Data-driven synthesis of broadband earthquake ground motions using artificial intelligence

Manuel Florez[1]; Michaelangelo Caporale[1]; Buabthong Pakpoom[1]; Zachary Ross[1]; Domniki Asimaki[1]; Meier Men-Andrin[1].

[1]California Institute of Technology

Abstract

Robust estimation of ground motions generated by scenario earthquakes is critical for many engineering applications. We leverage recent advances in Generative Adversarial Networks (GANs) to develop a new framework for synthesizing earthquake acceleration time histories. Our approach extends the Wasserstein GAN formulation to allow for the generation of ground-motions conditioned on a set of continuous physical variables. Our model is trained to approximate the intrinsic probability distribution of a massive set of strong-motion recordings from Japan. We show that the trained generator model can synthesize realistic 3-Component accelerograms conditioned on magnitude, distance, and Vs30. Our model captures most of the relevant statistical features of the acceleration spectra and waveform envelopes. The output seismograms display clear P and S-wave arrivals with the appropriate energy content and relative onset timing. The synthesized Peak Ground Acceleration (PGA) estimates are also consistent with observations. We develop a set of metrics that allow us to assess the training process's stability and tune model hyperparameters. We further show that the trained generator network can interpolate to conditions where no earthquake ground motion recordings exist. Our approach allows the on-demand synthesis of accelerograms for engineering purposes.

Methodology

Generative Adversarial Networks (GANs) are state-of-the-art generative models (Goodfellow et al., 2014). Advances in architecture and training techniques have enabled GANs to synthesize high-resolution realistic-looking images of human faces, audio and even video sequences. GANs are built using two networks that must be trained simultaneously: a discriminator network *D* and a generator network *G*. The discriminator model is continuously refined to tell apart real samples from those synthetically generated; while the generator is trained with respect to *D* to synthesize realistically looking samples.

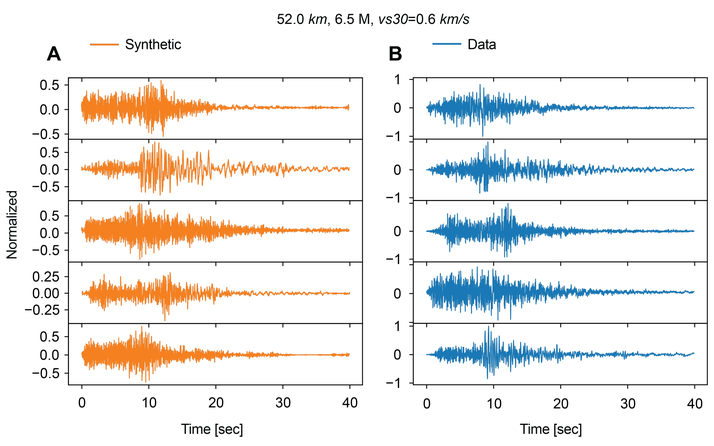

Figure 1: (A) Randomly sampled synthetic z-component accelerograms (orange) (B) Real z-component accelerograms, randomly sampled from the bin: 45.0-59.0 km, 6.4-6.6 M and 400-800 m/s (blue).

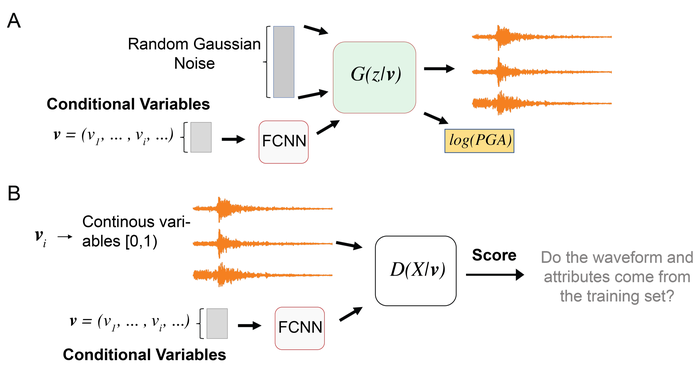

Figure 1: (A) Conditional Generator Model (B) Conditional Discriminator Model.

Data

We assemble a set of 260,764 strong-motion recordings from Japanese seismograph networks K-NET and KiK-Net, corresponding to 6,125 earthquakes. We only use ground surface stations. We use three conditional variables:

- Event-Station Distance, measured from the earthquake hypocenter
- Earthquake Magnitude
- Vs30 at the recording station (a proxy for site response)

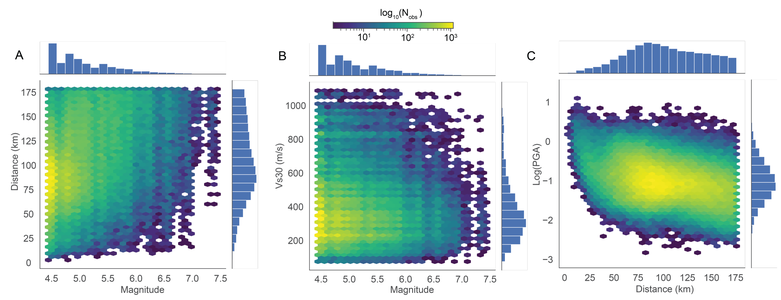

Figure 3: Cross-plots with conditional variable distributions. Hexagonal bins are used to visualize correlations between variables. Each bin is color-coded according to the logarithm of the number of observations that fall within the hexagon..

Model Validation in Time and Frequency Domains

We evaluate our model in the frequency domain (FD) by comparing the average Fourier amplitudes of real and synthetic seismograms. In the time domain (TD), we use average acceleration envelopes as our evaluation metric.

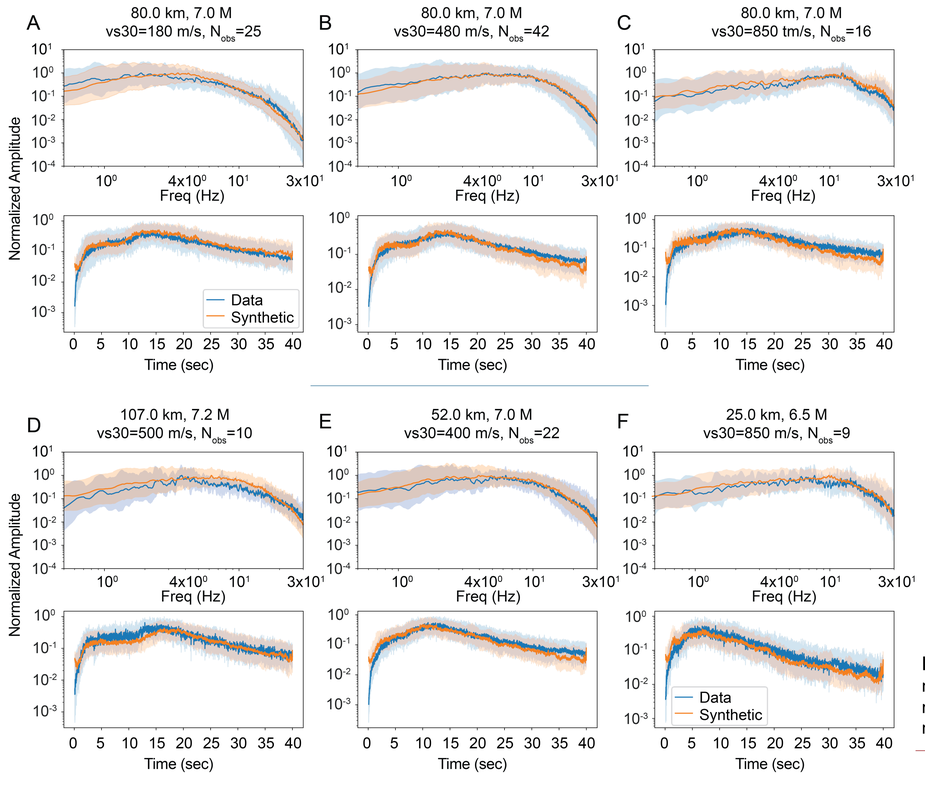

Figure 4: In each panel, average normalized amplitude spectra (upper pane) and average normalized acceleration envelopes (lower pane) are used to contrast real (blue) and synthetic (orange) acceleration time-histories. The title of each panel contains the bin-midpoints used to synthesize accelerograms. Nobs is the number of real observations used to compute statistical averages.

Scaling Properties

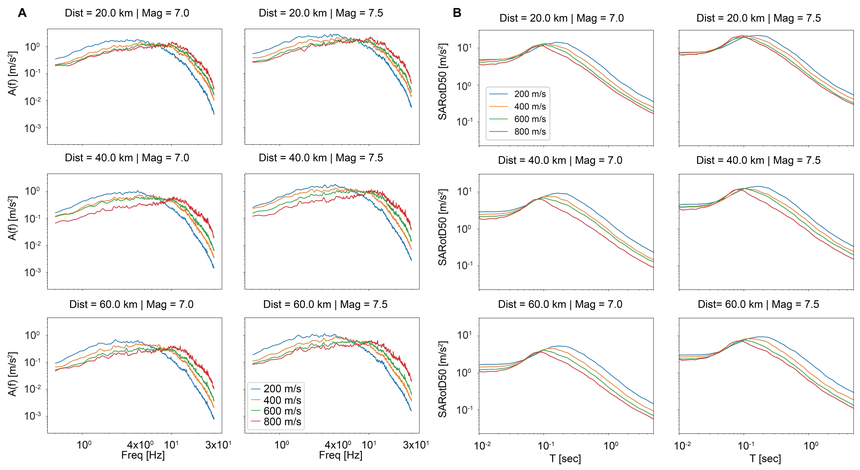

Figure 5: Illustration of our model's scaling properties. (A) In each panel, distance and magnitude are fixed, while Fourier amplitude spectra are displayed for increasing Vs30. (B) Response spectral acceleration from synthetic accelerograms, also for increasing values of Vs30.

Interpolation Experiments

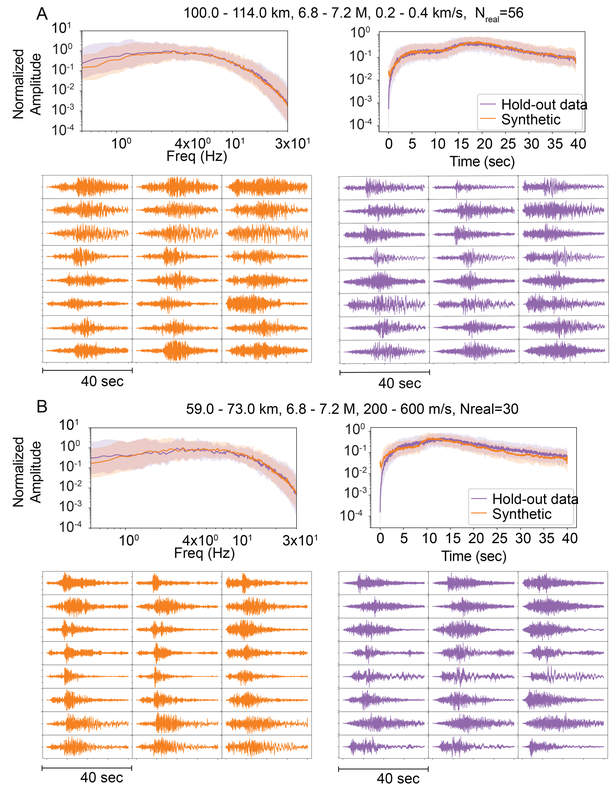

Figure 5: In each experiment, we take our initial training set and select a narrow distance-magnitude-vs30 bin. We remove all accelerograms belonging to it and retrain our model. The removed data is held out for subsequent testing. Once the generator model is trained, we use the bin mid-points to generate N = 256 synthetic seismograms.

Conclusions

- We are able to synthesize realistic 40-second long acceleration time-histories, sampled at 100 Hz, that could be used for engineering purposes.
- Our model directly captures site effects, as parametrized by Vs30.

References

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems 27. Curran Associates, Inc.

Arjovsky, M., Chintala, S., & Bottou, L. (2017, July 17). Wasserstein generative adversarial networks, In International conference on machine learning. ISSN: 1938-7228.