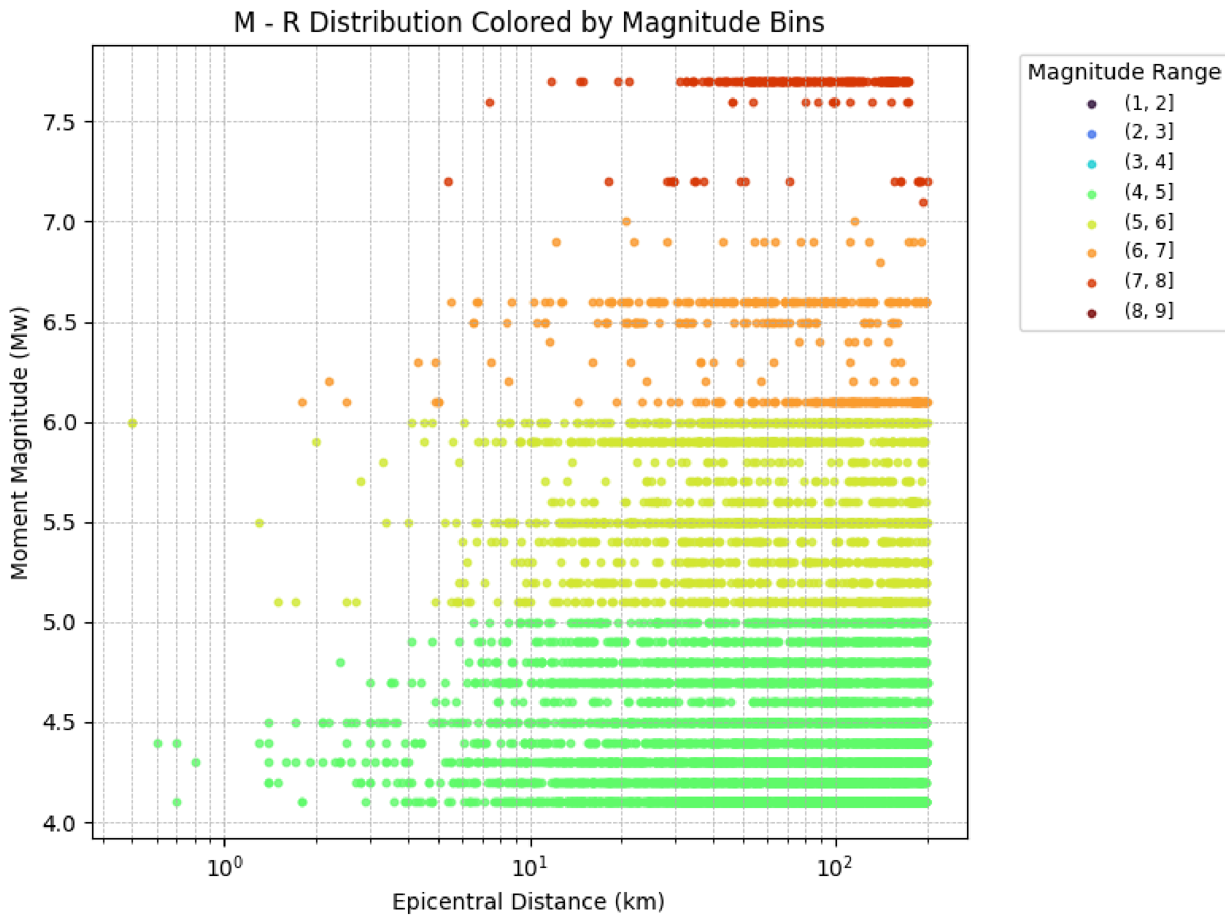
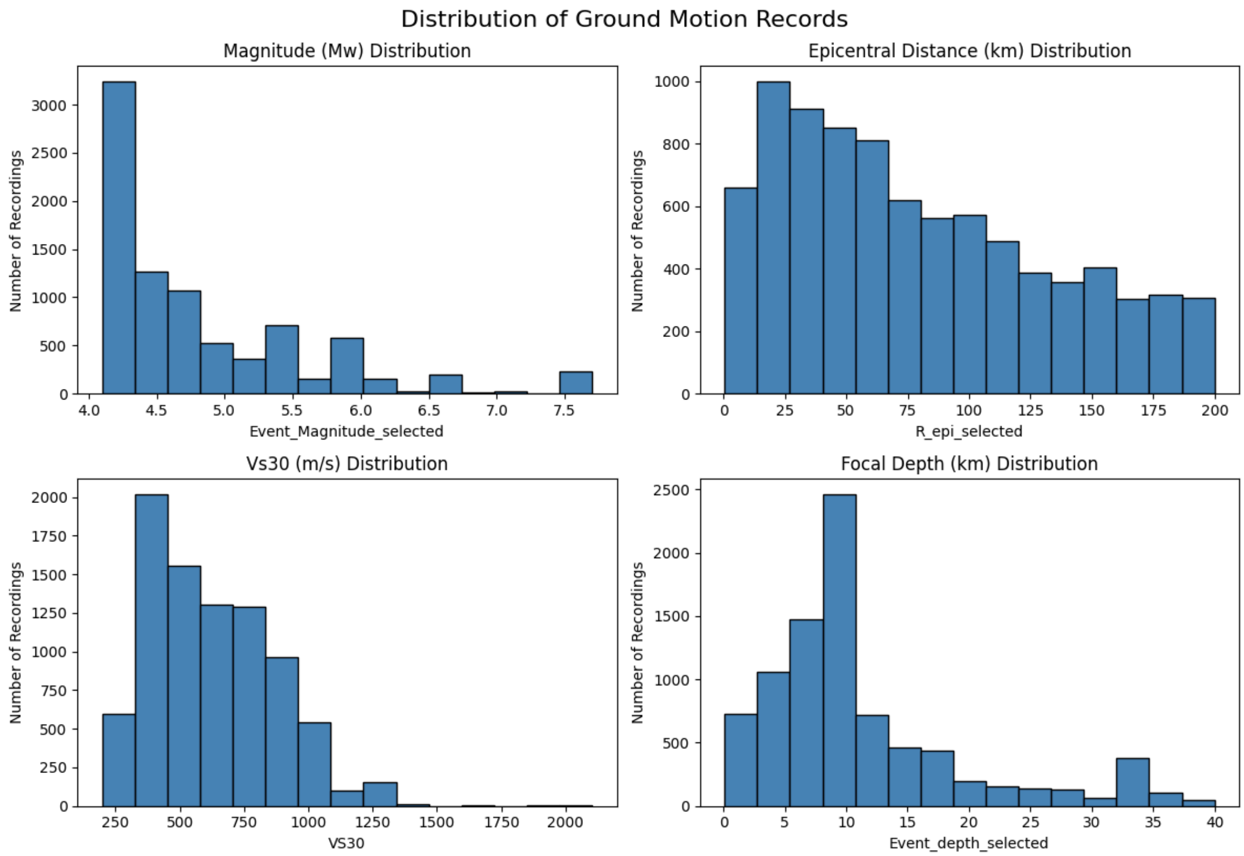
Conditional Variational Autoencoders

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	Event_depth_selected
min	4.1000	0.5000	198.0000	0.1000
max	7.7000	199.9000	2104.2000	40.0000
mean	4.8315	79.5987	623.9816	11.1482
std	0.8117	53.8051	242.8279	8.2397
skewness	1.5924	0.5351	0.6112	1.5129
kurtosis	2.4603	-0.8112	0.0107	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.

Table 2- target wise metrics

	Target Variable	R2	Inter-Std (τ)	Intra-Std (ϕ)	Total Std
0	T0pt010S	0.8820	0.6162	0.6739	0.9131
1	T0pt020S	0.8813	0.6196	0.6766	0.9175
2	T0pt030S	0.8795	0.6313	0.6834	0.9304
3	T0pt050S	0.8711	0.6628	0.7156	0.9754
4	T0pt075S	0.8581	0.6911	0.7693	1.0341
5	T0pt100S	0.8512	0.7037	0.8039	1.0683
6	T0pt150S	0.8549	0.6772	0.8027	1.0502
7	T0pt200S	0.8606	0.6598	0.7819	1.0231
8	T0pt300S	0.8741	0.6285	0.7247	0.9593
9	T0pt400S	0.8809	0.6216	0.6893	0.9282
10	T0pt500S	0.8827	0.6106	0.6730	0.9087
11	T0pt750S	0.8770	0.5920	0.6689	0.8933
12	T1pt000S	0.8715	0.5809	0.6777	0.8926
13	T1pt500S	0.8625	0.5433	0.6949	0.8821
14	T2pt000S	0.8616	0.5196	0.6959	0.8685
15	T2pt500S	0.8659	0.4981	0.6870	0.8486
16	T3pt000S	0.8709	0.4885	0.6774	0.8351
17	T3pt500S	0.8753	0.4883	0.6679	0.8274
18	T4pt000S	0.8805	0.4785	0.6526	0.8092
19	T5pt000S	0.8917	0.4642	0.6214	0.7756

Model Performance (R^2):

The R^2 values for all target variables are consistently high, ranging from 0.8512 to 0.8917. This indicates that the models account for a substantial portion of the variance in each target, reflecting strong predictive performance across all periods.

Uncertainty Measures:

- **Inter-Event Standard Deviation (τ):**
  τ values range from approximately 0.464 to 0.793, representing variability between different events. Lower values at longer periods

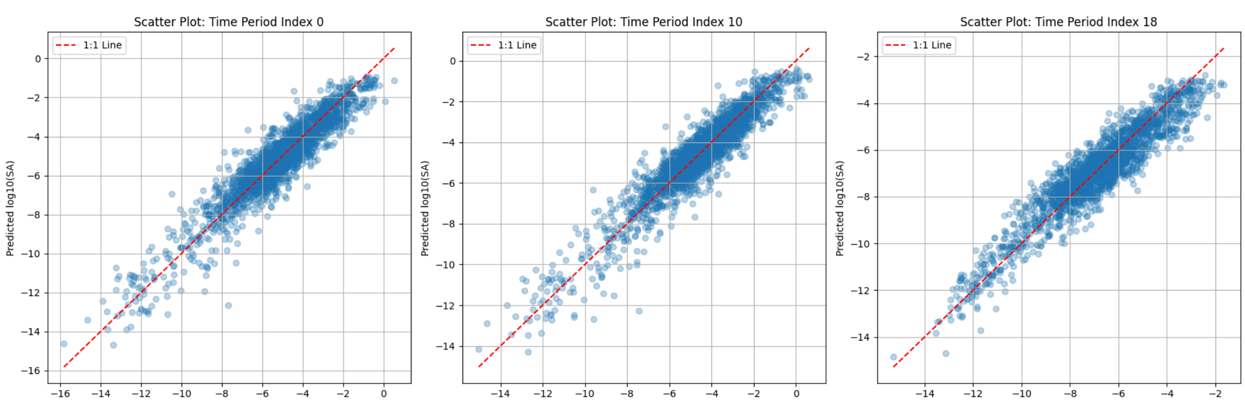
(e.g., T5pt000S) suggest reduced inter-event variability for those targets.

- **Intra-Event Standard Deviation (ϕ):**
  ϕ values fall between roughly 0.621 and 0.803, capturing variability within individual events. Similar to τ, intra-event variability tends to be lower at longer periods.

- **Total Standard Deviation:**
  This metric combines both inter- and intra-event variability. It is highest for shorter periods (e.g., T0pt100S, T0pt150S) and progressively decreases at longer periods (e.g., T5pt000S).

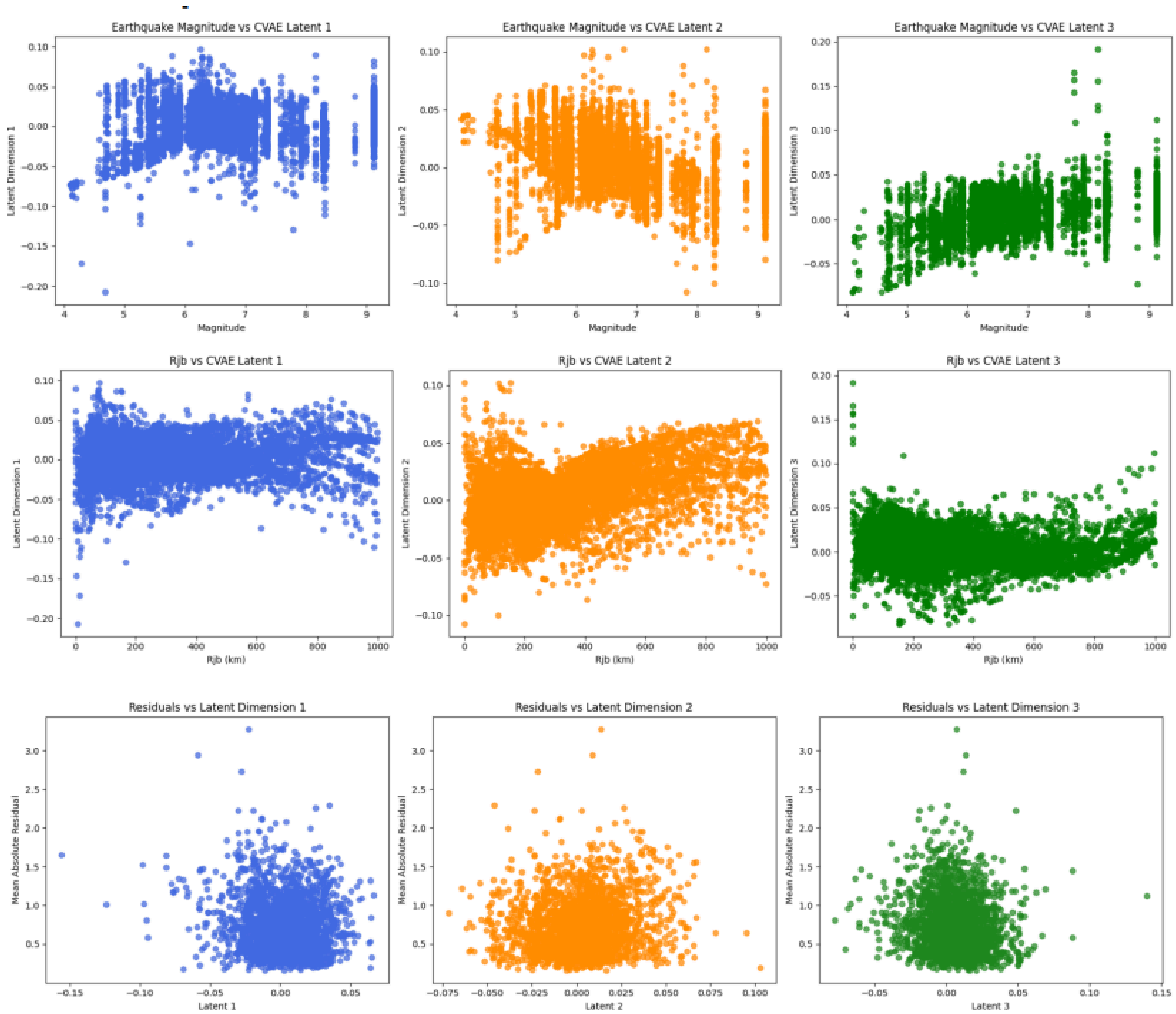
Trend Insights:
Both inter-event and intra-event standard deviations decline as the period increases, indicating greater stability and reduced variability in predictions at longer periods. Conversely, higher total standard deviations at shorter periods reflect increased uncertainty in those predictions.

4. Correlation plots:



We can see from the above image that the values are highly dense around the centerline.

5. Latent plots:



Inference from the Plots:

1. **Earthquake Magnitude vs. CVAE Latent Dimensions**

    ○ Latent Dimensions 1 and 2 exhibit a wide distribution across the entire range of magnitudes without any noticeable trend, implying these dimensions capture features unrelated to earthquake magnitude.

    - Latent Dimension 3 shows a modest upward trend as magnitude increases, suggesting it may be more responsive to changes in earthquake size.

2. **Rjb (Distance) vs. CVAE Latent Dimensions**

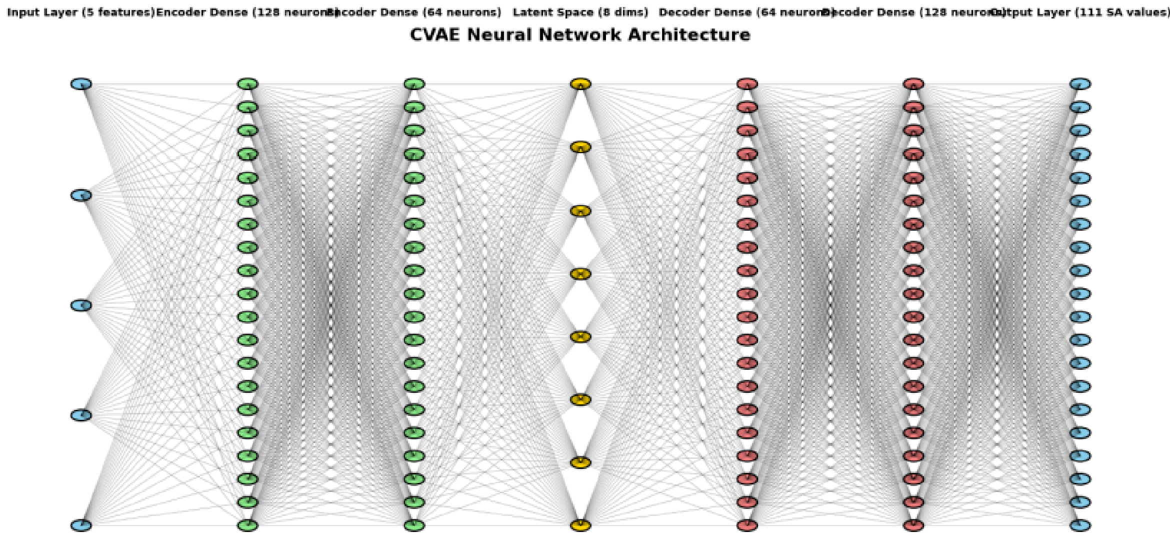
    - All three latent dimensions are broadly spread across different Rjb values, lacking any clear linear relationship. This indicates that distance (Rjb) is not directly or dominantly represented in the latent space.

3. **Residuals vs. Latent Dimensions**

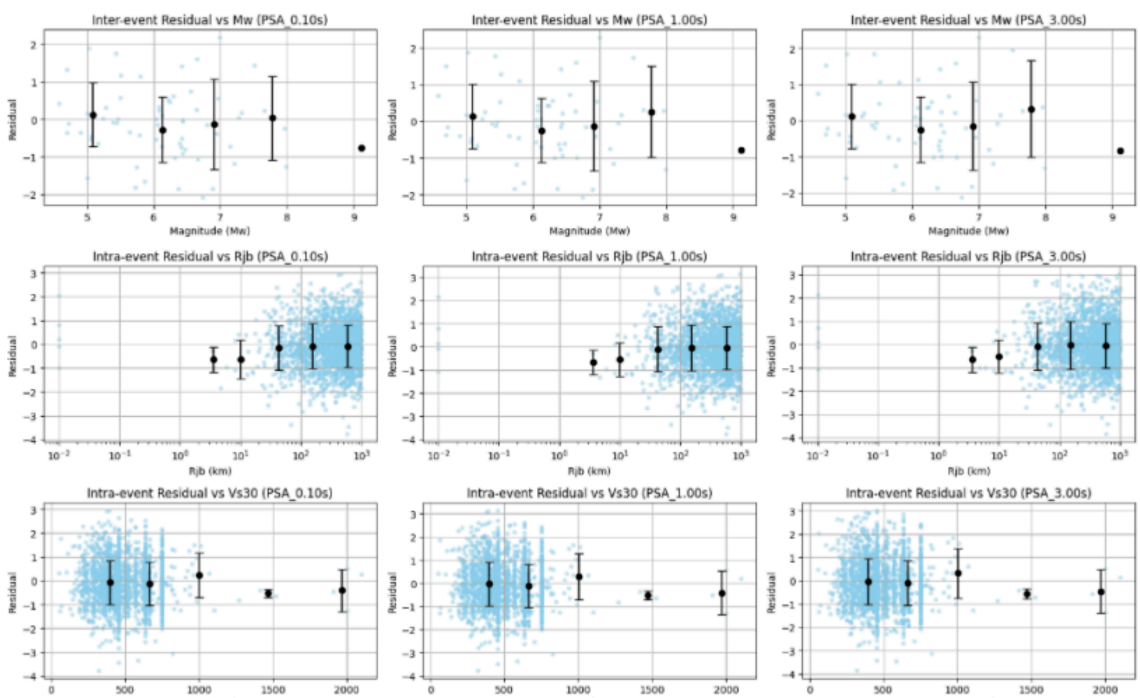
    - The residuals appear evenly scattered around zero for all latent dimensions, with no discernible patterns or correlations. This suggests that the latent space does not reflect prediction error in a straightforward way.

Summary:
The CVAE latent space captures a rich, multi-dimensional representation of the earthquake data. While Latent Dimension 3 may have some connection to magnitude, the remaining dimensions encode other complex patterns not directly tied to magnitude or distance. The absence of strong correlations with residuals further indicates that the latent dimensions represent diverse and disentangled features, rather than simply encoding model performance.

6. Network architecture



7. Residual Plots



Inference from the Residual Plots:

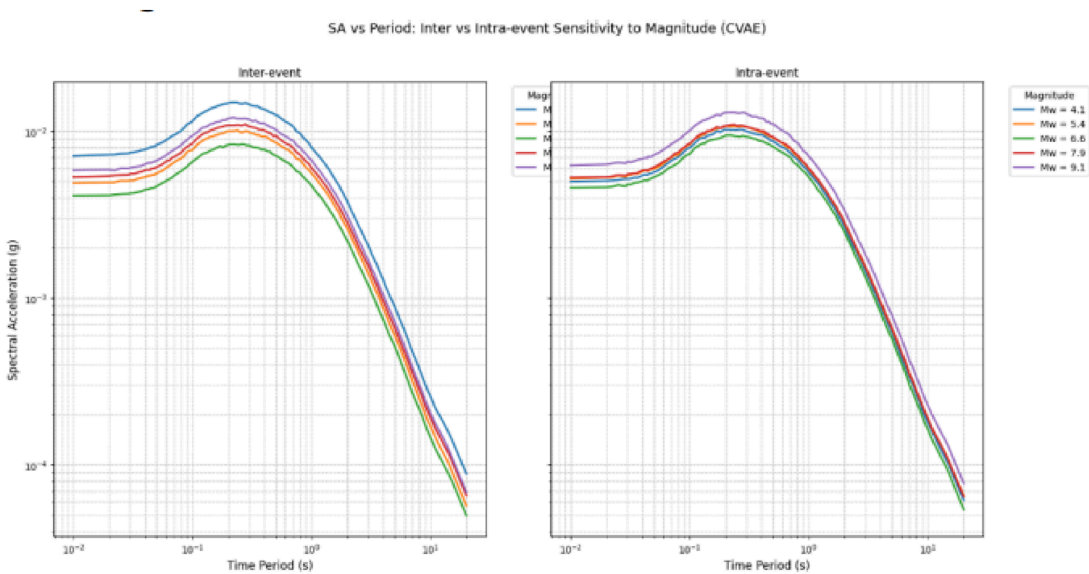
1. **Inter-event Residuals vs. Magnitude (Mw):**
   - Across all three spectral periods (0.1s, 1.0s, 3.0s), there is no clear trend between inter-event residuals and magnitude.
   - The average residuals within each magnitude bin are close to zero, and overlapping error bars (standard deviations) indicate no significant magnitude-related bias.
2. **Intra-event Residuals vs. Rjb (Distance):**
   - Intra-event residuals are consistently centered around zero across the full range of Rjb values for all periods.
   - There is no observable trend or shift in the mean residuals, suggesting that the model's performance is consistent and unbiased with respect to distance.
3. **Intra-event Residuals vs. Vs30 (Site Condition):**
   - Residuals remain close to zero across all Vs30 values and periods, showing no apparent pattern or bias.
   - This suggests the model effectively accounts for site condition variations without systematic over- or under-prediction.

Summary:

Residual analyses indicate that the model's predictions are generally unbiased across key parameters—magnitude, distance (Rjb), and site condition (Vs30)—for all examined spectral periods. The absence of systematic trends or significant biases in the residuals supports the robustness and reliability of the model's predictive performance.

8. Ground Motion Physics

a. Spectral Acceleration vs Time Period at different magnitudes



Inference from the Plots:

1. **Magnitude Sensitivity:**

    - Both inter-event and intra-event spectral acceleration (SA) plots display distinct separation between different magnitudes. Larger earthquakes (Mw 9.1, 7.9) consistently yield higher SA values across all spectral periods compared to smaller events (Mw 4.1, 5.4, 6.6), indicating strong magnitude-dependent behavior.

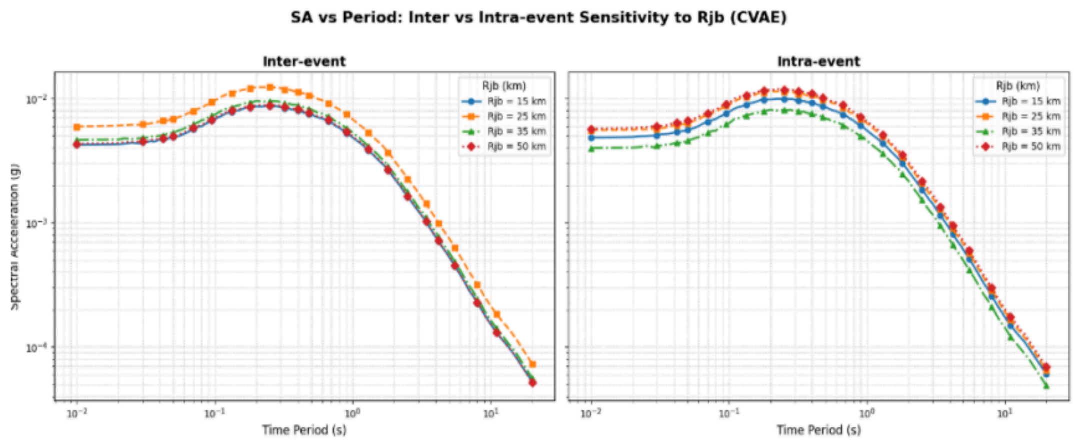
2. **Period Dependence:**

    - For all magnitudes, SA values peak in the 0.2–0.5 second range and decline at longer periods. This pattern aligns with established ground motion characteristics, where short periods typically exhibit higher accelerations.

3. **Inter-event vs. Intra-event Variability:**
    - Magnitude separation is more distinct in the inter-event plot, suggesting that differences between events are strongly influenced by magnitude. In contrast, intra-event curves are more tightly clustered, indicating that within-event variability is less sensitive to magnitude changes.

Summary:
The CVAE model effectively captures spectral acceleration behavior across varying earthquake magnitudes and periods. It demonstrates strong sensitivity to magnitude, particularly in inter-event variability, and reflects realistic period-dependent trends. This indicates the model's capacity to represent both broad magnitude scaling and the nuanced differences between and within seismic events.

b. Spectral Acceleration vs Time Period at different Joyner-Boore distance



Inference from the Plots:

1. **Distance Effect:**
    - Spectral acceleration (SA) consistently decreases with increasing Rjb (distance from the fault) in both inter-event and intra-event plots. SA curves for greater distances (35 km, 50 km) are lower than those for shorter distances (15 km, 25 km) across all spectral periods, reflecting expected attenuation behavior.

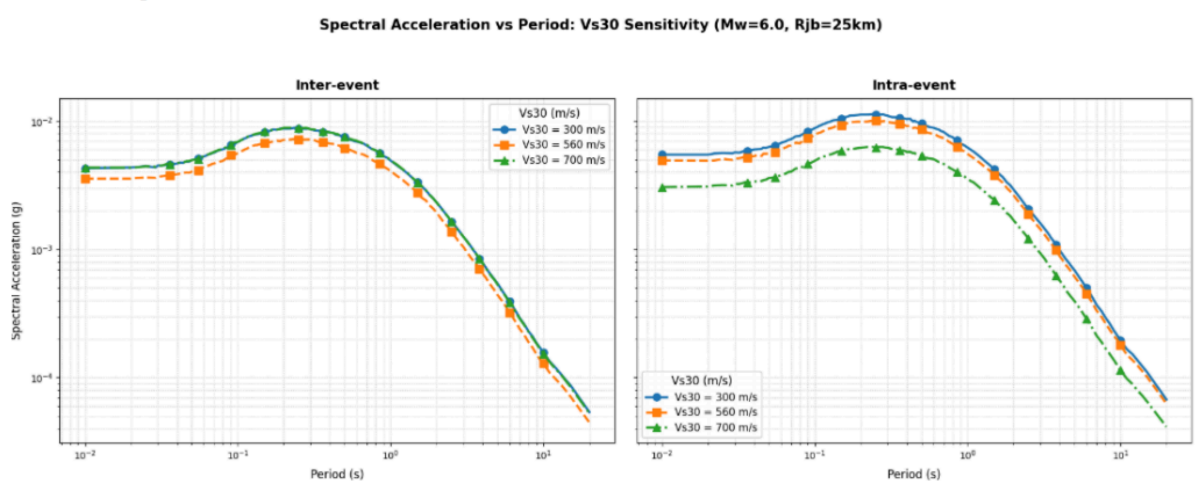
2. **Period Dependence:**
    - Across all distances, SA peaks at short periods (approximately 0.2–0.5 seconds) and diminishes at longer periods. This trend aligns with typical ground motion characteristics, where shorter periods generally show stronger shaking.

3. **Inter-event vs. Intra-event Variability:**
    - The distance-based separation is slightly more distinct in the intra-event plot, suggesting that SA variation within individual events is more sensitive to changes in distance than variation

between events.

Summary:
The CVAE model successfully captures key physical trends in ground motion: spectral acceleration decreases with distance from the fault, and this pattern is consistently represented in both inter-event and intra-event variability across all periods. The model reflects realistic attenuation behavior and distinguishes well between within-event and between-event distance effects.

c. Spectral Acceleration vs Time Period at different Average Shear-Wave Velocities of the ground within the top 30 meters



Inference from the Plots:

1. **Vs30 Sensitivity (Site Condition):**

- Both inter-event and intra-event plots show a clear decrease in spectral acceleration (SA) with increasing Vs30. Sites with lower Vs30 values (e.g., 300 m/s – softer soils) consistently exhibit higher SA across all periods, while sites with higher Vs30 (e.g., 700 m/s – stiffer soils) show reduced SA. This behavior aligns with established site response theory, where softer soils amplify ground motion more.

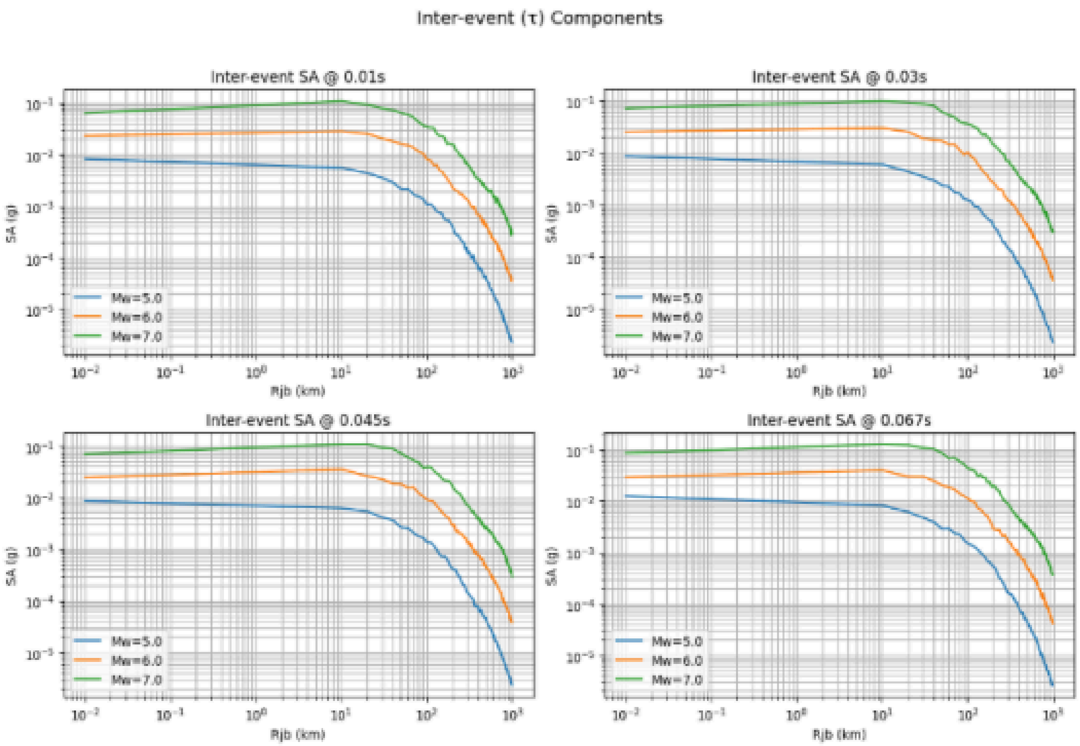
2. **Period Dependence:**
    - The influence of Vs30 on SA is most prominent at short to intermediate periods (0.1–1.0 seconds), where site amplification effects are typically strongest. At longer periods, the SA curves for different Vs30 values converge, indicating diminished site response impact.

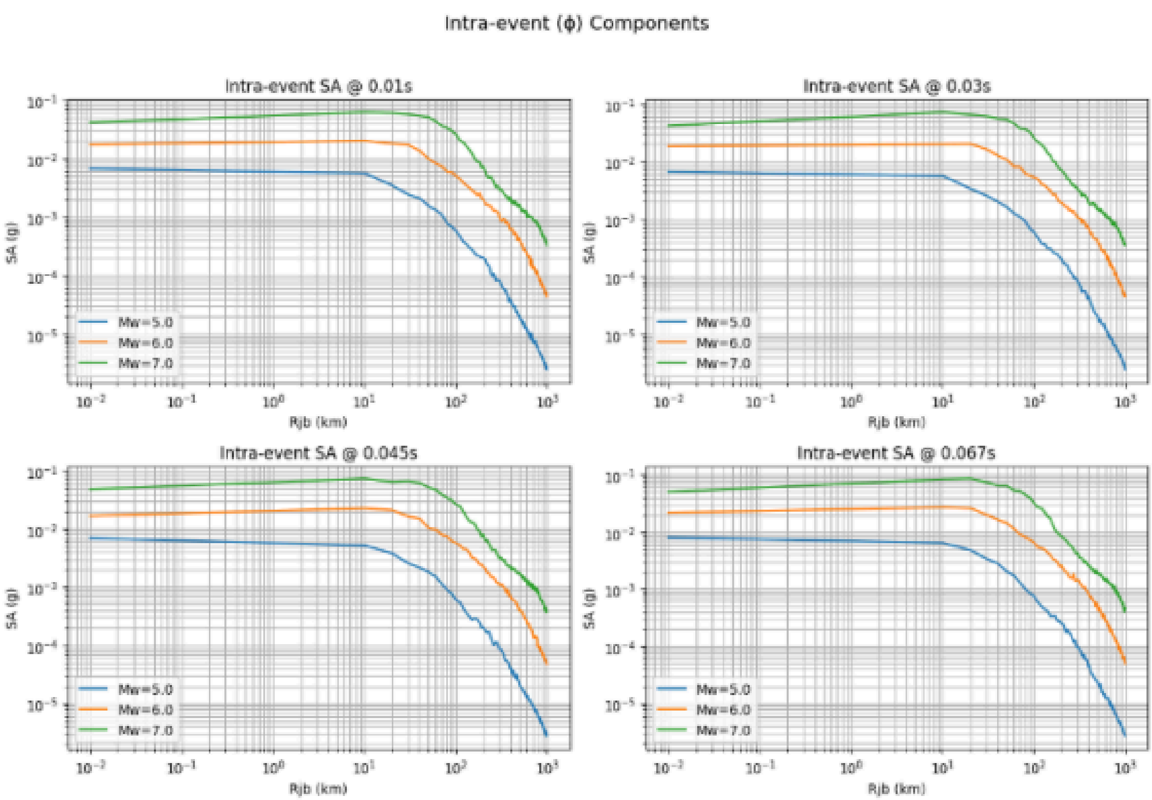
3. **Inter-event vs. Intra-event Variability:**
    - The intra-event plots show slightly more pronounced separation between Vs30 curves compared to inter-event plots. This suggests that within a single event, variability in ground motion is more sensitive to site conditions than variability across different events.

Summary:

The CVAE model effectively reproduces the expected relationship between site conditions and ground motion: softer soils (lower Vs30) lead to higher spectral accelerations, particularly at short to intermediate periods. This site amplification effect is visible in both inter- and intra-event variability, with a slightly stronger influence observed within events.

9. Ground Motion Physics wrt rjb





Inference from the Plots :

1. **Magnitude Effect:**
   - Spectral acceleration (SA) increases with earthquake magnitude for both inter-event (τ) and intra-event (ϕ) components. Higher-magnitude events (e.g., Mw 7.0) consistently result in greater SA across all short periods compared to lower magnitudes (Mw 5.0, 6.0), reflecting expected magnitude scaling behavior.

2. **Distance Attenuation:**
   - SA decreases with increasing Rjb (fault distance) across all plots, magnitudes, and periods. This attenuation trend is more pronounced at greater distances, consistent with typical ground motion decay.

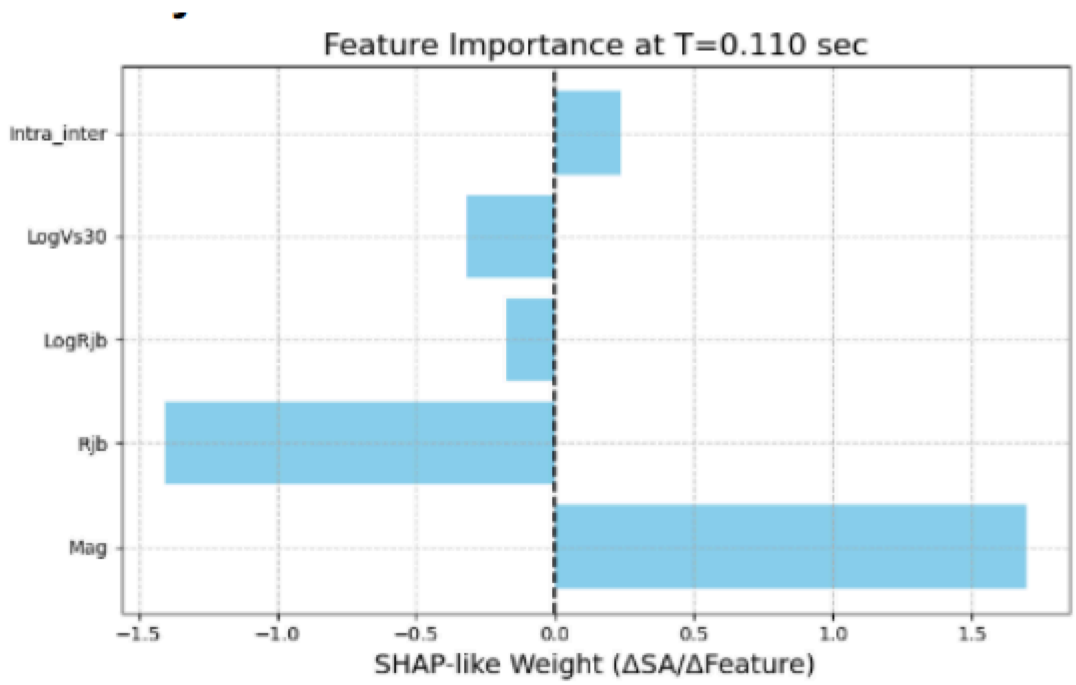
3. **Period Dependence:**
   - The separation between curves for different magnitudes is maintained across all short periods examined (0.01s, 0.03s, 0.045s, 0.067s), indicating that the model reliably captures magnitude scaling and distance attenuation over a range of short-period ground motions.

4. **Inter-event vs. Intra-event Variability:**
   - Both inter-event and intra-event plots show similar trends. However, the intra-event curves tend to converge more at shorter distances, suggesting slightly reduced variability within events in close proximity to the source.

Summary:

The CVAE model accurately reflects key physical ground motion behaviors: spectral acceleration increases with earthquake magnitude and decreases with distance from the source. These trends are consistently represented across both inter-event and intra-event components and are robust across short spectral periods.

10. SHAP Analysis



Inference from the Feature Importance Plot :

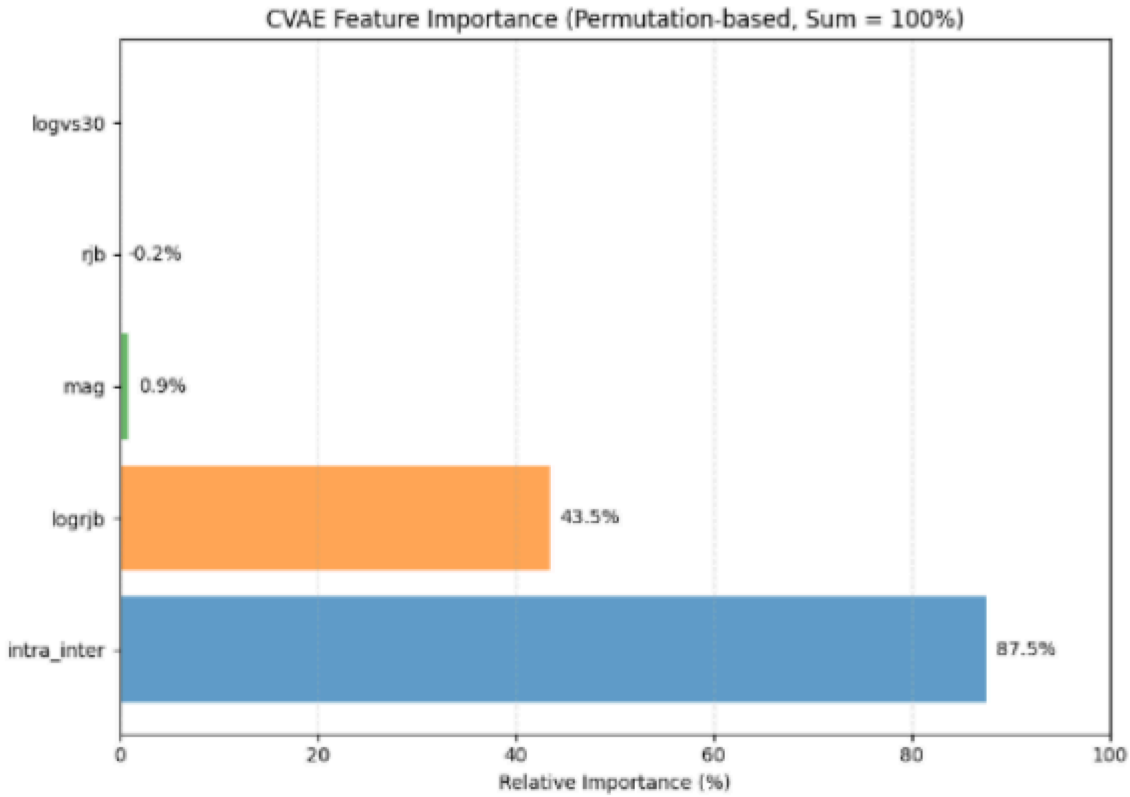
1. **Dominant Features:**
   - At a spectral period of **T = 0.110 sec**, **earthquake magnitude (Mag)** and **distance from the fault (Rjb)** emerge as the most influential predictors of spectral acceleration (SA). Magnitude has a strong positive effect, meaning larger earthquakes lead to higher SA, while Rjb has a strong negative effect, reflecting the

expected attenuation of ground motion with distance.

2. **Secondary Features:**
   - Features such as log-transformed distance (LogRjb), log-transformed site condition (LogVs30), and the intra/inter-event indicator (Intra_inter) exhibit much lower SHAP-like values. This indicates they have relatively minor influence on SA predictions at this short period.
3. **Interpretation:**
   - At **T = 0.110 sec**, the model's SA predictions are primarily driven by **earthquake magnitude and distance**, consistent with physical ground motion behavior. Site conditions and event-type distinctions contribute less significantly at this spectral period.

Summary:
The feature importance analysis shows that for short-period ground motion (T = 0.110 sec), **magnitude and distance** are the primary drivers of spectral acceleration in the CVAE model, while **site effects and event type** have a comparatively limited impact.

11. Relative importance plot



Inference from the Feature Importance Plot:

1. **Most Important Features:**
   - The **intra_inter** indicator contributes **87.5%**, and **log-transformed distance (logrjb)** contributes **43.5%** to the CVAE model's predictions at this spectral period. These are the dominant factors influencing the model's output.

2. **Minor Contributions:**
   - **Earthquake magnitude (mag)** contributes only **0.9%**, **raw distance (rjb)** just **0.2%**, and **log-transformed site condition (logvs30)** has a **negligible** effect on the predictions.

3. **Interpretation:**

    - At this particular spectral period, the model places the most emphasis on whether the observation is **intra-event or inter-event**, followed by the **logarithmic distance to the fault**. In contrast, **magnitude**, **raw distance**, and **site condition** contribute minimally to the output, indicating a shift in feature relevance depending on the period or context.

Summary:
For this spectral period, the CVAE model's predictions are primarily driven by **event classification (intra vs. inter)** and **log-distance**, while traditional seismic predictors like **magnitude** and **site condition** have limited influence. This highlights the model’s context-dependent weighting of input features.