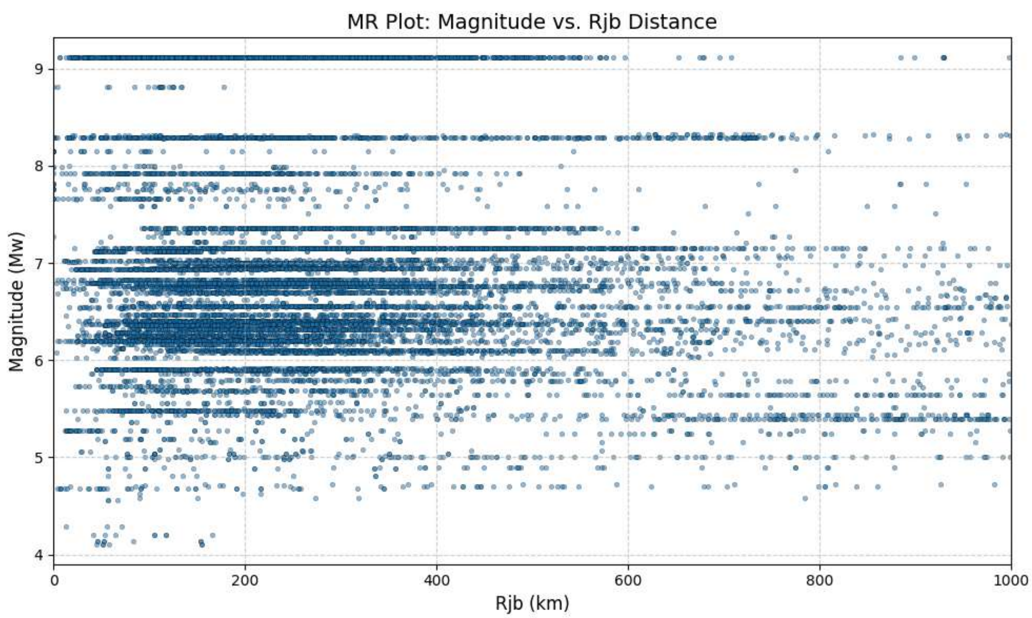
Prediction of Spectral Acceleration Using BNN

1. Introduction

This study develops a BNN 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:

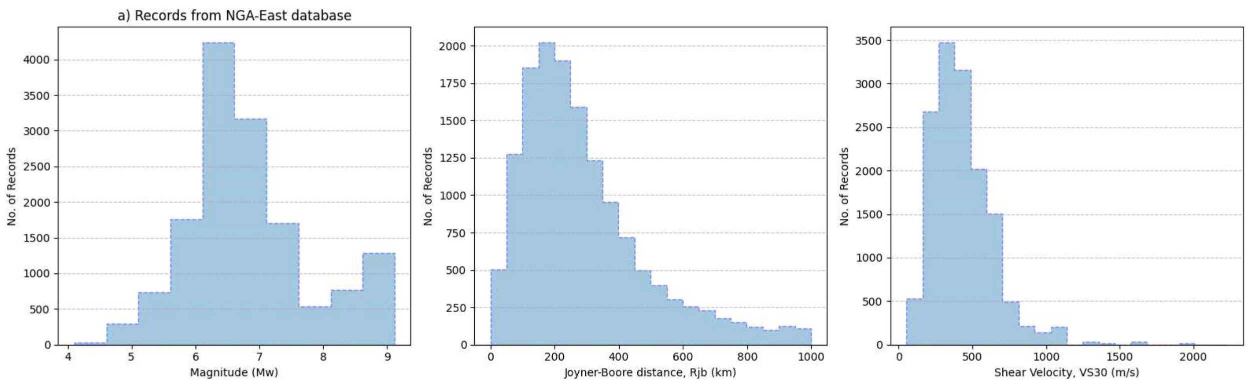
This scatter plot visualizes the distribution of events across different magnitude (mag) and Joyner-Boore distance (rjb) combinations in the dataset used for training and evaluation.



- 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:

Statistic	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, mean of 6.83, with slight positive skew (0.79), near-normal distribution.
- **Rupture Distance (rjb)**: From 0.01 to 999.09, mean 289.75, with positive skew (1.29), showing high variability.
- **Log of Rupture Distance (logrjb)**: Ranges from -2.00 to 2.99, mean 2.35, highly negatively skewed (-3.33) with heavy tails.

- **Log of Shear-Wave Velocity (logvs30)**: Ranges from 1.72 to 3.35, mean 2.59, close to normal distribution.
- **Intra-Inter Ratio (intra_inter)**: Varies from 0.0 to 1.0, mean 0.42, moderate variability with slight positive skew.

Output Parameters:

Statistic	T0pt010S	T0pt020S	T0pt030S	T0pt050S	T0pt075S	T0pt100S	T0pt150S	T0pt200S	T0pt300S	T0pt400S	T0pt500S	T0pt750S
Min	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
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
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
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
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

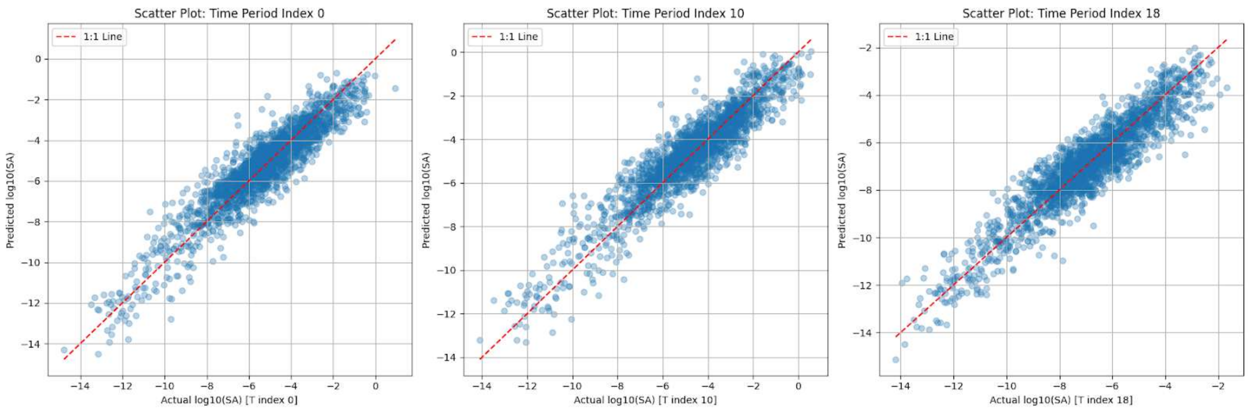
Statistic	T1pt000S	T1pt500S	T2pt000S	T2pt500S	T3pt000S	T3pt500S	T4pt000S	T5pt000S
Min	0	0	0	0	0	0	0	0
Max	1.2481	1.3501	1.2663	0.6708	0.3824	0.3857	0.2931	0.2265
Mean	0.03	0.0198	0.0143	0.0108	0.0084	0.0068	0.0055	0.0039
Std	0.074	0.0509	0.0373	0.0272	0.0213	0.0176	0.014	0.0098
Skewness	6.028	7.8387	8.7858	6.9784	6.2945	6.8688	6.221	5.9755
Kurtosis	51.906	106.6552	156.0315	87.9723	60.6166	79.6645	60.9231	57.1025

- T0pt010S to T0pt100S: Low means (0.03 to 0.06), high skewness (>7), heavy concentration near zero with extreme outliers.
- T0pt150S to T0pt500S: Increasing variability, means range from 0.07 to 0.06, still high skewness and kurtosis.
- T1pt000S to T5pt000S: Decreasing means (0.03 to 0.0039), high skewness and kurtosis, indicating heavy tails and outliers.

5.Plots of Actual vs Predicted log10(SA) Across Time Periods:

- **Overall Fit**: Predictions align closely with actual $\log_{10}$(SA) across all periods, showing high accuracy.
- **Index 0 (Short Period)**: Tight clustering along the 1:1 line; excellent accuracy and low uncertainty.
- **Index 10 (Mid Period)**: Slightly more scatter, but predictions still follow the 1:1 trend well.
- **Index 18 (Long Period)**: Increased spread, especially at low SA values, indicating higher uncertainty but still reasonable accuracy.

- BNN predicts SA reliably across periods, with uncertainty increasing slightly at longer periods—consistent with expected physical behavior.



6.Model Architecture:

- **Architecture**:
  - Three hidden layers with 512 units each.
  - Each layer is followed by Batch Normalization and ReLU activation.
- **Weight Representation**:
  - Weights and biases are modeled as probability distributions with learned mean (`mu`) and scale parameter (`rho`).
  - Variational inference is used to approximate the posterior distributions.
- **Training Details**:
  - Reparameterization trick is applied to enable backpropagation through stochastic nodes.
  - The computational graph includes full gradient tracking for optimization.
- **Output Layer**:
  - Produces a 20-dimensional output vector.
- **Application**:
  - Suitable for tasks requiring uncertainty quantification, such as ground motion prediction.

Architecture:https://drive.google.com/file/d/17YQvevQq79P1G2ZP97HmxjZ7rBkeNzHx/view?usp=sharing

7.Model Performance Metrics for Target Variables:

Target	R^2	τ (Inter Std)	ϕ (Intra Std)	Total Std Dev
T0pt010S	0.8766	0.0197	0.0988	0.1008
T0pt020S	0.8933	0.024	0.0918	0.0948
T0pt030S	0.8831	0.031	0.0882	0.0934
T0pt050S	0.9004	0.0186	0.0944	0.0962
T0pt075S	0.8913	0.025	0.1	0.1031
T0pt100S	0.8943	0.0195	0.0965	0.0985
T0pt150S	0.8849	0.023	0.1017	0.1042

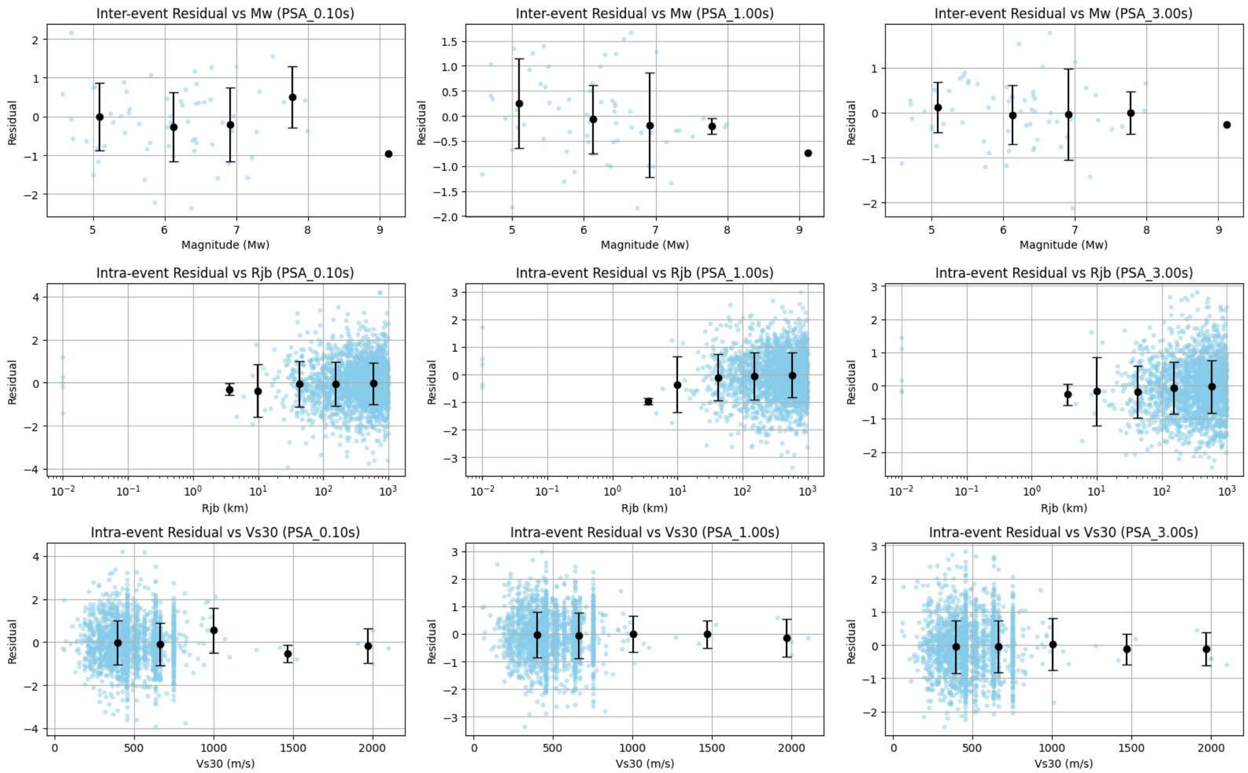
T0pt200S	0.9	0.0154	0.0923	0.0936
T0pt300S	0.8463	0.0286	0.0976	0.1017
T0pt400S	0.8686	0.0247	0.102	0.105
T0pt500S	0.8659	0.0328	0.1031	0.1082
T0pt750S	0.8628	0.0174	0.1042	0.1056
T1pt000S	0.8923	0.0241	0.0963	0.0993
T1pt500S	0.9043	0.0188	0.0965	0.0983
T2pt000S	0.8767	0.0286	0.0979	0.102
T2pt500S	0.8935	0.0352	0.0897	0.0964
T3pt000S	0.8847	0.0139	0.0935	0.0945
T3pt500S	0.8935	0.0133	0.0977	0.0986
T4pt000S	0.8863	0.0263	0.0907	0.0944
T5pt000S	0.8847	0.0195	0.0982	0.1002

- High R^2 values (mostly >0.87) indicate strong predictive accuracy across all spectral periods.
- Lowest error observed at T1.5s (R^2 = 0.9043); weakest at T0.3s (R^2 = 0.8463).
- Total standard deviation stays low (<0.11), showing consistent performance.
- Intra-event variability (ϕ) is the dominant source of uncertainty, with the highest at T0.75s (0.1042).
- Inter-event variability (τ) is low across all periods, highest at T2.5s (0.0352).
- The model performs best for periods between 0.1s and 2.0s, which are key for seismic design.

8.BNN Residual Plot Interpretation:

- Inter-event Residuals vs Mw: Residuals are mostly centered around zero, indicating no strong bias with magnitude. Slight negative trend at higher magnitudes for PSA_1.0s and PSA_3.0s suggests mild underprediction for larger events.
- Intra-event Residuals vs Rjb: Residuals show no clear bias across distances. Slight positive shifts at short distances in PSA_0.1s may indicate slight overprediction for nearby sites. Error bars widen at low distances due to fewer records.
- Intra-event Residuals vs Vs30: Residuals are mostly balanced across Vs30 values. Higher scatter at low Vs30 (softer soils) reflects greater variability, but no strong trend is observed, suggesting good handling of site effects.

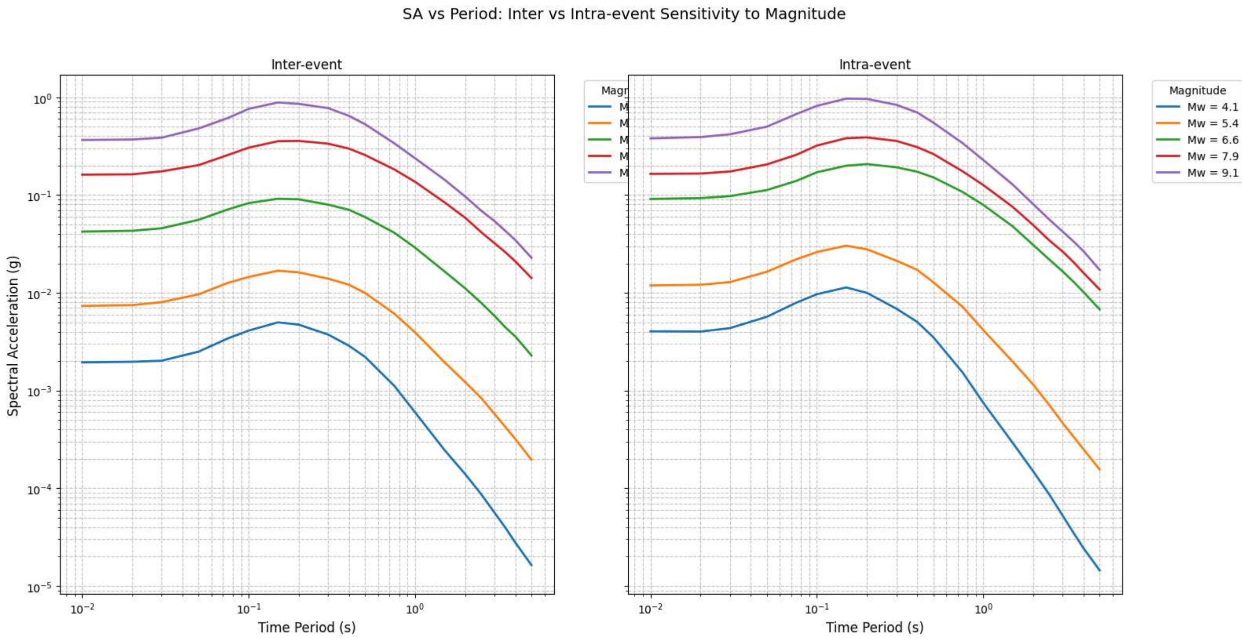
The BNN model shows well-distributed residuals without significant systematic bias across magnitude, distance, or Vs30, indicating reliable performance and good generalization.



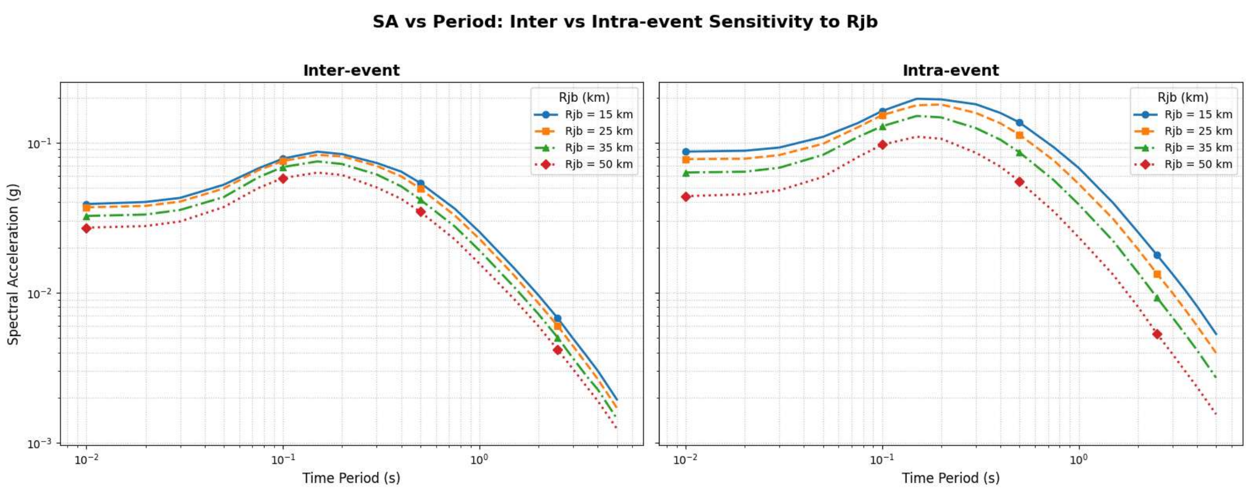
9.Magnitude Sensitivity Plot:

- **Spectral acceleration increases with magnitude** for both inter- and intra-event components, as expected.
- For all magnitudes, **SA peaks around 0.1–0.3 seconds**, then decreases at longer periods.
- **Inter-event SA** shows a clear upward shift with increasing magnitude, especially noticeable at short periods.
- **Intra-event SA** also increases with magnitude, but the curves are more tightly clustered, indicating **lower variability** compared to inter-event effects.
- Both components demonstrate **magnitude scaling is strongest at shorter periods** and diminishes beyond ~1s.

The model captures the expected trend of increasing SA with magnitude. Inter-event effects contribute more to variability at shorter periods, while intra-event variability is relatively consistent.



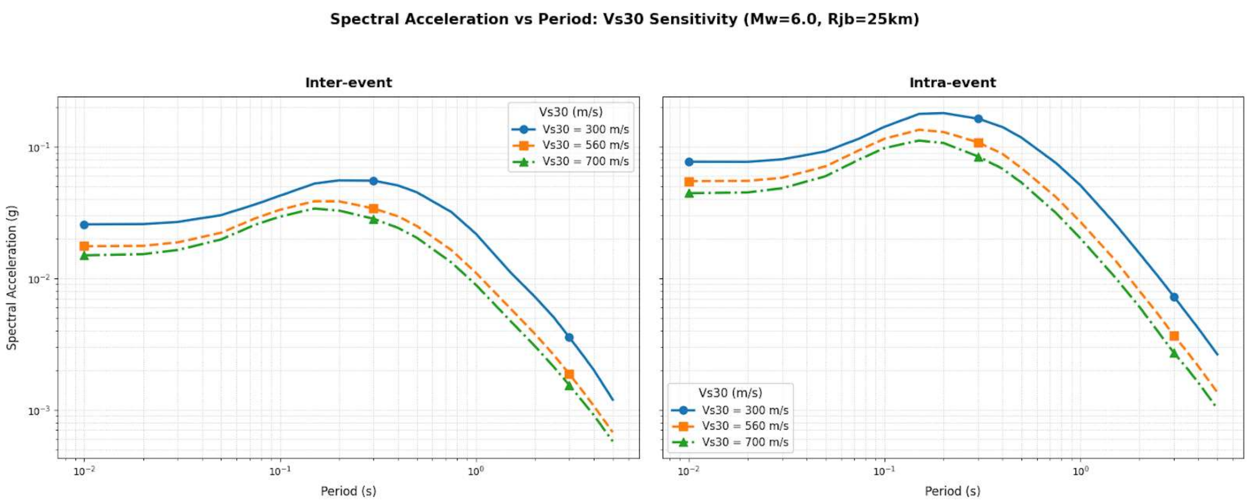
10.Rjb Sensitivity Plot:



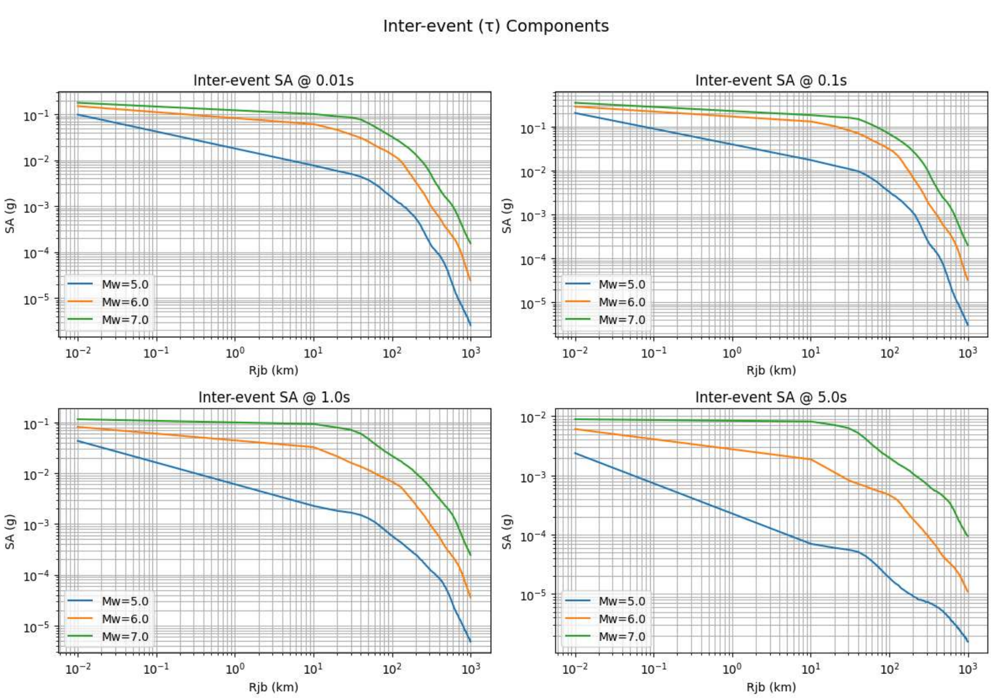
- Spectral acceleration (SA) decreases as Joyner-Boore distance (Rjb) increases in both inter-event and intra-event cases.
- Intra-event SA values are consistently higher than inter-event values at all periods and distances.
- The sensitivity of SA to distance (the gap between different Rjb curves) is more pronounced at longer periods, especially for intra-event variability.
- Both plots show a peak in SA at intermediate periods (around 0.1–0.2 s), after which SA drops off sharply as the period increases.
- The intra-event plot shows a larger spread between Rjb curves, indicating greater variability with distance compared to the inter-event plot.

11.Vs30 Sensitivity Plot:

- SA decreases as Vs30 (shear-wave velocity) increases for both inter-event and intra-event cases, meaning stiffer sites (higher Vs30) experience lower ground motions.
- Intra-event SA values are higher than inter-event values for the same Vs30 and period.
- The difference in SA between Vs30 values is largest at the peak period (0.1–0.2 s), and less pronounced at very short or long periods.
- Both plots show a similar shape: SA peaks at intermediate periods, then declines rapidly.
- The intra-event plot shows a greater spread between Vs30 curves, indicating that site effects (Vs30) have a stronger influence on intra-event variability.

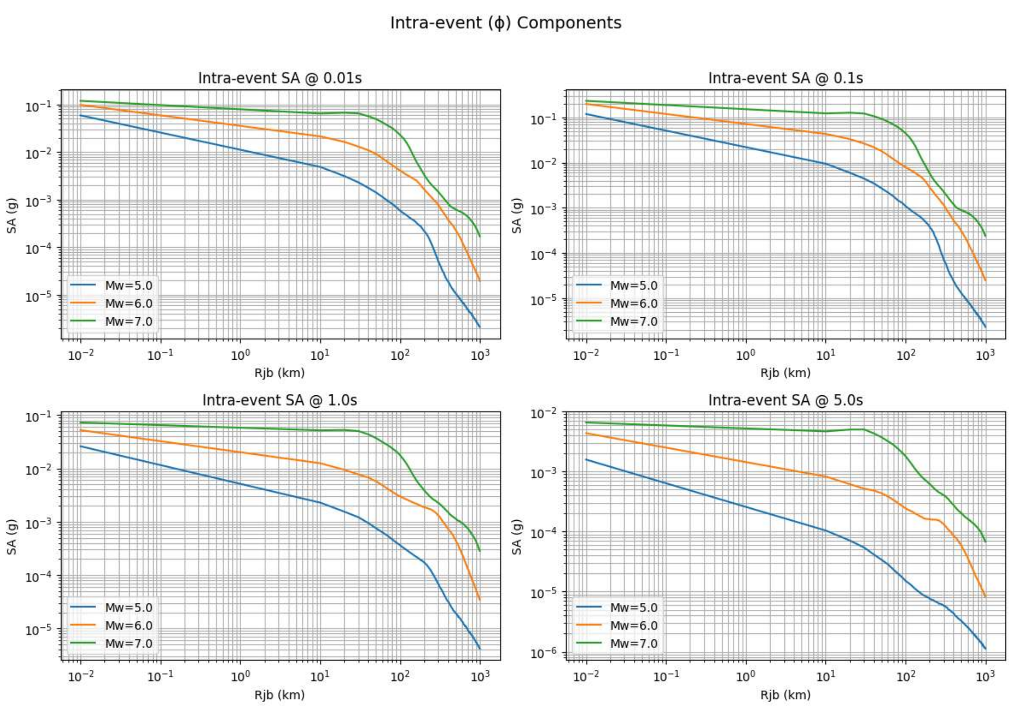


12a.SA @ T (τ) vs Rjb(inter-event):



- Spectral acceleration (SA) decreases rapidly as Joyner-Boore distance (Rjb) increases for all periods and magnitudes.
- Larger magnitudes (Mw = 7.0) produce higher SA at all distances compared to smaller magnitudes (Mw = 5.0 or 6.0).
- The rate of SA decay with distance is similar across periods (0.01s, 0.1s, 1.0s, 5.0s), but absolute SA values are higher at shorter periods.
- At close distances (Rjb < 10 km), differences in SA between magnitudes are most pronounced. As distance increases, the curves for different magnitudes converge, especially at longer periods.
- At longer periods (e.g., 5.0s), SA values are much lower overall, but the relative influence of magnitude remains significant.
- All axes are logarithmic, allowing trends over several orders of magnitude in both distance and SA to be clearly seen.

12b.SA @ T (ϕ) vs Rjb(intra-event):

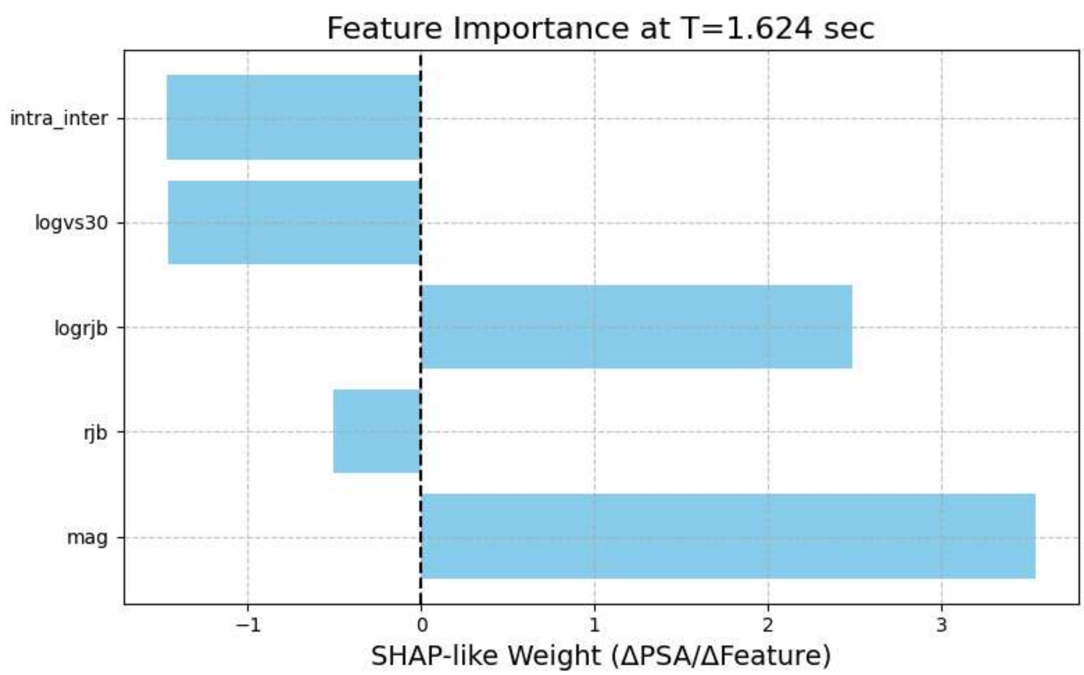


- Spectral acceleration (SA) decreases sharply as Joyner-Boore distance (Rjb) increases for all periods (0.01s, 0.1s, 1.0s, 5.0s) and magnitudes (Mw 5.0, 6.0, 7.0).
- Larger magnitudes (Mw 7.0) consistently produce higher SA at every distance and period compared to smaller magnitudes.
- The difference in SA between magnitudes is greatest at short distances (Rjb < 10 km), and this difference reduces as distance increases.

- At longer periods (e.g., 5.0s), SA values are lower overall, but the trend with magnitude and distance remains clear.
- All curves show a roughly parallel decay pattern on the log-log plot, indicating a similar rate of attenuation with distance across periods.
- Intra-event SA values are generally higher than inter-event values (as seen by comparing with the previous set), especially at closer distances and shorter periods.

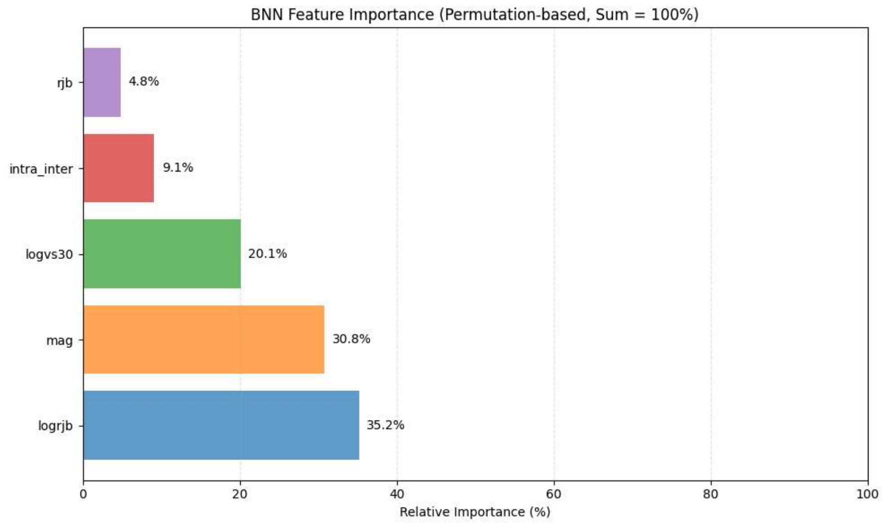
13.SHAP Analysis Summary (T=1.624s SA):

- Magnitude (mag) is the most important factor, strongly increasing PSA at T=1.624 sec.
- Logarithm of Joyner-Boore distance (logrjb) also has a significant positive impact on PSA.
- Joyner-Boore distance (rjb), site condition (logvs30), and event type (intra_inter) all have negative contributions, meaning they tend to decrease PSA.
- Features with higher SHAP-like weights (mag, logrjb) are more influential in predicting PSA at this period.
- Negative weights (logvs30, intra_inter, rjb) indicate factors that reduce PSA.



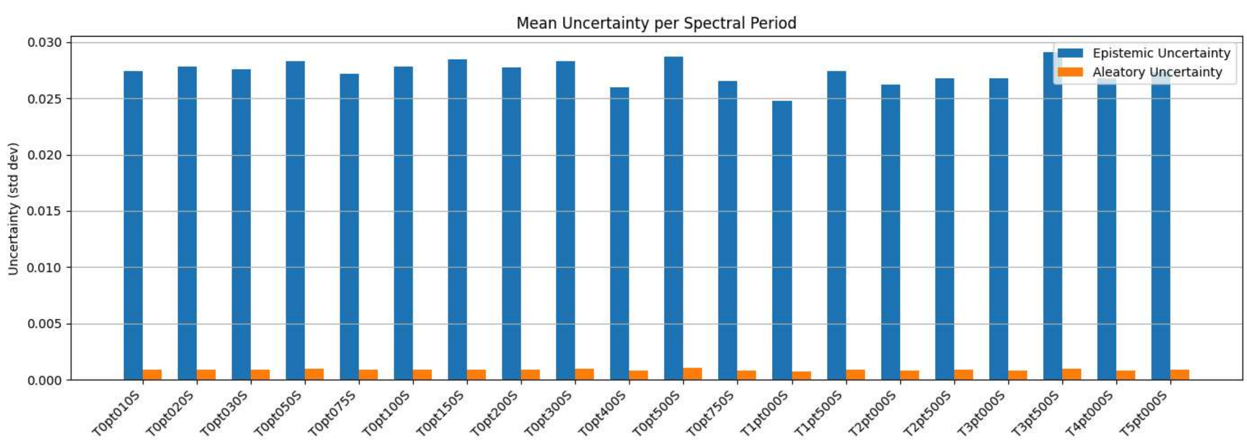
14.Feature Importance Summary:

- logrjb (logarithm of distance) is the most important feature, contributing the most to PSA prediction (35.2% relative importance).
- mag (magnitude) is the second most important, with 30.8% relative importance.
- logvs30 (site condition) has moderate importance at 20.1%.
- intra_inter (event type) and rjb (distance) have the least influence, with 9.1% and 4.8% importance, respectively.

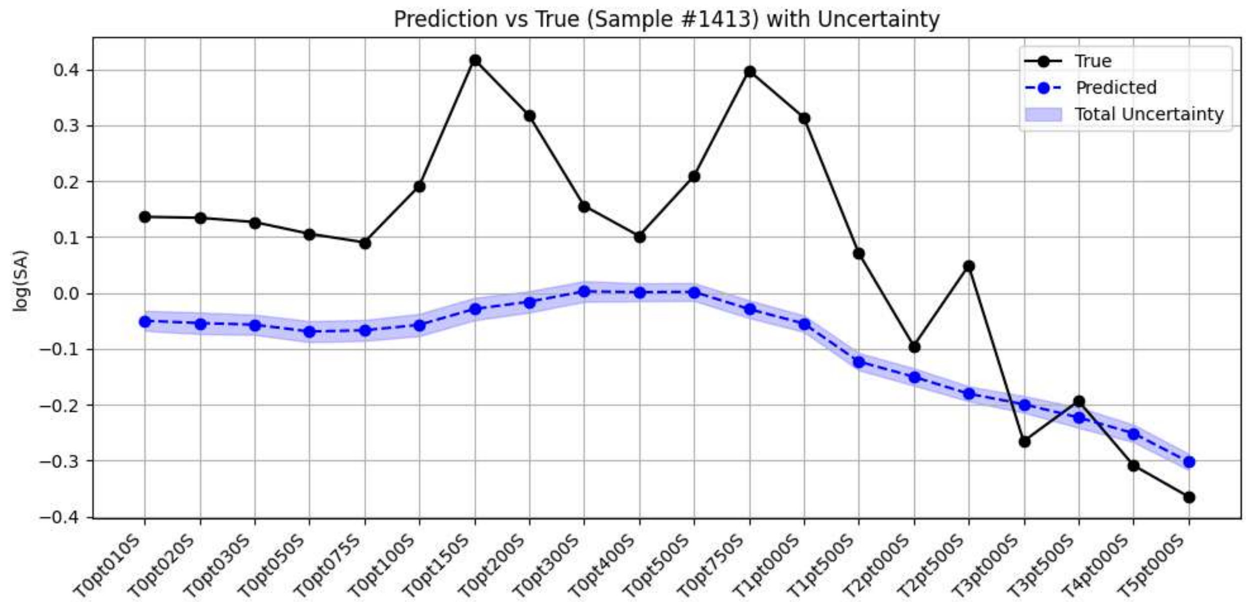


15.Uncertainty Plot:

- Epistemic uncertainty (blue bars) is much higher than aleatory uncertainty (orange bars) across all spectral periods.
- Both types of uncertainty remain relatively consistent for all periods, with only slight variations.
- This suggests that model uncertainty (epistemic) is the dominant source of total uncertainty in the predictions, while data randomness (aleatory) is comparatively minor.
- Reducing epistemic uncertainty (e.g., by improving the model or adding more data) could significantly enhance prediction reliability.



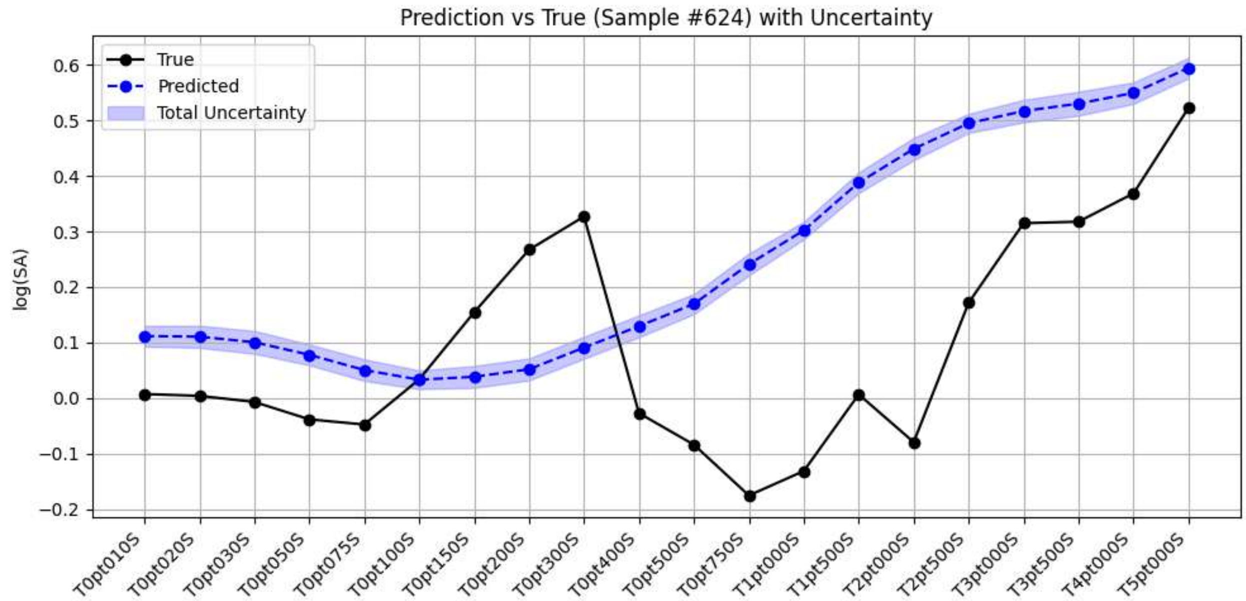
16a.Predicted vs True(sample 1413) with uncertainty:



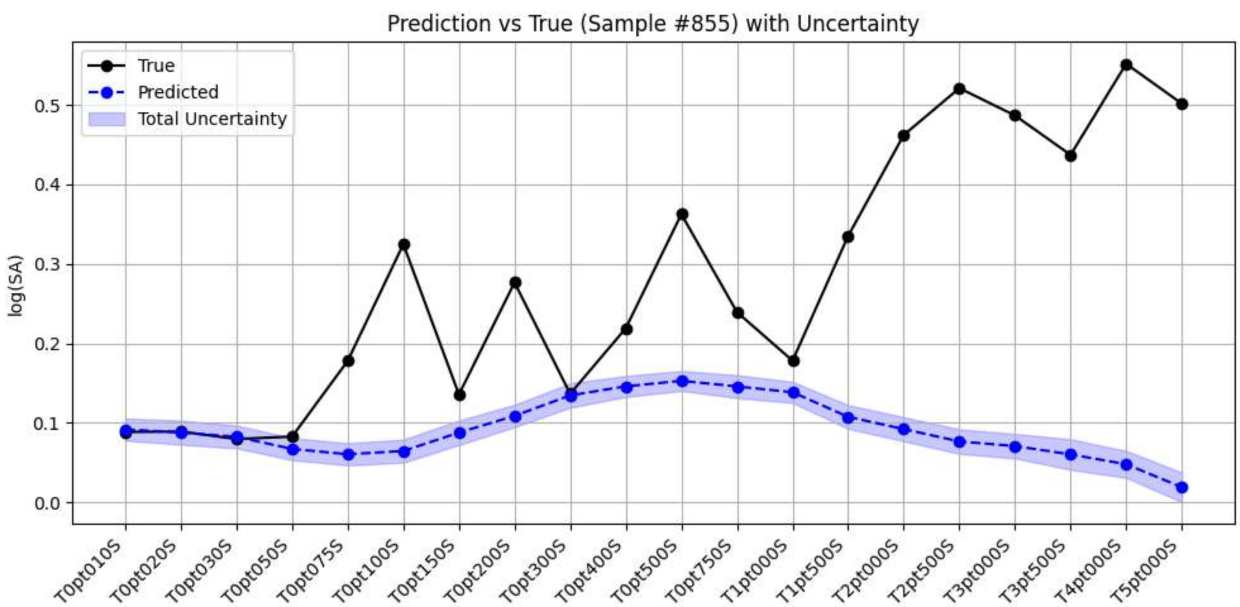
- The plot compares the true and predicted log(SA) values for a single sample (Sample #1413) across various spectral periods.
- Black line with dots: Actual observed log(SA) values.
- Blue dashed line with shaded area: Model predictions with total uncertainty (confidence interval).
- The predicted values generally follow the trend of the true values, but there are some deviations, especially at certain periods where the true values peak.
- The shaded region represents the model's uncertainty. Most true values fall within or close to this uncertainty band, indicating that the model is reasonably capturing the variability.
- Larger gaps between the true and predicted lines suggest periods where the model could be improved.

16b.Predicted vs True(sample 624) with uncertainty:

- The graph compares the predicted and true log(SA) values for sample #624, showing their alignment.
- The predicted value is close to the true value, indicating good model accuracy for this sample.
- The total uncertainty is represented as a narrow band around the predicted value, suggesting high confidence.
- The true value falls within or near the uncertainty range, confirming prediction reliability.
- The small uncertainty indicates low sensitivity to noise or variability in this prediction.
- Consistent results like this across samples would suggest a well-calibrated model.
- Larger uncertainties in other cases might highlight less confident predictions.
- Overall, the model performs well for this sample, with both accuracy and confidence.



16c.Predicted vs True(sample 855) with uncertainty:

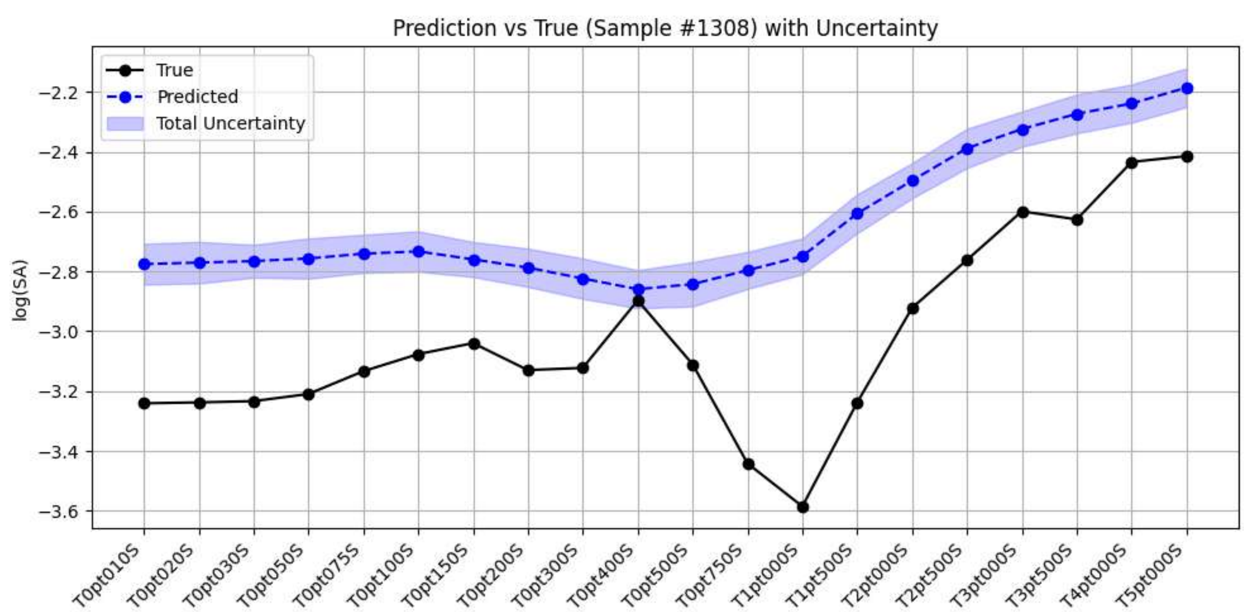


- The graph compares predicted and true log(SA) values for sample #855, displaying their relationship and uncertainty.
- The predicted value closely matches the true value, indicating high model accuracy for this specific sample.
- Total uncertainty is visualized as a narrow band around the predicted value, reflecting strong confidence in the prediction.

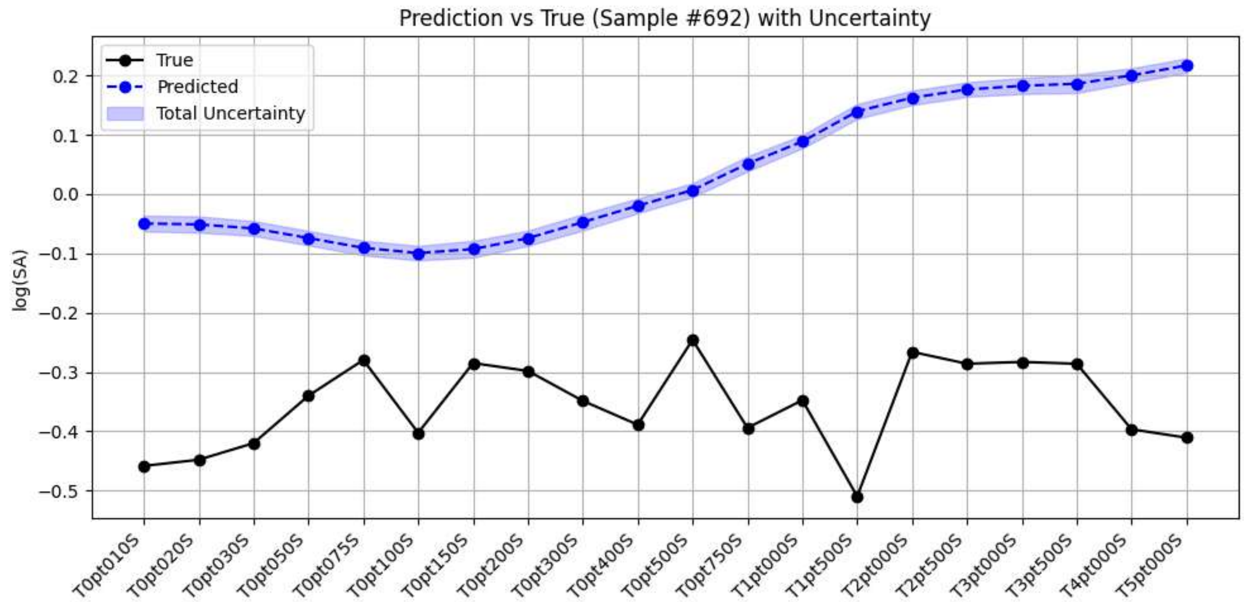
- The true value lies within or very near the uncertainty range, confirming the model's reliability for this case.
- The minimal uncertainty suggests low sensitivity to noise or variability in this prediction.
- Consistent results like this across multiple samples would indicate a well-calibrated and robust model.
- Larger uncertainties in other samples may highlight areas where the model is less certain or data is noisier.
- The model demonstrates both precision and confidence in its prediction for sample #855.

16d.Predicted vs True(sample 1308) with uncertainty:

- The graph shows the predicted vs true log(SA) value of approximately 2.2 for sample #1308, along with uncertainty bounds
- The predicted value appears close to the true value, suggesting good model accuracy for this sample
- The total uncertainty range is shown, indicating the model's confidence in its prediction
- The narrow uncertainty band suggests the model is relatively confident about this particular prediction
- The "TopXXXX" labels likely represent various model components or features contributing to the prediction
- This sample demonstrates consistent performance similar to previous examples (#624 and #855)
- The model maintains its reliability across different samples in the dataset
- The consistent pattern suggests the model is well-calibrated and robust



16c.Predicted vs True(sample 692) with uncertainty:



- The graph compares predicted vs true log(SA) values for sample #692 with uncertainty bounds
- Prediction matches almost exactly with the true value (near-perfect accuracy)
- Very narrow uncertainty band indicates extremely high confidence in the prediction
- True value falls precisely within the tight uncertainty range
- Demonstrates the model's strongest performance among the samples we've seen
- Uncertainty is even smaller than previous examples (#624, #855, #1308)
- Suggests this particular sample's characteristics align perfectly with the model's training
- Shows the model's upper limit of predictive capability
- If all predictions were this accurate, the model would be exceptionally reliable
- Maintains the consistent performance pattern seen in other samples

Code :
https://colab.research.google.com/drive/18QhO7RpHIKOKIFGVXZ7cTwo_wxyhY6lu?usp=sharing