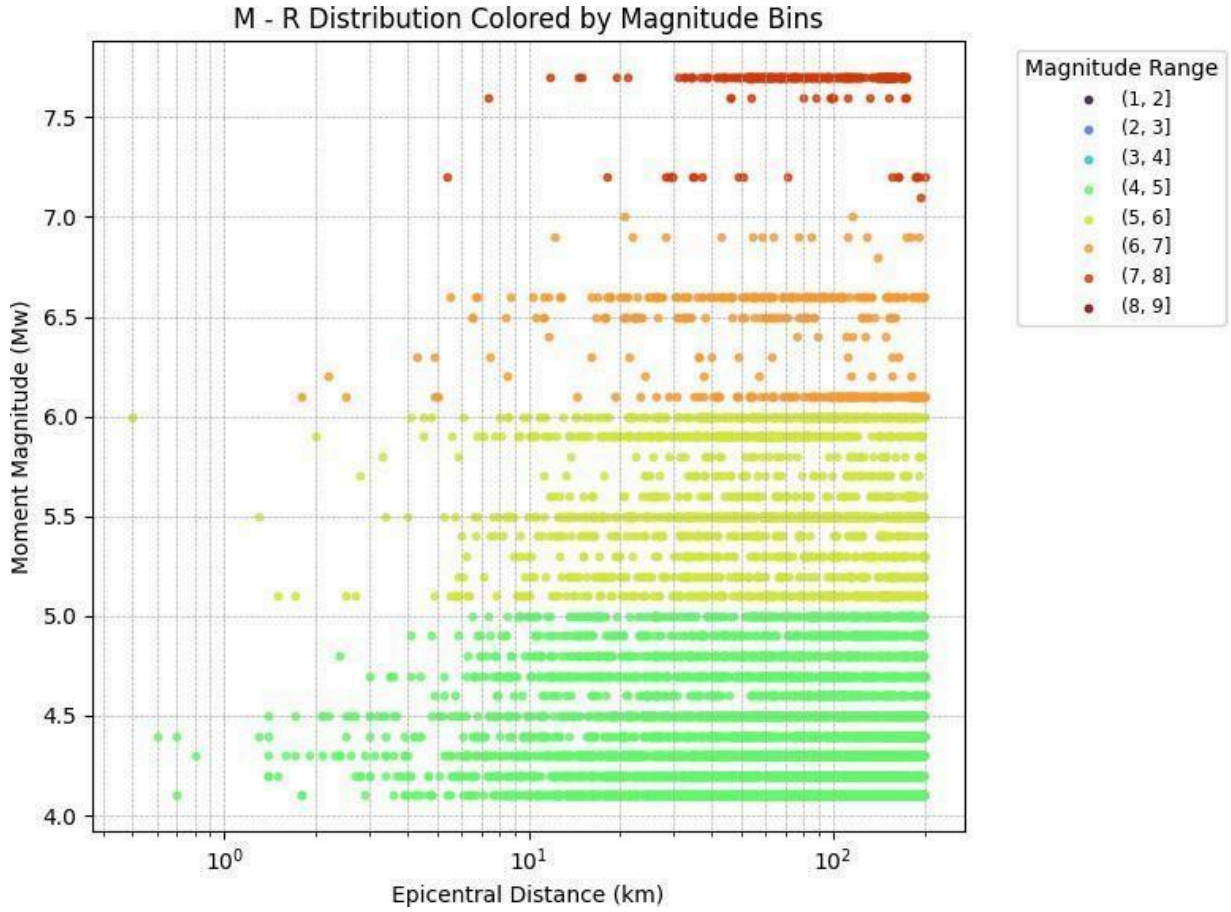
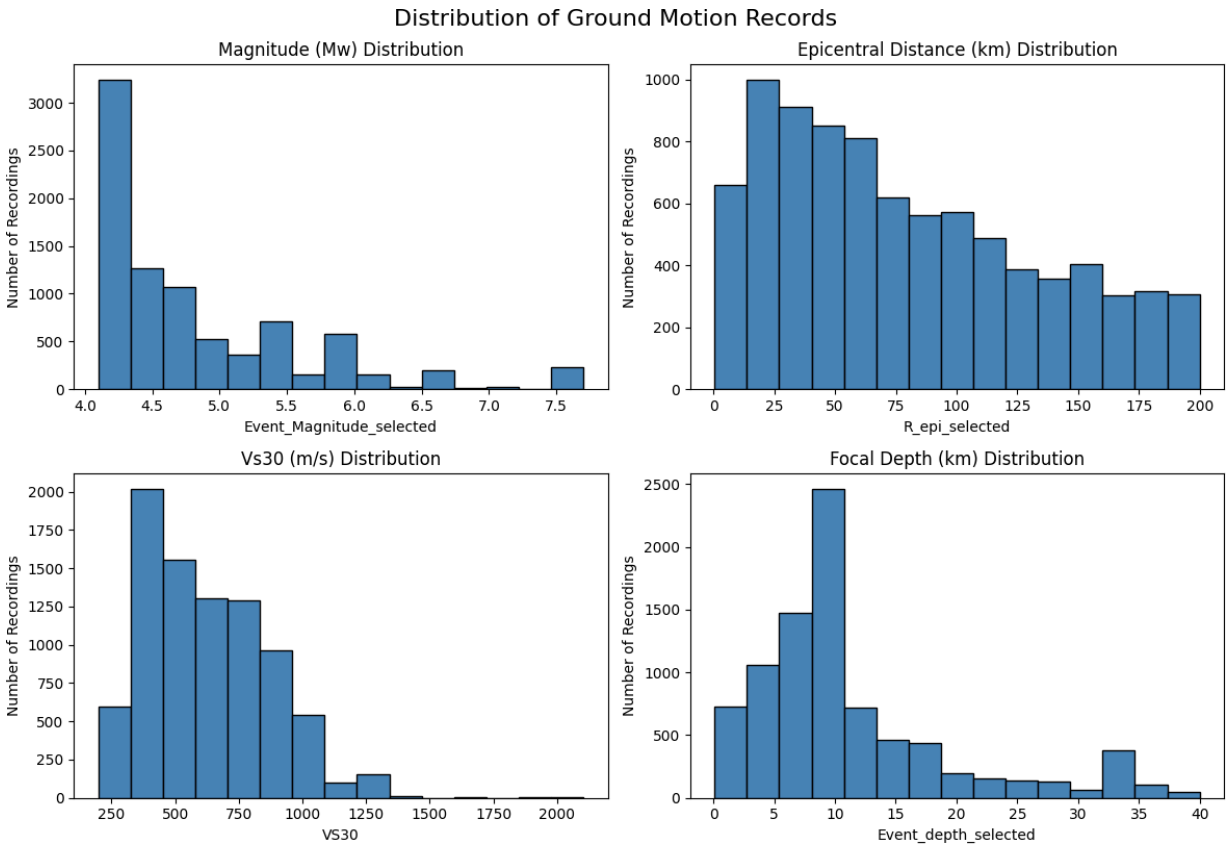
NLPCA (Nonlinear principal component analysis)

1. Magnitude vs Joyner-Boore Distance



- Earthquake magnitudes in the dataset range from about 4 to 8 Mw, with a higher concentration of events between magnitudes 4 and 7.
- There is no clear trend or correlation between magnitude and R epi distance; earthquakes of all magnitudes occur across the full range of distances up to 1000 km.
- The data points are densely clustered at higher R epi distances (100 - 1000 km), indicating most recorded events are relatively away to the reference point, but significant events are also observed at closer distances.

2. Histogram of each input parameter



- **Magnitude (Mw):**

Most earthquake records have magnitudes between 4.0 and 5.0, with a sharp peak around 4 Mw. There are fewer records for both lower (>6.5) and higher (>7.5) magnitudes, indicating the dataset is dominated by slight to moderate earthquakes.

- **R epi distance:**

The majority of records are at distances between 25 km and 75 km, peaking near 20 km. The number of records decreases steadily at larger distances, showing that most recordings are made relatively close to the earthquake source.

- **Shear Velocity (VS30):**

Most records correspond to sites with VS30 values between 200 m/s and 600 m/s, peaking around 400 m/s. This suggests that the data predominantly represent sites with soft to moderately stiff soil conditions, with fewer records from very stiff or rock sites (VS30 > 800 m/s).

3. Table1

Table 01 - Input Parameters:

	Event_Magnitude_selected	R_epi_selected	VS30
min	4.1000	0.5000	198.0000
max	7.7000	199.9000	2104.2000
mean	4.8315	79.5987	623.9816
std	0.8117	53.8051	242.8279
skewness	1.5924	0.5351	0.6112
kurtosis	2.4603	-0.8112	0.0107

	Event_depth_selected
min	0.1000
max	40.0000
mean	11.1482
std	8.2397
skewness	1.5129
kurtosis	1.7629

Table 01 - Output Parameters:

	U_target_selected_1	U_target_selected_2	U_target_selected_3
min	0.0000	0.0000	0.0000
max	4836.5170	4302.4922	3097.6693
mean	10.0283	25.3087	30.2395
std	93.1098	162.5251	154.4057
skewness	24.2859	12.0879	9.2424
kurtosis	974.0350	185.9719	108.9522

	U_target_selected_4	U_target_selected_5	U_target_selected_6
min	0.0000	0.0001	0.0002
max	4299.3010	4713.7618	4196.1082
mean	32.3795	36.3600	37.8630
std	162.0937	175.5214	171.9430
skewness	11.3707	10.8141	10.1882
kurtosis	196.7534	180.0279	151.6814

	U_target_selected_7	U_target_selected_8	U_target_selected_9
min	0.0003	0.0004	0.0006
max	4130.3150	6415.0268	6400.9357
mean	40.1494	42.6016	45.3880
std	172.5549	188.0650	200.8263
skewness	8.3731	11.1739	12.4133
kurtosis	97.7752	218.8951	271.2209

Input Parameters:

- The input variables (mw, r_epi, vs30, depth) show a range of values, with means and standard deviations indicating moderate spread.
- Skewness and kurtosis values suggest that most input parameters are moderately skewed (either positive or negative) and have distributions close to normal.

Output Parameters:

- All output parameters have minimum values of 0 or close, indicating possible zero or censored data.
- The means are generally low compared to their maximums, suggesting that most data points are clustered near the lower end of the range.
- High skewness and kurtosis values across almost all output parameters indicate highly skewed distributions with heavy tails. For example, U_target_selected_8 has skewness of 11.79 and kurtosis of 218.19, showing extreme outliers or rare large values.
- Standard deviations are often close to or larger than the mean, reinforcing the presence of outliers or a wide range of values.

4. NLPCA Model

The model is structured as an autoencoder:

- **Encoder**: Projects input variables (moment magnitude, rupture distance, Vs30, etc.) into a reduced latent space.

- **Decoder**: Reconstructs the full SA spectrum from the encoded representation.

- **Training Pipeline**:

    - Log transformation of SA for numerical stability.

    - Input standardization.

    - Mean squared error (MSE) as the loss function.

    - Adam optimizer with learning rate scheduling and early stopping for convergence control.

5. Latent Space Structure

To interpret the latent representation:

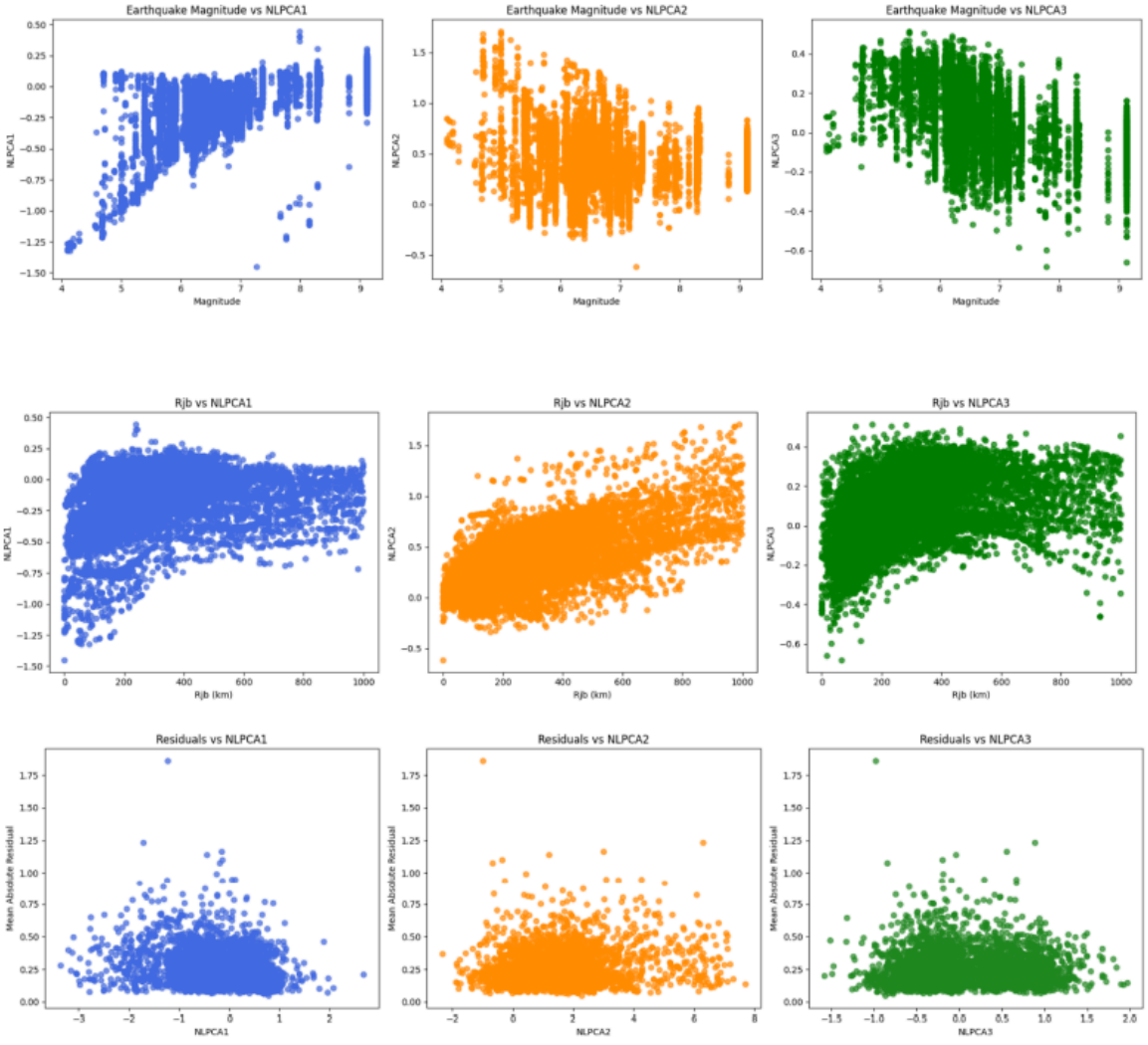
- **Principal Components**: PCA was applied to the encoded features, yielding three dominant latent axes (PC1, PC2, PC3).

- **Observed Trends**:

    - **PC1** correlates strongly with earthquake magnitude.

    - **PC2** captures distance-related attenuation effects.

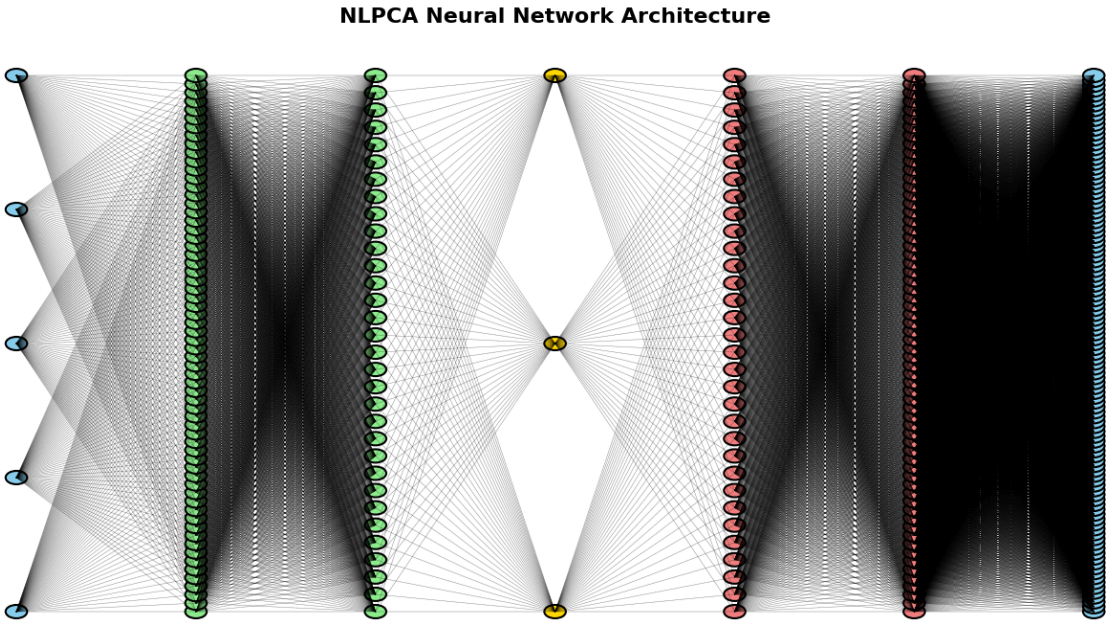
    - **PC3** is loosely associated with site condition variability (Vs30), though noisier.



Aspect	NLPCA1	NLPCA2	NLPCA3
Magnitude Correlation	Strong positive correlation → increases with magnitude	Weak or no clear correlation	Weak or no clear correlation
Distance (Rjb)	Clustered at short distances; spreads with distance → may reflect attenuation	Positive trend → encodes distance-dependent effects	Clustered at short distances; more spread at large distances
Residual Behavior	Residuals uniformly distributed → no systematic prediction bias	Residuals uniformly distributed	Residuals uniformly distributed

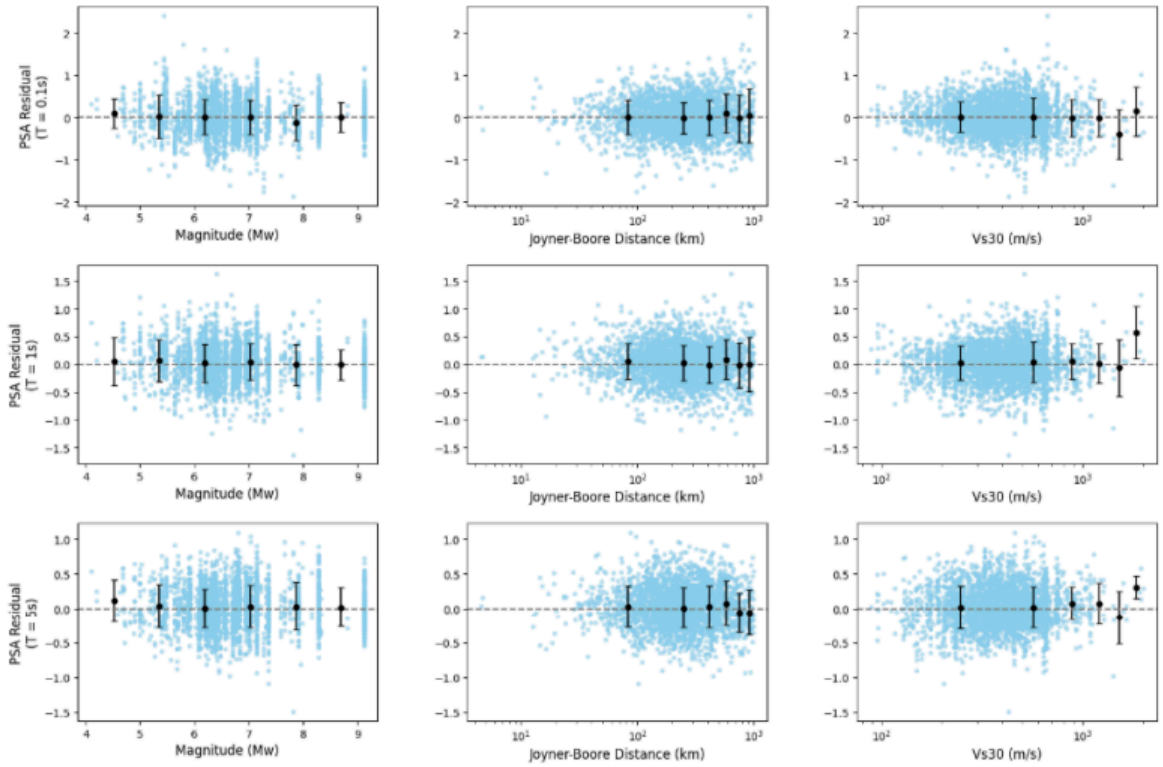
6 Architecture

Component	Configuration
Encoder Layers	64 → 32 → 3 (bottleneck)
Encoder Activation	ReLU → ReLU → Linear
Decoder Layers	3 → 32 → 64 → Output
Decoder Activation	ReLU → ReLU → Linear
Loss Function	Mean Squared Error (MSE)
Optimizer	Adam
Regularization	Early stopping on validation loss (patience = 10 epochs)



7. Residual Plots

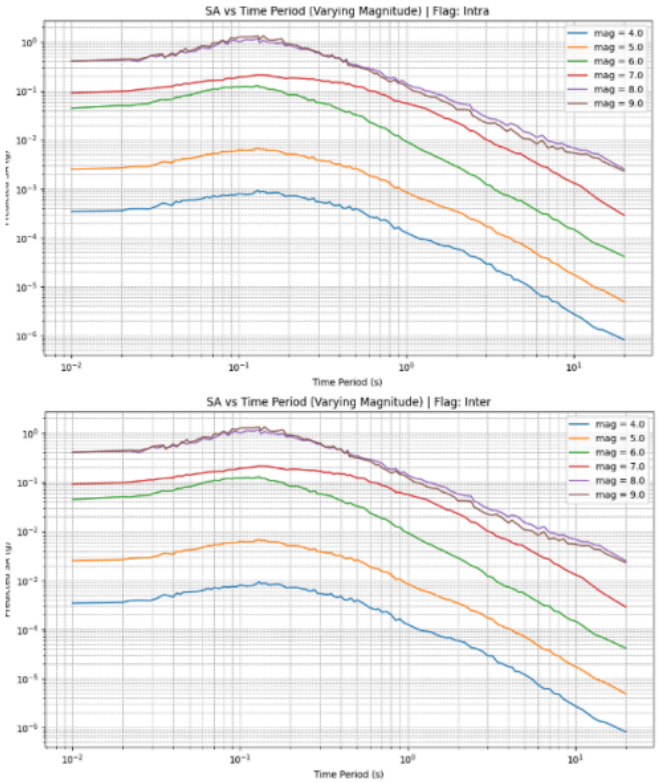
Parameter	Observation	Inference
Magnitude (Mw)	PSA residuals show no significant trend across periods (0.1s, 0.5s, 1.0s). Residuals remain close to zero for all bins.	The model does not systematically over- or under-predict ground motions for different earthquake sizes.
Joyner-Boore Distance	Residuals are centered around zero, with increased scatter at larger distances.	No clear systematic bias with distance; uncertainty increases with distance, likely due to data sparsity.
Vs30 (Shear-wave Velocity)	PSA residuals remain near zero across the range of Vs30 values.	The model adequately accounts for site effects represented by Vs30.



The ground motion model's residuals are unbiased with respect to magnitude, distance, and Vs30. There is no evidence of systematic error across the examined parameters.

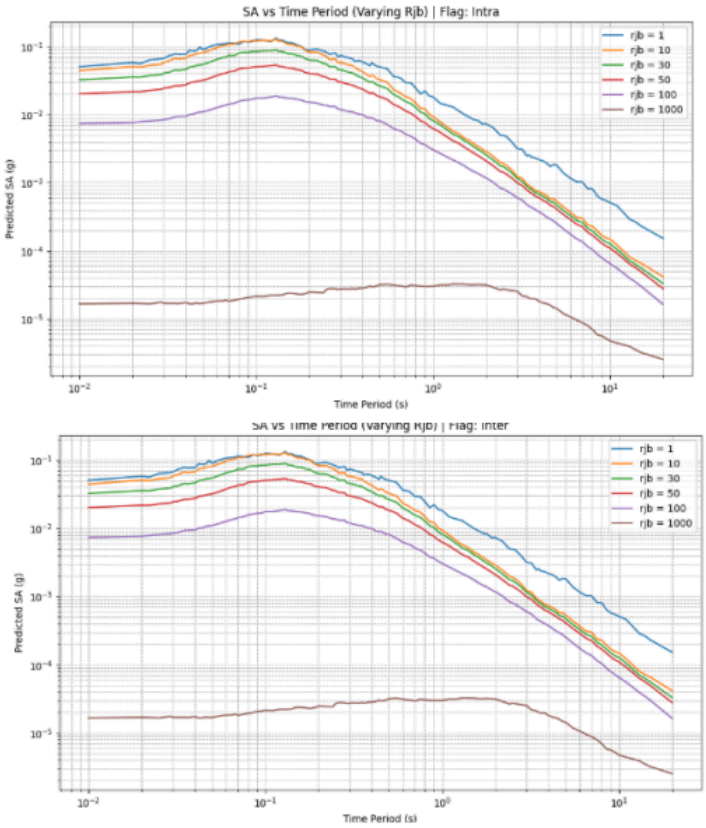
8. Spectral Acceleration vs Time Period at Different Magnitudes

Magnitude	Trend with Time Period	Summary
Mag 4.0, 5.0	Lower Spectral Acceleration (SA) values across all periods.	Larger magnitude earthquakes generate higher spectral accelerations.
Mag 8.0, 9.0	Higher Spectral Acceleration (SA) values across all periods.	Max SA occurs at short to intermediate periods (0.1-1.0s).
All Magnitudes	SA generally peaks in the 0.1s–1.0s range, then decreases.	The period-dependent trend is consistent for intra-event and inter-event.



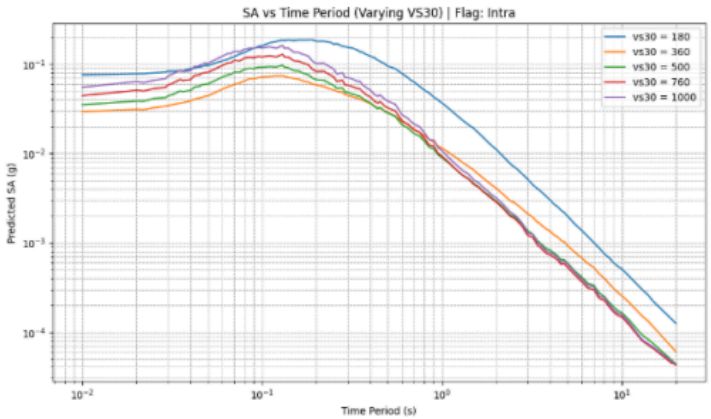
9. Spectral Acceleration vs Time Period at Different Joyner-Boore Distances (Rjb)

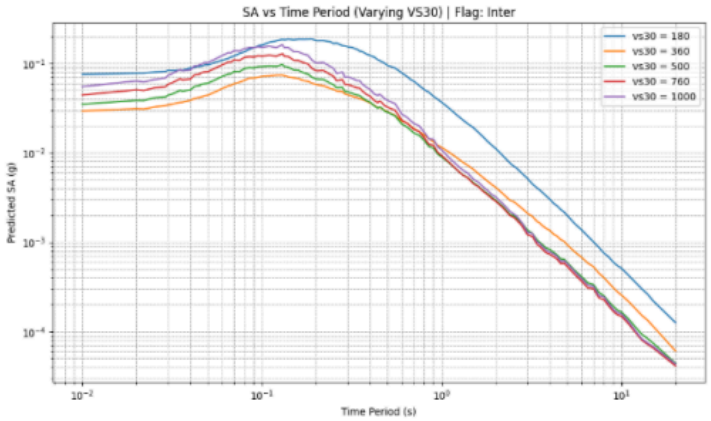
Distance (Rjb)	Trend with Time Period	Summary
1 km	Highest SA values, consistent decrease with distance.	Spectral acceleration decreases as distance increases across all periods.
Larger Distances	Parallel curves across all periods, with decreasing SA values.	The attenuation pattern is similar across the response spectrum.
All Distances	SA peaks at short to intermediate periods (0.1-1.0s), then declines.	Maximum SA occurs at short to intermediate periods for all distances.



10. Spectral Acceleration vs Time Period at Different Vs30 (Shear-Wave Velocity) Values

Vs30 (m/s)	Trend with Time Period	Summary
180 m/s (Softer Soil)	Significantly higher SA across all periods.	Lower Vs30 values (softer soils) amplify ground motions more than higher Vs30 (stiffer soils).
1000 m/s (Stiffer Soil)	Lower SA values across all periods.	Ground motion amplification is most pronounced at short to intermediate periods (0.1-1.0s).
All Vs30 Values	SA peaks at short to intermediate periods, then declines.	Consistent trends for both intra-event and inter-event cases.

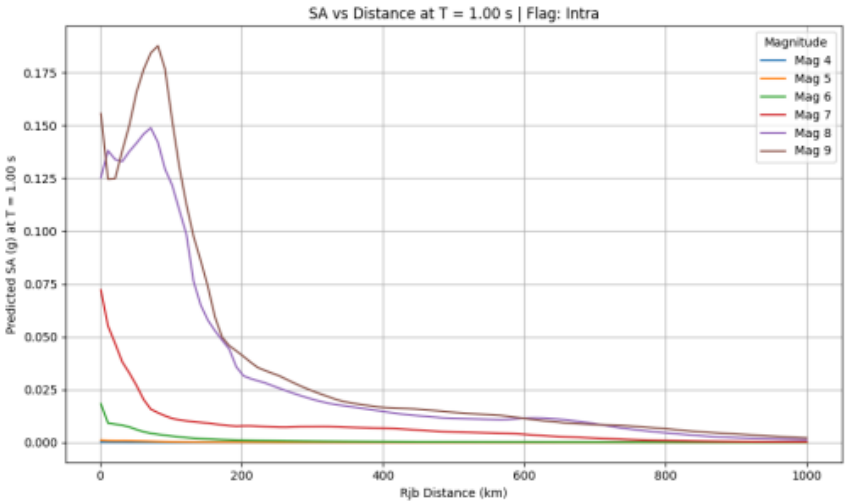




11. Ground Motion Physics with Respect to Rjb

Effect	Trend with Rjb	Summary
Magnitude Effect	At 1.0s period, higher magnitude earthquakes produce higher SA.	Larger magnitude earthquakes generate significantly stronger ground motions.
Distance Attenuation	SA decreases rapidly as distance (Rjb) increases; flattening of curves at larger distances.	Ground motion attenuates quickly with distance for all magnitudes.
Magnitude Separation	Magnitude separation is most pronounced at close distances, narrows with increasing distance.	The effect of magnitude on SA is strongest near the source.

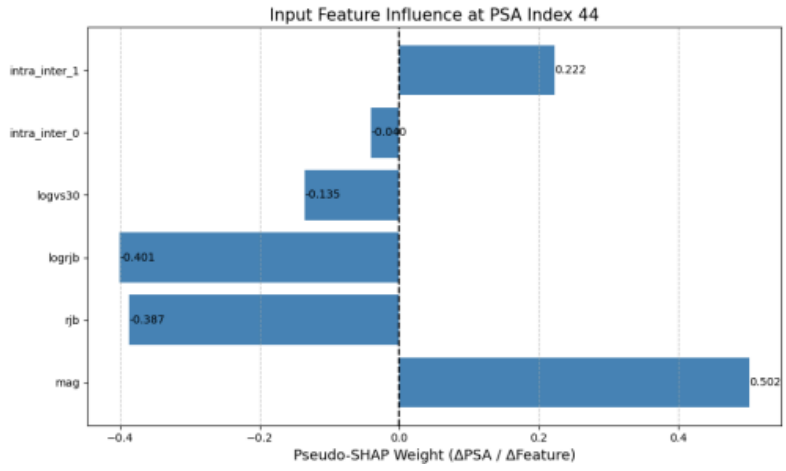
Larger earthquakes generate stronger ground motions, especially near the source. Ground motion decays quickly with distance, and the magnitude effect reduces with distance.



12. SHAP Analysis

Feature	Influence on PSA	Summary
Magnitude (mag)	Strong positive influence (0.5)	Magnitude has the most substantial positive effect on PSA.
Distance (rjb, logrjb)	Strong negative influences (−0.39 and −0.4, respectively)	Greater distance from the source leads to a significant decrease in PSA.
Site Condition (logvs30)	Moderate negative influence (−0.14)	Stiffer sites (higher Vs30) lead to lower PSA values.
Event Type (intra_inter_1, intra_inter_0)	Intra-event: moderate positive (0.2); Inter-event: small negative (−0.04)	Intra-event conditions slightly increase PSA compared to inter-event conditions.

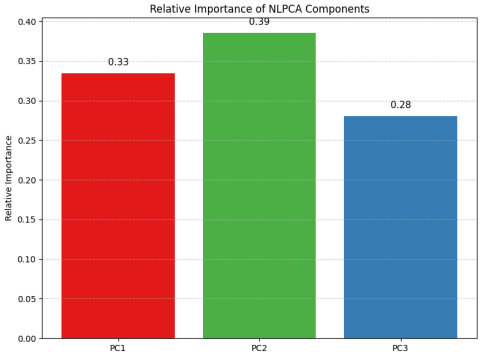
Magnitude is the most influential factor; distance and site conditions reduce PSA. Intra-event conditions have a minor positive influence on PSA.

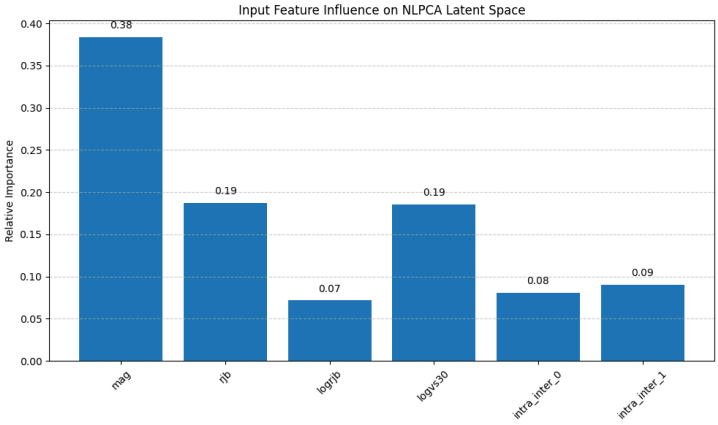


13. Relative Importance of NLPCA Components and Input Features

Component/Feature	Relative Importance/Influence	Summary
NLPCA Components	PC2: 0.39, PC1: 0.33, PC3: 0.28	PC2 is the most important NLPCA component, followed by PC1 and PC3.
Magnitude (mag)	0.38	Earthquake magnitude is the most influential input feature in the latent space.
Distance (rjb, logrjb)	0.19	Distance measures (rjb, logrjb) also play a significant role in the latent space.
Site Condition (logvs30)	0.19	Site conditions (logvs30) are important, though slightly less influential.
Event Type (intra/inter-event)	≤0.09	Intra/inter-event indicators have minimal influence on the latent space.

PC2 is the most important NLPCA component, and magnitude is the key feature influencing the latent space. Distance and site conditions are also significant, but to a lesser degree.


Relative Importance of NLPCA Components
Relative Importance
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.33
0.39
0.28
PC1
PC2
PC3


Input Feature Influence on NLPCA Latent Space
Relative Importance
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.38
0.19
0.07
0.19
0.08
0.09
mag
rjb
logrjb
logvs30
intra_inter_0
intra_inter_1