Final Report
on
GROUND MOTION TIME HISTORY: ANALYSIS AND PREDICTION

Submitted by
SAMARTH K J
CE22B099

Under the guidance of

Raghukanth

DR. STG RAGHUKANTH

for
Undergraduate Research (CE4902)

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
CHENNAI 600 036, INDIA

May 2024

Abstract

Earthquake engineering problems involve the use of many functions, such as Response Spectrum, Fourier Amplitude Spectrum, and Evolutionary Power Spectral Density (EPSD), all of which are useful functions derived from the data extracted from ground motion time histories; these functions are used in several civil engineering problems related to earthquakes.

Deterministic and stochastic methods can be used to synthesize earthquake time histories; however, stochastic methods are much more efficient as the time histories rely on many parameters and exhibit inherent stochasticity. Also, stochastic methods allow the coverage of a wider range of frequencies. Several methods were undertaken to generate earthquake time histories and response spectra from a few input parameters.

The initial data visualization and analysis depicted meaningful relationships between earthquake parameters and the ground motion time histories, depicting the tractability of devising statistical methods and machine learning algorithms to perform the mapping between the earthquake parameters and ground motion time histories & response spectra.

Conditional Generative Adversarial Networks (CGAN) were used to generate response spectra, along with Bayesian Neural Networks (BNNs) and Forward Neural Networks (FFNNs), which were used to generate ground motion parameters such as Peak Ground Acceleration (PGA) values and Response Spectra (RS). The R^2 values obtained from these models were >0.75, depicting such models' usability in predicting ground motions.

1. Introduction

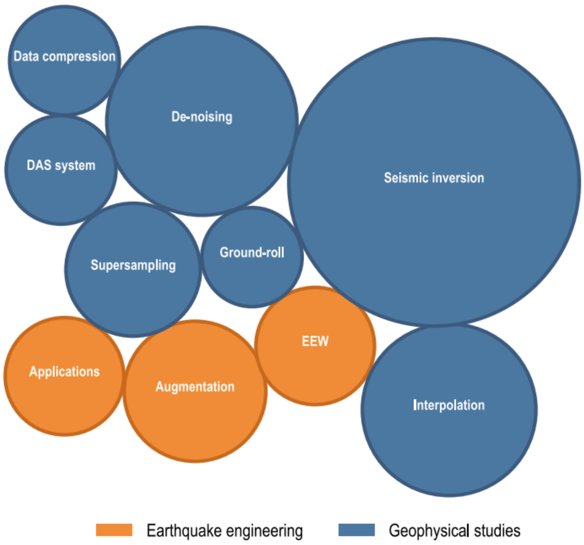
Acceleration time histories are records of ground motion accelerations over time, usually significant during earthquakes. They provide information about the amplitude, frequency and duration of ground motion. They play a vital role in mitigating the impacts of earthquakes and establishing standard design codes for adequate structural resilience. Response Spectra, Fourier Amplitude Spectra and other such techniques for engineering seismic analysis rely on time histories for derivation.

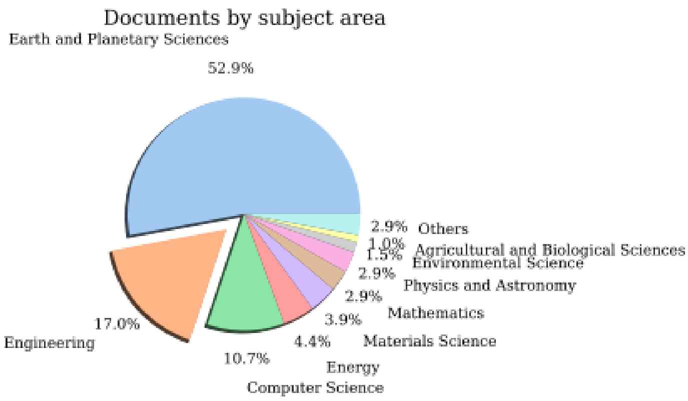
Obtaining a large number of real data can be challenging for many reasons. Usage of machine learning techniques in seismology, geology, or civil and structural engineering is a great way to generate synthetic data.

In this report, data was visualized and interpreted before an attempt was made to use the following machine learning techniques: cGANs, BNNs, and FFNNs, with special emphasis on conditional GANs to predict ground motion during Earthquakes via prediction of response spectra.

2. Literature Review

Giuseppe et al. (2023) reviewed 138 papers related to seismology and GAN. Out of the 138 papers, quite a small fraction corresponded to earthquake engineering based applications and problems; most of the applications in earthquake engineering problems were relatively recent. The paper discussed several variations of GANs, including WGANs, RGANs, RCGANs, conditional DCGANs, and many more alterations of the traditional GAN architecture for different applications within earthquake engineering. Within earthquake engineering, GANs are used for structural health monitoring (SHM), and structural damage detection using various piezometric and optical sensors. Machine learning algorithms, especially anomaly detection techniques, are also used for post-earthquake damage analysis and victim detection using aerial and satellite images. Relatively very few papers had predictions of earthquake ground motion time histories or parameters related to it. The paper also makes the major pitfalls evident, which are (i) lack of high-quality data available to train such models, causing an inherent upper-cap to exist for the quality of models that can be built (ii) analytical or numerical synthetic seismic data generated, with traditional methods may be unable to capture all the real-world aspects (Kaur et al. 2020; Ovcharenko et al. 2021; Wang et al. 2022). Thus, the models trained on synthetic data have an inherent bias, a problematic limitation. (iii) A well-trained GAN is virtually able to capture the probability distribution of the input data successfully; its limited extrapolation capabilities represent another possible drawback. (iv) lack of interpretability of GANs, and virtually any ML model is another drawback.





Yinjun et al. (2020) used conditional GANs to predict the aftershocks of earthquakes, arguing their importance in the seismic design of structures. They obtained promising results when the ground motion prediction equation by Abrahamson et al. (ASK14) was selected to compare with the CGAN model, and it was seen the CGAN model matches the

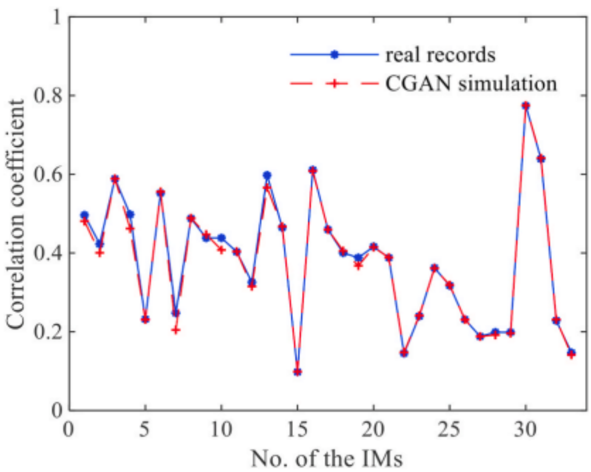
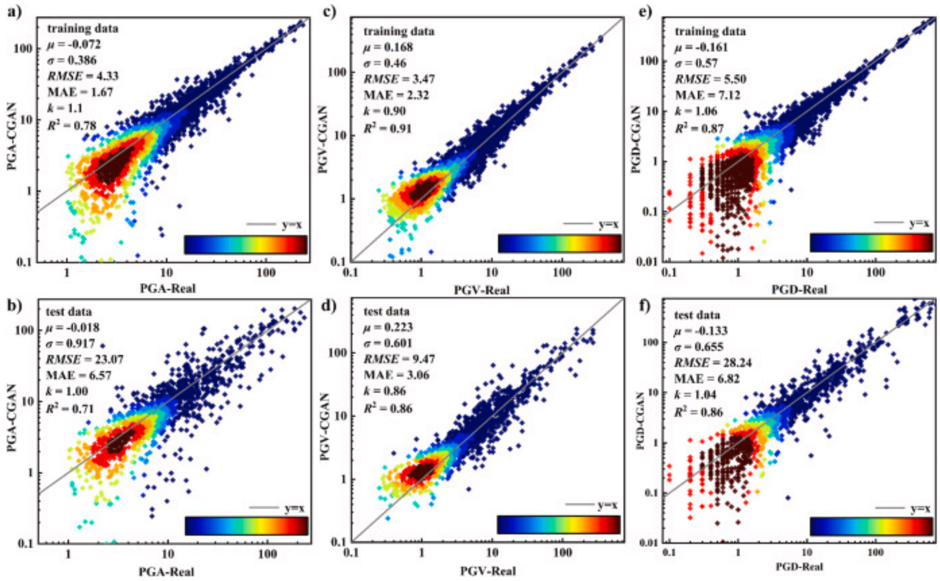


Fig. 5. Linear correlation coefficient of the MS–AS IMs.

Intensity Measures (IMs) of ASs better than the former.

Shuqian et al. (2024) used the following intensity measures: peak ground acceleration (PGA), peak ground velocity (PGV), and peak ground displacement (PGD) as their target variables (variables which were to be predicted) for training a conditional GAN. The trained model performed better than the pre-existing ground motion prediction equations, depicting the usefulness of cGANs in ground motion prediction. The CGAN model they trained had a higher predictive accuracy for intensity measures with larger values, suggesting that when the intensity parameter values are small, their impact on the intensity measure diminishes, causing the model to overlook the corresponding errors during the automatic parameter optimization process. Adjusting the data distribution range can enhance prediction results.



3. Knowledge Gaps and Research Needs

From the literature review, the following knowledge gaps and research needs were identified:

a. There is a significant lack of data available for earthquake engineering purposes. Thus, there is a growing need for synthetic data that is true to the real data (their distributions need to match)
b. There has been meager research concerning the usage of machine learning techniques to improve the field of earthquake engineering. Very few attempts have been made to predict intensity measures of earthquakes using machine learning algorithms like cGANs.
c. cGAN and other machine learning techniques have showcased promising results in other applications of seismic parameter prediction.

4. Objective and Scope

a. Pre-process earthquake time history data using methods such as cleaning, transforming, and reducing the data for suitable analysis.

b. To visualize and analyze the processed data to derive meaningful inferences and interpretations concerning different data features.
c. To produce synthetic time history data by producing artificial intensity measures like response spectra and PGA values, utilizing the response spectra and PGA values obtained from past earthquakes in different regions using machine learning techniques such as FFNNs, BNNs, and primarily CGANs.
d. To investigate the effectiveness of these models, with emphasis on cGAN, and suggest improvements to generate more accurate data.

5. Work Done and Key Findings

A. Data Pre-Processing

Preprocessing data is crucial for enhancing model performance and interpretability.

Preprocessing ensures data consistency, removes noise, and reduces computational complexity. It also helps in feature engineering by extracting relevant information and reducing dimensionality.

a. Data Cleaning

All the data with missing values of important parameters were eliminated.

b. Data Transformation

The data was normalized by eliminating NaN values and exceptionally large values to the required extent by employing adequate methodologies such as interpolation and elimination.

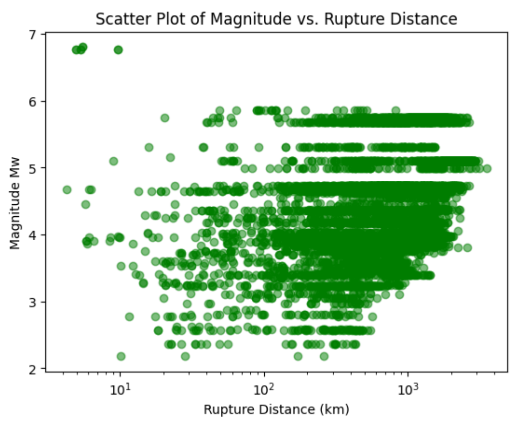
c. Data Reduction

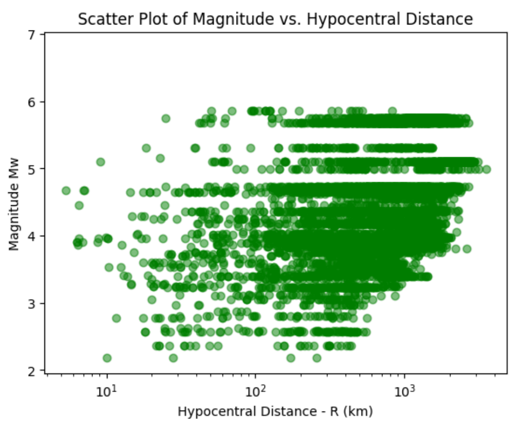
Parameters that were not useful for the training of models were eliminated. These include parameters like Earthquake ID, location, etc.

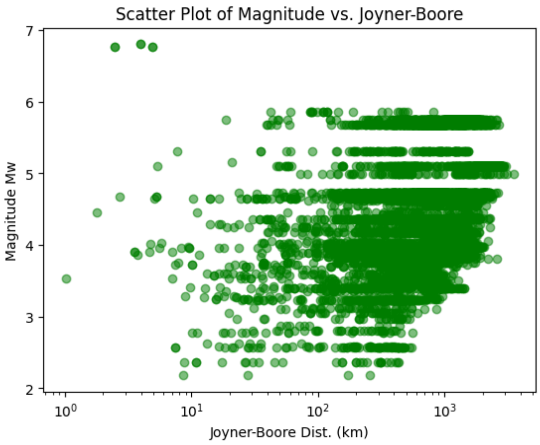
B. Data Visualisation

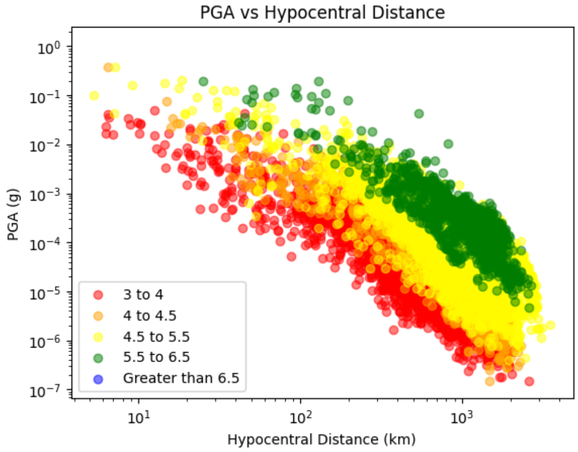
The following are some plots obtained from analyzing different features of the dataset:

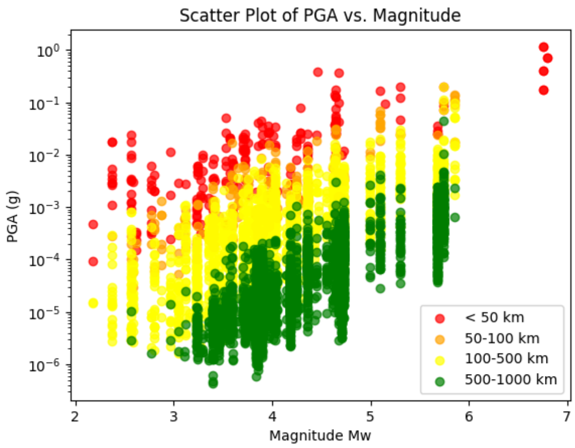
a. Scatter Plots:



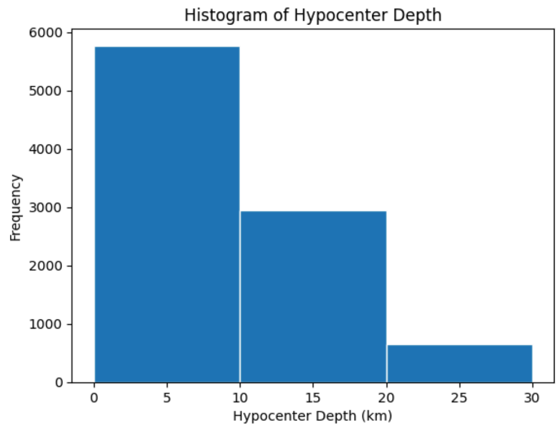


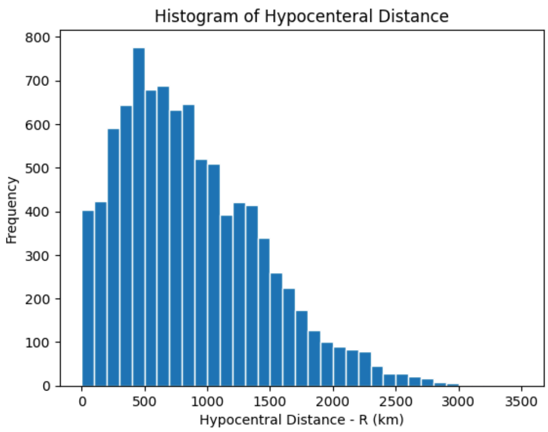


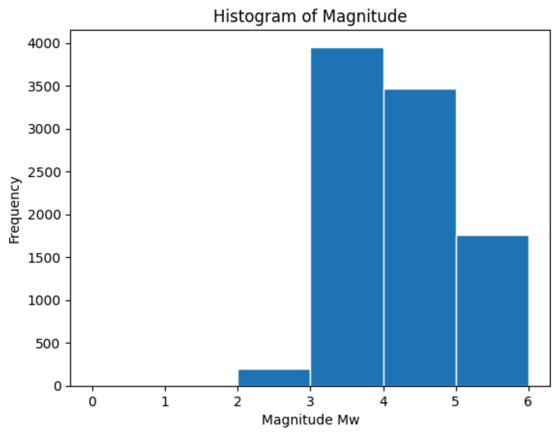




b. Histograms





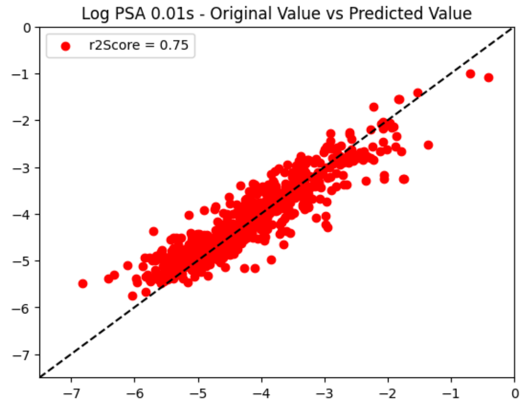


C. Prediction of PGA with FFNN

A Feed Forward Neural Network processes information in a one-way flow, unlike RNNs or LSTMs, from input nodes through hidden layers to output nodes.

Input nodes receive data, hidden layers perform computations using weighted sums and activation functions to introduce non-linearity, and output nodes produce final predictions.

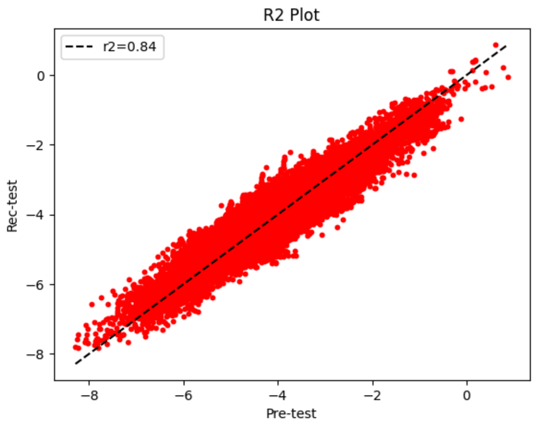
Weights and biases govern the strength of connections between nodes, while activation functions enable the network to approximate complex relationships. During training, backpropagation adjusts weights and biases to minimize prediction errors.



To train the model, 5 input parameters were used: Earthquake Magnitude, Joyner-Boore distance, the logarithm of Joyner-Boore distance, Mechanism Based on Rake Angle, Preferred shear wave velocity (VS30)

Predicting the logarithm of pseudo-absolute-acceleration spectra (PSA) (an intensity measure) using an FFNN, an R^2 value of 0.75 was obtained.

D. Prediction of PGA using BNN



Bayesian neural networks (BNNs) are a type of neural network that incorporates Bayesian inference into the training and prediction process. In traditional neural networks, parameters (weights and biases) are usually deterministic values learned from data using techniques like gradient descent. In contrast, BNNs treat parameters as probability distributions rather than fixed values.

When BNNs are used to predict PGA values from Joyner-Boore distance, the logarithm of Joyner-Boore distance, the magnitude of the earthquake, Mechanism Based on Rake Angle, and shear wave velocity (VS30), an R^2 value of 0.84 was obtained.

E. Prediction of Response Spectra using cGAN

A Generative Adversarial Network consists of two components, which can be any two models (mathematical/statistical/deep). It is an ingenious way to train computers into producing/generating data from scratch. The two components of a GAN are:

- **a. Generator:** It takes random noise (often sampled from a simple distribution like Gaussian) as input and transforms it into data samples that resemble real data.
- **b. Discriminator/Adversary:** The discriminator acts as a critic that evaluates the realism of the generated samples. It takes both real data samples and generated data samples as input and learns to distinguish between them.

Simultaneously, the generator aims to fool the discriminator into classifying its generated samples as real.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right]$$

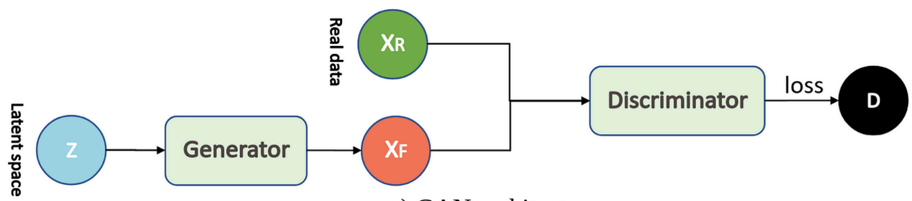
This is the loss function for GANs. It is defined such that the generator tries to increase the loss while the discriminator tries to reduce the loss. This process drives the generator and the discriminator to improve significantly over training iterations. This is also why GANs take significantly longer to train and generate good results.

The response spectra of earthquakes were generated using a conditional GAN.

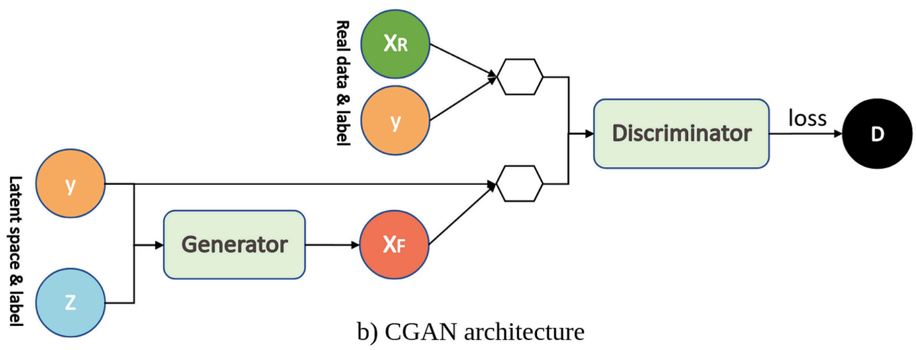
Conditional GANs (cGANs) improve upon traditional GANs by incorporating additional conditioning information, such as class labels, which guides the generation process. This allows for targeted generation of specific classes or attributes, enhancing the model's control over output.

cGANs produce more structured and meaningful results compared to traditional GANs, making them suitable for tasks like response spectra generation, where conditioning information is crucial for generating output that is required for specific earthquake parameters.

The difference in architecture between GAN and CGAN is as follows:



a) GAN architecture



b) CGAN architecture

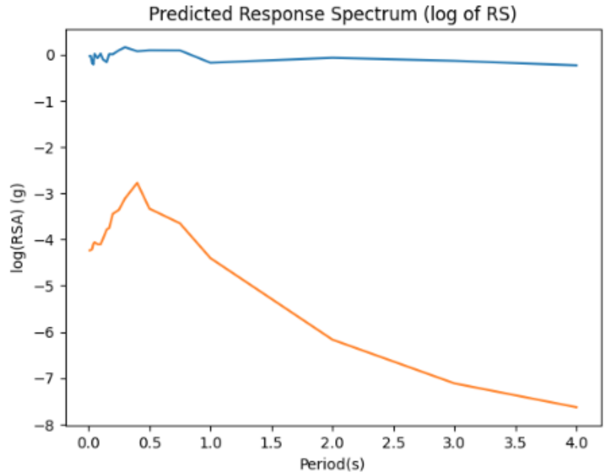
We will pass in 6 earthquake parameters: hypocentral Joyner-Boore distance, logarithm of Joyner-Boore distance (even though these parameters are related, they help create better models), shear wave velocity (VS30), focal mechanism, magnitude of the earthquake and depth.

Using these inputs, the aim is to generate the corresponding response spectrum.

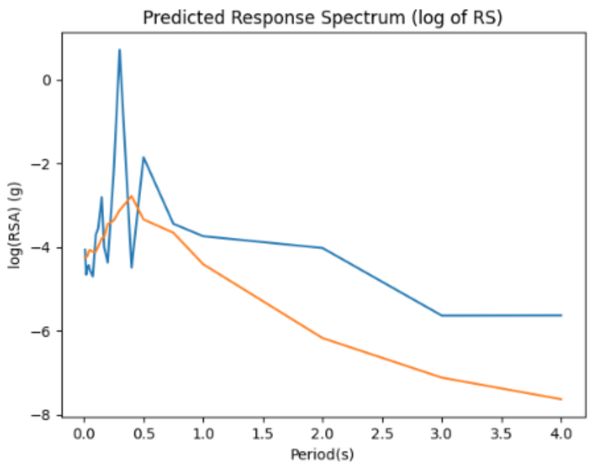
It generally takes >1000 epochs to generate a good conditional GAN; since there wasn't enough computational power, the current GAN was trained for 60 epochs.

The following were some of the results obtained. The yellow line pertains to the generic shape of the response spectrum. It is to be noted that the algorithm isn't trying to match the exact distribution in yellow, but it serves as a reference to what the predicted response spectrum should look like.

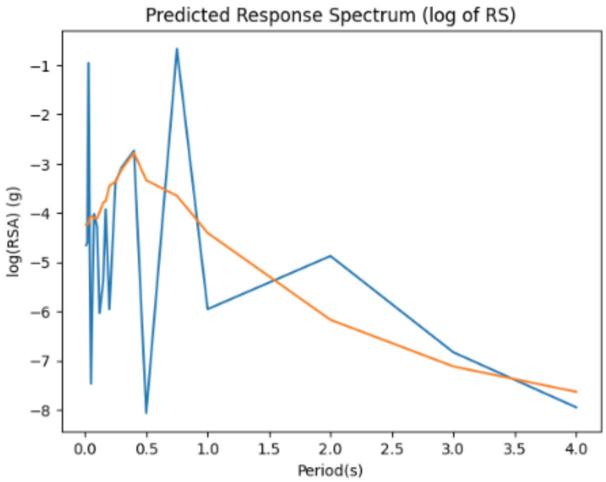
A general response spectrum starts low, shoots up, stabilizes, and decreases conclusively; there can be sharp peaks and troughs throughout the spectrum.



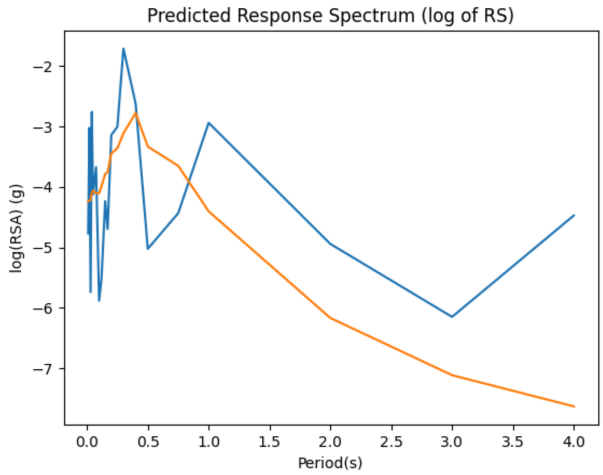
Epoch -1



Epoch - 10

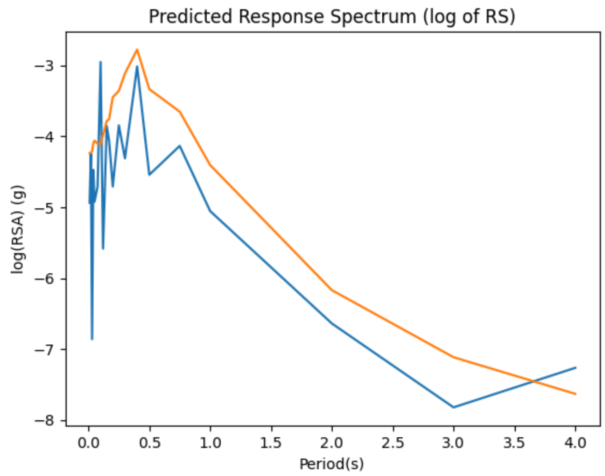


Epoch - 20

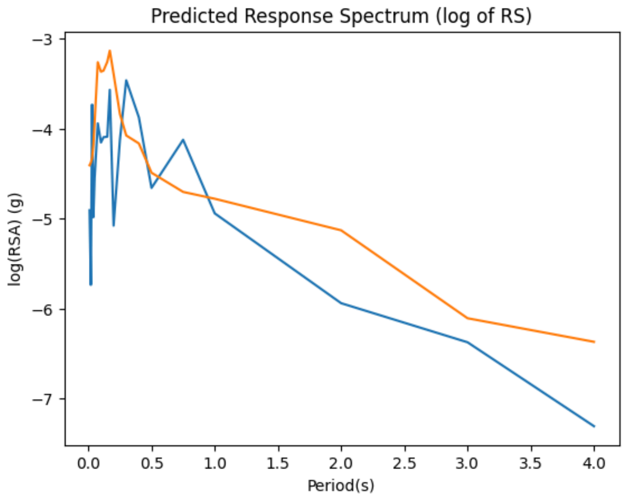


Epoch - 30

Random response spectra generated by the generator are in blue, the shape of which gets progressively resemblant to the actual shape of the response spectra.



Epoch - 40



Epoch - 50

6. Future Work

The cGAN can be further improved upon by the following techniques:

- Training the model for much longer, to at least 500 epochs, using GPUs.
- Using convolutional layers instead of linear layers as cGANs perform exceptionally well with multidimensional data, according to the papers reviewed.
- Using more input parameters for the model to derive more information to generate more accurate results.
- Altering model architectures to suit better the purpose, normalization of the training data and alteration of training procedure.

7. Learning Outcomes and Conclusions

From the data visualizations and analysis, the following can be deduced:

- Several data are invalid due to absent values, NaN values, and anomalously high values. These data need to be eliminated as they are not useful.
- PGA increases with an increase in earthquake magnitude and decreases with an increase in hypocentral distance, regardless of the earthquake's magnitude.
- The histograms depict the distribution of all the data collected:
  - Most data are of earthquakes from magnitude 3-4 $M_{w.}$
  - The hypocentral distances of all the earthquakes recorded lie in the 0-1000m range.
  - Most earthquakes recorded have hypocenter depth up to 10km.

  Data cleaning, transformation, and reduction were performed using these results to train machine learning models.

  The ML techniques cGAN, BNN, and ANN showcased promising results:

- FFNN had an R^2 value of 0.75 for PSA prediction.

- BNN had an R^2 value of 0.84 for PGA prediction.
- cGAN showed response spectrum values based on input parameters resembling the real response spectrum values.
    - R^2 is not used to evaluate cGANs as the R^2 metric is used for continuous outputs, while cGANs generate images or sequential outputs.

8. References

Sreejaya, K.P., Basu, J., Raghukanth, S.T.G. et al. Prediction of Ground Motion Intensity Measures Using an Artificial Neural Network. Pure Appl. Geophys. 178, 2025–2058 (2021). https://doi.org/10.1007/s00024-021-02752-9

Meenakshi Y, Sreenath V, STG R. Ground motion models for Fourier amplitude spectra and response spectra using Machine learning techniques. Earthquake Engng Struct Dyn. 2024; 53: 756–783. https://doi.org/10.1002/eqe.4036

Meenakshi Y, Vemula S, Alne A, Raghukanth S. Ground motion model for Peninsular India using an artificial neural network. Earthquake Spectra. 2023;39(1):596-633. doi:10.1177/87552930221144330

Dhanya, J., Raghukanth, S.T.G. Ground Motion Prediction Model Using Artificial Neural Network. *Pure Appl. Geophys.* **175**, 1035–1064 (2018). https://doi.org/10.1007/s00024-017-1751-3

Marano, G.C., Rosso, M.M., Aloisio, A. et al. Generative adversarial networks review in earthquake-related engineering fields. Bull Earthquake Eng 22, 3511–3562 (2024). https://doi.org/10.1007/s10518-023-01645-7

Ding, Yinjun & Chen, Jun & Shen, Jiaxu. (2020). Conditional generative adversarial network model for simulating intensity measures of aftershocks. Soil Dynamics and Earthquake Engineering. 139. 106281. 10.1016/j.soildyn.2020.106281.

Duan, S., Song, Z., Shen, J., Xiong, J.\ 2024.\ Prediction for underground seismic intensity measures using conditional generative adversarial networks.\ Soil Dynamics and Earthquake Engineering 180. doi:10.1016/j.soildyn.2024.108619

Raghukanth