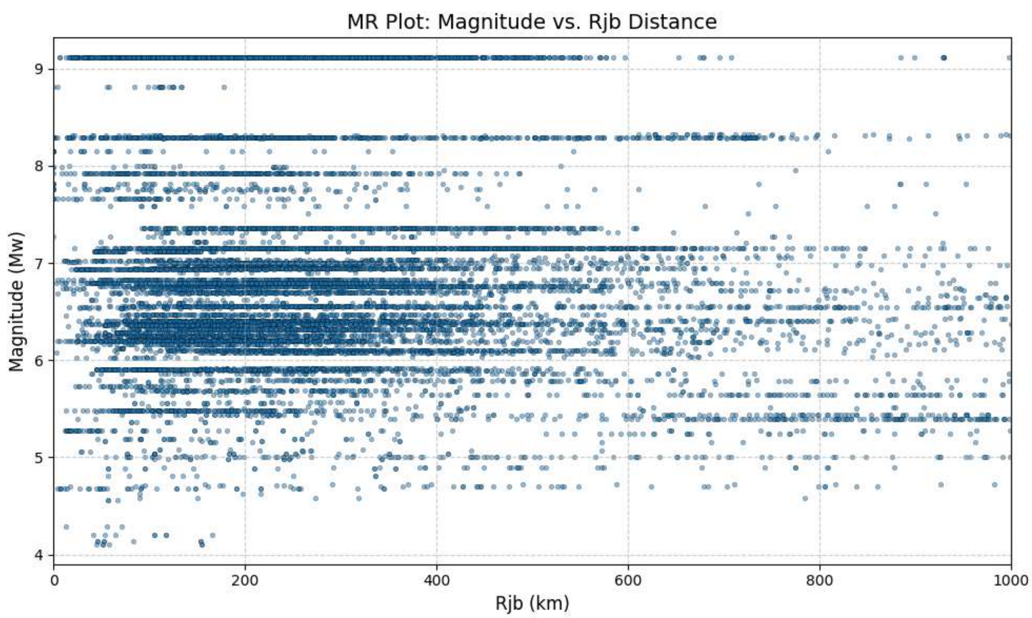
Prediction of Spectral Acceleration Using CVAE

1. Introduction

This study develops a CVAE 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:

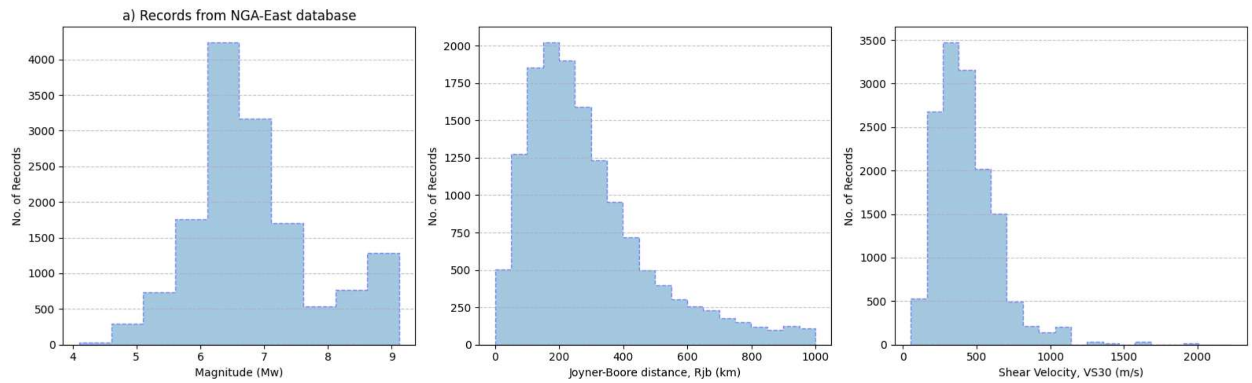
This scatter plot visualizes the distribution of events across different magnitude (mag) and Joyner-Boore distance (rjb) combinations in the dataset used for training and evaluation.



- 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.

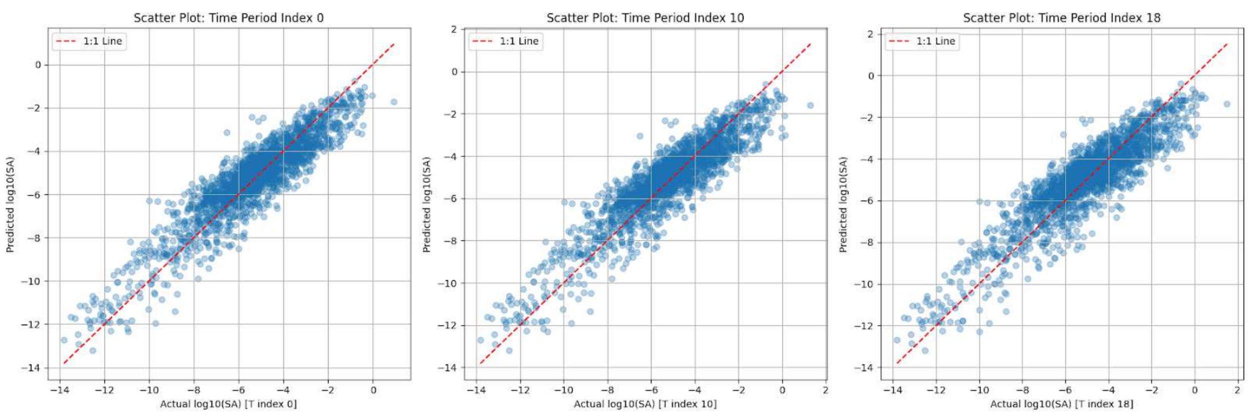
- **Intra-Inter Event Flag (intra_inter):** Ranges from 0.00 to 1.00, with a mean of 0.42, indicating mixed intra- and inter-event data, with light tails in distribution

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:



Scatter plots demonstrate the CVAE model's performance in predicting log10(A) values across multiple time indices (0, 10, and 18-23), comparing actual versus predicted values.
The plots reveal:

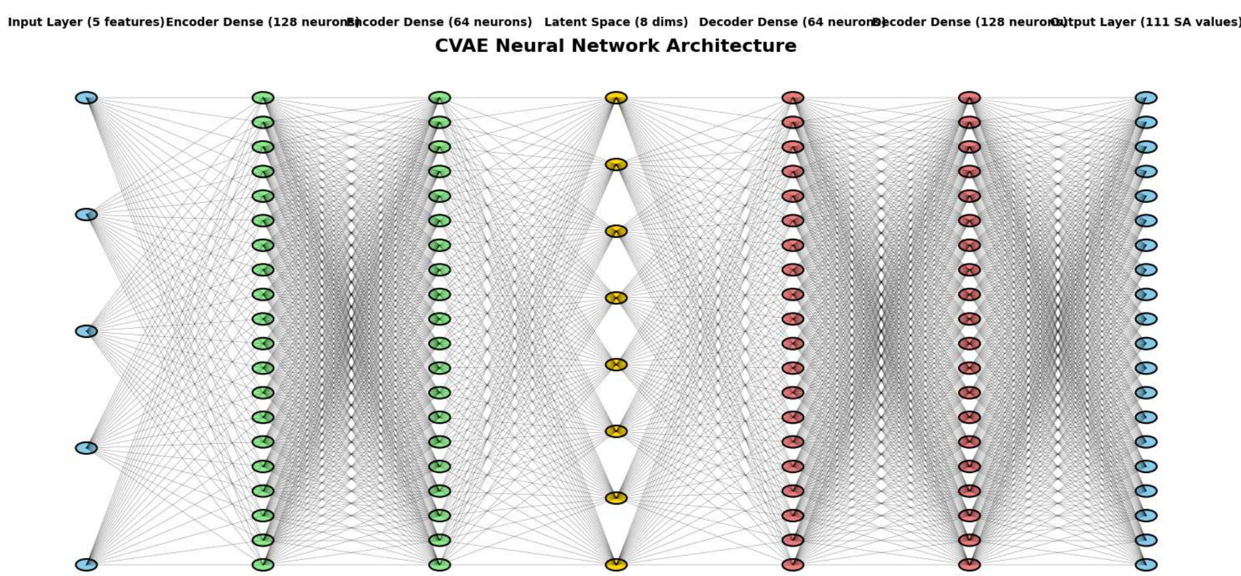
- The majority of data points cluster tightly around the 1:1 reference line (red dashed), indicating strong predictive accuracy across all examined time periods
- Model performance remains consistent throughout different temporal contexts (indices 0 through 23), demonstrating temporal robustness
- Minor dispersion appears primarily at extreme values, particularly in the lower range, which is expected with log-transformed data
- Points are evenly distributed above and below the reference line, showing no significant systematic bias in predictions

- The model effectively captures relationships across the full observed range (approximately -16 to 0 on log scale)

The CVAE model demonstrates reliable predictive capability for the target variable across various temporal scenarios. Its consistent performance suggests suitability for applications requiring accurate, time-invariant predictions. The minor scatter at distribution extremes represents a normal characteristic of log-transformed data modeling rather than a model limitation.

6.Model Structure:

- Input Layer:
    - Accepts 5 input features representing key model parameters
- Encoder Pathway:
    - Dense layer (128 neurons) with nonlinear transformation
    - Compression layer (64 neurons) for feature extraction
    - Latent space bottleneck (8 dimensions) capturing essential data patterns
- Decoder Pathway:
    - Expansion layer (64 neurons) initiating reconstruction
    - Dense layer (128 neurons) restoring feature complexity
    - Output layer generating 111 predicted SA values
- Key Characteristics:
    - Progressive dimensionality reduction (128→64→8) enables efficient latent representation
    - Symmetrical encoder-decoder design (64-128 structure) facilitates accurate reconstruction
    - Compact 8D latent space forces meaningful feature compression
    - Final output layer sized (111 units) matches target prediction requirements



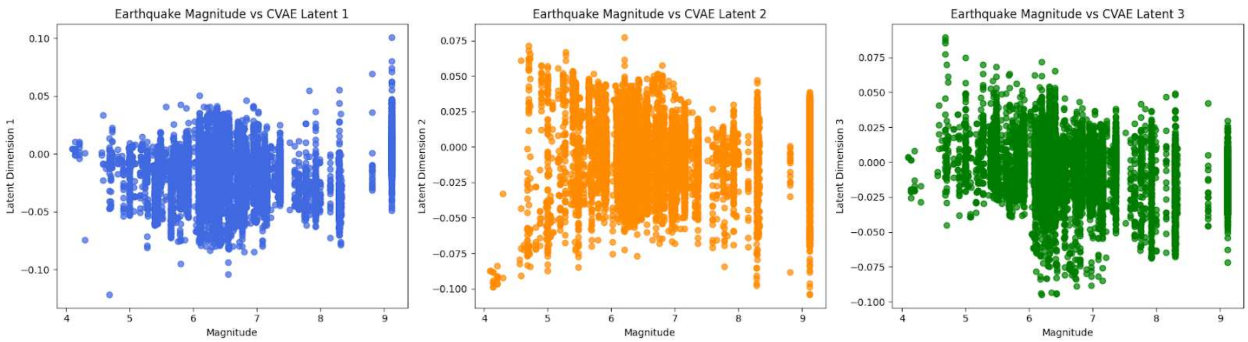
7.Model Performance Metrics for Target Variables:

- The model performs best on "T3pt000S" with an R^2 of 0.9005, low inter-event standard deviation (0.0231), and low intra-event standard deviation (0.0892), suggesting that the predictions for this target are both accurate and stable.
- On the other hand, "T0pt150S" has a lower R^2 of 0.8569 and a slightly higher total standard deviation (0.1057), indicating that the model's performance is slightly less reliable for this target.
- Overall, the model seems to perform well across all targets, with R^2 values above 0.85 for most of them and relatively low standard deviations, indicating robust predictions for spectral accelerations across different ground motion features.

CVAE Target-wise Metrics

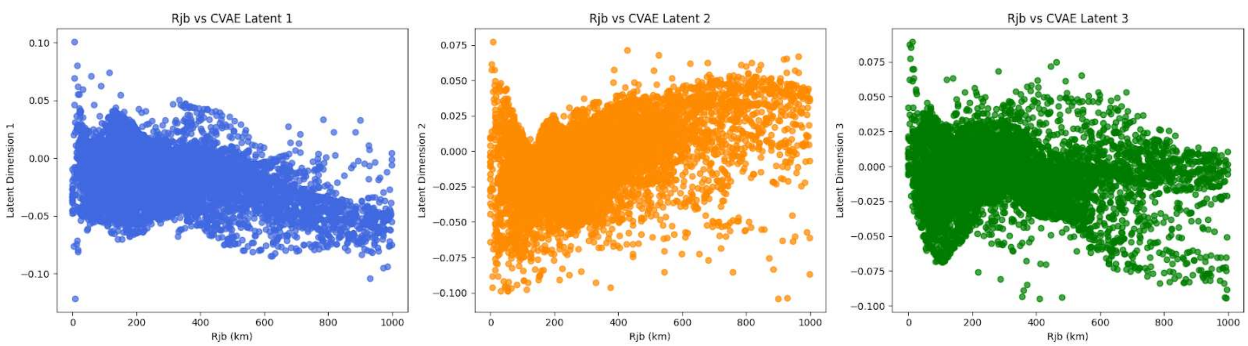
Target	R^2	τ (Inter Std)	ϕ (Intra Std)	Total Std Dev
T0pt010S	0.8817	0.0201	0.0933	0.0954
T0pt020S	0.8545	0.0208	0.1065	0.1085
T0pt030S	0.865	0.0159	0.1066	0.1077
T0pt050S	0.8637	0.0267	0.1014	0.1049
T0pt075S	0.8931	0.0263	0.0933	0.097
T0pt100S	0.8871	0.0149	0.0996	0.1007
T0pt150S	0.8569	0.0193	0.104	0.1057
T0pt200S	0.8878	0.0242	0.098	0.1009
T0pt300S	0.8807	0.0206	0.1002	0.1023
T0pt400S	0.8886	0.0265	0.0923	0.096
T0pt500S	0.8708	0.0258	0.0986	0.102
T0pt750S	0.9086	0.0172	0.0908	0.0924
T1pt000S	0.8757	0.0211	0.0995	0.1017
T1pt500S	0.8899	0.0288	0.0932	0.0976
T2pt000S	0.8881	0.0169	0.0922	0.0937
T2pt500S	0.8861	0.0357	0.0965	0.1029
T3pt000S	0.9005	0.0231	0.0892	0.0921
T3pt500S	0.8881	0.0135	0.0979	0.0989
T4pt000S	0.8844	0.0272	0.0953	0.0991
T5pt000S	0.8796	0.0184	0.1023	0.1039

8a.CVAE Latent Dimensions vs Earthquake Magnitude:



- Latent Dimension 1 (Range: ±0.10)
  - Minimal magnitude dependence
  - Likely encodes source characteristics or regional effects
- Latent Dimension 2 (Range: -0.10 to 0.075)
  - Shows magnitude sensitivity
  - May capture rupture complexity or attenuation
- Latent Dimension 3 (Range: -0.10 to 0.075)
  - Symmetric magnitude relationship
  - Potentially linked to frequency content

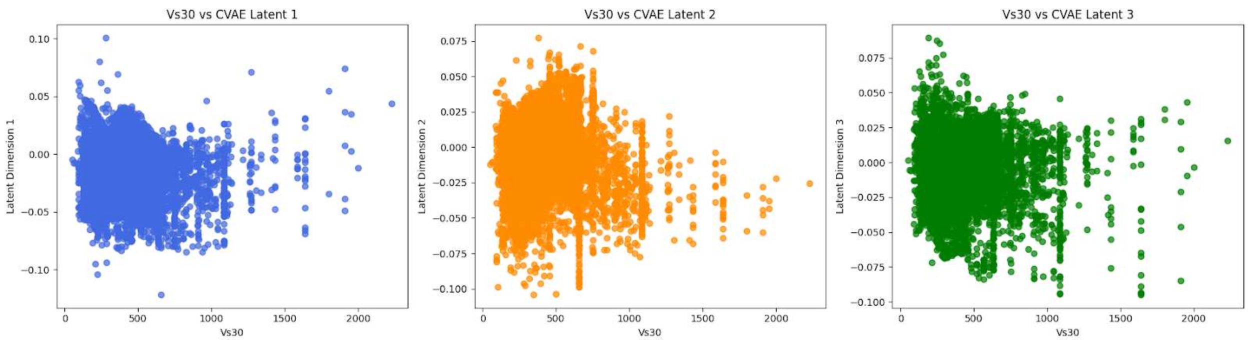
8b.CVAE Latent Dimensions vs Rjb:



- Latent Dimension 1 (Range: ±0.10)
  - Strong negative correlation with Rjb
  - Likely encodes distance-dependent effects (e.g., geometric spreading, attenuation)
  - Minimal influence from other factors
- Latent Dimension 2 (Range: -0.10 to 0.075)
  - Nonlinear relationship with Rjb (notably a "V" shape at short distances, increasing at larger distances)
  - May capture complex distance effects such as near-source saturation or regional path differences
- Latent Dimension 3 (Range: -0.10 to 0.075)
  - Weak negative trend and high scatter
  - Suggests minor or secondary dependence on distance

- May encode subtle path or site effects not captured by the first two dimensions

8c.CVAE Latent Dimensions vs Vs30:



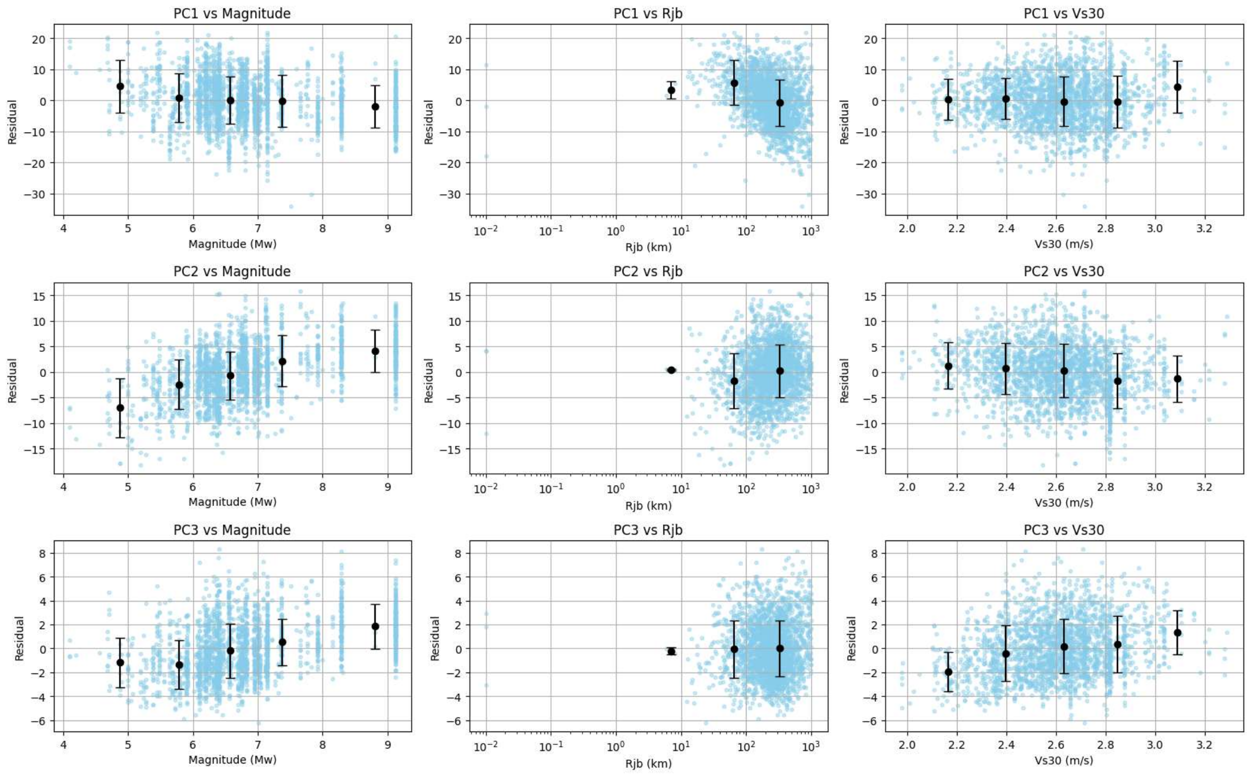
- Latent Dimension 1 (Range: ±0.10)
  - Shows minimal Vs30 correlation
  - Likely encodes source/path effects rather than site response
- Latent Dimension 2 (Range: -0.10 to 0.09)
  - Moderate Vs30 sensitivity
  - May capture nonlinear site amplification effects
  - Higher values could indicate softer soil sites
- Latent Dimension 3 (Range: -0.10 to 0.09)
  - Strongest Vs30 relationship
  - Negative values may correspond to hard rock sites (high Vs30)
  - Positive values likely represent soft soil conditions

8d.Residual Plots for Principal Components:

The residual plots show how the prediction errors for the first three principal components (PC1, PC2, PC3) relate to Magnitude (Mw), distance (Rjb), and Vs30:

- PC1: Residuals are generally centered around zero for all variables, with no strong trends, though there is a slight negative bias at larger Rjb.
- PC2: Residuals are also centered near zero, but show a small positive trend with increasing Magnitude, suggesting minor underprediction at higher magnitudes. No clear trends with Rjb or Vs30.
- PC3: Residuals are tightly clustered around zero for all variables, indicating good model performance and no systematic bias.

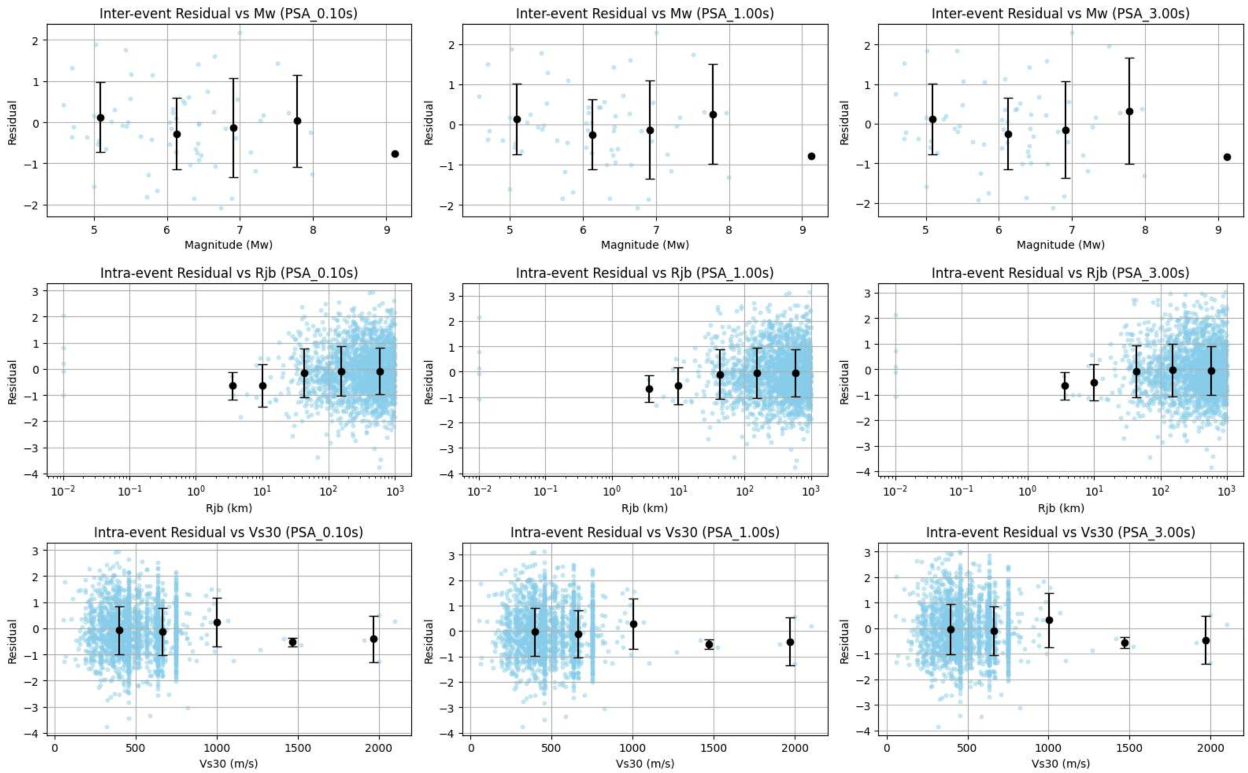
Overall, the model residuals do not show significant dependence on Mw, Rjb, or Vs30, indicating the model is not systematically biased with respect to these variables. Error bars (standard deviations) overlap zero, supporting this conclusion.



9. Inter- and Intra-event Residuals:

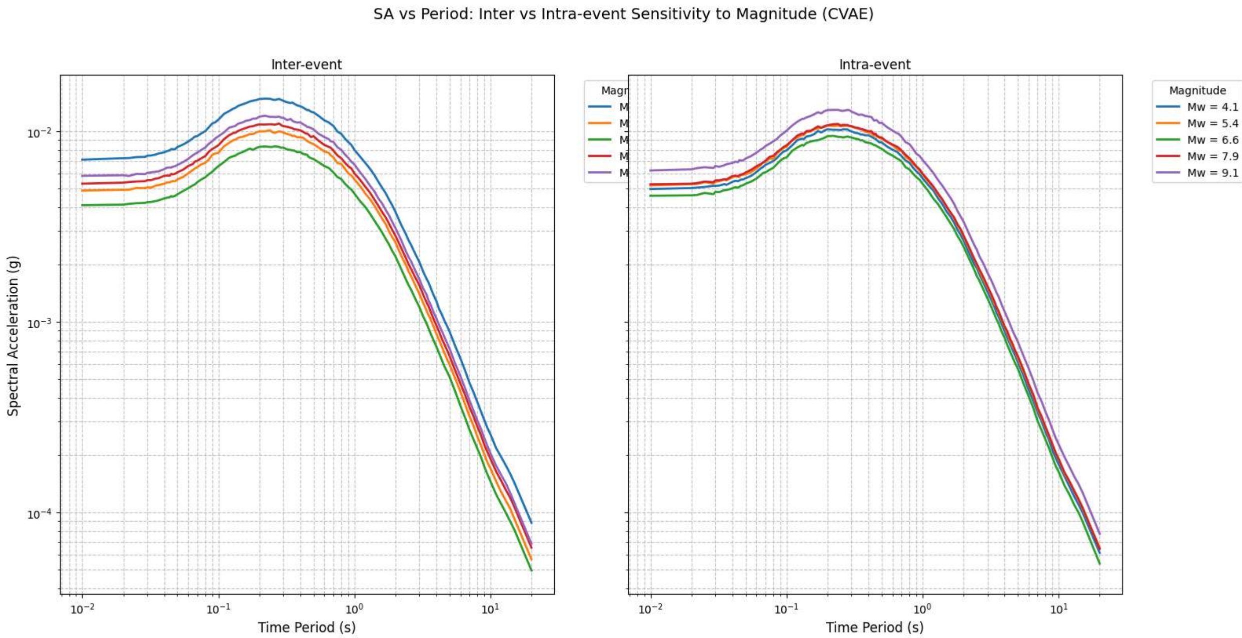
- Inter-event Residuals vs Mw:
  - Residuals are centered around zero for all magnitudes and periods, with no significant trend, indicating no event-size bias.
- Intra-event Residuals vs Rjb:
  - Residuals remain close to zero across all distances, with only slight increases in scatter at larger distances, showing no distance bias.
- Intra-event Residuals vs Vs30:
  - Residuals show no clear trend with Vs30, and mean values are near zero, indicating proper handling of site effects.

Across all plots, the CVAE model demonstrates robust and unbiased performance, with residuals showing no significant trends or systematic errors related to magnitude, distance, or site condition. This indicates the model reliably captures the key features of the data without missing major patterns.

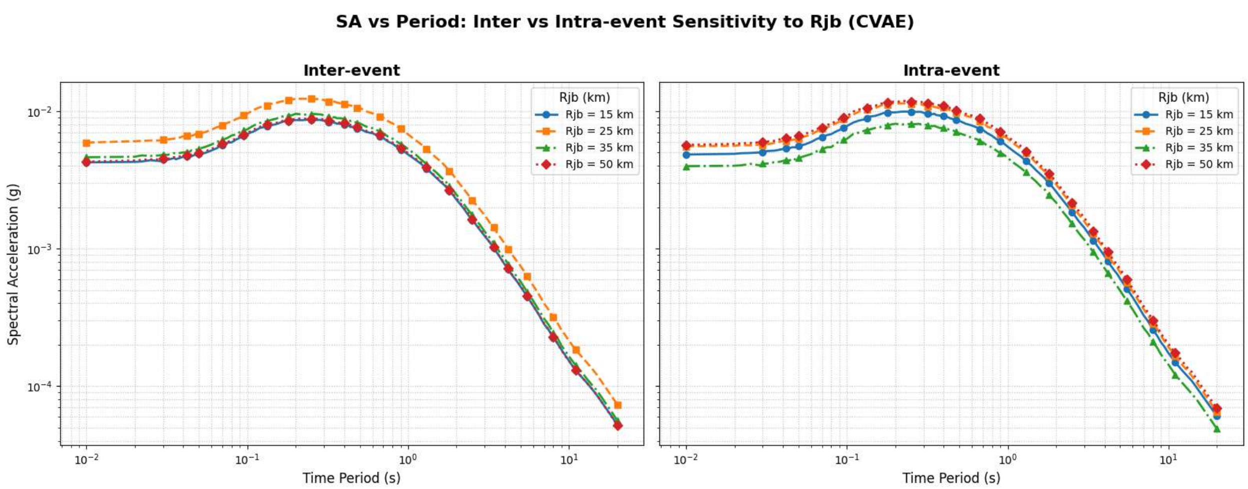


10.Magnitude Sensitivity Plot:

- **Magnitude Effect**:
  - Higher magnitudes (e.g., Mw 9.1) result in **increased spectral acceleration (SA)** across all periods, especially between **0.3–1 s**, due to greater energy release.
- **Period Dependence:**
  - SA differences between magnitudes are **most significant at mid-periods (0.3–1 s)**. At short (<0.3 s) and long periods (>1 s), the curves begin to **converge**.
- **Inter-event vs Intra-event:**
  - **Inter-event:** Shows **more variability**, with some unexpected trends (e.g., Mw 4.1 showing higher SA at short periods), indicating event-specific effects.
  - **Intra-event:** Displays **consistent SA increase with magnitude**, reflecting expected physical behavior.
- The CVAE model effectively captures the **increase in ground motion with magnitude**, especially in intra-event patterns, and handles complex inter-event variations.



11.Rjb Sensitivity Plot:

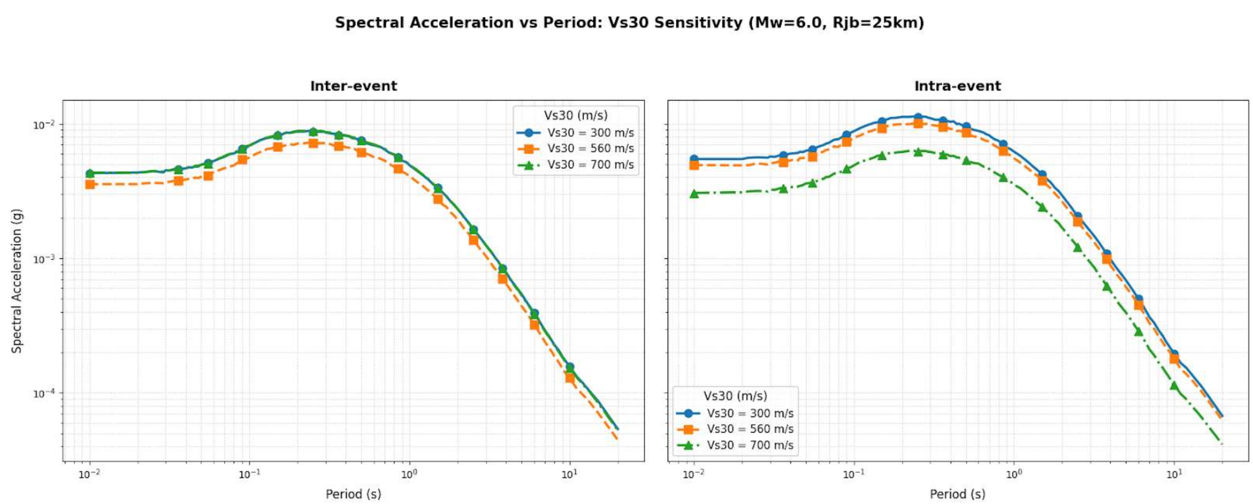


- Distance Dependence:
  - Lower Rjb values (e.g., 15 km) result in noticeably higher spectral accelerations (SA) compared to larger Rjb values (e.g., 50 km), particularly in the intra-event plot. This reflects the expected physical behavior where ground motion attenuation with distance due to energy dispersion and absorption through the medium.
- Period Dependence:
  - The SA difference between near-field and far-field sites is most significant at **short periods (< 0.3 s)**, where high-frequency content dominates the shaking. At **longer periods (> 1 s)**, SA curves for all distances converge, indicating that long-period (flexible) structures are less sensitive to distance-related attenuation.

- Inter-event vs. Intra-event:
  - **Inter-event** variability displays more **irregular distance trends**, with **Rjb = 25 km** unexpectedly showing higher SA than 15 km at peak periods, highlighting the influence of event-specific source characteristics.
  - **Intra-event** variability shows a **smooth, monotonic decrease** in SA with increasing Rjb, demonstrating the model's ability to capture consistent within-event distance attenuation.

The CVAE model effectively captures key ground motion trends—**strong near-source shaking**, **attenuation with distance**, and **period-dependent sensitivity**. It distinguishes between inter- and intra-event patterns, correctly showing that **short-period (rigid) structures are more affected by distance** than long-period (flexible) ones. Inter-event variability amplifies distance sensitivity, while intra-event trends are smoother and more predictable.

12.Vs30 Sensitivity Plot:



Vs30 Dependence :

- Lower Vs30 values (softer soils) lead to **higher spectral accelerations (SA)** across periods, with the greatest amplification at **~0.3–1 s.**

Period Dependence:

- The SA difference across Vs30 values is **most significant at mid-periods**, where site effects strongly amplify shaking. At very short and long periods, differences diminish.

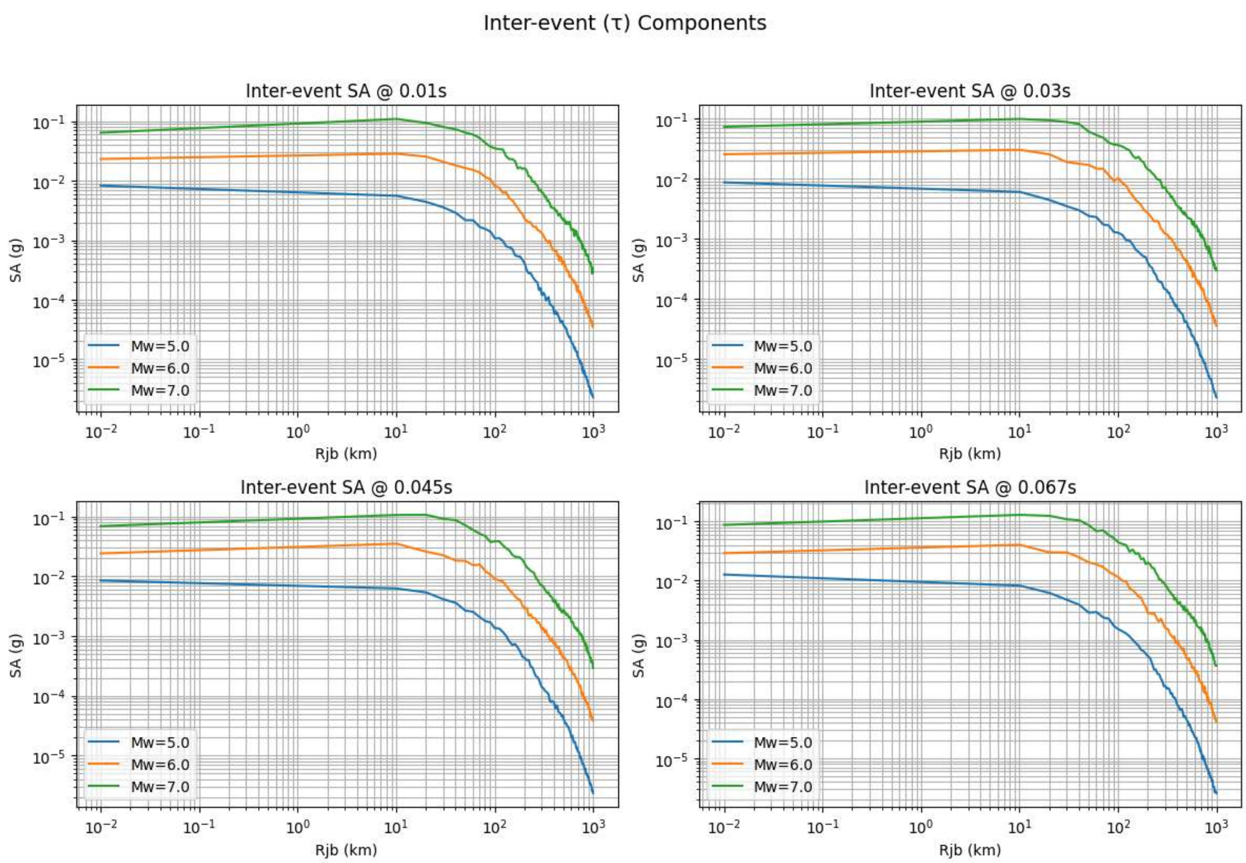
Inter-event vs. Intra-event:

- **Intra-event:** Shows **clearer Vs30 sensitivity**, with consistent SA reduction from soft (300 m/s) to stiff (700 m/s) sites.
- **Inter-event:** Patterns are **less consistent**, suggesting additional event-specific variability affecting site response.

The CVAE model captures **expected site amplification effects**, especially in intra-event predictions, aligning with established ground motion behavior.

13.SA @ T (τ) vs Rjb(inter-event):

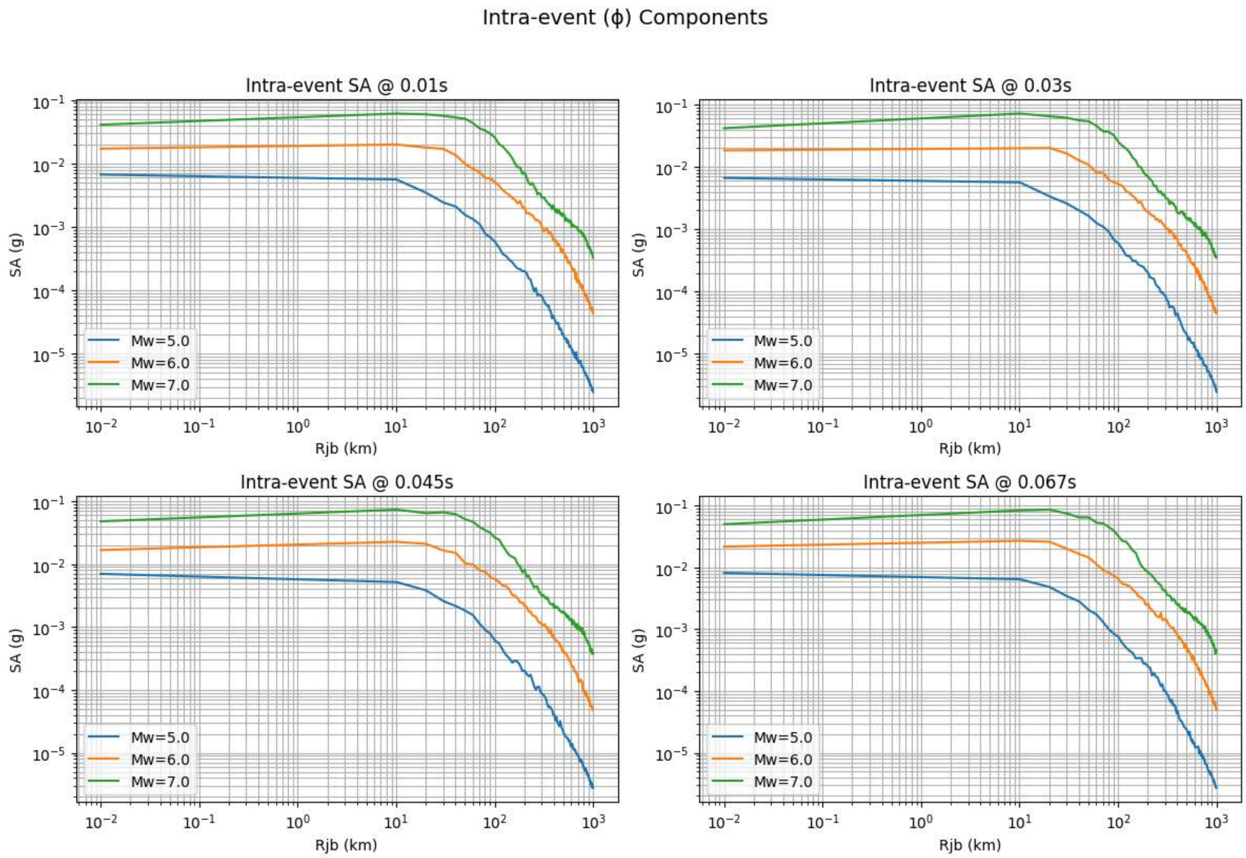
- **Short Distances**: At distances ≤ 10 km, SA is highest, indicating strong ground shaking due to proximity to the earthquake source.
- **Magnitude Influence**: Higher magnitudes (Mw = 7.0) consistently show greater SA compared to lower magnitudes, emphasizing the impact of earthquake strength.
- **Attenuation Trend**: With increasing distance, SA decreases sharply, demonstrating the attenuation of seismic waves. Longer time intervals (0.067s) show more convergence in SA values across magnitudes, indicating reduced sensitivity to earthquake strength at greater distances.



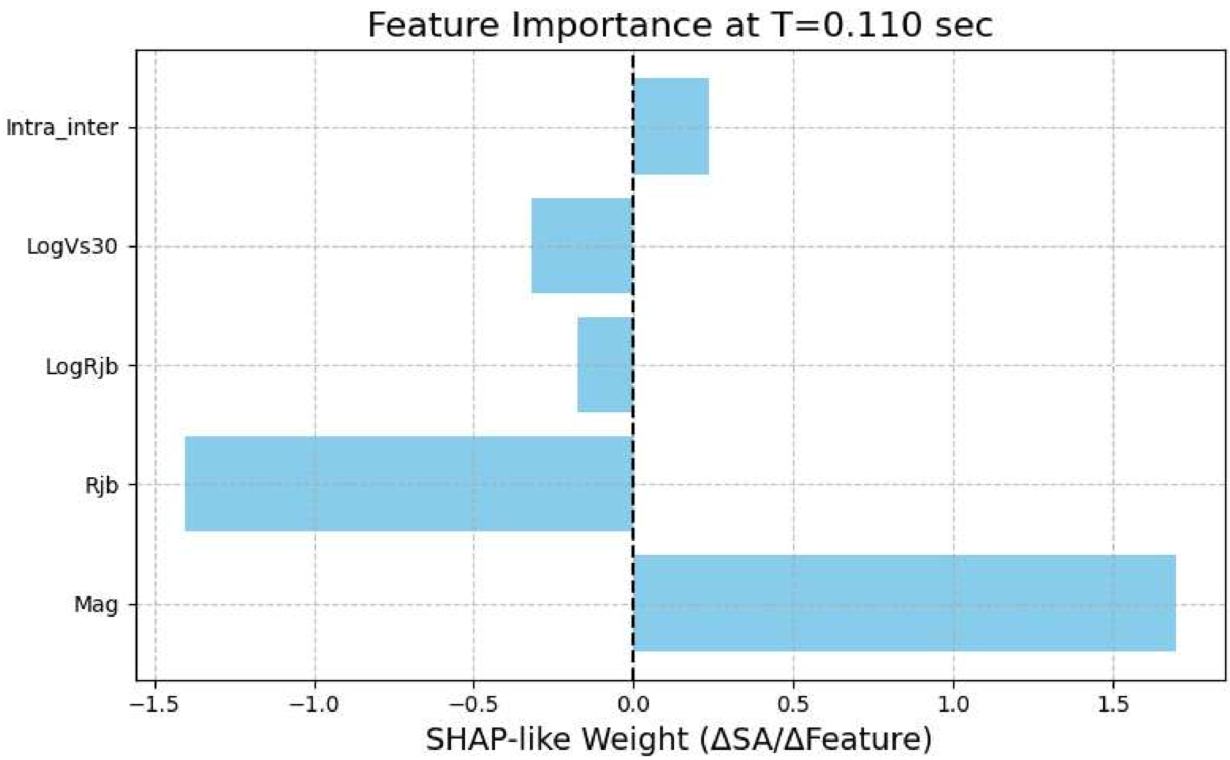
13b.SA @ T (ϕ) vs Rjb(intra-event):

- Intra-event SA (ϕ) decreases with distance (Rjb) for all periods and magnitudes.
- **Short Distances**: At distances ≤ 10 km, SA values are notably higher, indicating significant shaking near the earthquake source.
- **Magnitude Influence**: Higher magnitudes (Mw = 7.0) consistently yield higher SA across all distances, reinforcing the correlation between earthquake strength and ground shaking.

- **Attenuation Trend**: As distance increases, SA decreases, demonstrating the expected attenuation of seismic waves. The convergence of curves at longer time intervals (0.067s) suggests that the impact of magnitude diminishes with distance for longer periods.

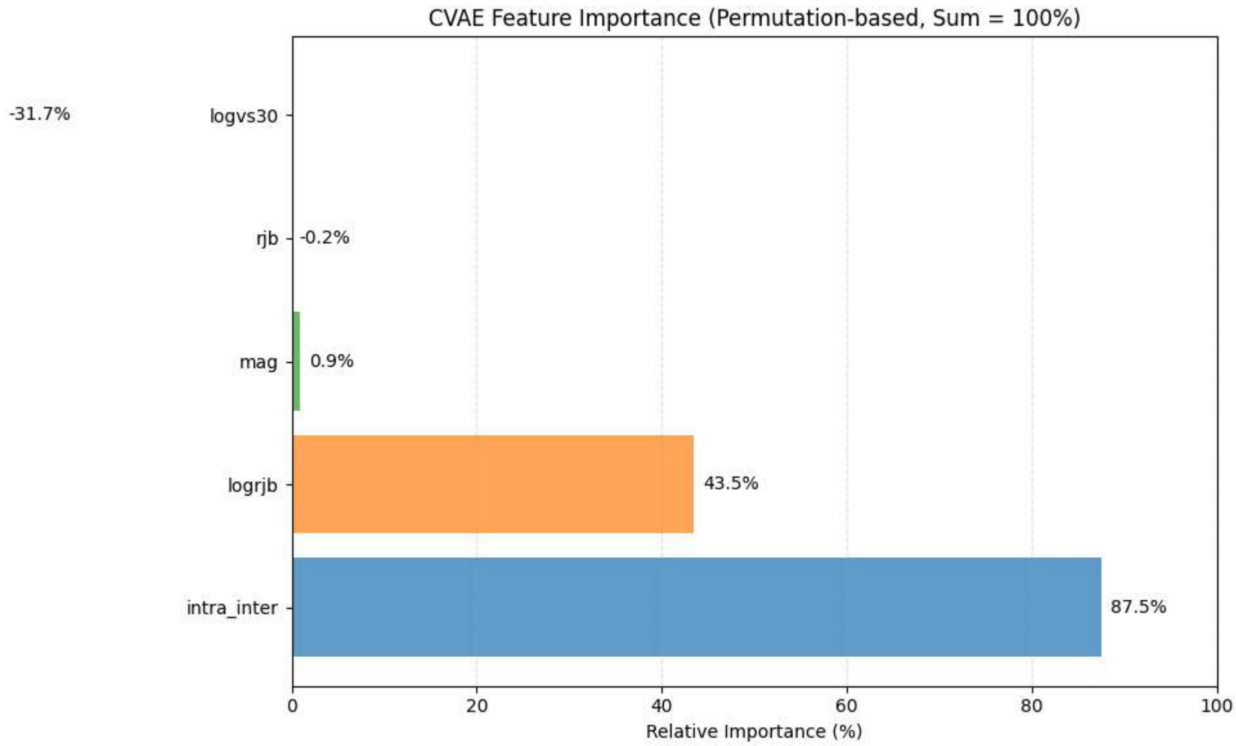


14.SHAP Analysis Summary (T=0.11s SA)



- Magnitude (Mag) is the most influential feature, with the highest positive SHAPED-like weight. This means changes in magnitude have the largest impact on the model's output at this time.
- Rjb (Joyner-Boore distance) is the second most important feature, but its SHAPED-like weight is negative, indicating that increases in Rjb reduce the predicted value.
- LogRjb and LogVs30 have smaller negative contributions, suggesting they have a moderate but negative impact on the prediction.
- Intra_inter has a small positive contribution, indicating a minor positive effect on the prediction.
- Features are ranked by their absolute importance, with the most impactful at the bottom (Mag) and the least at the top (Intra_inter).

15.Feature Importance Summary:



- Intra_inter is by far the most important feature, contributing 87.5% to the model's predictive power. This means the model's output is highly sensitive to whether the event is intra-plate or inter-plate.
- LogRjb is the second most important feature, with a relative importance of 43.5%. This indicates that the logarithm of the Joyner-Boore distance plays a significant role in the model's predictions.
- Magnitude (Mag) and Rjb have minimal contributions, with importances of 0.9% and -0.2%, respectively, suggesting they have little to no effect on the model's output.
- LogVs30 has a negative importance (-31.7%), which may indicate that including this feature introduces noise or reduces model performance, possibly due to multicollinearity or redundancy with other features.

Code: [CVAE Model code]