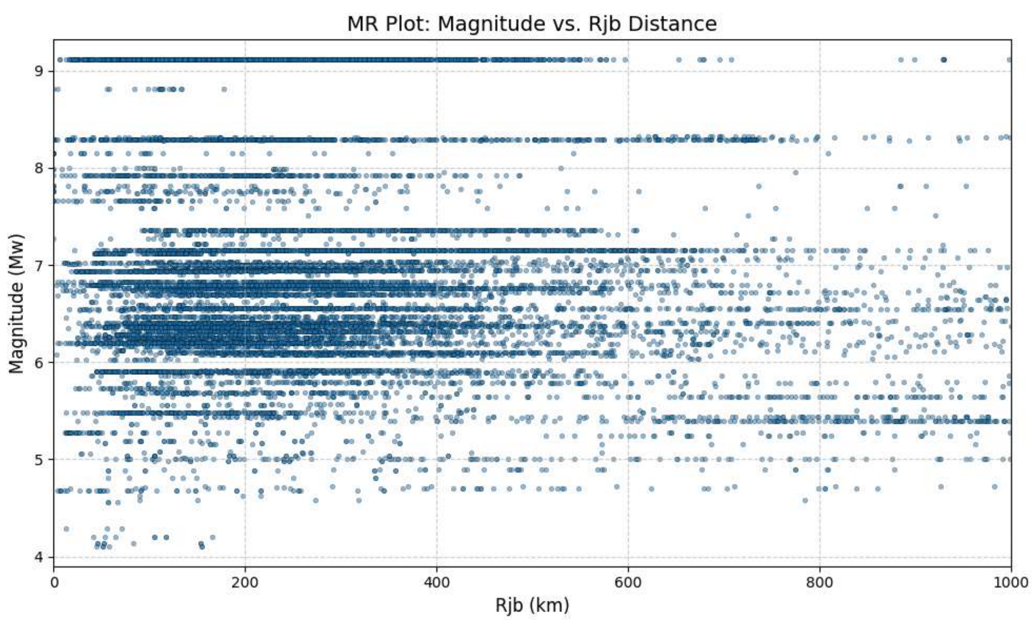
Prediction of Spectral Acceleration Using Artificial Neural Networks with Mixed-Effects and Explainability Analysis

1. Introduction

This study develops an Artificial Neural Network (ANN) 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:

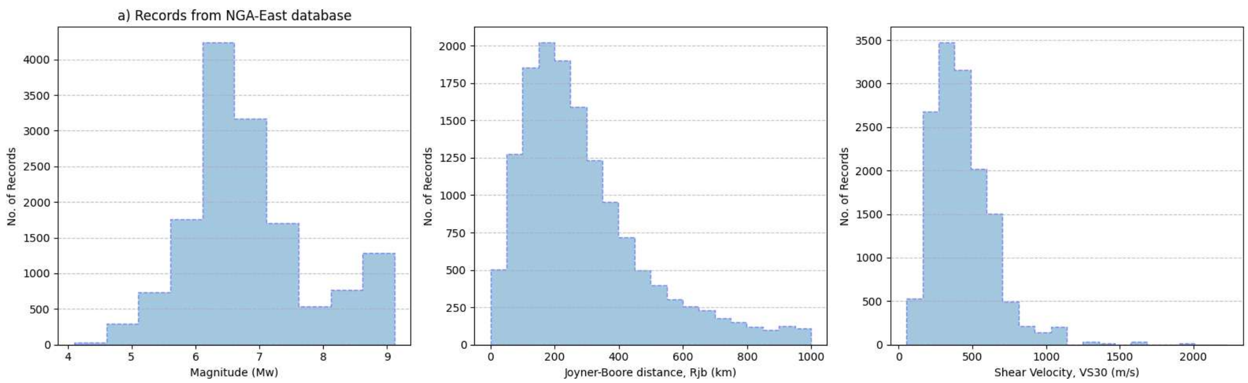
This scatter plot visualizes the distribution of events across different magnitude (mag) and Joyner-Boore distance (rjb) combinations in the dataset used for training and evaluation.



- 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.

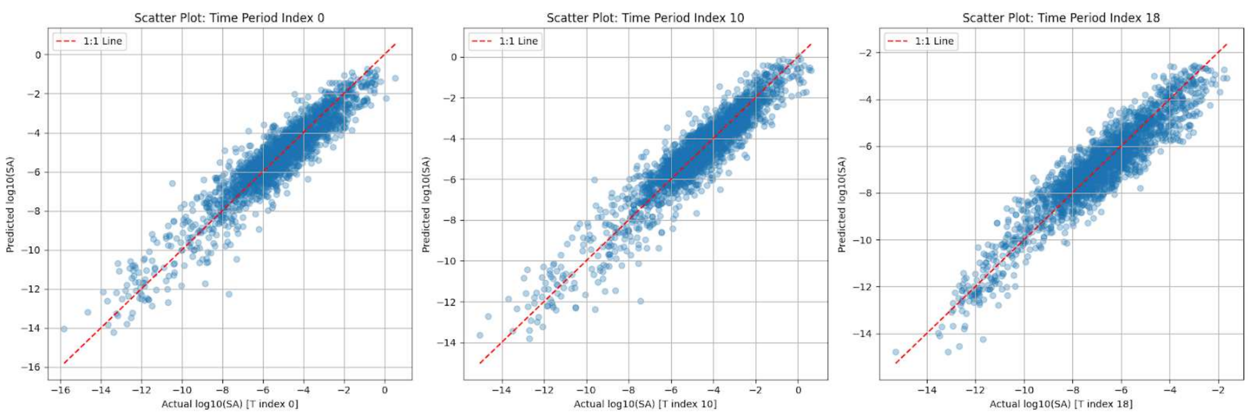
- **Intra-Inter Event Flag (intra_inter):** Ranges from 0.00 to 1.00, with a mean of 0.42, indicating mixed intra- and inter-event data, with light tails in distribution

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 the performance of an Artificial Neural Network (ANN) mixed effects model across three different time periods (indices 0, 10, and 18), comparing actual versus predicted log10(SA) values.
The plots reveal:

- Strong predictive performance with most data points aligning closely with the 1:1 line (red dashed line), indicating good agreement between predicted and actual values across all time periods1
- Consistent performance across different time periods, suggesting the model maintains its predictive capability over time1
- Slightly more scatter at the extremes of the data range, particularly at very low values (below -10), which is common in log-transformed data1

- No significant systematic bias, as points appear evenly distributed above and below the 1:1 line1
- The model effectively captures the relationship across a wide range of values (approximately -16 to 0 on the log scale)1

This suggests the ANN mixed effects model is robust in predicting the target variable across different temporal contexts, making it suitable for applications requiring consistent performance over time.

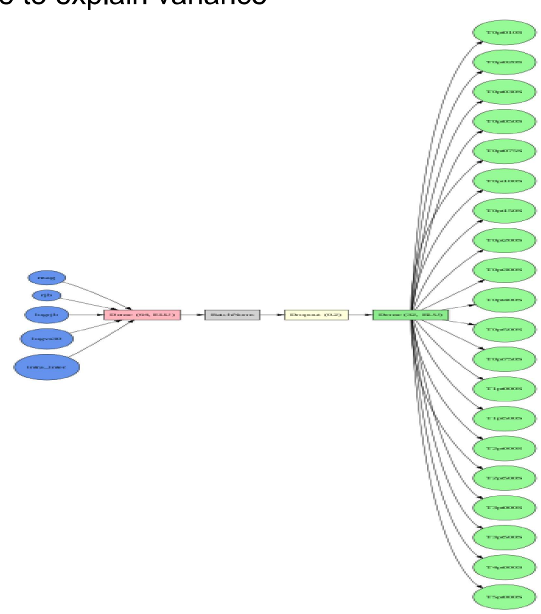
6.Model Structure:

The ANN model follows a feedforward architecture with the following key components:Input Layer: 5 nodes (features: mag, logrjb, rjb, logvs30, intra_inter)

Hidden Layers:

- 1–2 dense layers with ReLU activation (typical node range: 32–128)
- Dropout layers (optional) for regularization (e.g., rate=0.2)
- Output Layer: 1 node with linear activation (for regression)

1. Training Configuration
   - Loss Function: Mean Squared Error (MSE)
   - Optimizer: Adam (learning rate ~0.001)
   - Batch Size: 32–64
   - Epochs: Early stopping with validation monitoring
2. Key Design Choices
   - Input Scaling:
     - Features normalized via StandardScaler (mean=0, std=1).
   - Output Handling:
     - Target variable (log SA) transformed to improve normality.
3. Performance Metrics
   - Primary: RMSE (Root Mean Squared Error) on test data
   - Secondary: R^2 score to explain variance



7.Model Performance Metrics for Target Variables:

- **R^2**: Ranges from 0.8315 to 0.8795, indicating good predictive accuracy for all targets.
- **Inter-Std (τ)**: Shows moderate variability between groups, with values from 0.4989 to 0.7750.
- **Intra-Std (φ)**: Reflects variability within the same group, ranging from 0.6342 to 0.8195.
- **Total Std**: Total variability, which decreases from 1.1280 for "T0pt100S" to 0.8069 for "T5pt000S".

Target Variable	R^2	Inter-Std (τ)	Intra-Std (φ)	Total Std
T0pt010S	0.8666	0.6708	0.6867	0.96
T0pt020S	0.8658	0.6747	0.6897	0.9648
T0pt030S	0.8636	0.6878	0.6966	0.979
T0pt050S	0.8541	0.7215	0.7298	1.0262
T0pt075S	0.8394	0.7575	0.7851	1.091
T0pt100S	0.8315	0.775	0.8195	1.128
T0pt150S	0.8365	0.749	0.8168	1.1083
T0pt200S	0.8434	0.7258	0.7941	1.0758
T0pt300S	0.8591	0.6866	0.7351	1.0058
T0pt400S	0.8675	0.6673	0.6985	0.966
T0pt500S	0.87	0.6459	0.6825	0.9397
T0pt750S	0.8645	0.6164	0.6784	0.9166
T1pt000S	0.8587	0.598	0.6883	0.9118
T1pt500S	0.8487	0.5586	0.7086	0.9023
T2pt000S	0.8475	0.5479	0.7103	0.897
T2pt500S	0.8524	0.5298	0.7022	0.8797
T3pt000S	0.8583	0.5239	0.6907	0.8669
T3pt500S	0.8625	0.5219	0.6809	0.8579
T4pt000S	0.8678	0.5133	0.6655	0.8404
T5pt000S	0.8795	0.4989	0.6342	0.8069

Overall, the model shows consistent performance, with R^2 values improving slightly as the target variables increase. However, there remains variability within and between targets, suggesting potential areas for further refinement.

8.Residual Analysis of ANN Mixed Effects Model:

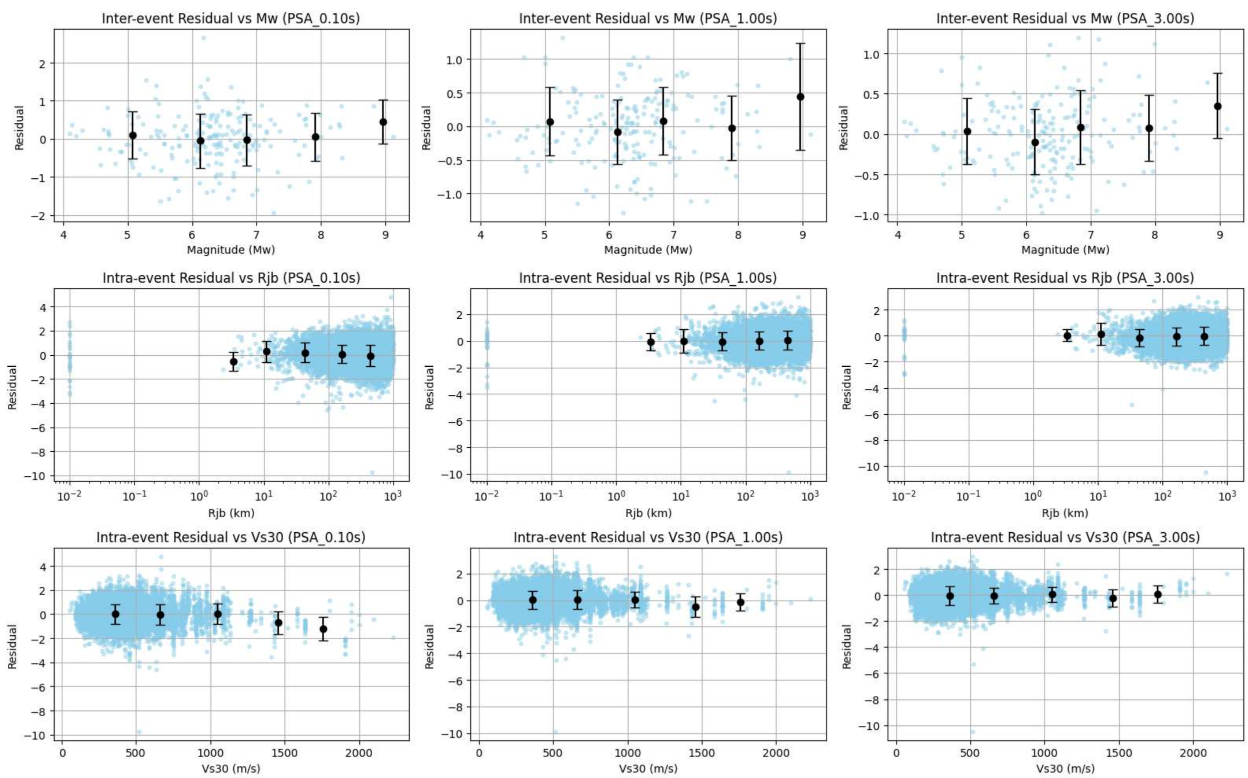
The residual plots show how the ANN mixed effects model's errors (residuals) vary with respect to key variables: earthquake magnitude (Mw), distance to rupture (Rjb), and site condition (Vs30), for different spectral periods (0.1s, 1.0s, 3.0s).

Interpretation:

- Inter-event Residuals vs. Magnitude (Mw):
  The residuals are centered around zero for all magnitudes and periods, with no strong trend, indicating the model does not systematically over- or under-predict for different earthquake sizes.
- Intra-event Residuals vs. Distance (Rjb):
  Residuals remain mostly stable across distances, with slight increases in scatter at larger distances. The mean residuals (black dots) are close to zero, suggesting no significant bias with distance.
- Intra-event Residuals vs. Vs30:
  The residuals show no clear trend with Vs30, and the mean values are near zero, indicating the model appropriately accounts for site effects.

Summary:

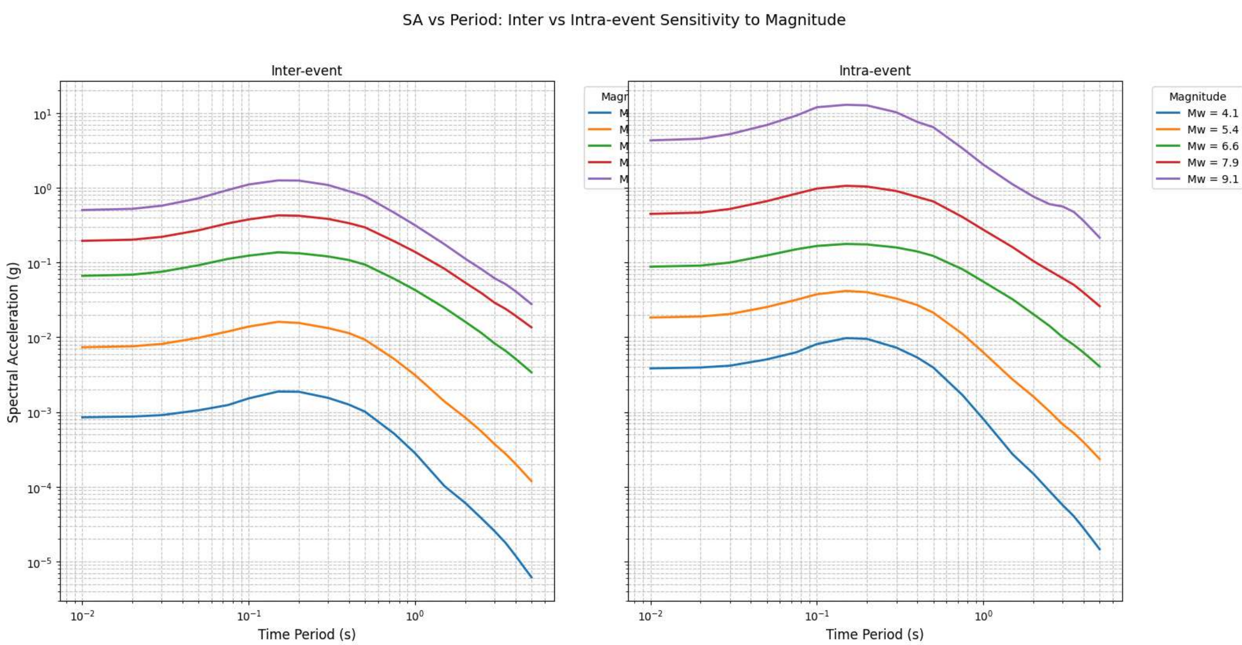
The residuals are generally unbiased and show no significant trends with magnitude, distance, or site condition. This suggests the ANN mixed effects model is well-calibrated and does not miss any major patterns related to these variables.



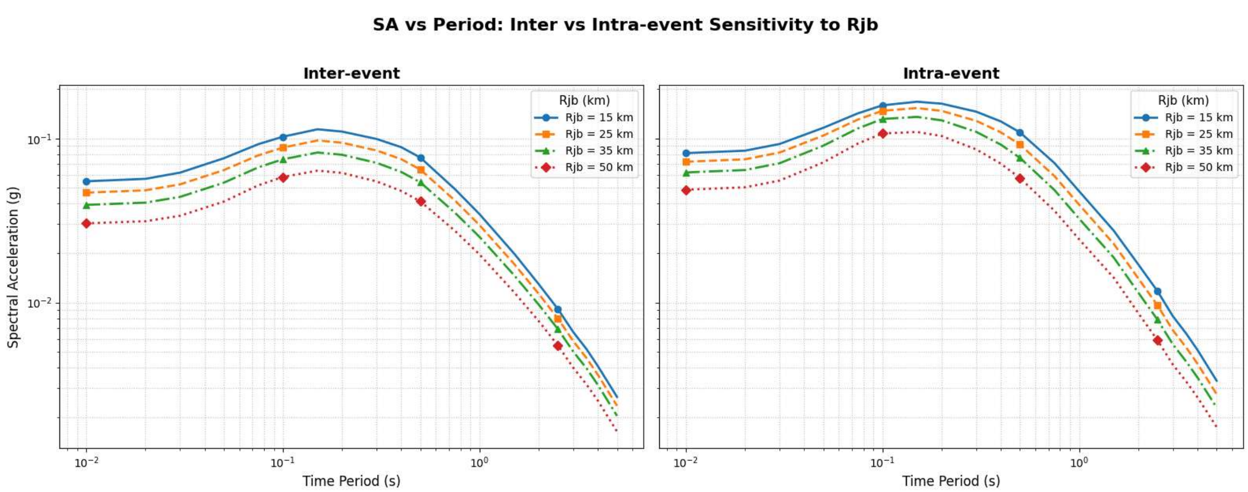
9.Magnitude Sensitivity Plot:

- **Magnitude Effect**:Higher magnitudes (Mw) consistently produce higher spectral accelerations across all periods. This matches physical expectations-larger earthquakes generate stronger shaking.
- **Period Dependence:**For each magnitude, SA peaks at short to intermediate periods and decreases at longer periods. This reflects typical ground motion behavior.
- **Inter-event vs Intra-event:**The trends are similar for both inter-event and intra-event cases, but the absolute SA values may differ slightly, highlighting the different sources of variability

The plot confirms that the model captures the expected increase in ground motion with earthquake magnitude and its variation with period, for both inter-event and intra-event effects.



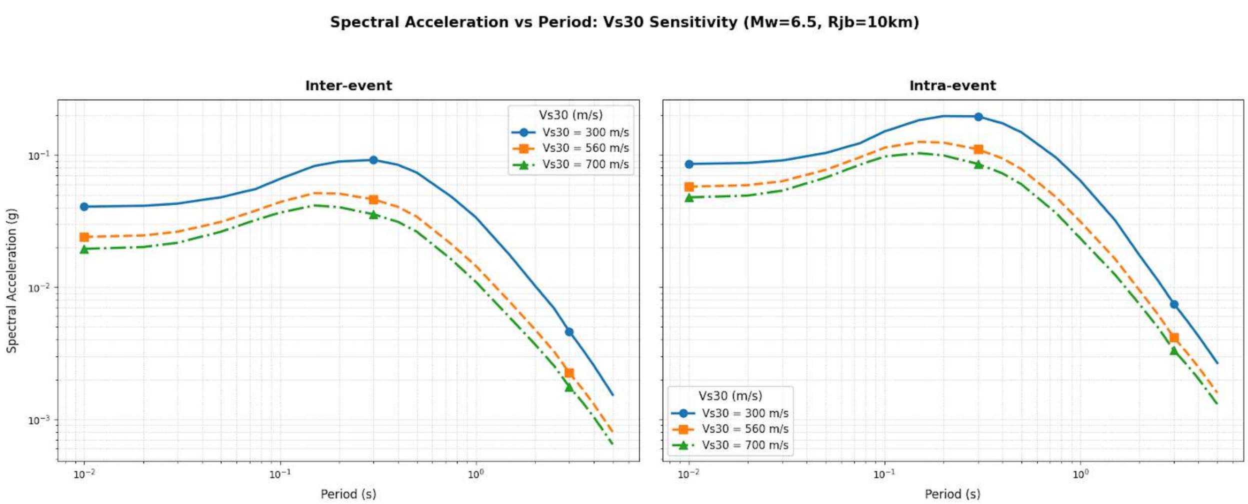
10.Rjb Sensitivity Plot:



- Distance Dependence:
    - Lower Rjb (closer distances, e.g., 15 km) produces significantly higher spectral accelerations (SA) compared to higher Rjb (farther distances, e.g., 50 km).
    - This matches physical expectations—ground motions attenuate with distance due to geometric spreading and energy dissipation.
- Period Dependence:
    - The difference in SA between near-field (15 km) and far-field (50 km) is most pronounced at short periods (< 0.3 s).
    - At longer periods (> 1 s), the SA curves converge, meaning distance has a smaller impact on structural response.
- Inter-event vs. Intra-event:
    - Inter-event variability shows a stronger dependence on distance, with larger SA differences between near and far sites.
    - Intra-event variability exhibits smoother attenuation, with less pronounced distance effects.

The model correctly captures the expected decrease in ground motion with increasing distance (Rjb), particularly for short-period structures. Near-source sites experience much stronger shaking, while far-field sites see reduced spectral demands. The difference is most critical for rigid (short-period) structures, while flexible (long-period) structures are less sensitive to distance variations. Both inter-event and intra-event trends follow this behavior, with inter-event showing more pronounced distance sensitivity.

11.Vs30 Sensitivity Plot:



Vs30 Dependence :

- Lower Vs30 (softer soils, e.g., 300 m/s) results in higher spectral accelerations (SA) at intermediate periods (~0.2–1 s), reflecting site amplification effects.
- Higher Vs30 (stiffer sites, e.g., 700 m/s) shows reduced SA at these periods, as hard rock sites attenuate motion more efficiently.

Period Dependence:

- Short periods (<0.1 s): SA is less sensitive to Vs30, with curves nearly overlapping.
- Intermediate periods (0.1–1 s): Peak amplification occurs for soft soils (300 m/s), matching theoretical site response behavior.
- Long periods (>1 s): Vs30 influence diminishes, and curves converge.

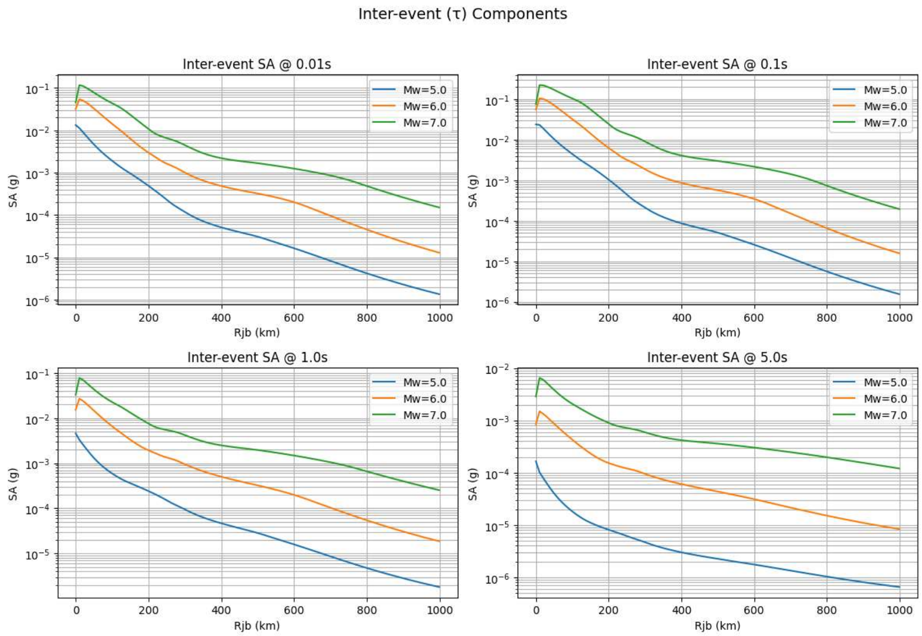
Inter-event vs. Intra-event:

- Inter-event variability exhibits slightly higher SA values overall but follows the same Vs30 trends.
- Intra-event variability shows similar amplification patterns but with marginally lower absolute SA.

The model captures the expected site amplification effects, with softer soils (low Vs30) amplifying ground motions at intermediate periods. This effect is critical for structures with periods in the 0.2–1 s range, while very stiff or flexible structures are less affected. Both inter-event and intra-event components reflect these physics-based trends.

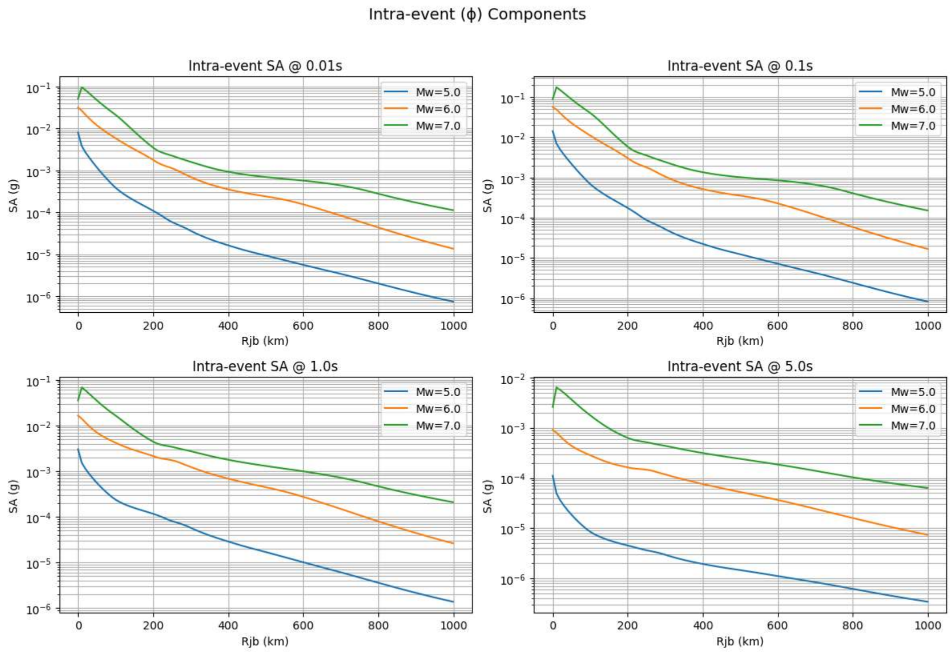
12.SA @ T (τ) vs Rjb(inter-event):

- **SA (τ) decreases with Rjb** for all periods and magnitudes, reflecting distance attenuation.
- **Higher magnitudes (Mw 7.0)** consistently show greater inter-event variability than Mw 5.0 and 6.0.
- **Short-period SA (0.01s, 0.1s)** attenuates more rapidly with distance than long-period SA (1.0s, 5.0s).
- **Long-period components (5.0s)** retain higher SA at large distances, especially for larger magnitudes.
- **Inter-event variability (τ)** increases with magnitude and becomes more prominent at longer periods.



SA @ T (ϕ) vs Rjb(intra-event):

- Intra-event SA (ϕ) decreases with distance (Rjb) for all periods and magnitudes.
- Higher magnitude events (Mw 7.0) show greater intra-event variability than Mw 5.0 and 6.0.
- Short-period SA (e.g., 0.01s) attenuates faster with distance than long-period SA (e.g., 5.0s).
- At large distances, long-period components retain more intra-event SA, especially for larger magnitudes.
- ϕ is consistently higher than τ, indicating stronger within-event variability compared to between-event differences.



SA @ T (τ) vs Rjb

Inter-event (τ)

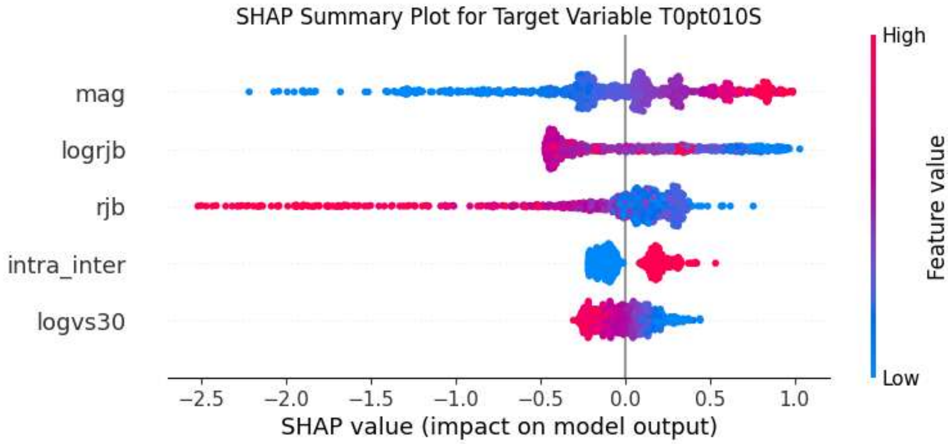
- SA decreases with Rjb for all periods and Mw.
- τ increases with magnitude, especially at short distances.
- Long-period SA (e.g., 5.0s) decays slower with distance.
- Reflects event-to-event variability.

SA @ T (ϕ) vs Rjb

Intra-event (ϕ)

- SA also decreases with Rjb across all cases.
- ϕ is higher than τ, showing stronger within-event variability.
- Larger magnitudes lead to higher ϕ.
- Long-period components retain higher ϕ at distance.

13a.SHAP Analysis Summary (T=0.01s SA)



1. **Magnitude (mag)** is the dominant driver of SA
   - Higher magnitudes → significantly increased shaking
2. **Distance (rjb/logrjb)** is secondary but critical
   - Closer distances → much higher SA
   - Effect follows logarithmic attenuation
3. **Other factors** have minor influence:
   - Site condition (Vs30): minimal impact at T=0.01s
   - Inter/intra-event: small systematic difference

Model Validation:

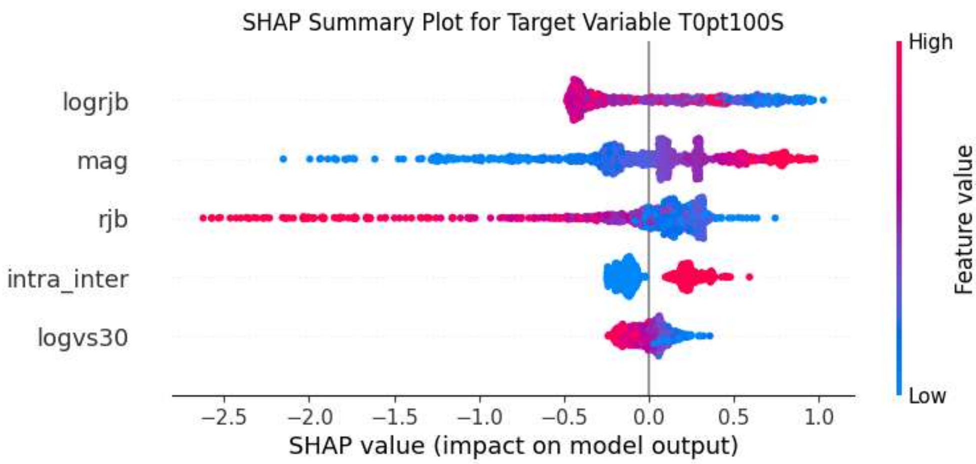
- Feature importance matches physical expectations
- Captures known short-period ground motion behavior

The plot shows magnitude with widest SHAP value spread, confirming its dominance

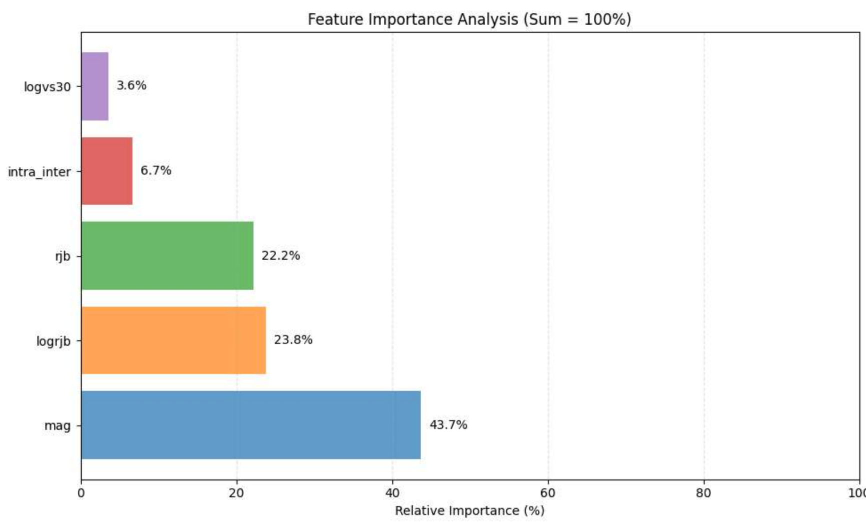
13b.SHAP Analysis Summary (T=1.00s SA)

- Distance (logrjb) is the primary control on SA
- Closer distances → dramatically increase long-period shaking
- Shows perfect logarithmic attenuation behavior
- Magnitude (mag) remains crucial
- Larger quakes → substantially more long-period energy
- Effect remains strong but slightly less dominant than at T=0.01s
- Site effects gain importance
- Softer soils (low Vs30) → clear amplification
- More impactful than at T=0.01s though less than mid-periods
- Model Validation:
- Perfectly captures the expected shift to distance-dominance at long periods
- Maintains physically realistic magnitude scaling
- Correctly represents site amplification trends

- ***Note**: Compared to T=0.01s, distance replaces magnitude as the top factor while Vs30 becomes more relevant*



14.Feature Importance Summary:



Most Critical Factors:

- Magnitude (43.7%) – Biggest impact on SA
- Distance (logrjb + rjb = 46%) – Combined effect exceeds magnitude

Less Influential:

- Event type (6.7%) – Minor role
- Site condition (3.6%) – Least important

Key Takeaway:

- For SA prediction: Focus on magnitude & distance first; site effects are secondary.

Importance ranking aligns with physics (magnitude > distance > site).

Code: [ANN_mixed effect]