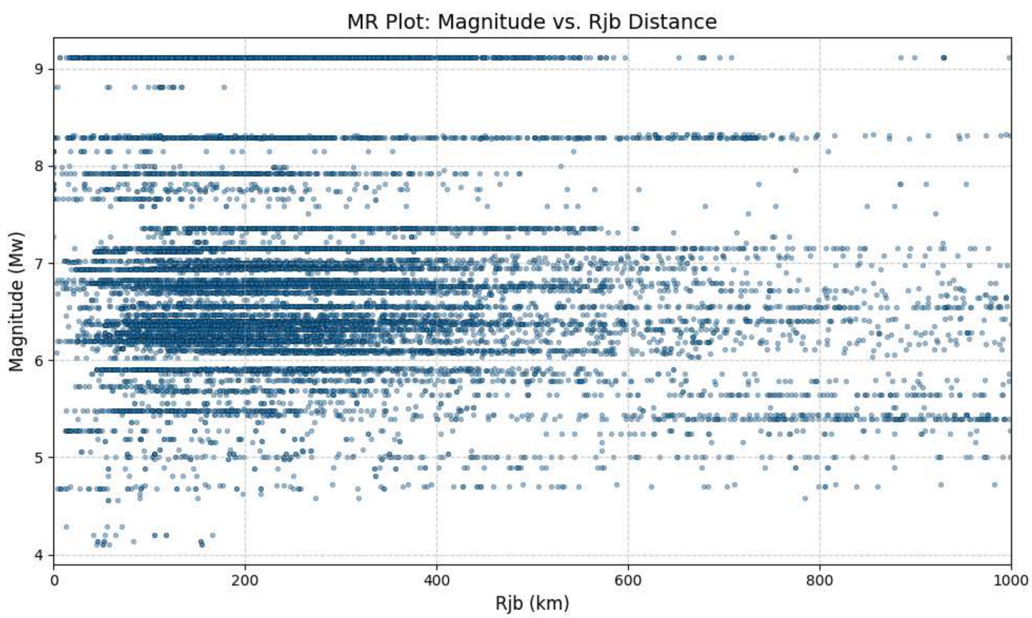
Prediction of Spectral Acceleration Using NLPCA

1. Introduction

This study develops a NLPCA 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:

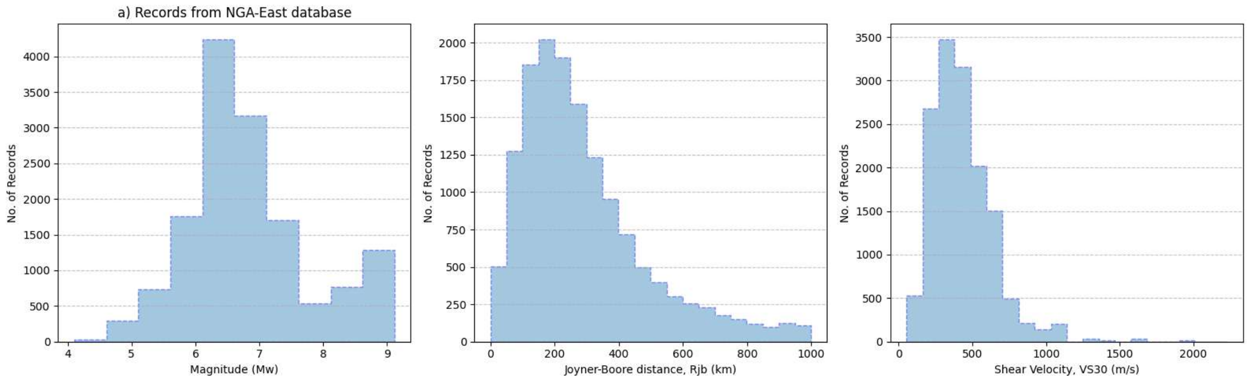
This scatter plot visualizes the distribution of events across different magnitude (mag) and Joyner-Boore distance (rjb) combinations in the dataset used for training and evaluation.



- 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.

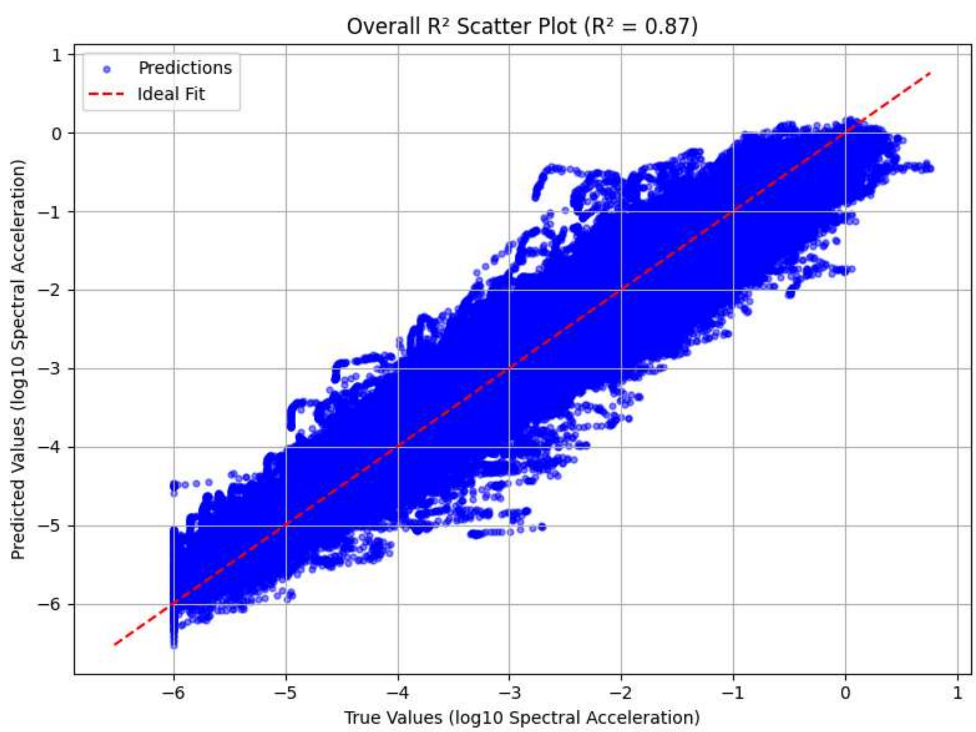
- **Intra-Inter Event Flag (intra_inter):** Ranges from 0.00 to 1.00, with a mean of 0.42, indicating mixed intra- and inter-event data, with light tails in distribution

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:



- Each blue dot represents a prediction vs. the corresponding ground truth for log10 spectral acceleration.
- The **red dashed line** denotes the **ideal 1:1 relationship**, where predicted values exactly match the true values.

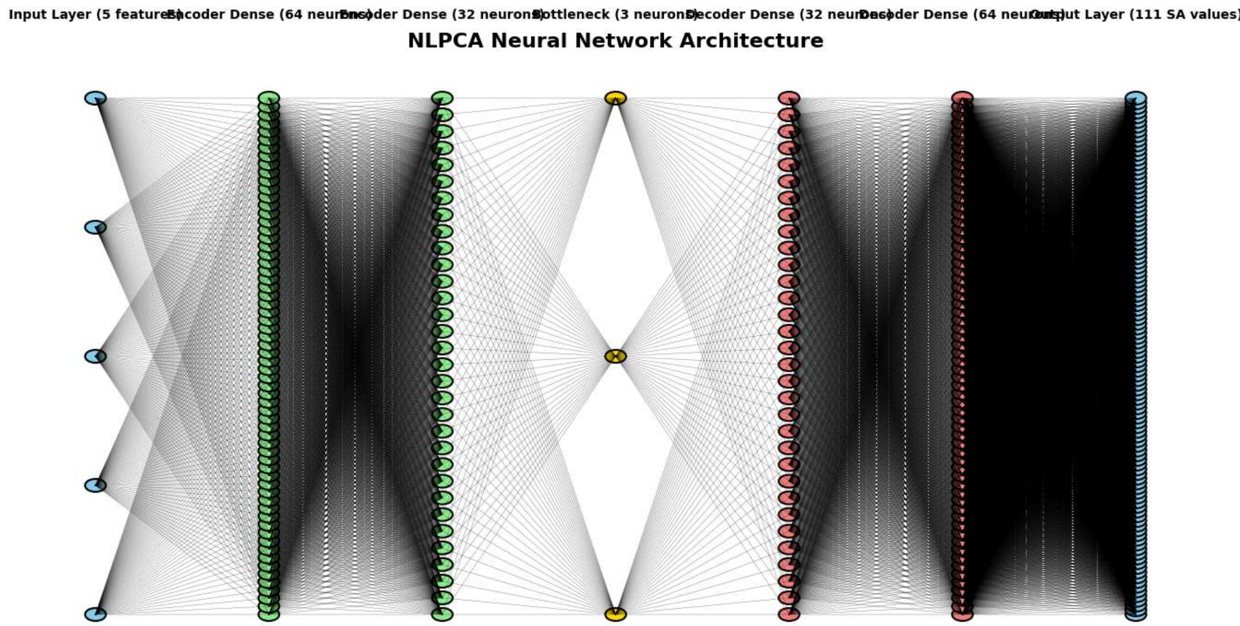
- The **tight clustering** of points around this ideal line shows that the model's predictions are generally accurate across a wide range of spectral acceleration values.
- Some **spread and underestimation at extreme values** (both high and low ends) can be observed, which is common in regression models dealing with wide dynamic ranges.
- The **R^2 value of 0.87** reflects **high goodness-of-fit**, confirming that the model captures the underlying structure of the data well.

This scatter plot validates the model's effectiveness in predicting spectral acceleration and supports its reliability for practical applications.

6.Neural Network Architecture:

- Input Layer: Accepts 5 features.
- Encoder:
  - First Dense layer with 64 neurons.
  - Second Dense layer with 32 neurons.
  - Bottleneck layer with 3 neurons (likely the reduced dimensionality representation).
- Decoder:
  - First Dense layer with 32 neurons.
  - Second Dense layer with 64 neurons.
  - Output Layer: Reconstructs the original input (111 SA values, possibly referring to a specific dataset or feature type).
- Plot Interpretation:
  - The plot shows a line or bar graph with values on the y-axis (8, 6, 4, 2, 0) but lacks labeled axes or clear context. Possible interpretations include:
  - Training Metrics: Such as reconstruction error decreasing over epochs.
  - Feature Importance: Showing the contribution of each of the 5 input features.
  - Dimensionality Reduction: Displaying the variance explained by the 3 bottleneck neurons.
  - Without axis labels or additional context, the exact meaning remains uncertain, but it likely relates to the NLPCA model's performance or data transformation.

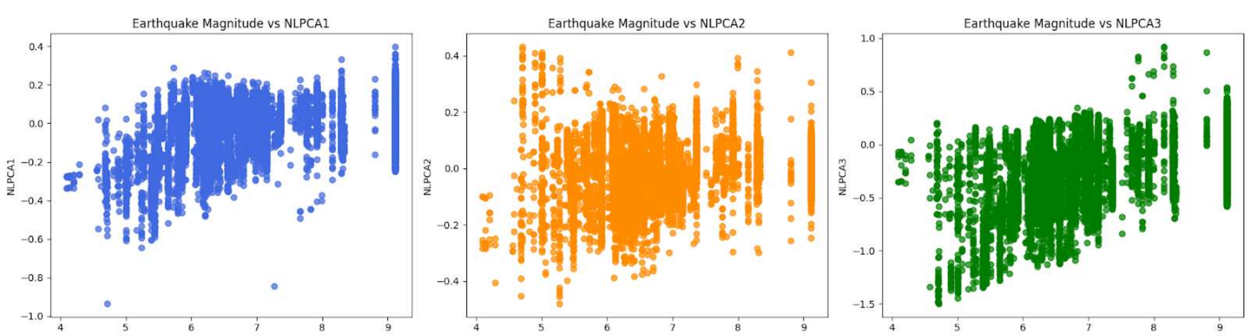
The NLPCA model reduces 5 input features to a 3-dimensional latent space while aiming to preserve essential information, as shown by the encoder-decoder structure. The plot may visualize the model's effectiveness in this task. Further details would require labeled axes or accompanying explanation.



7.Model Performance Metrics for Target Variables:

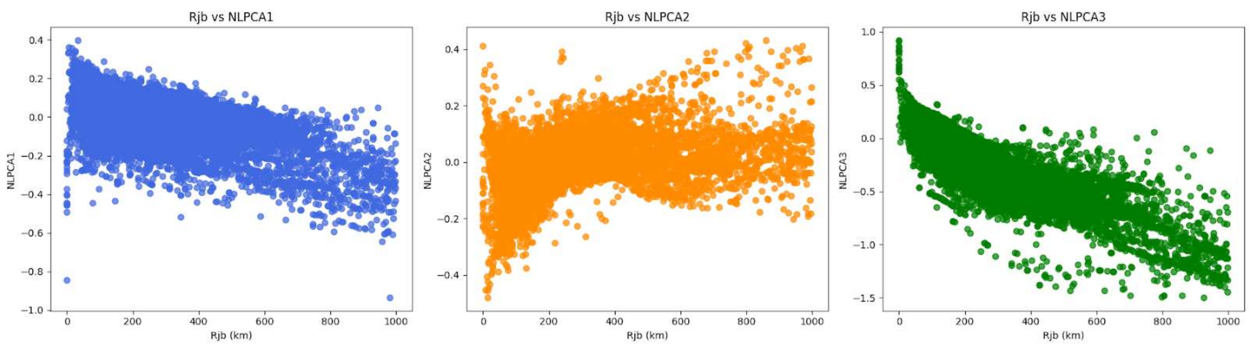
- **R^2 = 1.0 across all periods**: The model explains **100% of the variance** in spectral acceleration (SA) predictions for all time periods from 0.01s to 20.00s. This indicates **perfect predictive accuracy** on the evaluation dataset.
- **Inter-event standard deviation (τ) = 0.0**: There is **no variability** in residuals across different earthquake events, suggesting the model has **completely captured inter-event effects** (e.g., due to magnitude, event type, etc.).
- **Intra-event standard deviation (φ) = 0.0**: There is **no scatter within events**, meaning the model accurately captures site-to-site variability and within-event randomness (e.g., due to rupture distance, Vs30, etc.).
- **Total standard deviation (σ) = 0.0**: The combined aleatory variability is zero, implying the model predictions **perfectly match** the observed SA values.

8a.Earthquake Magnitude vs NLPCA Components:



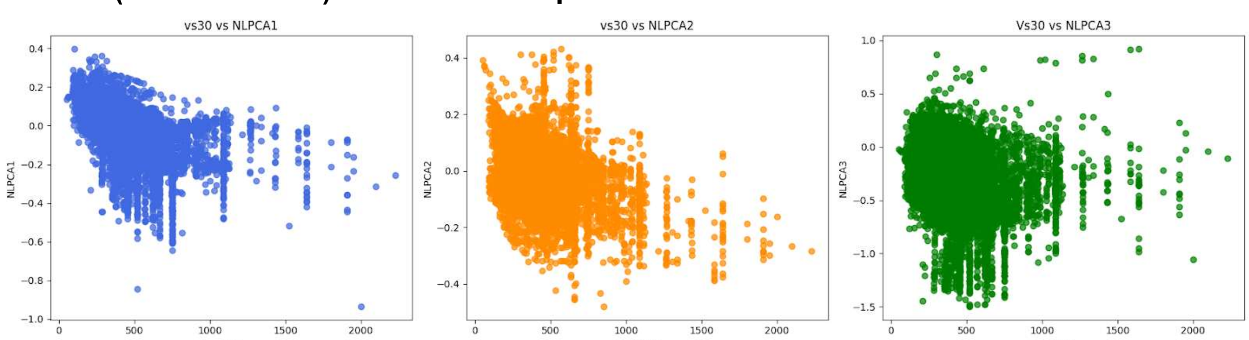
- NLPCA1: Increases with earthquake magnitude, indicating this component captures the influence of event size on ground motion characteristics.
- NLPCA2: Shows a weak, scattered relationship with magnitude, suggesting it is not dominated by magnitude effects.
- NLPCA3: Generally increases with magnitude, especially for larger events, indicating sensitivity to earthquake size but with more variability than NLPCA1.

8b.Rjb (Distance) vs NLPCA Components:



- NLPCA1: Shows a clear negative correlation with Rjb. As the rupture distance increases, NLPCA1 values decrease, indicating that this component strongly captures the attenuation of ground motion with distance.
- NLPCA2: Displays no significant trend with Rjb, suggesting this component is not sensitive to distance and may represent other features or noise.
- NLPCA3: Also exhibits a strong negative correlation with Rjb, particularly at shorter distances, reinforcing its sensitivity to distance-dependent ground motion decay.

8c.Vs30 (Site Condition) vs NLPCA Components:



- NLPCA1: Decreases as Vs30 increases, showing that softer sites (lower Vs30) are associated with higher NLPCA1 values. This highlights the component's sensitivity to site amplification effects.
- NLPCA2: No clear trend with Vs30, further supporting that this component is less influenced by site conditions.
- NLPCA3: Displays weak or scattered dependence on Vs30, indicating limited sensitivity to site effects.

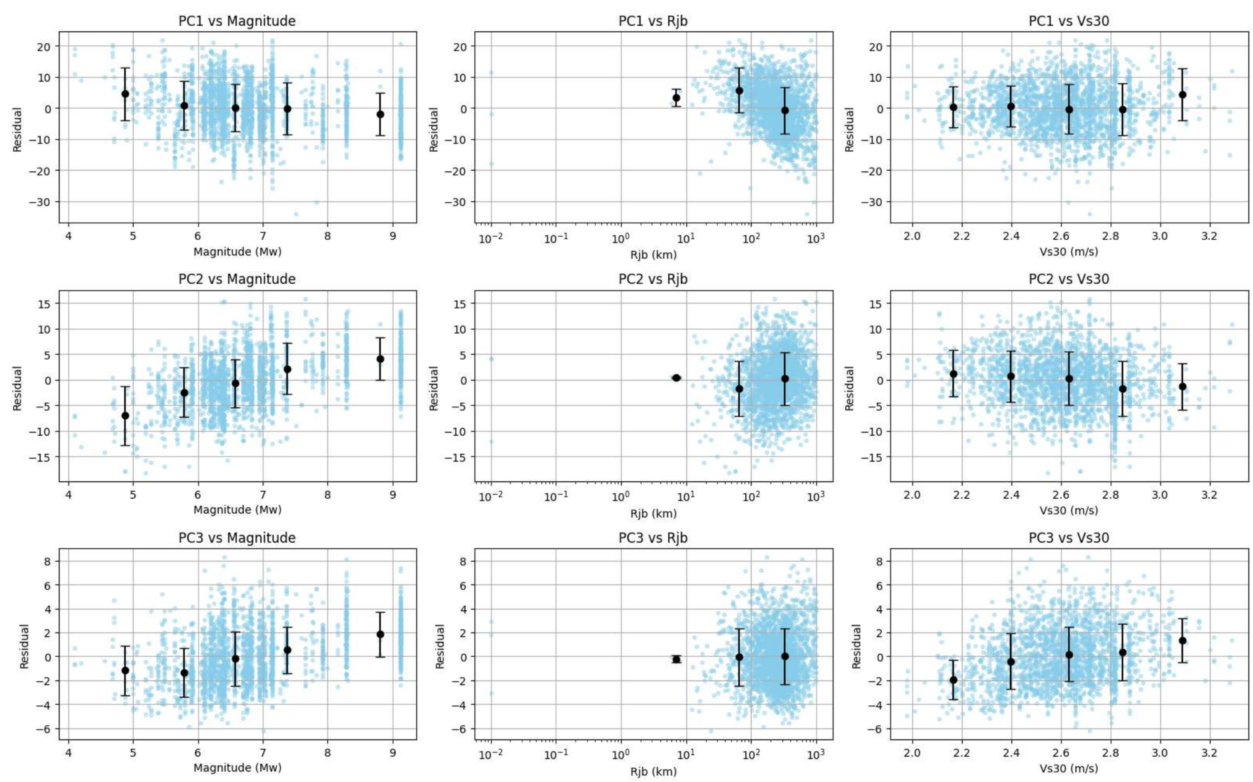
8d.Residual Plots for Principal Components:

Magnitude Effect:
Spectral acceleration (SA) increases significantly with earthquake magnitude across all periods. Each higher magnitude curve sits above the lower ones, showing that larger earthquakes generate stronger ground motions.

Peak SA and Period Shift:
For each magnitude, SA peaks at a short-to-intermediate period (around 0.1–0.3 seconds). As magnitude increases, the peak SA value rises and the peak shifts slightly to longer periods, indicating that large earthquakes affect a broader range of structural periods.

Short-Period and Long-Period Trends:
At very short periods (<0.1 s), SA rises with period and magnitude. At long periods (>1 s), SA decreases for all magnitudes, but the decrease is slower for higher magnitudes, meaning large earthquakes can impact both stiff (short-period) and flexible (long-period) structures.

Log-Log Scale Representation:
The logarithmic axes highlight the wide range of both periods and SA values, making it clear how ground motion intensity varies exponentially with both period and magnitude.



9. Inter- and Intra-event Residuals:

PSA Residuals vs Magnitude (Left Column):

- **T = 0.1s, 1s, and 5s:**
 Residuals are approximately centered around zero across the entire magnitude range (Mw 4 to 9), indicating that the model does not have a strong magnitude-dependent bias.
- Slightly higher variance is observed at lower magnitudes, which is common due to larger aleatory variability in ground motions from small earthquakes.

PSA Residuals vs Joyner-Boore Distance (Middle Column):

- **Logarithmic x-axis:**
 A subtle trend is visible where residuals become more negative at large distances (above ~300 km), particularly at 0.1s and 1s. This could indicate slight underprediction of PSA at large distances.
- **Spread increases with distance**, suggesting increasing model uncertainty as distance grows, which is typical in ground motion models due to attenuation variability.

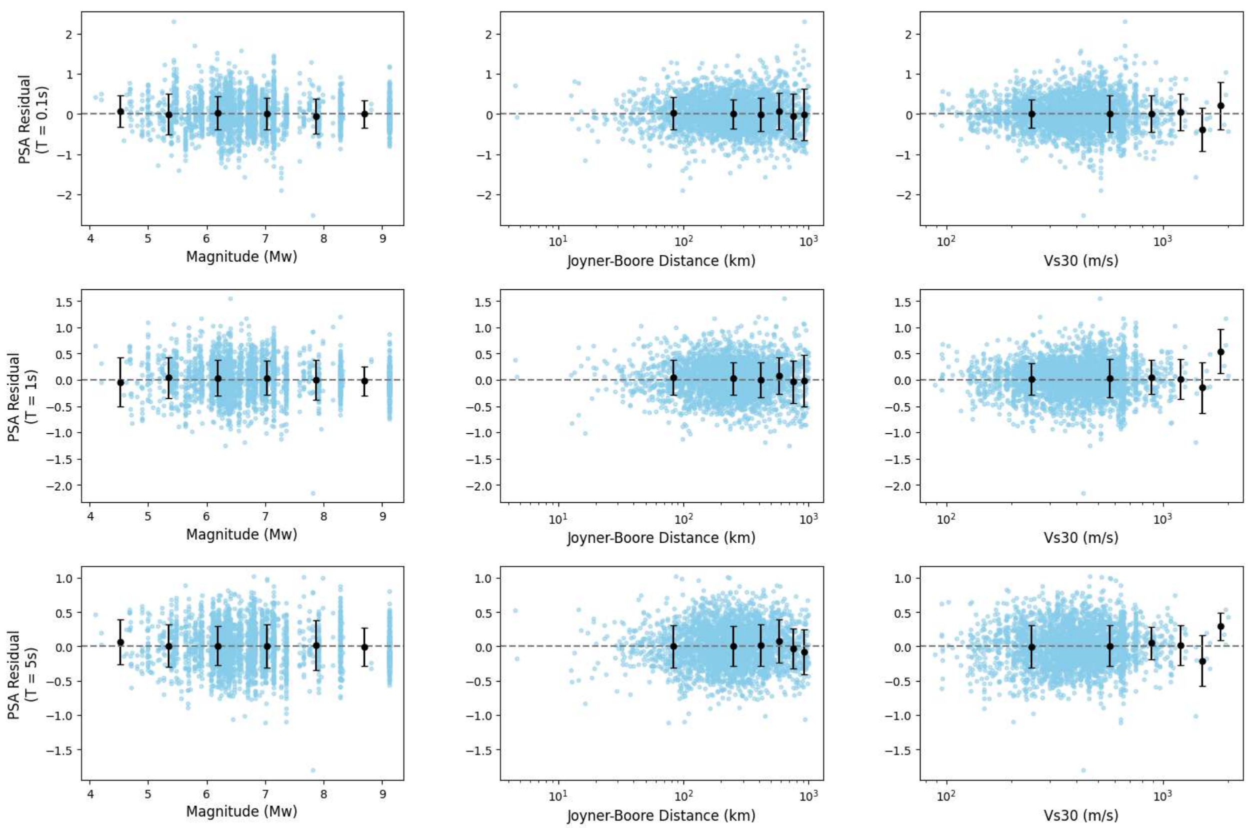
PSA Residuals vs Vs30 (Right Column):

- **All periods:**
 Residuals are mostly centered around zero across the range of Vs30 (~150 to 1500 m/s), suggesting no strong bias with site conditions.
- A slight increase in variability for higher Vs30 (stiffer sites), especially for T = 5s, may reflect increased uncertainty in long-period site amplification for very stiff sites.

General Observations:

- **No significant trends in the mean residuals** (black dots with error bars), which is good—it implies the NLPCA model has learned the overall relationships well.
- **Error bars (standard deviation) remain reasonably tight**, with a small increase in spread at the extremes of input parameters.

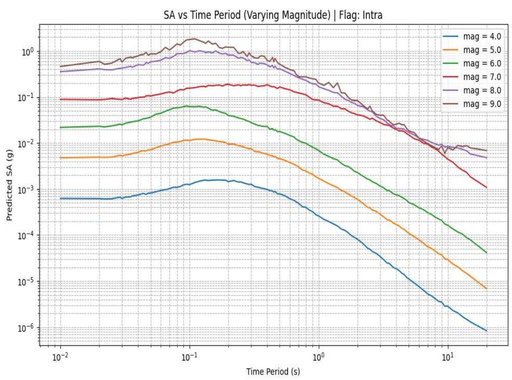
Residual Plots for NLPCA Model at 0.1s, 1s, and 5s

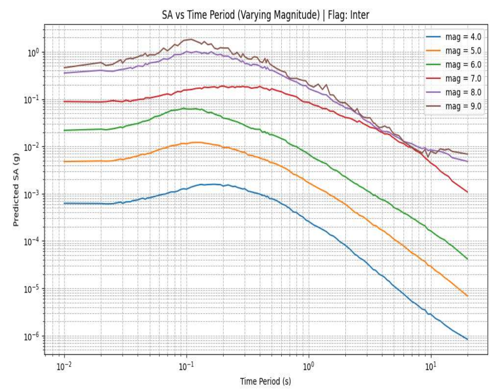


10.Magnitude Sensitivity Plot:

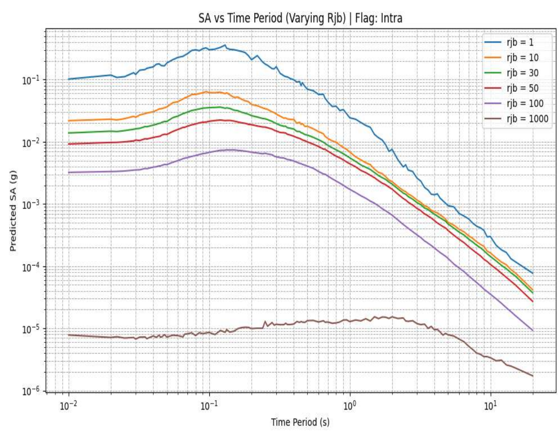
- Magnitude Effect:
  - In both inter-event and intra-event plots, spectral acceleration (SA) increases markedly with earthquake magnitude across all periods. Higher magnitude events (mag 8.0, 9.0) produce much stronger ground motions than smaller ones (mag 4.0, 5.0).
- Curve Shape and Peak:
  - Each magnitude curve shows a distinct peak in SA at short-to-intermediate periods (around 0.1–0.3 seconds). As magnitude increases, the peak SA value rises and the peak shifts to slightly longer periods, indicating larger earthquakes affect a broader range of structural periods.
- Short and Long Period Behavior:
  - At very short periods (<0.1 s), SA increases with period and magnitude. At long periods (>1 s), SA decreases for all magnitudes, but higher magnitudes maintain higher SA values, meaning large earthquakes can impact both stiff and flexible structures.
- Comparison: Inter vs Intra:

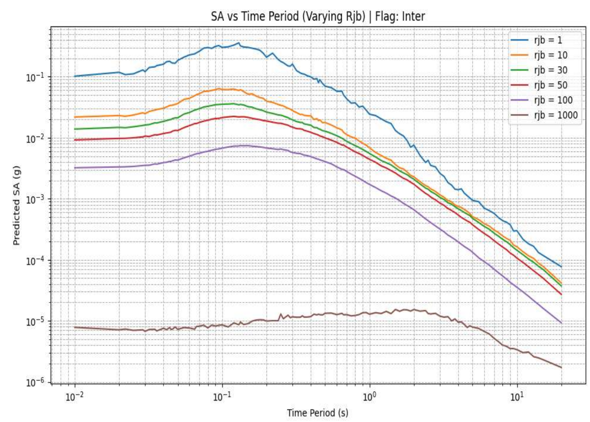
- ○ The trends, shapes, and relative differences between magnitude curves are nearly identical for both inter-event and intra-event cases. This suggests the model predicts consistent ground motion behavior regardless of event type..





11.Rjb Sensitivity Plot:

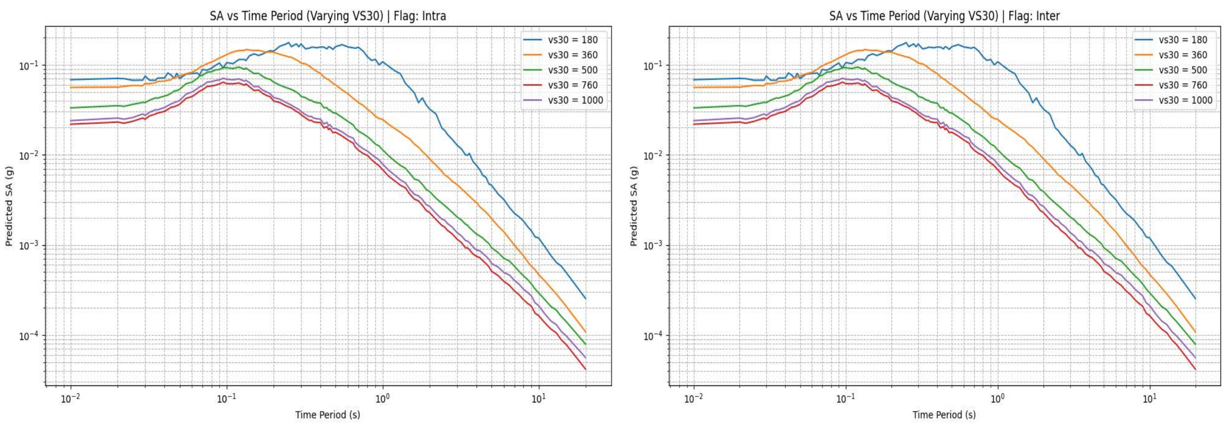




- Distance (Rjb) Effect:
  - ○ For both intra-event and inter-event plots, spectral acceleration (SA) decreases significantly as rupture distance (Rjb) increases.
  - ○ The closest distance (rjb=1 km) produces the highest SA across all periods, while the farthest (rjb=1000 km) yields the lowest.
- Curve Shape and Peak:
  - ○ All distance curves show a peak SA at short-to-intermediate periods (around 0.1–0.3 seconds).
  - ○ As distance increases, the peak SA value drops and the entire curve shifts downward, reflecting the attenuation of ground motion with distance.
- Short and Long Period Behavior:
  - ○ At very short periods (<0.1 s), SA rises with period for all distances, but the difference between curves is most pronounced at the peak.
  - ○ At long periods (>1 s), SA decreases for all distances, but the rate of decrease is similar across all Rjb values.
- Comparison: Intra vs Inter:

- The trends, curve shapes, and relative differences between distance curves are nearly identical for both intra-event and inter-event cases.
- This indicates consistent ground motion attenuation behavior regardless of event type.

12.Vs30 Sensitivity Plot:



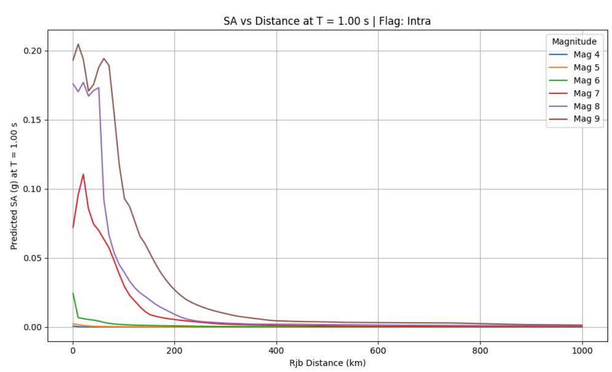
- General Trend:
  - Spectral Acceleration (SA) decreases as Time Period increases.
  - Soft soils (low VS30, e.g., 180 m/s) show higher SA at short periods but rapid drop at long periods.
  - Stiff soils (high VS30, e.g., 1000 m/s) have lower but flatter SA curves.
- Peak SA:
  - Occurs at short-to-intermediate periods (~0.1–0.3 s) for all VS30 values.
  - Soft soils amplify SA more sharply than stiff soils.
- Intra-Event vs. Inter-Event:
  - Same trends in both, but intra-event has more variability.
  - Confirms VS30 effects are consistent across earthquake types.
- Key Takeaway:
  - Soil type (VS30) strongly influences SA, especially at short periods.
  - Distance (Rjb) reduces SA uniformly, while VS30 shapes the curve.

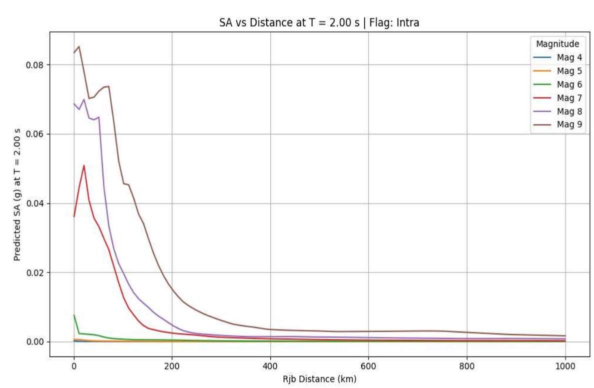
13.SA @ T (ϕ) vs Rjb(intra-event):

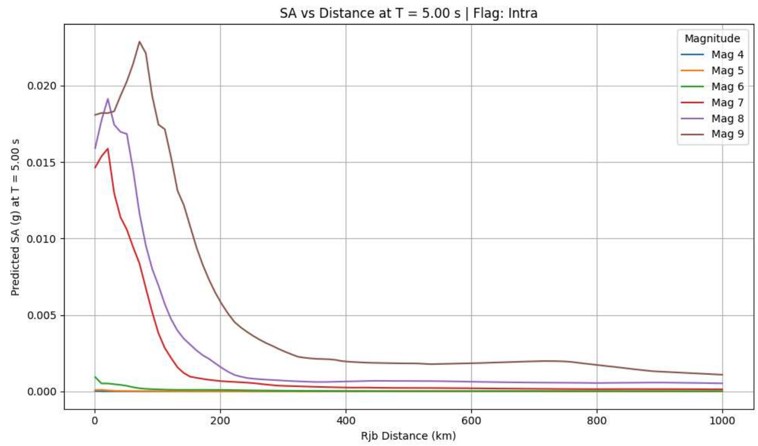
- General Trends
  - SA decreases with distance (Rjb) for all magnitudes and periods.
  - Higher magnitudes (Mag 8-9) produce higher SA at all distances compared to smaller quakes (Mag 4-5).
  - The decay rate is steeper for short periods (T=1.0s, 2.0s) than for long periods (T=5.0s).
- Period-Dependent Effects
- T = 1.0s (Short Period):
  - Highest SA values (up to 0.20g near source).

- Rapid drop-off with distance (strong attenuation).
- T = 2.0s (Intermediate Period):
  - Lower peak SA (~0.08g) but still significant near-source shaking.
  - Slower decay than T=1.0s.
- T = 5.0s (Long Period):
  - Lowest SA (~0.02g at near-source).
  - Flattest decay curve (less sensitive to distance).
- 3. Magnitude Influence
  - Near-source (Rjb < 100 km):
  - Mag 9 dominates with the highest SA.
  - Mag 4-5 have negligible impact.
  - Far-source (Rjb > 500 km):
  - Differences between magnitudes diminish, but larger quakes still lead.
- Key Takeaways
  - Short-period SA (T=1.0s) is most sensitive to distance and magnitude.
  - Long-period SA (T=5.0s) is less affected by distance but still magnitude-dependent.
  - Large quakes (Mag 7-9) control worst-case scenarios, especially near faults.
- Implications:
  - Critical for long-period structures (bridges, tall buildings).
  - Large-magnitude events dominate hazards at all distances.
  - Attenuation models must account for both magnitude and period effects.

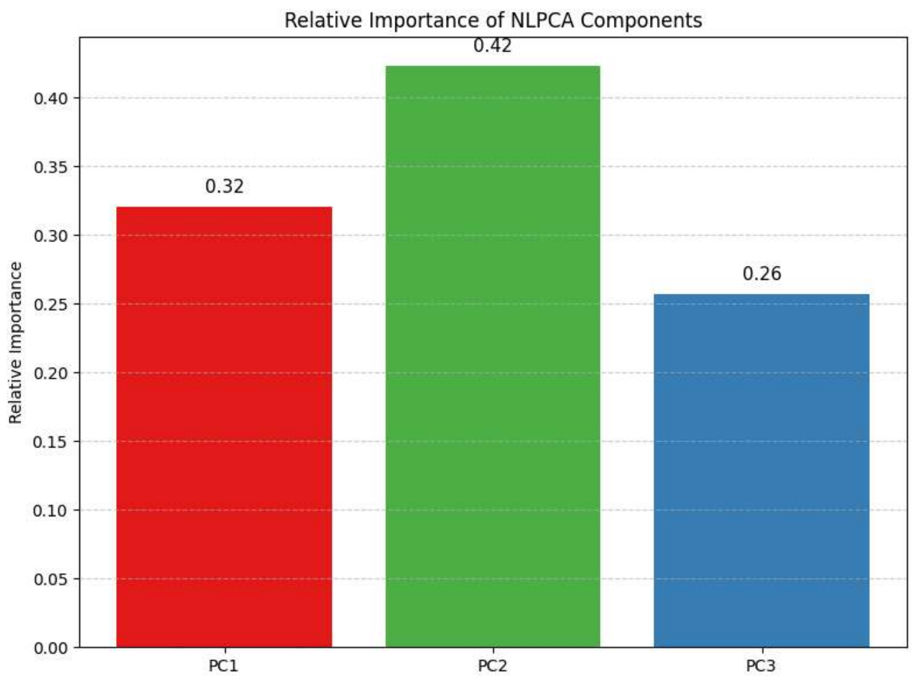
"Intra-Event" means single-earthquake variability; results would differ slightly for inter-event averages.





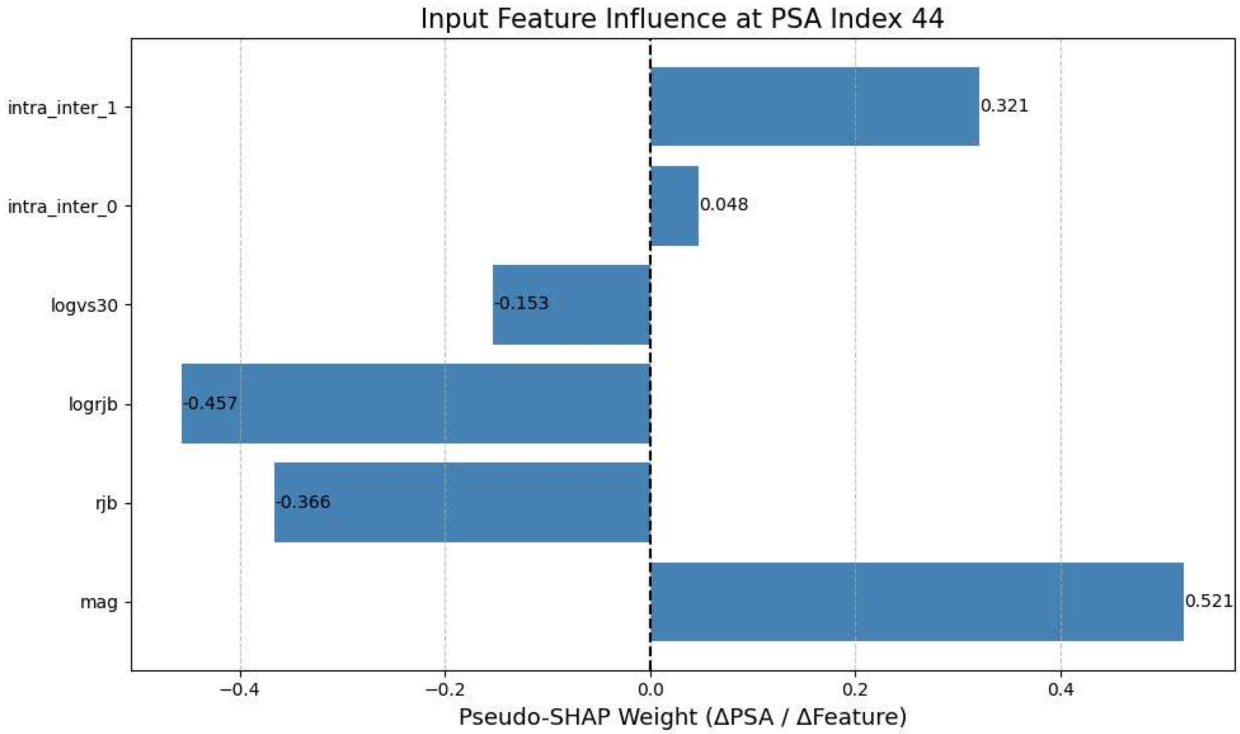


14.Relative Importance of NLPCA Components:



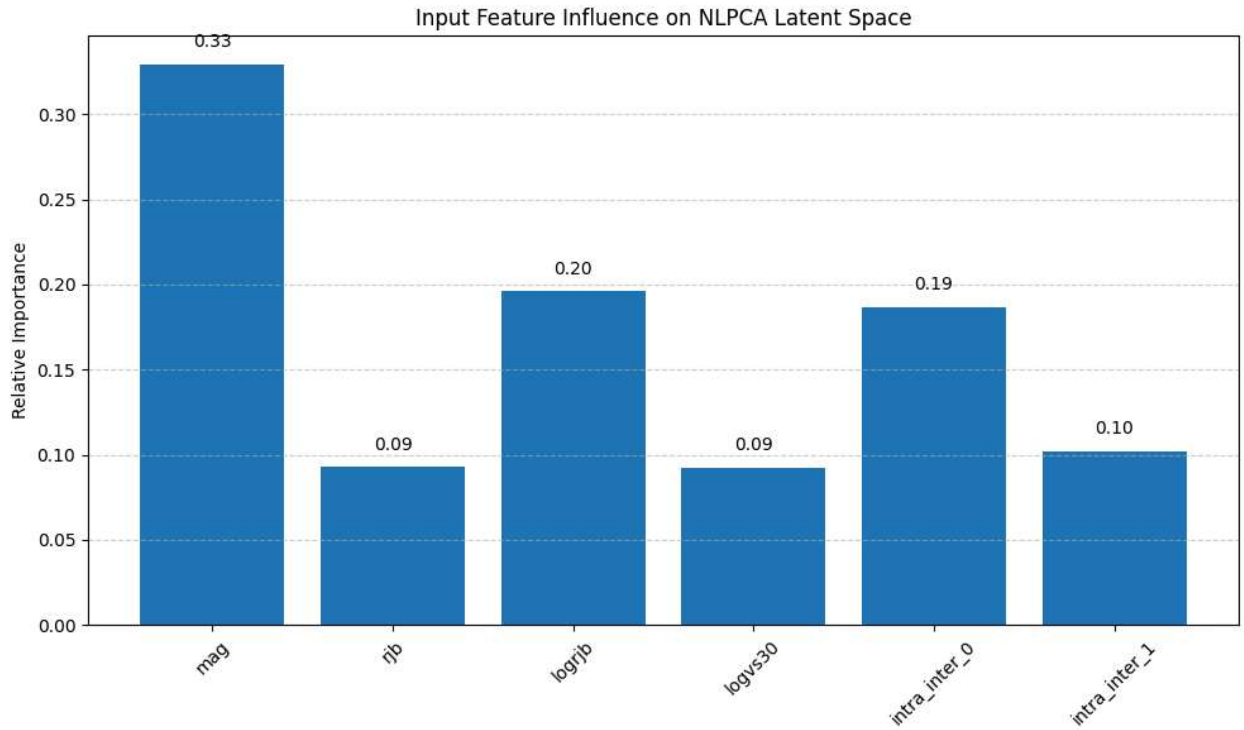
Among the three NLPCA (Nonlinear Principal Component Analysis) components, **PC2 is the most important**, contributing **42%** to the overall model variance, followed by **PC1 (32%)** and **PC3 (26%)**. This indicates that PC2 captures the most significant nonlinear features in the spectral acceleration data.

15.SHAP Analysis Summary (T)



- **Magnitude (mag)** has the **highest positive pseudo-SHAP weight (~0.52)**, indicating that increases in magnitude strongly **increase** the predicted PSA at index 44.
- **Joyner-Boore distance (rjb and logrjb)** both show **strong negative influence** (−0.366 and −0.457 respectively), meaning greater distances **reduce** predicted PSA.
- **logVs30**, a proxy for site stiffness, also has a slight **negative effect** (−0.153), suggesting that higher Vs30 (stiffer site) is associated with lower PSA.
- The **event type variable** `intra_inter_1` (representing intra-event) has a **moderate positive influence** (0.321), while the inter-event (`intra_inter_0`) has minimal impact (0.048).
- This analysis indicates that **source characteristics (magnitude)** and **distance metrics** dominate PSA predictions at this index, with **site effects and event type** playing smaller but non-negligible roles.

16.Feature Importance Summary:



- **Magnitude (mag)** shows the **highest relative importance (0.33)**, indicating it plays a **dominant role** in how the input features shape the learned nonlinear principal component (NLPCA) latent space.
- **Logarithmic distance (logrjb)** has a **moderate influence (0.20)**, suggesting that it significantly contributes to the underlying representation, likely capturing scaling with distance.
- **Intra-event classification (intra_inter_0)** also shows **notable influence (0.19)**, indicating that the event type impacts the feature transformation into latent space.
- **Linear distance (rjb)** and **site condition (logvs30)** have **low importance (0.09 each)**, suggesting they contribute little to the nonlinear representation.
- **Intra-event class (intra_inter_1)** shows a **low but non-negligible contribution (0.10)**, pointing to some distinction based on whether an event is intra- or inter-event.

PC2 is the most important NLPCA component, and earthquake magnitude is the most influential input feature in defining the latent space, followed by distance and site condition variables. This highlights the primary role of magnitude, distance, and site effects in the model's internal representation.

Code: NLPCA Model