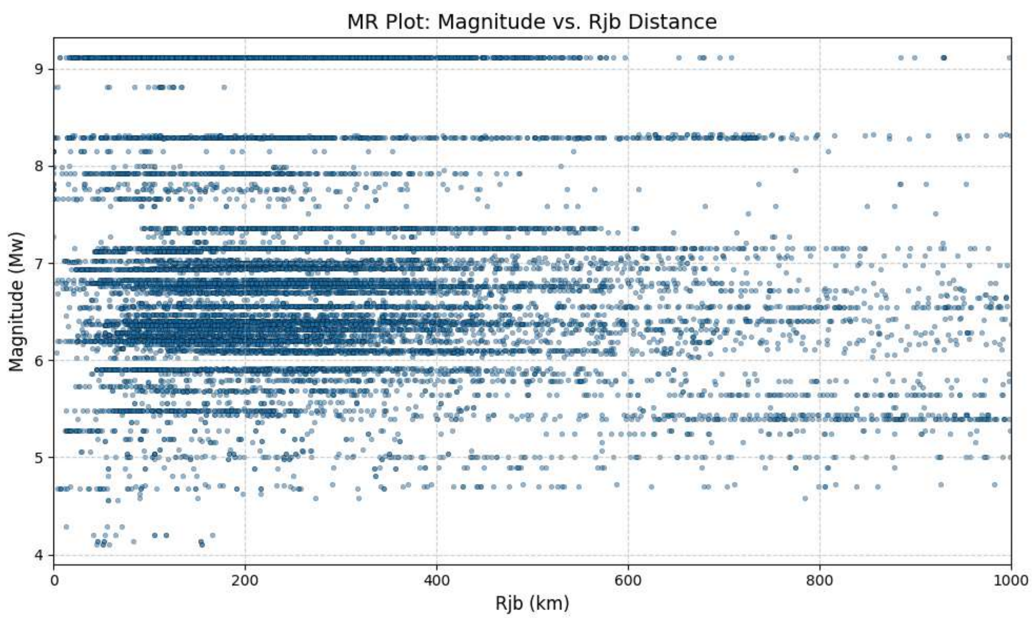
Prediction of Spectral Acceleration Using LSTM

1. Introduction

This study develops an LSTM 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:

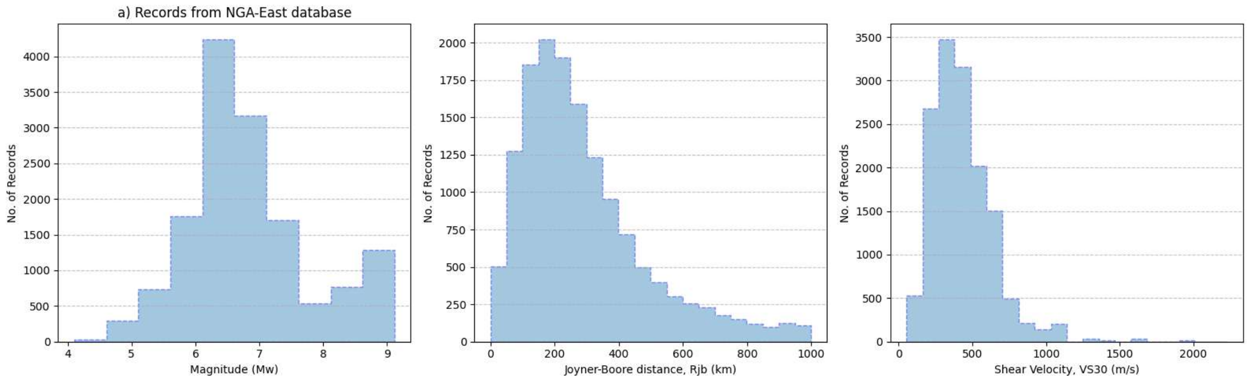
This scatter plot visualizes the distribution of events across different magnitude (mag) and Joyner-Boore distance (rjb) combinations in the dataset used for training and evaluation.



- 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:

Statistic	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, mean of 6.83, with slight positive skew (0.79), near-normal distribution.
- **Rupture Distance (rjb)**: From 0.01 to 999.09, mean 289.75, with positive skew (1.29), showing high variability.
- **Log of Rupture Distance (logrjb)**: Ranges from -2.00 to 2.99, mean 2.35, highly negatively skewed (-3.33) with heavy tails.

- **Log of Shear-Wave Velocity (logvs30)**: Ranges from 1.72 to 3.35, mean 2.59, close to normal distribution.
- **Intra-Inter Ratio (intra_inter)**: Varies from 0.0 to 1.0, mean 0.42, moderate variability with slight positive skew.

Output Parameters:

Statistic	T0pt010S	T0pt020S	T0pt030S	T0pt050S	T0pt075S	T0pt100S	T0pt150S	T0pt200S	T0pt300S	T0pt400S	T0pt500S	T0pt750S
Min	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
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
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
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
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

Statistic	T1pt000S	T1pt500S	T2pt000S	T2pt500S	T3pt000S	T3pt500S	T4pt000S	T5pt000S
Min	0	0	0	0	0	0	0	0
Max	1.2481	1.3501	1.2663	0.6708	0.3824	0.3857	0.2931	0.2265
Mean	0.03	0.0198	0.0143	0.0108	0.0084	0.0068	0.0055	0.0039
Std	0.074	0.0509	0.0373	0.0272	0.0213	0.0176	0.014	0.0098
Skewness	6.028	7.8387	8.7858	6.9784	6.2945	6.8688	6.221	5.9755
Kurtosis	51.906	106.6552	156.0315	87.9723	60.6166	79.6645	60.9231	57.1025

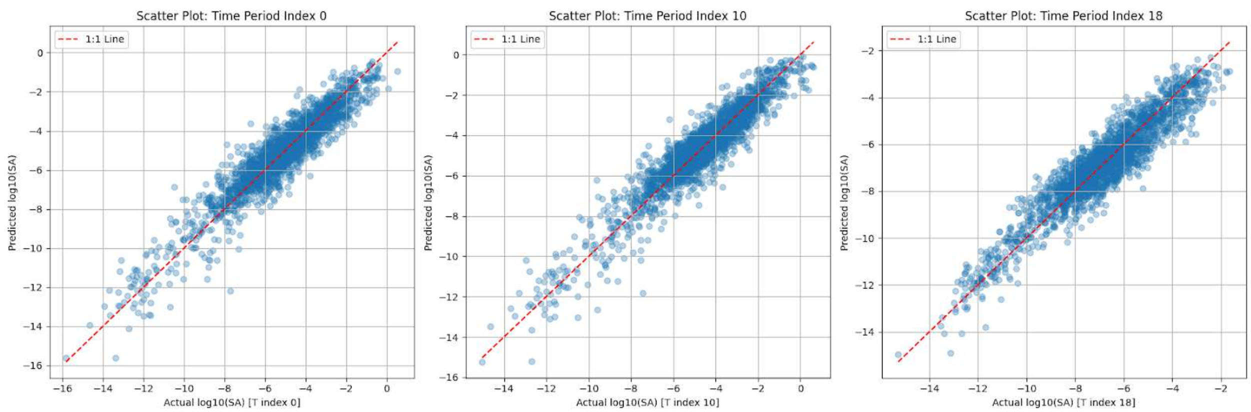
- T0pt010S to T0pt100S: Low means (0.03 to 0.06), high skewness (>7), heavy concentration near zero with extreme outliers.
- T0pt150S to T0pt500S: Increasing variability, means range from 0.07 to 0.06, still high skewness and kurtosis.
- T1pt000S to T5pt000S: Decreasing means (0.03 to 0.0039), high skewness and kurtosis, indicating heavy tails and outliers.

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

- The LSTM model demonstrates **strong predictive accuracy**, with most points closely following the 1:1 line across all time indices.
- **Consistent performance** is evident across time periods, indicating the model's ability to generalize well temporally.
- A **slight increase in scatter** is noted at the extremes (especially for very low values below -10), which is expected in log-transformed seismic data.
- There is **no clear systematic bias**, as deviations appear evenly distributed above and below the 1:1 line.

- The model effectively captures `log10(SA)` over a wide dynamic range (approx. -16 to 0), reinforcing its suitability for ground motion prediction tasks.

Overall, the plots suggest the LSTM mixed-effects model is **robust and reliable**, maintaining accurate predictions across different spectral periods.



6.Model Structure:

The figure illustrates the architecture of an LSTM-based model designed to predict 20 spectral acceleration values (log10(SA)) from ground motion input features.

Input Features (Blue Nodes)

- The model takes in 6 key seismic parameters:
  - `mag` (magnitude)
  - `rjb` (Joyner-Boore distance)
  - `vs30` (shear-wave velocity)
  - `logrjb`, `logvs30` (log-transformed features)
  - `intra_inter` (event type indicator)

Model Layers

1. Reshape Layer: Adjusts input shape to match LSTM expectations (sequence format with 1 timestep).
2. LSTM Layer (64 units): Captures temporal and sequential dependencies in the data.
3. Dense Layer (20 units, ReLU): Expands the learned representation to match the number of target outputs.
4. Final Dense Layer (20 units, Linear): Outputs predicted `log10(SA)` values corresponding to 20 different spectral periods.

Output (Green Nodes)

- The model predicts 20 separate `log10(SA)` values, representing spectral accelerations across a range of periods.

This architecture effectively combines temporal modeling through LSTM with dense transformations, making it well-suited for complex ground motion prediction tasks.

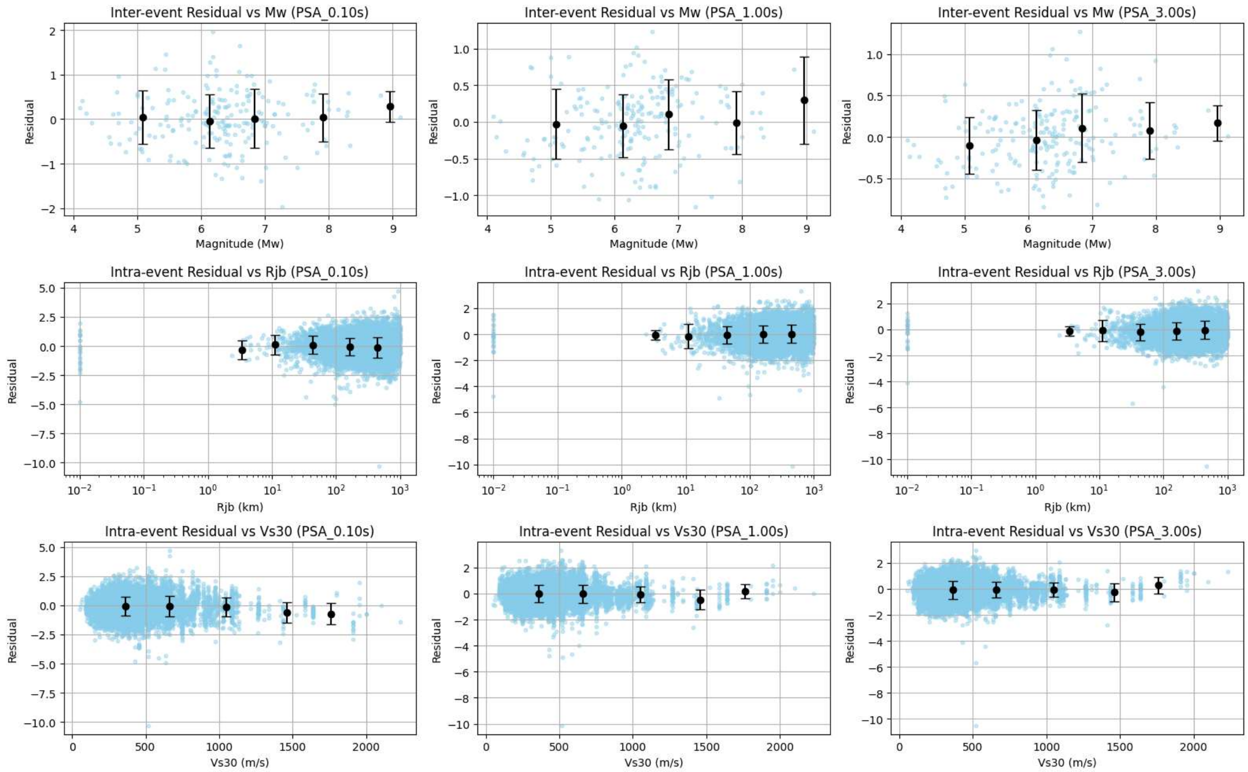
Neural Architecture:https://drive.google.com/file/d/1SHmyYxrGoh8tKIEfhSvuADJcLWUPXKF6/view?usp=sharing

7.Model Performance Metrics for Target Variables:

Target Variable	R^2	Inter-Std (τ)	Intra-Std (φ)	Total Std
T0pt010S	0.8811	0.6239	0.6777	0.9212
T0pt020S	0.8804	0.6278	0.6806	0.9259
T0pt030S	0.8785	0.6422	0.6873	0.9407
T0pt050S	0.8701	0.6789	0.7194	0.9891
T0pt075S	0.857	0.7119	0.7732	1.051
T0pt100S	0.8498	0.7237	0.8085	1.0851
T0pt150S	0.8536	0.6912	0.8069	1.0625
T0pt200S	0.8594	0.6665	0.7862	1.0307
T0pt300S	0.8729	0.6314	0.7297	0.965
T0pt400S	0.8799	0.6199	0.6947	0.9311
T0pt500S	0.8818	0.6075	0.6776	0.91
T0pt750S	0.8757	0.5887	0.6722	0.8935
T1pt000S	0.8697	0.5722	0.6809	0.8894
T1pt500S	0.8611	0.5311	0.6971	0.8763
T2pt000S	0.8594	0.5127	0.6991	0.8669
T2pt500S	0.8639	0.4968	0.6899	0.8502
T3pt000S	0.8687	0.4887	0.6808	0.838
T3pt500S	0.8729	0.4874	0.6702	0.8287
T4pt000S	0.878	0.4769	0.6552	0.8104
T5pt000S	0.8888	0.4594	0.6252	0.7758

- **T0pt010S to T0pt100S**: High R^2 values (0.88 to 0.85), indicating strong model performance. Inter- and Intra-Standard Deviations gradually increase from T0pt010S (0.62, 0.68) to T0pt100S (0.72, 0.81), reflecting growing variability and uncertainty in predictions.
- **T0pt150S to T0pt500S**: R^2 values range from 0.85 to 0.88, suggesting consistent performance. Standard deviations increase slightly, with values for intra-standard deviation reaching 0.81 in T0pt150S and 0.68 in T0pt500S, indicating moderate variability.
- **T1pt000S to T5pt000S**: R^2 decreases slightly from 0.87 (T1pt000S) to 0.89 (T5pt000S), with lower inter- and intra-standard deviations. This suggests more stable predictions for higher target values, with decreasing uncertainty, especially in T5pt000S where total standard deviation is lowest (0.78).

8.LSTM Residual Plot Interpretation



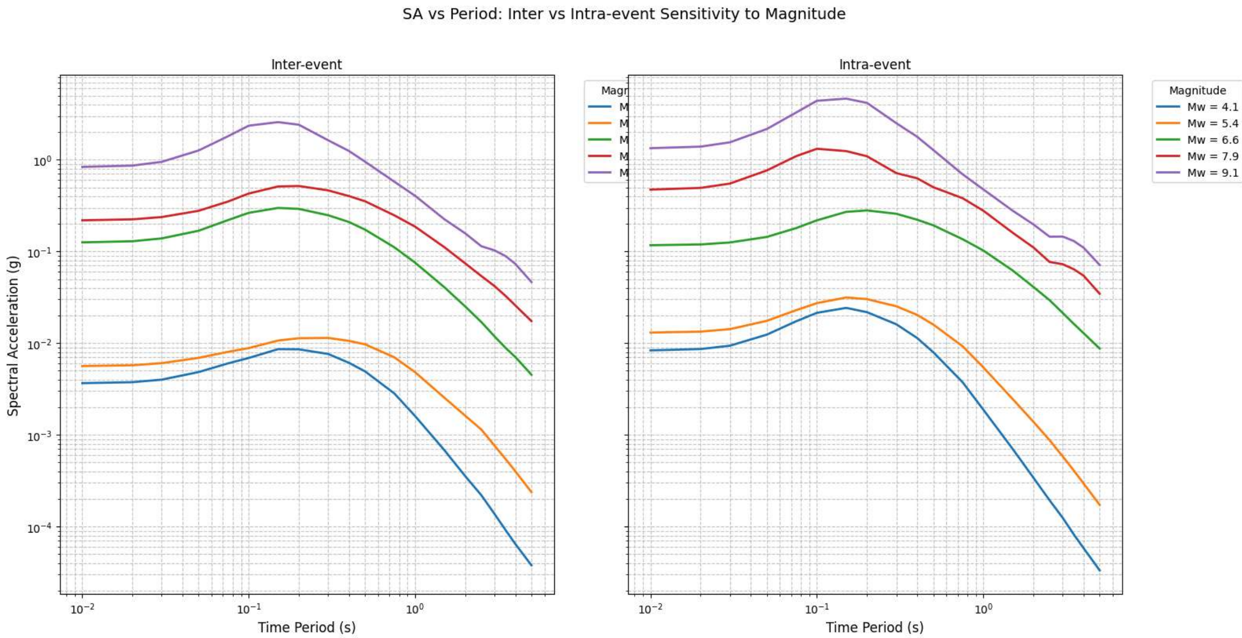
- **Inter-event vs Mw**: Residuals centered around zero with no clear trend—indicates good model performance across magnitudes.
- **Intra-event vs Rjb**: Slight negative bias at short distances; wider spread at long distances is expected and acceptable.
- **Intra-event vs Vs30**: Mostly unbiased; minor negative bias on soft soils, but overall consistent across site conditions.

No major bias or trends—LSTM model is robust across magnitude, distance, and site effects.

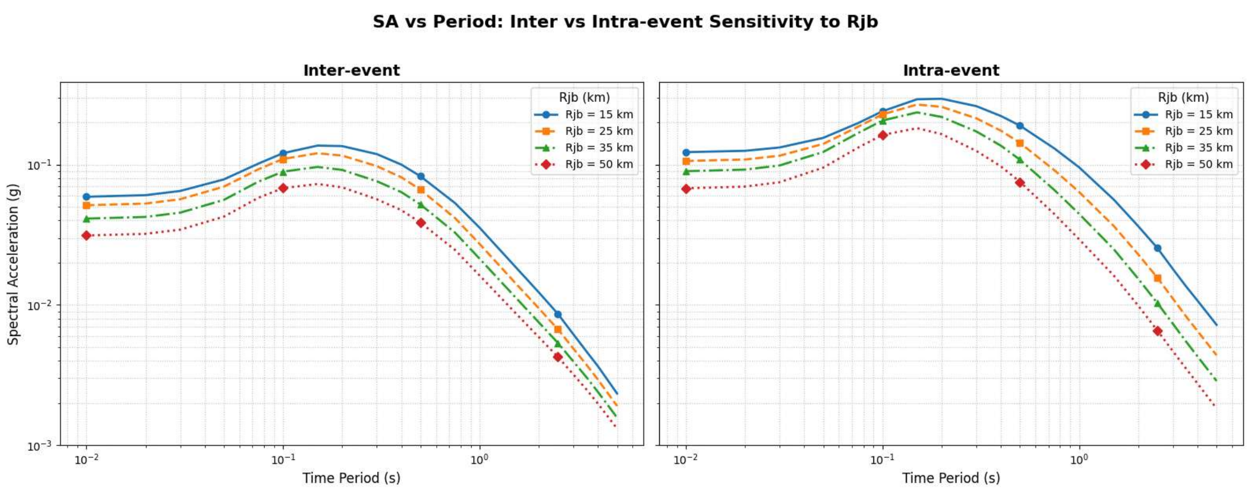
9.Magnitude Sensitivity Plot:

- **Spectral acceleration (SA)** increases with magnitude across all periods, both for **inter-event** and **intra-event** components.
- Peak SA occurs around **0.2–0.4s**, decreasing at longer periods.
- **Larger magnitudes (Mw ≥ 7.9)** show significantly higher SA, especially in inter-event terms, indicating stronger source effects.
- Trends are consistent between inter- and intra-event curves, confirming magnitude-dependent scaling in both components.

SA predictions show strong sensitivity to magnitude across periods, captured well by both inter- and intra-event terms.



10.Rjb Sensitivity Plot:



This plot shows how spectral acceleration (SA) varies with period for different rupture distances (Rjb) in **inter-event** (left) and **intra-event** (right) scenarios.

- **SA decreases as Rjb increases** across all periods, with a stronger drop at longer periods (>1s).
- **Intra-event SA** is slightly higher than inter-event SA at short periods, indicating more variability within events.
- Both plots peak around **0.1–0.2s**, typical of moderate earthquakes.
- The **impact of distance is more pronounced in intra-event residuals**, especially from 0.1–1s.

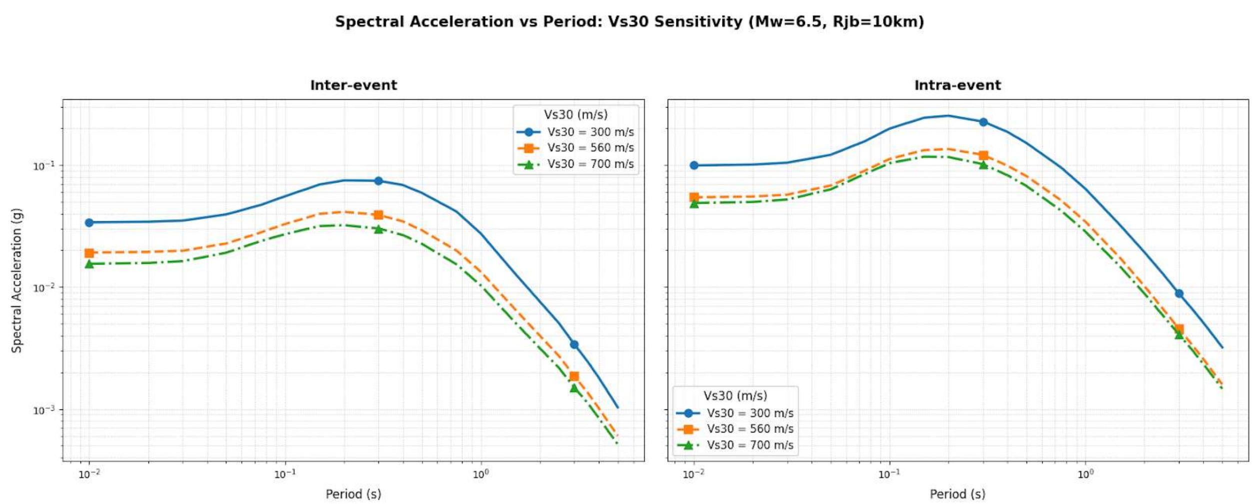
This highlights the importance of considering both inter- and intra-event variability and distance effects in seismic design.

11.Vs30 Sensitivity Plot:

This plot shows how spectral acceleration (SA) varies with period for different Vs30 values, under both inter-event and intra-event conditions.

- Lower Vs30 (softer soils) yields higher SA at all periods, especially near the peak (~0.2s).
- SA decreases with increasing Vs30, reflecting the site amplification effect.
- The trend is consistent in both inter- and intra-event plots, with slightly more separation in inter-event results.

This highlights the strong dependence of ground motion on local site conditions.



12a.SA @ T (τ) vs Rjb(inter-event):

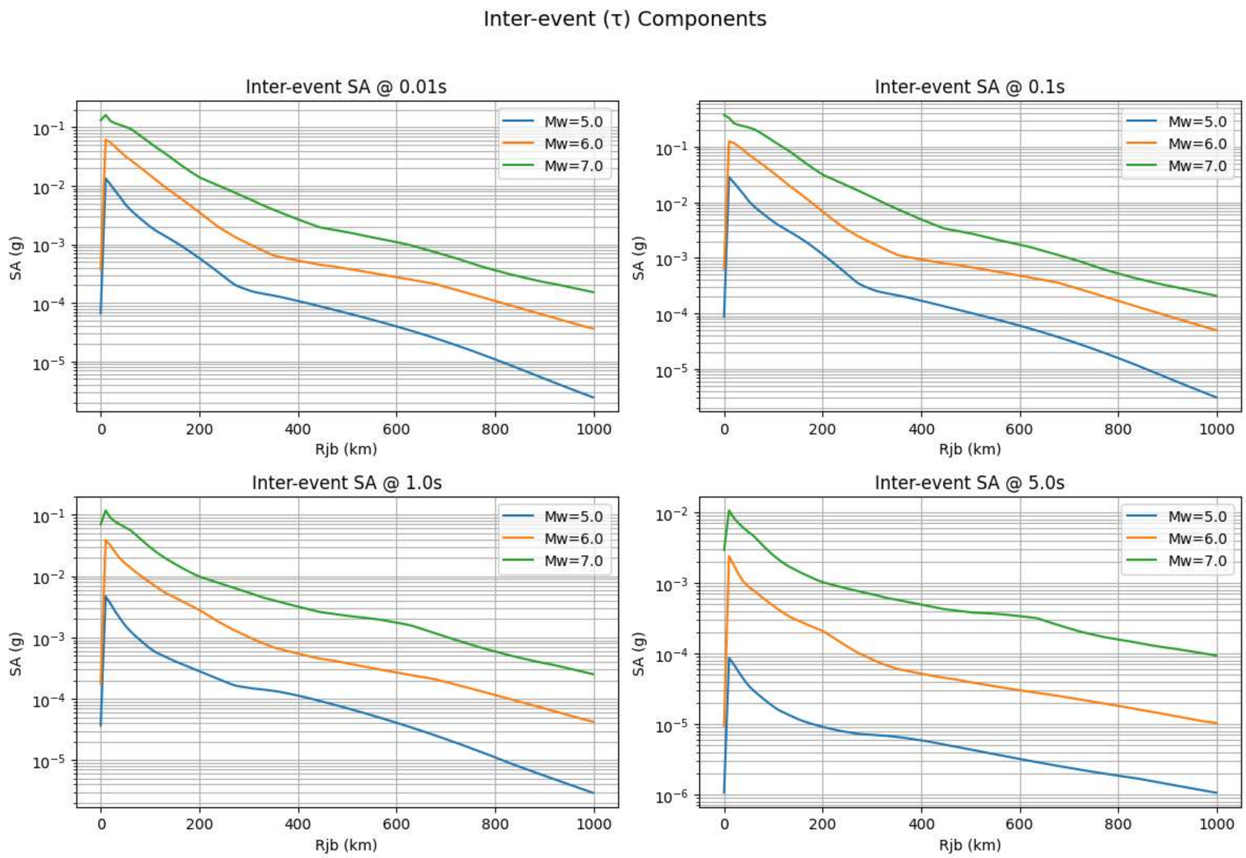
This plot shows **inter-event spectral acceleration (SA)** as a function of rupture distance (Rjb) for various magnitudes (Mw = 5.0, 6.0, 7.0) at four periods: 0.01s, 0.1s, 1.0s, and 5.0s.

- SA decreases with increasing Rjb for all magnitudes and periods.

- Higher magnitudes consistently produce higher SA across all distances and periods.

- The **rate of decay** is sharper at short distances and lower periods.

- Long-period ground motions (1.0s and 5.0s) retain higher SA over distance for large magnitudes, reflecting their deeper penetration and lower attenuation.



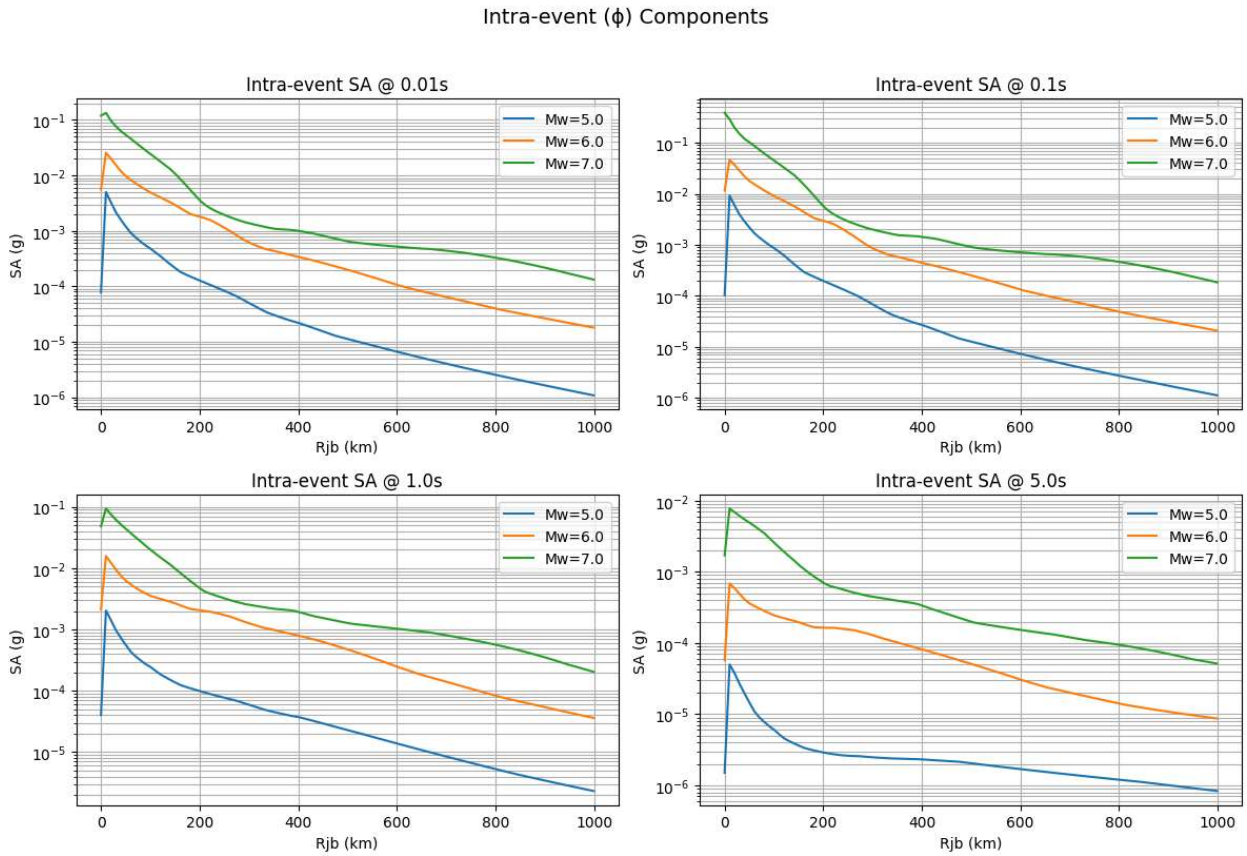
12b.SA @ T (ϕ) vs Rjb(intra-event):

- **Distance decay:** Intra-event SA decreases with increasing Rjb for all magnitudes and periods, consistent with attenuation of ground motions with distance.

- **Magnitude scaling:** Higher magnitude events produce larger intra-event variability in SA, particularly at short distances.

- **Period dependence:** Variability is generally higher at shorter periods (0.01s, 0.1s), but even at long periods (5.0s), large-magnitude events retain higher SA, indicating persistent variability.

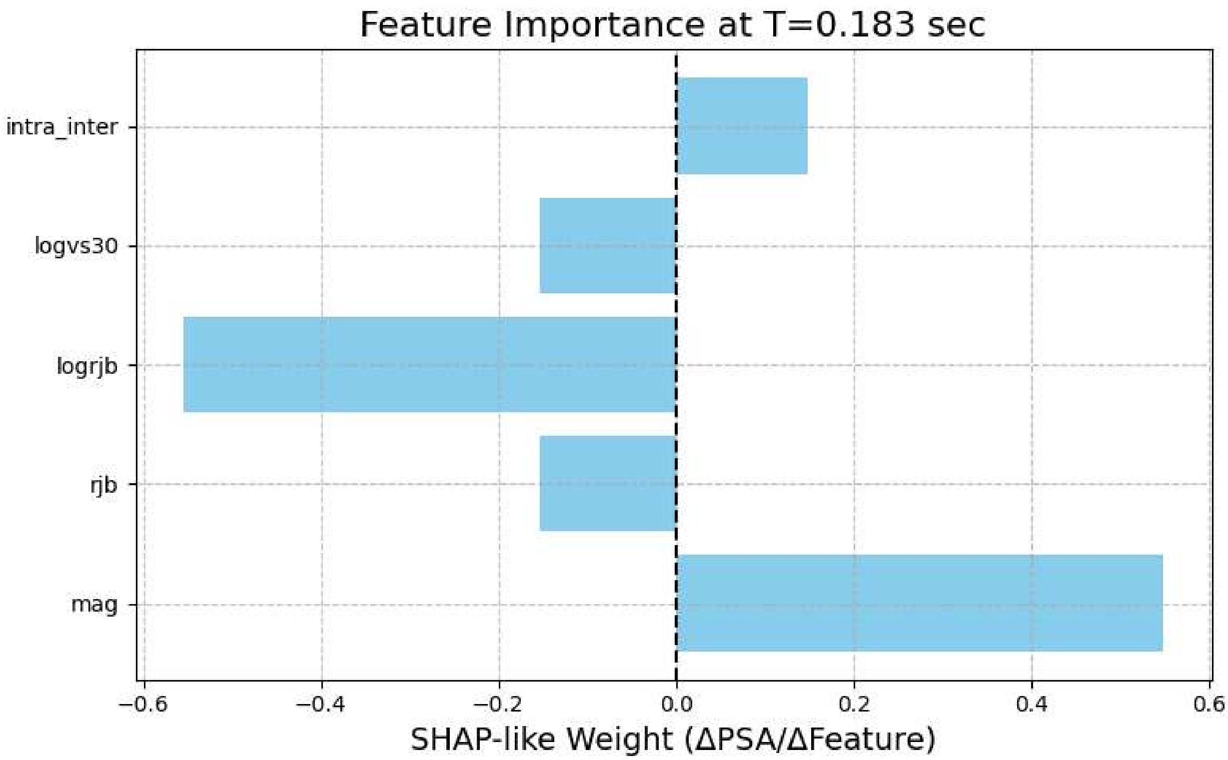
- **Log-scale insights:** The log-log scale emphasizes the wide dynamic range in intra-event ground motion variability across distances and periods.



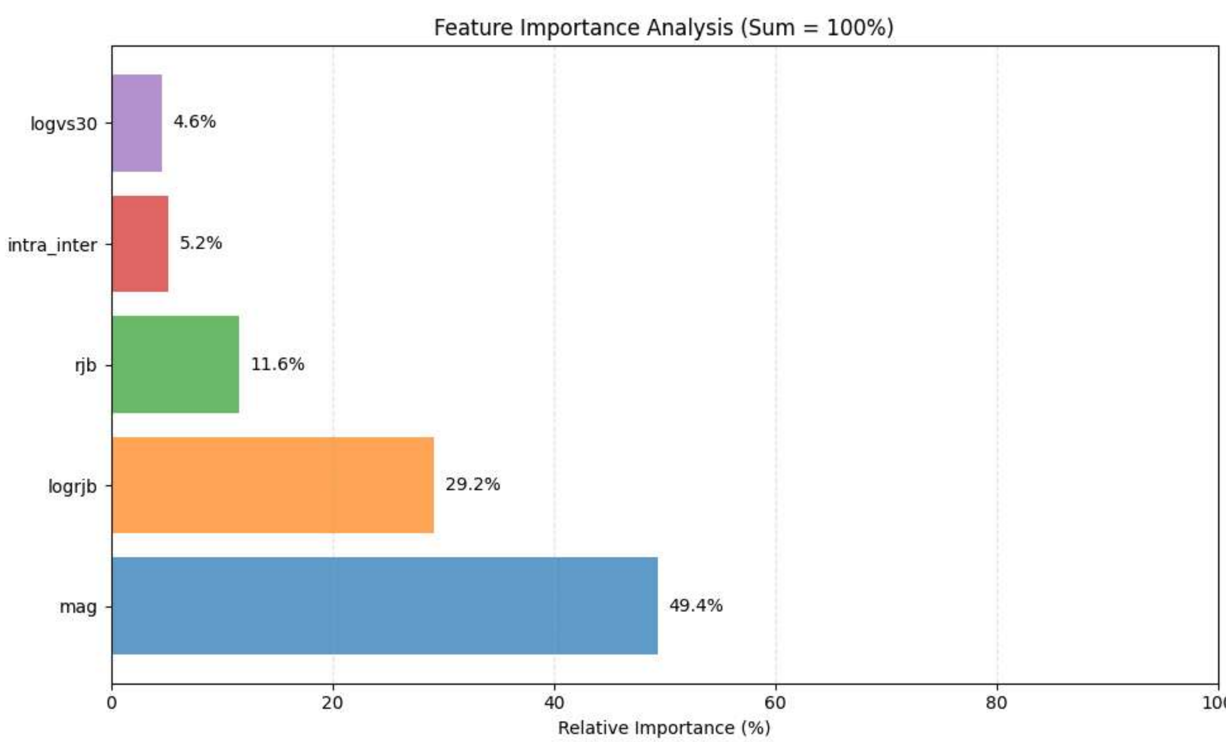
- Intra-event variability (ϕ) generally contributes **more to total aleatory uncertainty**, especially at short distances and high frequencies.
- Inter-event variability (τ) becomes more relevant for **regional and probabilistic seismic hazard analysis (PSHA)**, where many events are considered.
- Site-specific studies need to account more for ϕ, whereas regional hazard maps must factor in τ variability across sources.

13b.SHAP Analysis Summary (T=0.183s SA):

- mag has the strongest positive effect → higher magnitude increases predicted ground motion.
- logrjb & rjb show negative effects → greater distance reduces ground motion.
- logvs30 has moderate negative impact → stiffer soils (high Vs30) decrease shaking.
- intra_inter has minimal influence → earthquake type matters less at this period (T=0.183s).



14.Feature Importance Summary:



- Magnitude is most critical (49%) - Drives nearly half of predictions

- Log distance matters 2.5× more than raw distance (29% vs 12%)
- Other factors have minor impact (<10% combined)
- Model follows physics - Properly prioritizes key seismic parameters
- This shows the model correctly weights the most important earthquake characteristics.

Code : **https://colab.research.google.com/drive/1VIO5nbQ7ip2FdLnDT4KOIX95fXzMhifr?usp=sharing**