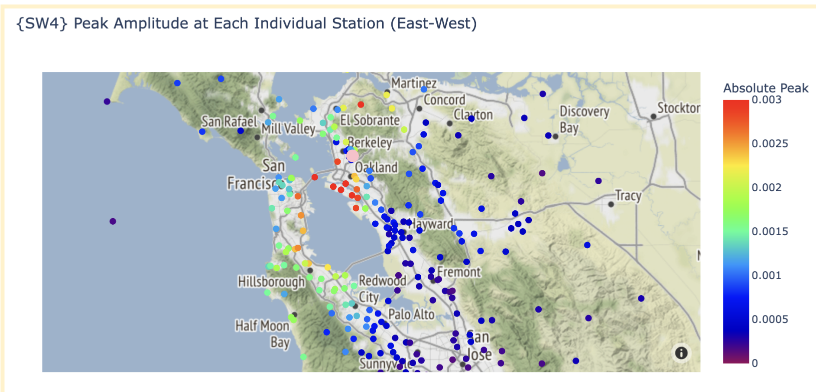
Results

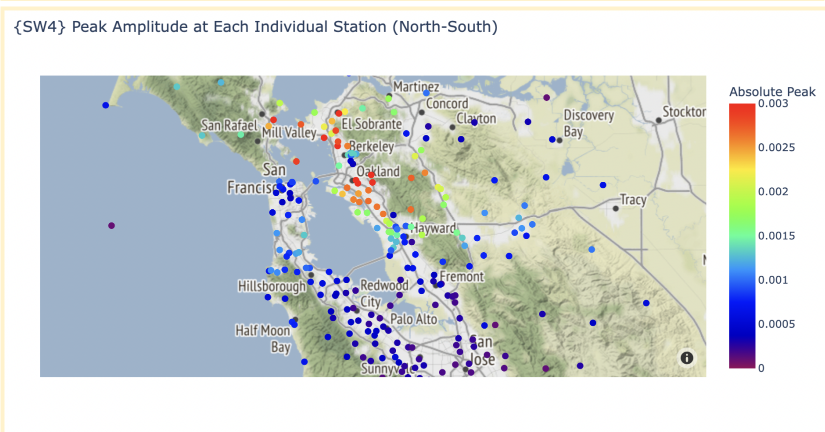
Visualizing Peaks and Duration for Each Direction:

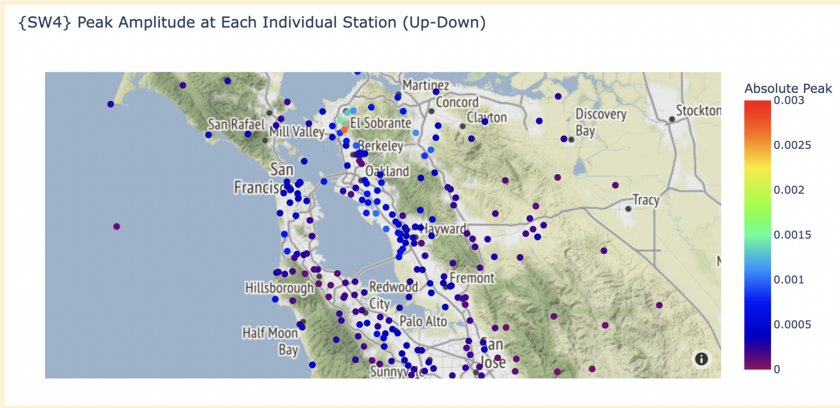
In addition to categorizing each point as a 3D vector point, we were able to obtain the values in different directions based on the simulated data provided.

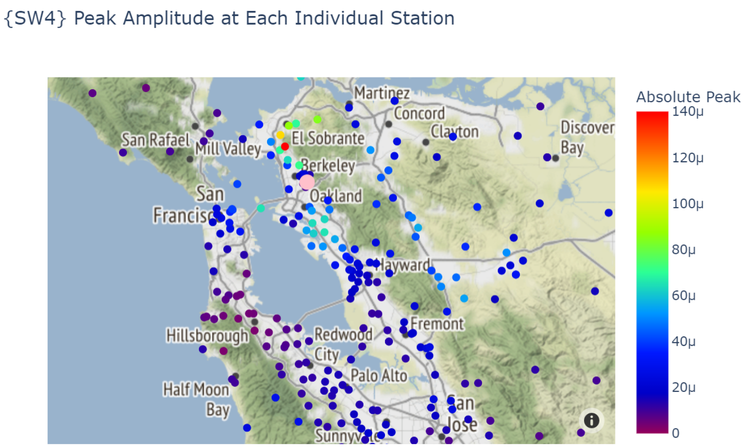
Letter at the end of Station Point	Direction
e	east/west
n	north/south
u	up/down

We decided to take a deeper look at the up/down direction and see if there were any significant differences to the 3D vector data. Up-down movement seemed to be slightly weaker than the other directions, and this trend was present in the observed data as well.







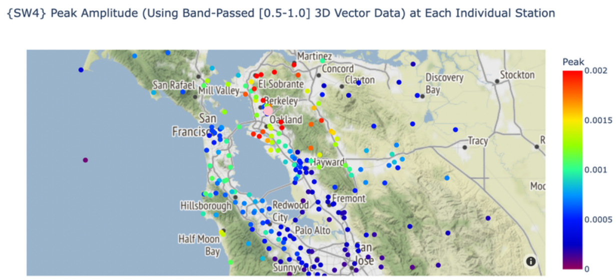


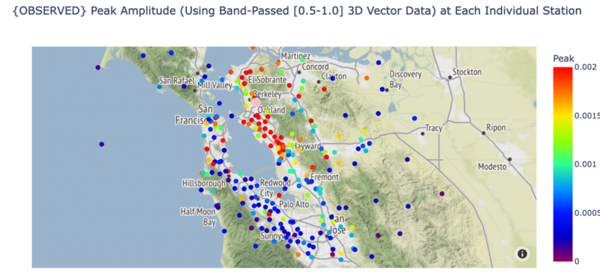
Visualized Data as 3D Vector Data:

In addition to observing each point direction, we combined all 3 points into a single 3D vector for simplicity and overall trends. We tested different ranges for our bandpass filter in order to see if there was a difference in the values produced.

We notice that there are more stations to the West of the Hayward fault with a higher peak amplitude (0.001) for the SW4 data than the observed values when the bandpass is filtered between [0.5 and 1.0]. However, the observed data has an overall higher peak amplitude average as can seen in the amount of red dots in the plot to the right below.

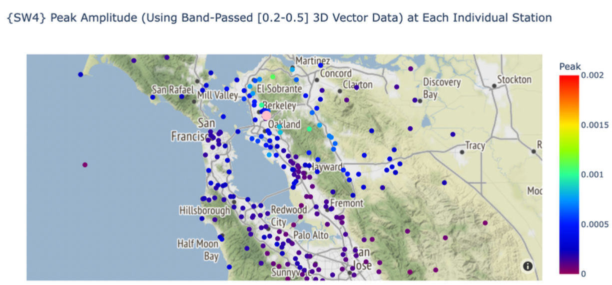
Peak Amplitude [0.5-1.0]

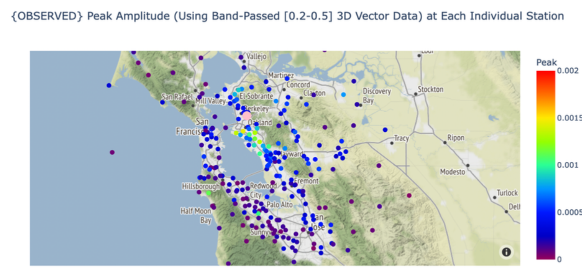




There is a significant difference in peak amplitude in both the SW4 and observed stations when using a lower range for the bandpass filter as shown below. In this range, there is only a small number of stations that reach 0.001 in peak amplitude or above.

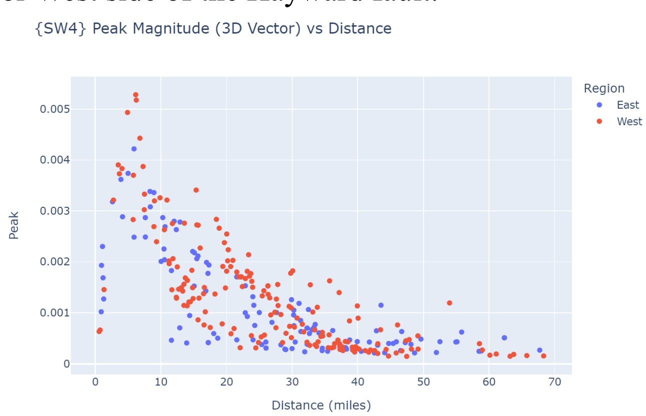
Peak Amplitude [0.2-0.5]

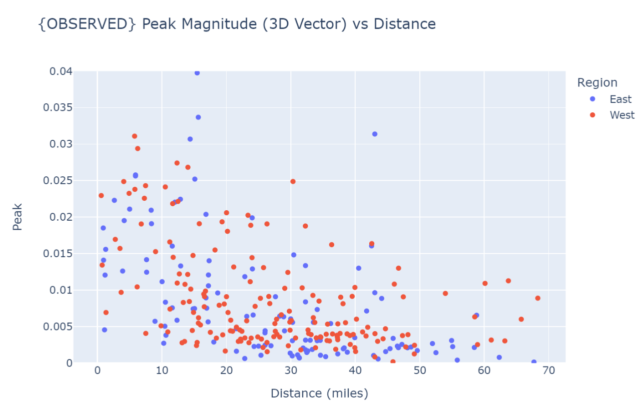


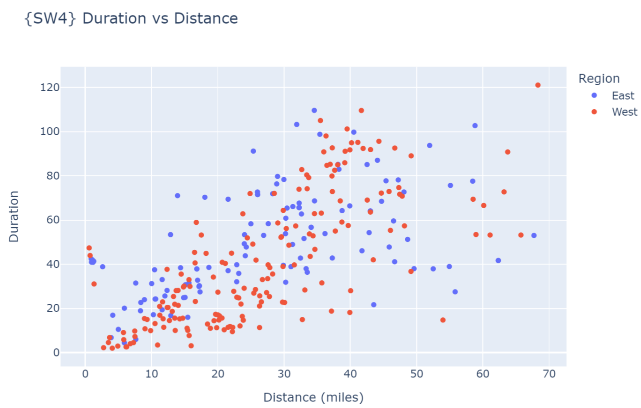


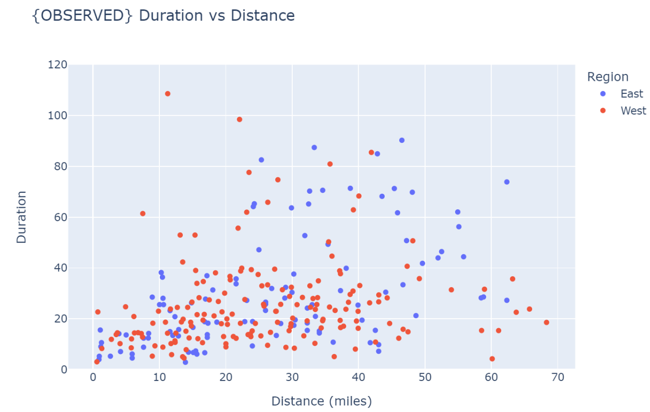
SW4 vs Observed - Peak Magnitude vs Distance and Duration vs Distance:

One can observe that the SW4 model produces a negative exponential relationship between peak magnitude and distance. This makes logical sense since the frequency of the source should decrease as we get further away. However, the observed data does not have a general trend for either peak magnitude nor duration. There does not seem to be any significant relationship between the East or West side of the Hayward fault.







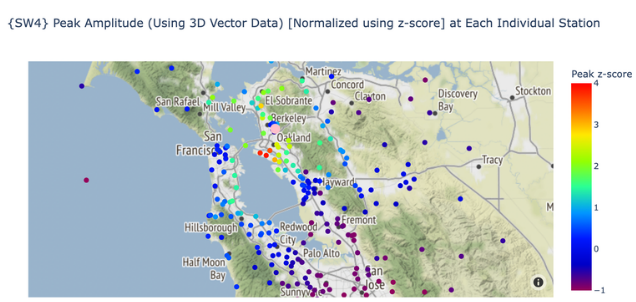


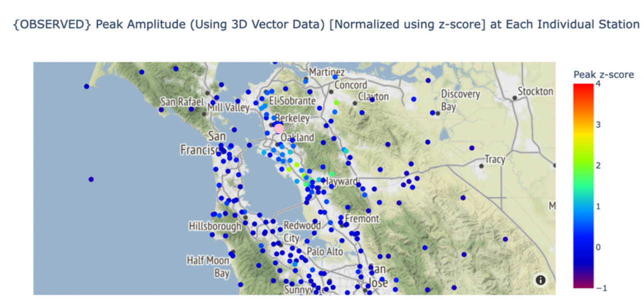
Rescaling Data:

Given that the observed data has significantly higher actual amplitudes than the simulated data, we tried rescaling the data through the z-score. The z-score tells us how many standard deviations a station lies above or below the mean. Initially, we attempted to calculate the ratio of the observed and simulated values. However, we did not find any conclusive results leading to these two new methods.

Normalized Peak Amplitude:

We notice that there is more variability in peak amplitude for the SW4 data than the observed data. We see a general similar peak z-score for stations in the same radius from the source. There are a few outliers in the observed stations near Hayward where the peak z-score is higher.

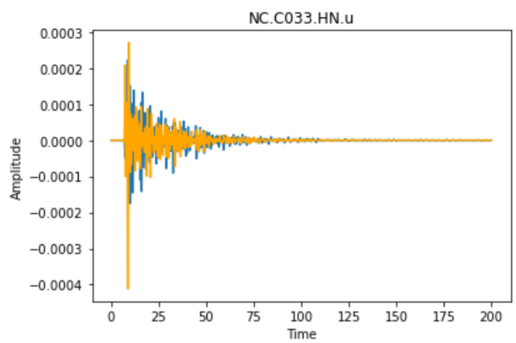


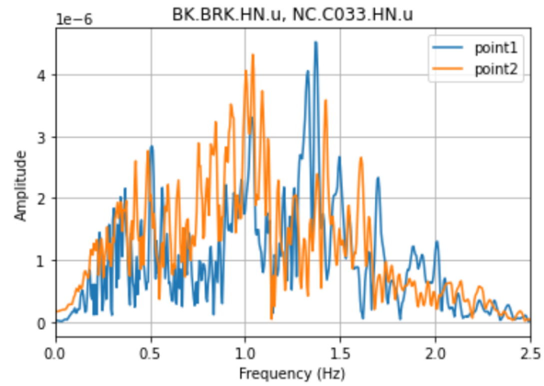


Discussion:

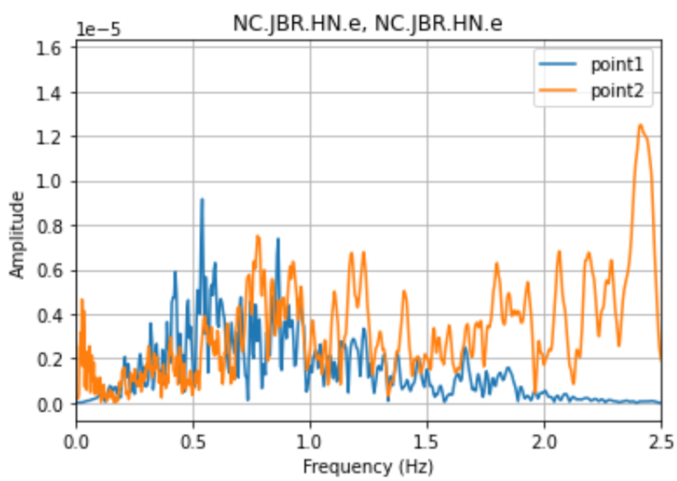
Comparison:

We compared NC.C033 and BK.BRK since it seemed that the peak amplitude for BK.BRK was an extremely small value despite being approximately as close to the source as NC.C033. Here is the amplitude plot as a function of time and frequency for both stations in the up/down direction.





NC.C033 is found to have a higher peak and more volatility at the beginning. Interestingly, BK.BRK has relatively low amplitudes despite being so close to the source. Taking a closer look at another station: “NC.JBR.HN.e”, we check to see if the model is accurate in representing the observed data. We see that the amplitudes are significantly different between the SW4 model and the observed data which is strange. This could be due to an incorrect model for calculating the peak amplitude for the SW4 data or an extremely large amount of error in the observed measurements.



Conclusion:

In conclusion, we utilized different python libraries to clean and visualize the simulated SW4 and observed seismic data. We created a rough estimate of the East and West sides of a local earthquake in Berkeley to determine if there were significant differences in peak magnitude or duration. In addition, we filtered the data with multiple frequency ranges in order to reduce noise and found a big difference in peak amplitude between [0.2, 0.5] and [0.5, 1.0]. There seems to be a general trend in peak amplitude and duration in relation to distance from the source for the SW4 but not the observed data. There is room for further research in modeling the peak amplitude of the SW4 data since it does not align with the observed data.

References

"Reading Seismograms." *Reading Seismograms - ObsPy 1.3.0 Documentation*, 10 Mar. 2022, https://docs.obspy.org/tutorial/code_snippets/reading_seismograms.html.

Rodgers, Arthur J., et al. "The Effect of Fault Geometry and Minimum Shear Wavespeed on 3D Ground-Motion Simulations for an MW 6.5 Hayward Fault Scenario Earthquake, San Francisco Bay Area, Northern California." *Bulletin of the Seismological Society of America*, vol. 109, no. 4, 2019, pp. 1265–1281., https://doi.org/10.1785/0120180290.

Rodgers, Arthur J., et al. "Broadband (0–4 Hz) Ground Motions for a Magnitude 7.0 Hayward Fault Earthquake with Three‐Dimensional Structure and Topography." *Geophysical Research Letters*, vol. 45, no. 2, 2018, pp. 739–747., https://doi.org/10.1002/2017gl076505.

Rodgers, Arthur J., et al. "Broadband (0–5 Hz) Fully Deterministic 3D Ground‐Motion Simulations of a Magnitude 7.0 Hayward Fault Earthquake: Comparison with Empirical Ground‐Motion Models and 3D Path and Site Effects from Source Normalized Intensities." *Seismological Research Letters*, vol. 90, no. 3, 2019, pp. 1268–1284., https://doi.org/10.1785/0220180261.