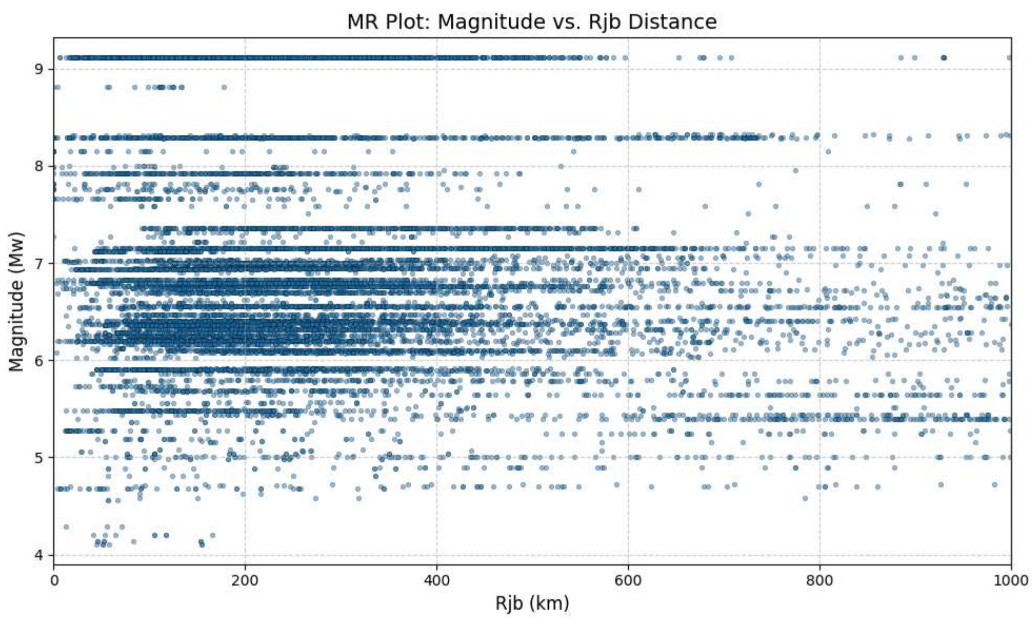
Prediction of Spectral Acceleration Using CGAN

1. Introduction

This study develops a CGAN model to predict 20 spectral acceleration (SA) values based on five input ground motion features: magnitude (mag), rupture distance (rjb), logrjb , logvs30 , and event type (inter-intra). The model includes a careful preprocessing pipeline, model training with early stopping, residual decomposition using mixed-effects modeling, Residual analysis,Ground motion physics,Importance,SHAP analysis for explainability.

2.Magnitude vs Rjb Scatter Plot:

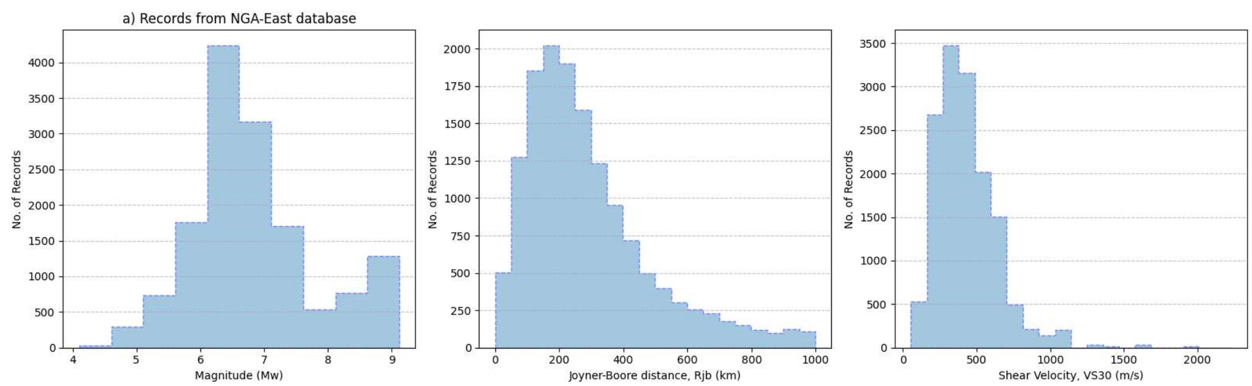
This scatter plot visualizes the distribution of events across different magnitude (mag) and Joyner-Boore distance (rjb) combinations in the dataset used for training and evaluation.



- The plot shows a dense cluster of data points for **moderate magnitudes (5.0–6.5)** and **short-to-moderate distances (0–100 km)**, which is typical of recorded ground motion datasets like NGA.
- Fewer data points appear at **larger distances (>200 km)** or for **larger magnitudes (>7.0)**, consistent with the relative rarity of such records.
- The coverage ensures that the model is well-trained across the critical near-field range but may have increased uncertainty for predictions at far distances or large magnitudes due to data sparsity.

3.Histograms of Input Features:

This figure presents histograms of three key input parameters—Moment Magnitude (Mw), Joyner-Boore distance (Rjb), and Shear-wave velocity at 30 m depth (Vs30)—from the NGA-East database used in this study.



- **Magnitude (Mw)** is concentrated around 6.0–6.5, reflecting a dataset dominated by moderate earthquakes.
- **Rjb** is right-skewed, with most recordings within 0–300 km, ensuring good coverage of near-field motions.
- **Vs30** peaks around 300–500 m/s, indicating a prevalence of stiff soil and soft rock sites in the data.

4.Summary Statistics of Input and Output:

Input Parameters:

Parameter	mag	rjb	logrjb	logvs30	intra_inter
min	4.1	0.01	-2	1.7243	0
max	9.12	999.0898	2.9996	3.3483	1
mean	6.8318	289.7475	2.352	2.5906	0.4232
std	1.0028	196.9747	0.3695	0.2032	0.4941
skewness	0.7859	1.2926	-3.3307	-0.087	0.3107
kurtosis	0.3906	1.535	33.8885	0.1169	-1.9035

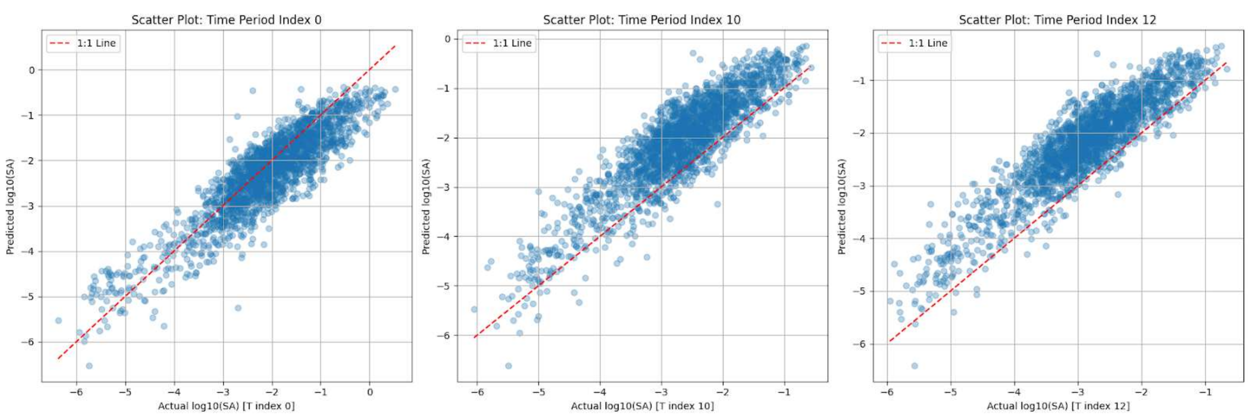
- **Magnitude (mag):** Ranges from 4.1 to 9.12, with a mean of 6.83, showing variability in seismic event intensity. Slight positive skew (0.79) and near-normal distribution.
- **Rupture Distance (rjb):** Varies widely from 0.01 to 999.09, with a mean of 289.75, showing high variability and positive skew (1.29).
- **Log of Rupture Distance (logrjb):** Range from -2.00 to 2.99, mean of 2.35, with a highly negative skew (-3.33) and heavy-tailed distribution (high kurtosis).
- **Log of Shear-Wave Velocity (logvs30):** Ranges from 1.72 to 3.35, with a mean of 2.59, close to normal distribution.
- **Intra-Inter Event Flag (intra_inter):** Ranges from 0.00 to 1.00, with a mean of 0.42, indicating mixed intra- and inter-event data, with light tails in distribution

Output Parameters:

Parameter	T0pt010S	T0pt020S	T0pt030S	T0pt050S	T0pt075S	T0pt100S	T0pt150S	T0pt200S	T0pt300S	T0pt400S	T0pt500S	T0pt750S	T1pt000S	T1pt500S	T2pt000S	T2pt500S	T3pt000S	T3pt500S	T4pt000S	T5pt000S
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
max	2.5801	2.7391	3.5567	4.9801	5.9791	3.6631	5.8752	6.2565	5.252	4.234	3.0608	2.27	1.2481	1.3501	1.2663	0.6708	0.3824	0.3857	0.2931	0.2265
mean	0.0304	0.0311	0.033	0.0396	0.0499	0.0608	0.0715	0.0738	0.0678	0.0591	0.0515	0.0382	0.03	0.0198	0.0143	0.0108	0.0084	0.0068	0.0055	0.0039
std	0.085	0.0884	0.099	0.128	0.1542	0.1829	0.2191	0.2259	0.2036	0.1681	0.1412	0.0969	0.074	0.0509	0.0373	0.0272	0.0213	0.0176	0.014	0.0098
skewness	8.2602	8.5575	10.0383	11.3257	10.0677	7.6269	8.5839	8.8185	8.9577	7.8704	6.9268	6.591	6.028	7.8387	8.7858	6.9784	6.2945	6.8688	6.221	5.9755
kurtosis	120.298	128.9357	190.275	242.2775	208.1898	82.9457	117.0851	124.5151	131.3643	99.5985	68.6172	68.6205	51.906	106.6552	156.0315	87.9723	60.6166	79.6645	60.9231	57.1025

Most parameters show high skewness (>7) and heavy kurtosis, suggesting significant outliers and concentrated distributions around low values. Parameters like **T0pt010S to T0pt100S** have lower mean values, while others (e.g., **T0pt150S to T0pt500S**) show increasing variability.

5.Plots of Actual vs Predicted log10(SA) Across Time Periods:



The scatter plots show actual vs. predicted log10(SA) for the Gradient Boosting model at time period indices 0, 10, and 18. Across all periods, predictions align well with the 1:1 line, indicating strong model performance.

Index 0:

- Predictions tightly clustered around 1:1 line.
- Indicates **high accuracy** and **low error**.

Index 10:

- Slightly more scatter around 1:1 line.
- **Moderate accuracy**, with some **under-prediction**.

Index 12:

- Increased scatter and deviation from 1:1 line.

- **Higher prediction error** and **variability**, especially at low SA values.

Overall:

- CGAN performs well across periods.
- Accuracy **decreases with increasing period index.**

6.Model Architecture:

Generator

- Input: 10 features (5 noise + 5 ground motion inputs)
- Layers:
  - Linear(10 → 16) → ReLU → Dropout(0.2)
  - Linear(16 → 32) → ReLU → Dropout(0.2)
  - Linear(32 → 20) → Output: 20 log(PSa) values
- Purpose: Generates PSa values conditioned on input features.

Discriminator

- Input: 25 features (20 PSa + 5 conditions)
- Layers:
  - Linear(25 → 16) → LeakyReLU(0.2) → Dropout(0.2)
  - Linear(16 → 1) → Sigmoid
- Purpose: Classifies PSa as real or fake, given conditions.

Highlights

- Conditioning: Achieved by concatenating input features with noise/PSa.
- Regularization: Dropout prevents overfitting.
- Activation: ReLU for generator, LeakyReLU + Sigmoid for discriminator.

7.Model Performance Metrics for Target Variables:

Total Standard Deviation (σ):

- Ranges from **0.329 to 0.471**.
- Peaks around **Period Index 5**, indicating higher uncertainty in mid-period spectral accelerations.

Intra-event Standard Deviation (ϕ):

- Dominates the total σ across all periods (**ϕ ≈ σ**).
- Suggests the model captures **site-specific and path-related variability** effectively.

Inter-event Standard Deviation (τ):

- Remains consistently low (**0.016 to 0.093**).
- Indicates the model captures **very limited event-to-event variation**.
- May imply **insufficient use of event-level features** (e.g., magnitude, fault type) in the conditioning.

Period Index	Total Std Dev (σ)	Inter-event Std Dev (τ)	Intra-event Std Dev (ϕ)
0	0.396	0.045	0.395
1	0.398	0.083	0.393
2	0.423	0.03	0.422
3	0.412	0.027	0.412
4	0.442	0.029	0.441
5	0.471	0.073	0.468
6	0.456	0.036	0.455
7	0.457	0.093	0.452
8	0.423	0.063	0.42
9	0.411	0.079	0.407
10	0.401	0.084	0.397
11	0.392	0.03	0.391
12	0.401	0.068	0.398
13	0.391	0.034	0.39
14	0.414	0.016	0.414
15	0.365	0.071	0.362
16	0.371	0.082	0.367
17	0.366	0.082	0.361
18	0.353	0.063	0.35
19	0.329	0.021	0.328

8.Residual Analysis:

1. Inter-event Residual vs. Magnitude (Mw)

- **Subplots**: Top row (PSa at 0.1s, 1.0s, 3.0s).
- **Interpretation**:
  - Residuals are centered around zero across all magnitude bins.
  - No significant bias observed—indicates **stable prediction across magnitudes**.
  - Error bars remain consistent, suggesting **uniform performance for small and large events**.

2. Intra-event Residual vs. Distance (Rjb)

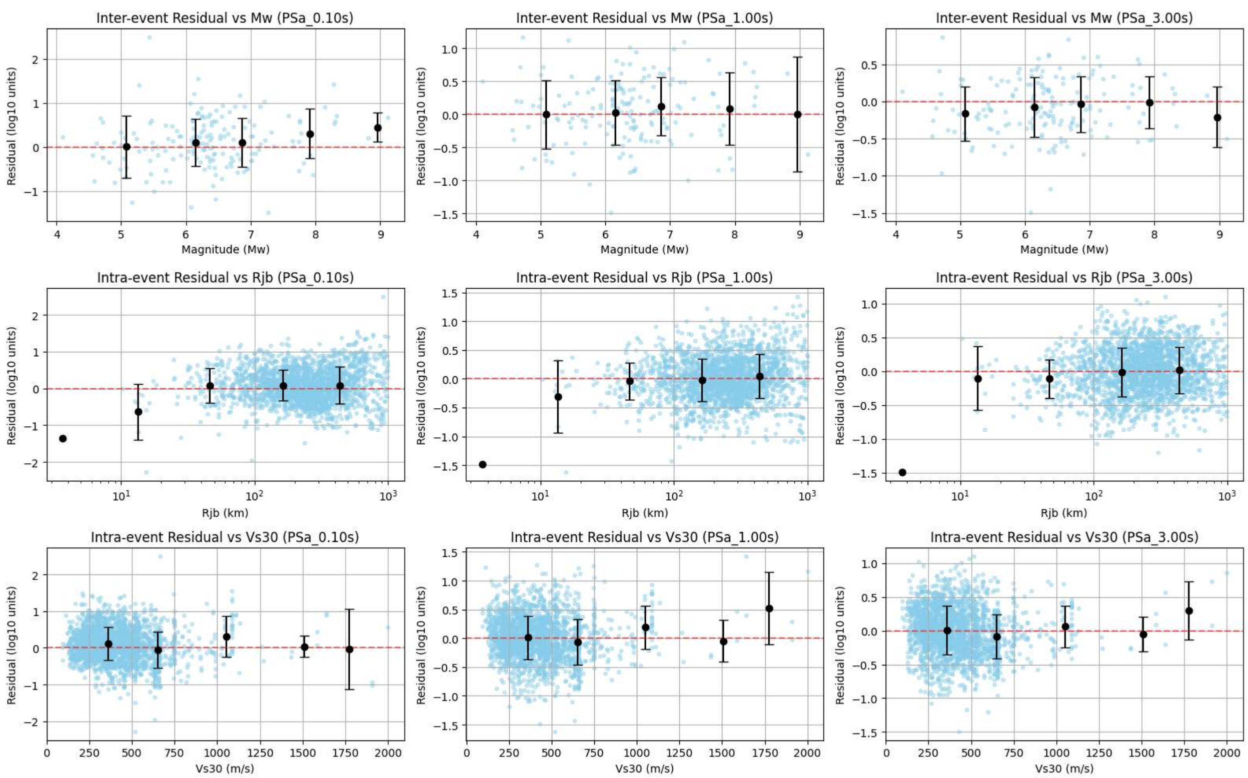
- **Subplots**: Middle row (PSa at 0.1s, 1.0s, 3.0s).
- **Interpretation**:
  - Residuals generally cluster around zero, with no major systematic bias.
  - Slight increase in spread at large distances, especially at short periods (0.1s), likely due to **signal attenuation and data sparsity**.
  - Overall, the CGAN model maintains **robust distance scaling**.

3. Intra-event Residual vs. Vs30 (Site Condition)

- **Subplots**: Bottom row (PSa at 0.1s, 1.0s, 3.0s).
- **Interpretation**:
    - Residuals are centered with no strong trend, indicating **no bias related to site condition**.
    - Slightly more variability at high Vs30 (very stiff/rocky sites), but within acceptable range.
    - Suggests the model **handles site amplification effects effectively**.

Overall Summary

- The CGAN model shows:
    - **No systematic bias** across magnitude, distance, or site condition.
    - **Consistent residual spread**, with expected increases at large distances and extreme Vs30.
    - **Good generalization and physical consistency**, supporting its reliability in ground motion prediction.



9.Magnitude Sensitivity Plot:

1. PSA vs. Time Period Trend:

- For each magnitude (Mw = 3, 4, 5), PSA rises with increasing period, peaks around 0.1–0.3s, then gradually declines.

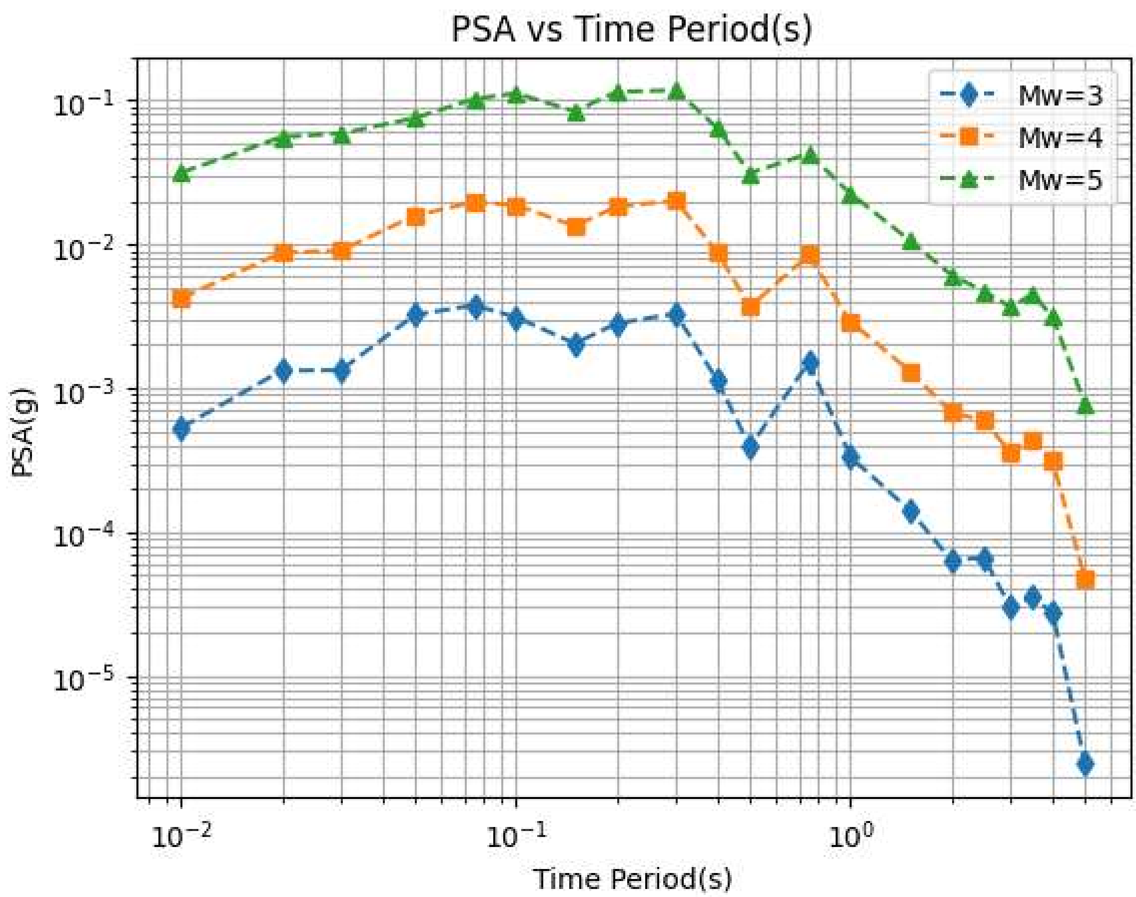
- This bell-shaped trend is typical and reflects the natural frequency response of the ground.

2. Effect of Magnitude:

- At all time periods, PSA is higher for larger magnitudes:
    - Mw = 5 (green) > Mw = 4 (orange) > Mw = 3 (blue).
- Larger events generate more energy, leading to higher spectral accelerations.

3. Log-Log Scale Observations

- Both axes use logarithmic scaling to clearly show variations over wide ranges.
- PSA differences become more pronounced at short periods.



10.Rjb Sensitivity Plot:

1. PSA vs. Time Period Behavior

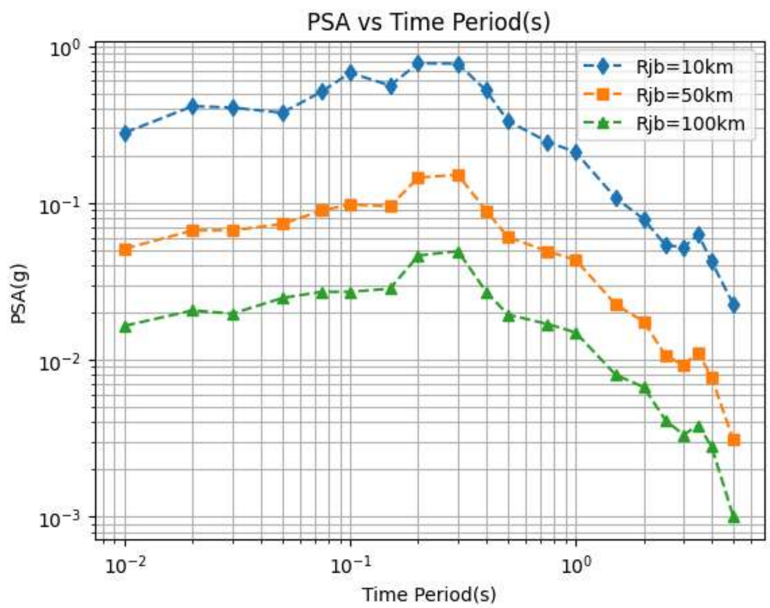
- All curves show a rise in PSA up to ~0.1–0.3s (typical of resonance range), followed by a decline at longer periods.
- This trend reflects frequency-dependent energy distribution in seismic waves.

2. Impact of Distance (Rjb)

- PSA is **highest at Rjb = 10 km**, followed by 50 km, then 100 km.
- This reflects **attenuation of seismic energy** with distance due to geometric spreading and anelastic damping.

3. Log-Log Scale Use

- The log-log scale allows clear visualization of PSA across wide period and amplitude ranges.
- Distance-based separation is more pronounced at shorter periods (<0.3s), indicating **greater attenuation of high-frequency energy** with distance.



11.Vs30 Sensitivity Plot:

1. PSA vs. Time Period

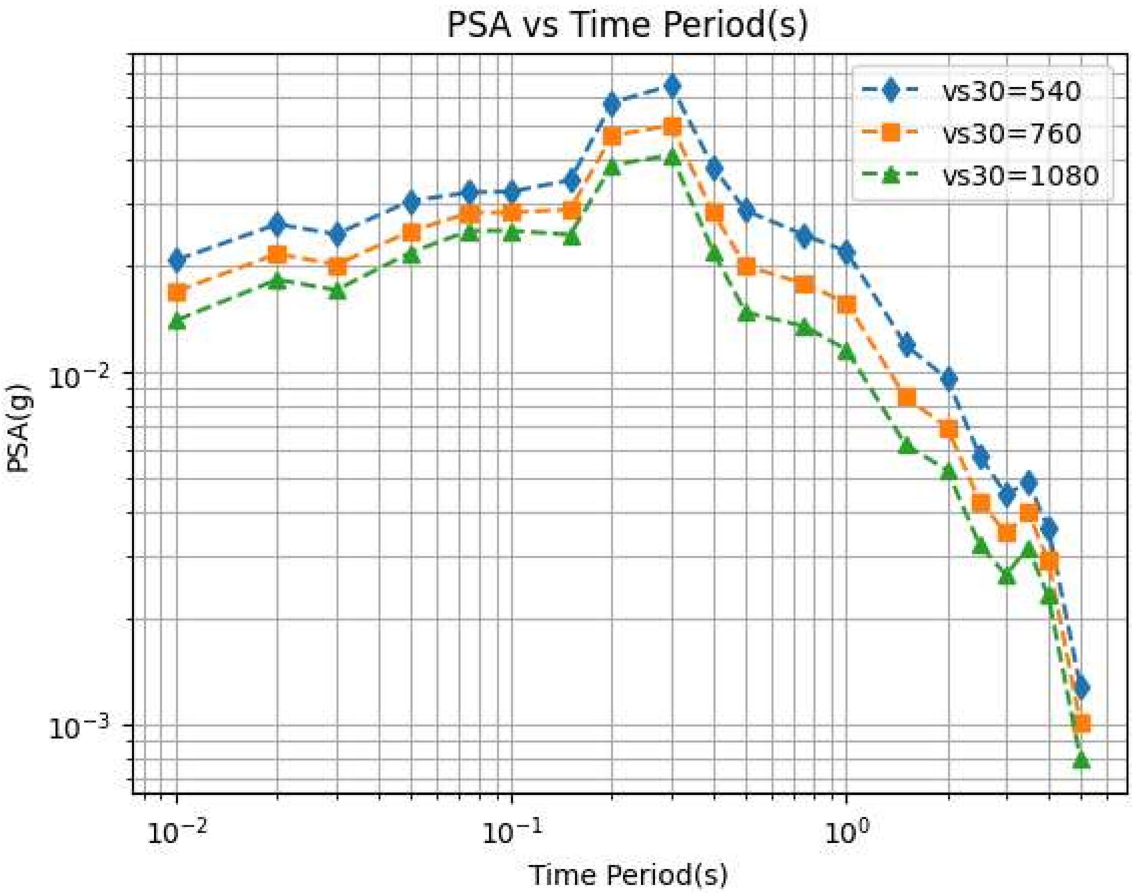
- All curves rise to a peak PSA around 0.2–0.3s, followed by a steady decline.
- This reflects the amplification of certain frequency ranges based on site properties.

2. Influence of Vs30

- Vs30 = 540 m/s (softer soil) shows the highest PSA, while Vs30 = 1080 m/s (stiffer/rocky site) shows the lowest.
- Lower Vs30 corresponds to greater amplification of seismic waves, especially at short to mid periods.

3. Log Scale Insights

- The log-log scale helps illustrate subtle differences in PSA behavior across periods and site types.
- Differences in PSA are more evident at shorter periods.



12.SA @ T vs Rjb:

- **Inter-event (τ) Components: SA vs Rjb**

  1. General Trend

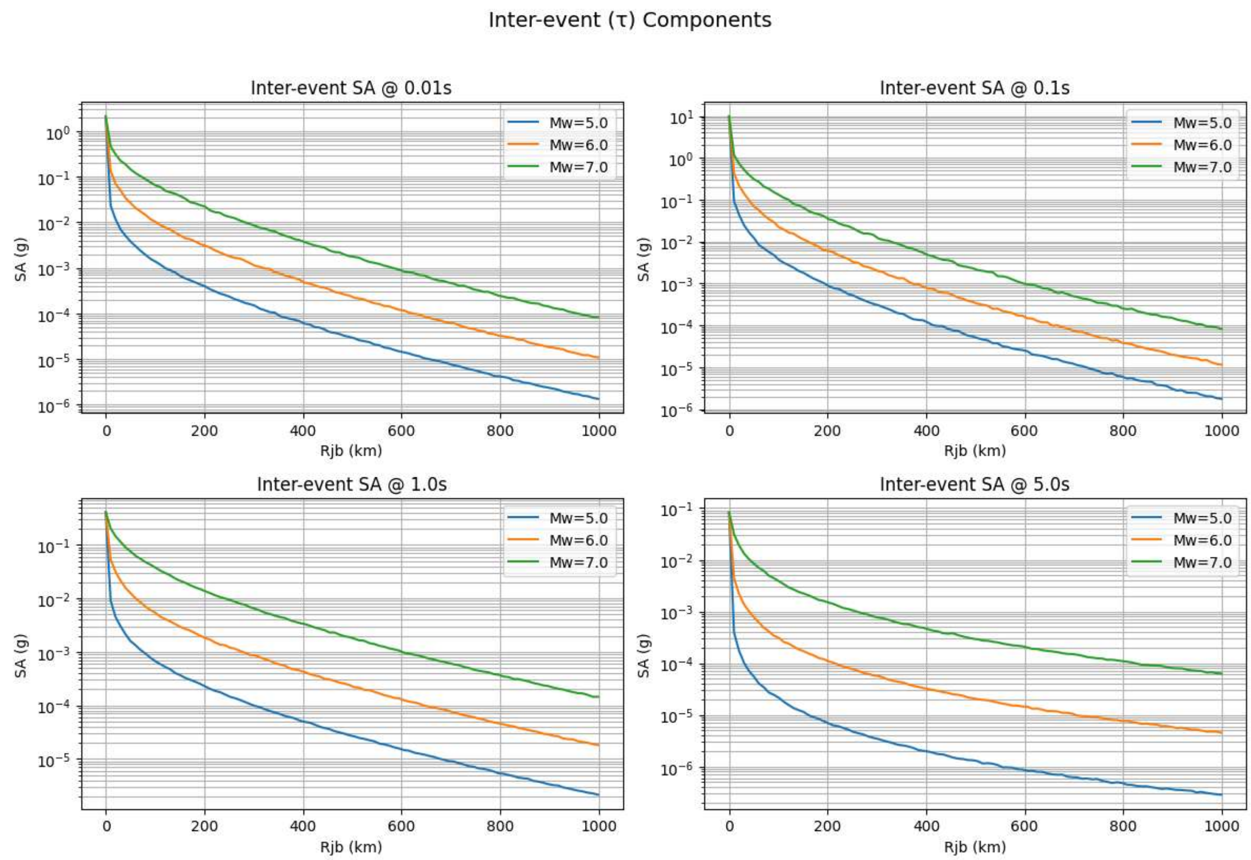
  - All plots show a **rapid decay of SA with increasing Rjb** (distance from the rupture), which aligns with attenuation behavior in ground motion.
  - This decay is sharper at **short distances**, especially below 100 km.

  2. Effect of Magnitude (Mw)

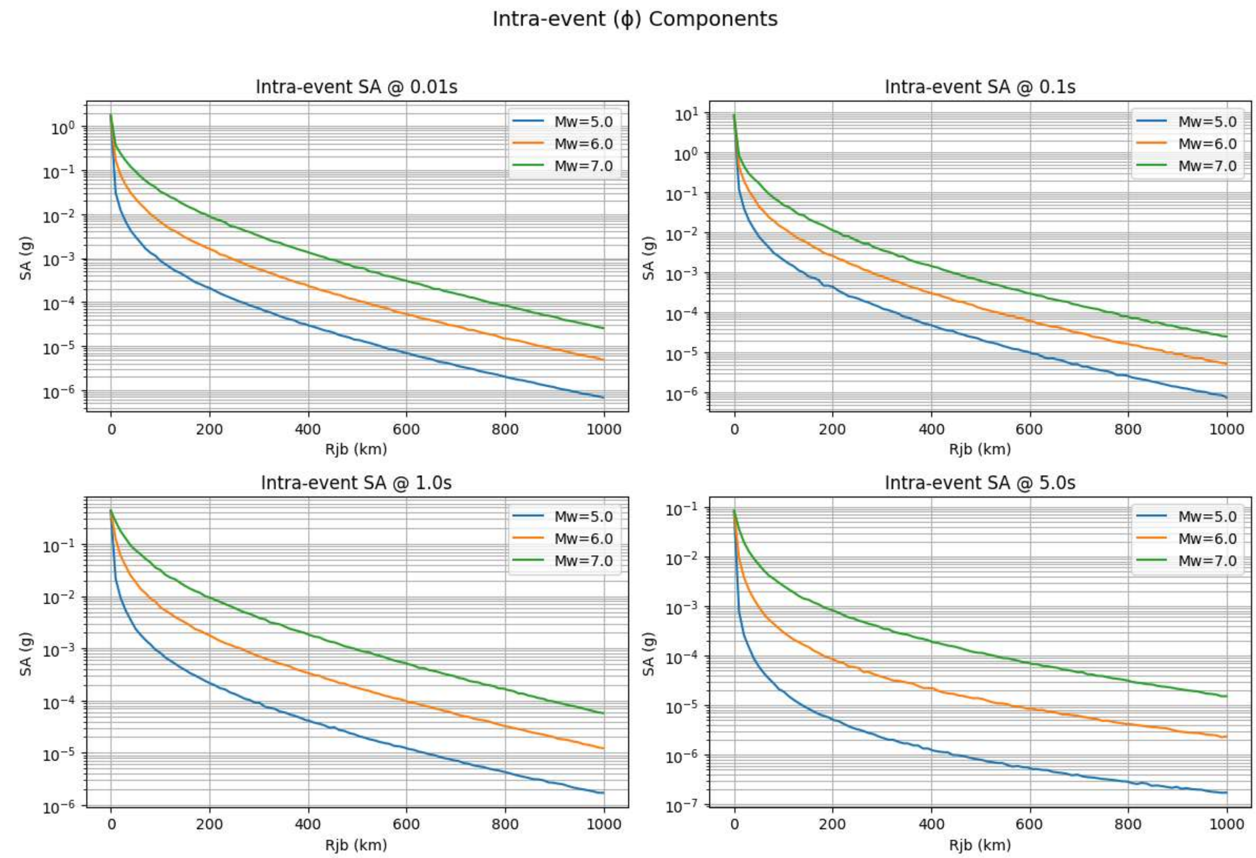
  - For each time period (0.01s, 0.1s, 1.0s, 5.0s), SA is consistently **higher for larger Mw** (Mw=7 > Mw=6 > Mw=5).
  - Larger earthquakes release more energy, hence **greater inter-event SA**.

  3. Period Dependency

- At **short periods (0.01s, 0.1s)**: Ground motion values are **higher overall**, indicating more energy at high frequencies.
- At **long periods (1.0s, 5.0s)**: SA values are lower, and differences between magnitudes are still visible but less steep.



- **Intra-event Components: SA vs Rjb**
  - The **ϕ component** represents the **variability within a single event**, such as site effects or path effects not captured by deterministic models.
  - Its decay with distance suggests that **local site conditions** dominate variability closer to the rupture.
  - At large distances, intra-event variability reduces, likely due to wave dispersion and attenuation smoothing out localized differences.

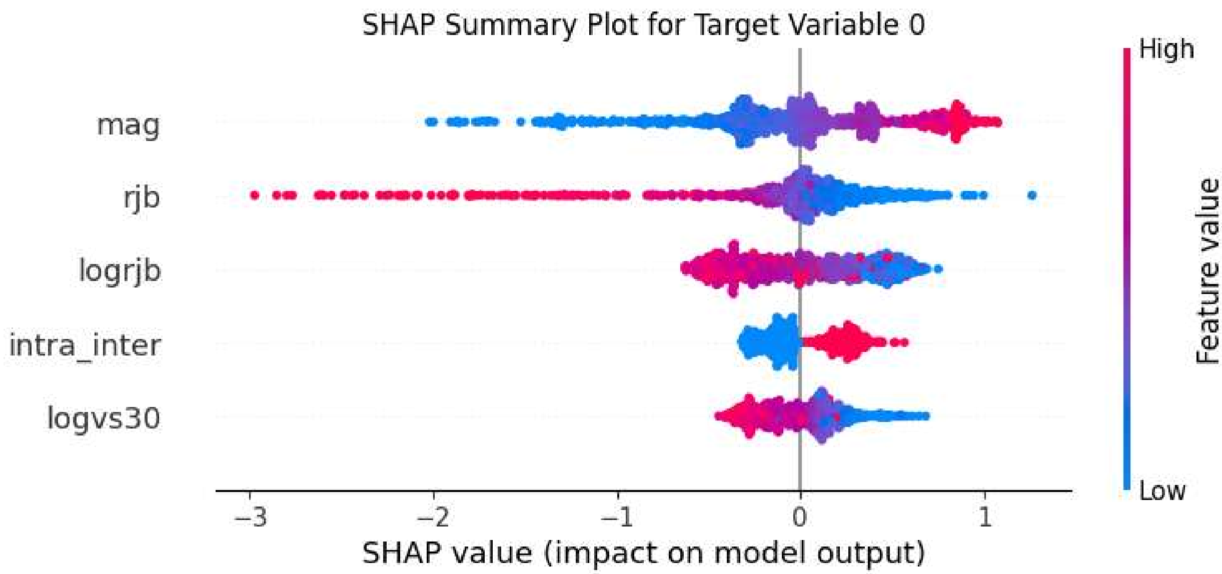


8. SHAP Analysis:

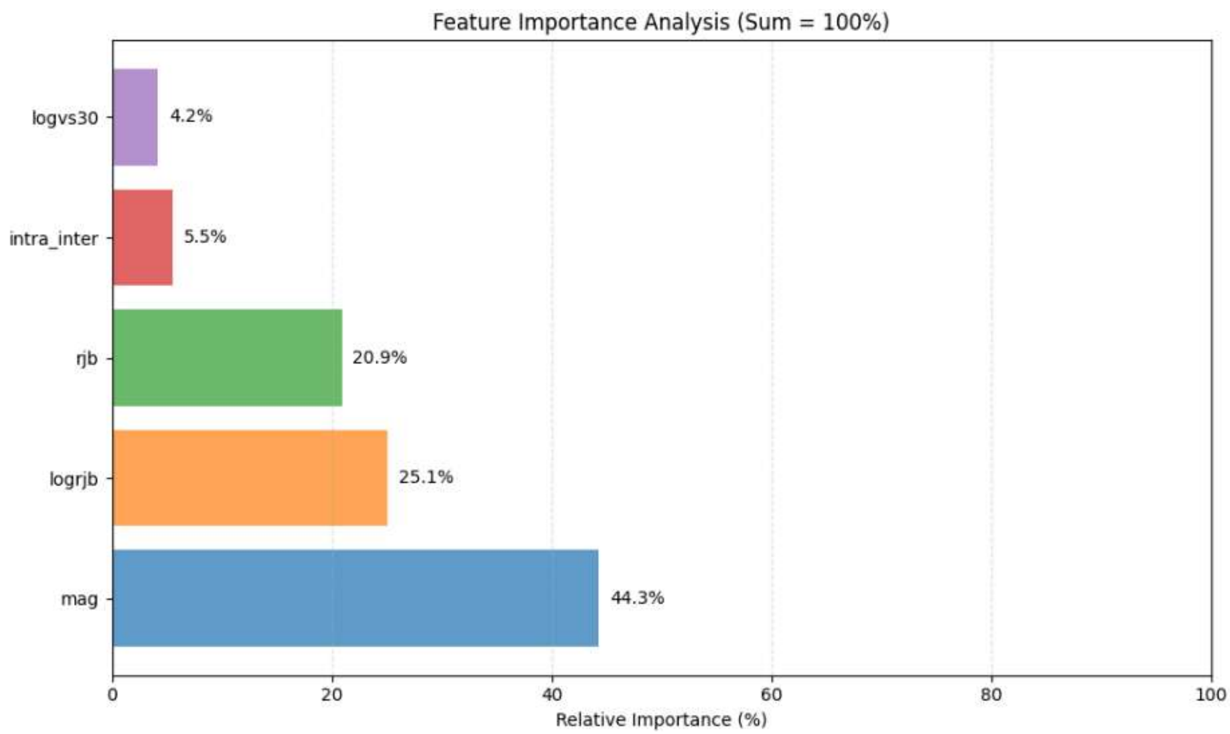
Inference from SHAP Summary Plot

- Magnitude (mag): High magnitude values (red) have a strong positive impact on the model output, while low magnitudes (blue) tend to decrease the output.
- Distance (rjb, logrjb): High values of rjb/logrjb (greater distance from source) have a strong negative impact, reducing the model output. Low values (closer distances) increase the output.
- Intra/Inter-event (intra_inter): This variable shows moderate impact, with high values slightly increasing the output.
- Site Condition (logvs30): Low logvs30 values (softer soil) increase the output, while high values (stiffer sites) decrease it.

Conclusion: Magnitude and distance are the most influential features: larger magnitudes and shorter distances increase the model output, while stiffer sites (higher logvs30) and greater distances decrease it. The SHAP plot confirms these variables' physical relevance and their direction of influence in the model.



9. Relative importance plot



Inference from Feature Importance Analysis

- Magnitude (mag) is the most influential feature, accounting for 44.3% of the model's predictive power. This highlights that earthquake size is the primary driver of the output.
- Distance metrics (logrjb at 25.1% and rjb at 20.9%) together contribute nearly as much as magnitude, showing that proximity to the earthquake source is also critical.
- Intra_inter (5.5%) and logvs30 (4.2%) have much lower importance, indicating that intra/inter-event variability and site condition (soil stiffness) play a relatively minor role in this model.

Conclusion: Magnitude and distance are the dominant factors influencing the model, while site effects and intra/inter-event variability have limited impact on the output.

Code : [CGAN Code]