

Author Comment (AC)
Response to Referee #2 (All comments)

Manuscript: “A ground motion prediction model for the Italian region based on a mixture of experts framework”

Referee #2 – Comment 1 (Validation strategy and event-wise independence)

The referee expresses concern that the current training/validation framework is based on random record-wise data partitioning, which may allow recordings from the same earthquake event to appear across different subsets. Because such records are not fully independent, this strategy may not provide a sufficiently rigorous assessment of the model’s ability to generalize to unseen earthquake events. The referee therefore suggests adding an event-based partition experiment, analogous to the final independent-event testing stage, in order to strengthen the evaluation of event-level generalization.

Author response

We sincerely thank the reviewer for this insightful comment. We agree that, because records from the same earthquake event may be statistically correlated, a record-based partition in the training/validation stage may overestimate the model’s ability to generalize to completely unseen events. This is an important point, and we will revise the manuscript accordingly.

The original record-based setting was adopted mainly to maintain strict consistency with the benchmark framework used in the previous GPR-based study, thereby enabling a direct and fair comparison under the same experimental conditions. However, we agree that such a setting should not be regarded as the primary basis for assessing event-level generalization.

In the revised manuscript, we will therefore introduce an event-wise grouped cross-validation scheme as the main evaluation protocol for generalization. Using `event_id` as the grouping variable, all recordings from a given earthquake event will be kept in the same fold, ensuring strict separation of events between training and testing in every iteration. For each fold, the model will be trained on all remaining events and evaluated only on the fully held-out events.

Based on this revised protocol, we will recompute and report the relevant predictive and residual statistics, including RMSE, correlation/bias diagnostics, residual dispersion, and the variability components (τ , ϕ , and σ , where applicable). The final results will be summarized across folds, and the corresponding dispersion across folds will also be reported to reflect robustness and uncertainty.

At the same time, we will retain the current independent-earthquake analysis as a benchmark-aligned case study for comparison with the previous study, but we will revise the manuscript text

to make clear that the main claims regarding generalization to unseen earthquakes are supported by the new event-wise evaluation. We believe this revision will substantially strengthen the methodological completeness of the study and provide a more rigorous assessment of the MoE model's event-level generalization performance.

Referee #2 – Comment 2 (Feature importance and model interpretability)

The referee encourages the authors to include an analysis of input feature importance. Without feature importance analysis or sensitivity evaluation, the proposed model may appear as a black-box ensemble structure rather than a predictive tool supported by physical and statistical interpretation. In earthquake engineering, where interpretability is increasingly important, reporting only error metrics may be insufficient. At minimum, the referee suggests providing feature importance rankings, partial dependence relationships, or response trends under input perturbations to demonstrate that the model behavior is consistent with established ground-motion attenuation patterns.

Author response

We thank the referee for this valuable suggestion. We fully agree that, in addition to predictive accuracy, interpretability is essential for demonstrating that the proposed model is not merely a black-box ensemble, but a physically and statistically meaningful predictive framework. In the context of ground-motion modeling, this is particularly important because the credibility of a model depends not only on its overall error performance, but also on whether its behavior is broadly consistent with established attenuation mechanisms and engineering understanding.

To address this point, we will add a dedicated global interpretability analysis for the proposed MoE-XGB model based on SHAP (SHapley Additive exPlanations). Specifically, we will use standard SHAP summary visualizations, including SHAP beeswarm plots / summary plots, to present the global contribution and relative importance of the input features. This addition will allow us to quantify which predictors most strongly influence the model output and to show the direction and magnitude of their effects across the dataset.

In the revised manuscript, we will further discuss the resulting feature-importance patterns in relation to established knowledge in ground-motion prediction, so that the behavior of the MoE-XGB model can be interpreted in a physically meaningful way rather than only through aggregate error metrics. This revision will therefore improve both the transparency and the engineering interpretability of the proposed method.

Planned revisions

We will add a new subsection in the revised manuscript to present a SHAP-based global interpretability analysis of the MoE-XGB model. In particular:

1. We will compute SHAP values for the trained MoE-XGB model to quantify the contribution of each input feature to the predicted spectral acceleration.
2. We will present standard SHAP summary visualizations (e.g., beeswarm plots / summary plots) to provide a clear ranking of global feature importance and to illustrate how feature values influence the prediction output.

3. We will include a brief interpretation of the most influential predictors and discuss whether the observed response patterns are consistent with established ground-motion attenuation behavior and engineering expectations.

Through this revision, the manuscript will provide not only predictive-performance comparisons, but also an explicit interpretability analysis that helps demonstrate the physical plausibility and practical credibility of the proposed model.

Referee #2 – Comment 3 (Clarification of Bayesian optimization and tuned hyperparameters)

The referee notes that, in Section 3.2, the manuscript states that Bayesian optimization was used for hyperparameter tuning, but it does not clearly specify which hyperparameters were actually included in the optimization procedure. The referee therefore suggests that the manuscript should provide a clear list of the optimized hyperparameters, together with a brief description of the corresponding search ranges or optimization settings, so that readers can better understand the model training process.

Author response

We thank the referee for this helpful comment. We agree that the current description of the Bayesian optimization procedure in Section 3.2 is too concise and does not provide sufficient detail on the specific hyperparameters included in the tuning process. This omission may indeed make it difficult for readers to fully understand and reproduce the model training workflow.

In the revised manuscript, we will expand this section to explicitly describe the Bayesian optimization setup used for the MoE-XGB model. In particular, we will provide a clear list of the hyperparameters optimized during training, including the main XGBoost-related parameters such as the learning rate, maximum tree depth, and regularization-related parameters, together with the corresponding search ranges or candidate domains considered in the optimization process. We will also briefly clarify the main optimization settings so that the tuning procedure is more transparent and reproducible.

We appreciate this suggestion, as it will improve both the methodological completeness and the reproducibility of the manuscript.

Referee #2 – Comment 4 (Clarification of SA period-range selection)

The referee notes that the manuscript does not clearly explain the basis for the selection of SA period ranges. Since the segmentation of SA periods is closely related to the model output structure and its engineering interpretation, the referee suggests clarifying whether these period intervals were determined according to engineering standards, statistical characteristics of the dataset, or configurations adopted in existing Italian ground-motion models. The referee also notes that, if different period ranges correspond to different submodels or shared parameter structures, this should be further explained.

Author response

We sincerely thank the referee for this valuable comment. We agree that the basis for the selection of SA period ranges should be described more clearly, since it is directly related to both the engineering relevance of the model outputs and the completeness of the methodological presentation.

In the present study, the selection and grouping of SA periods were guided primarily by established practice in previous ground-motion studies, including NGA-related research and similar ground-motion modeling frameworks, where spectral periods are commonly analyzed and discussed in comparable ranges. Our focus was placed particularly on the shorter-to-intermediate periods, such as $T = 0.01$ s, 0.3 s, and 1 s, because these periods are of greatest practical relevance for earthquake engineering applications. In particular, the natural periods of many engineering structures fall within this range, making these SA values especially important for seismic demand characterization and engineering interpretation.

From a structural perspective, the fundamental period of a building is often approximately related to its height or number of stories, and in practical engineering this commonly leads to dominant periods that are well represented by the selected short and intermediate spectral periods. For this reason, $T = 0.01$ s, 0.3 s, and 1 s can be considered representative of a broad range of typical engineering structures. Although higher-period results are also presented in the manuscript, the available data in these ranges are relatively sparser, and their engineering relevance is comparatively more case-dependent.

In the revised manuscript, we will clarify that the chosen SA periods were selected mainly based on their widespread use in prior ground-motion studies and their strong engineering significance, especially for representing the dynamic characteristics of typical structures. We will also further explain how these period ranges relate to the presentation of the model outputs, so that the rationale for the selected spectral periods is more transparent to readers.

Referee #2 – Comment 5 (Notation, formatting consistency, and readability of figures/tables)

The referee points out that the manuscript still contains some inconsistencies in notation and formatting, including the use of hyphens and dashes, italicization of variables, and the definition and use of abbreviations. The referee also notes that the font sizes in some figures and tables appear relatively small, which may affect readability, and encourages the authors to standardize these aspects in the final version.

Author response

We sincerely thank the referee for this careful comment. We agree that consistency in notation and formatting is important for the overall clarity and professionalism of the manuscript. We also appreciate the referee's observation regarding the readability of some figures and tables.

In the revised manuscript, we will carry out a thorough check of the full text to standardize notation and formatting throughout. In particular, we will carefully review the use of hyphens and dashes, ensure consistent italicization of mathematical variables and symbols, and unify the definition and subsequent usage of abbreviations across the manuscript. In addition, we will revise the figures and tables to improve readability by increasing font sizes where necessary and ensuring that labels, legends, and annotations are presented more clearly.

We appreciate this suggestion, as these revisions will improve the presentation quality and readability of the manuscript.

Planned revisions

In the revised version, we will:

1. perform a full consistency check of notation and formatting throughout the manuscript;
2. standardize the use of hyphens/dashes, variable italicization, and abbreviation definitions;
3. enlarge and adjust font sizes in figures and tables where needed to improve readability and visual clarity.

These revisions will be incorporated throughout the manuscript to ensure a more consistent and reader-friendly final version.

Concluding note

We sincerely thank the referee for the careful reading of our manuscript and for providing this thoughtful and constructive overall assessment. We are grateful for the referee's positive recognition of the motivation, experimental logic, and potential application value of the proposed MoE-based framework for spectral acceleration prediction. In particular, we appreciate the acknowledgment that the study design—combining benchmark-aligned comparison, broader dataset validation, and an additional independent-event test—has a clear and reasonable methodological structure.

At the same time, we fully appreciate the referee's suggestions regarding methodological clarity, interpretability, and completeness of presentation. We agree that these aspects are important for strengthening the manuscript and improving its rigor, transparency, and readability. Following the referee's comments, we will revise the manuscript accordingly by:

- (i) introducing a stricter event-wise evaluation framework for assessing generalization to unseen earthquakes;
- (ii) adding a SHAP-based interpretability analysis to provide global feature-importance information for the proposed MoE-XGB model;
- (iii) clarifying the details of the Bayesian hyperparameter optimization procedure, including the optimized parameters and their search settings;
- (iv) explaining more clearly the rationale for the selected SA period ranges and their engineering relevance; and
- (v) conducting a full revision of notation, formatting, and figure/table readability to improve presentation consistency.

We believe that these revisions will directly address the referee's concerns and will substantially improve the methodological completeness and overall clarity of the manuscript. We sincerely appreciate the referee's constructive feedback, which has been very helpful in guiding us toward a stronger revised version.