

**Author Comment (AC)**  
**Response to Referee #1 (All comments)**

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

**Referee #1 – Follow-up on Comment 1 (Validation strategy and event-wise independence)**

The referee acknowledges that most issues raised in the initial review have been satisfactorily addressed, but emphasizes that the validation strategy still requires a fully event-wise independent evaluation. Specifically, the referee clarifies that strict event-wise validation in ground-motion modeling should be implemented through leave-one-event-out (LOEO) or an equivalent grouped-by-event cross-validation scheme, where the model is trained on all events except the excluded event(s) and performance metrics/residual statistics are computed exclusively on the fully excluded event(s). Final metrics should then be reported as averages over all cross-validation iterations. The referee notes that (i) Track A (single held-out event) is informative as a case study/benchmark comparison but not a robust unseen-event generalization assessment, and (ii) Track B, as previously formulated, does not unambiguously enforce strict event-wise independence.

**Author response**

We thank the referee for this additional clarification and for recognizing the improvements in our planned revisions. We fully agree that this point concerns experimental design rather than wording, and that a robust assessment of generalization to unseen earthquakes requires at least one evaluation that enforces strict event-wise independence in the sense defined by the referee. In the revised manuscript, we will therefore incorporate an explicit grouped-by-event cross-validation evaluation and report all key performance and residual-variability metrics under this event-wise independent framework.

At the same time, we will maintain our benchmark-oriented evaluation (previous Track A) strictly for its intended purpose—namely, direct comparability with the reference study used in our manuscript—while clearly delimiting its interpretive scope.

**Planned revisions (updated evaluation design)**

**(1) Track A retained as benchmark-aligned case study (scope clarified)**

- Purpose: We will retain the single held-out earthquake evaluation to preserve strict comparability with the benchmark study used in our manuscript.

- Scope: We will explicitly frame Track A as a benchmark-aligned case study, informative for reproducibility and direct comparison, but not as a standalone, robust quantification of generalization to unseen earthquakes.

**(2) Track B revised to strict event-wise independence via grouped-by-event cross-validation (new primary generalization assessment)**

We will replace the previous record-wise evaluation in Track B with an explicit **grouped-by-event cross-validation scheme**, implemented with **group = event\_id**:

- Event-wise split rule: All splits will be performed by event, such that recordings from the same earthquake cannot appear in both training and testing data in any iteration.
- Cross-validation protocol: For each fold, the model will be trained on all events outside the held-out fold, and all residuals/metrics will be computed exclusively on the fully excluded event-fold.
- Reported metrics under event-wise CV: We will recompute and report the full set of performance and residual-variability statistics under this framework, including (as applicable) RMSE, correlation/bias diagnostics, residual dispersion measures, and variability decomposition metrics ( $\tau$ ,  $\phi$ ,  $\sigma$ ), computed fold-wise on excluded events.
- Aggregation across folds: Final reported values will be summarized as averages across folds (and we will also report the dispersion across folds, to convey stability/uncertainty of the estimates).
- Interpretation: Claims regarding generalization to unseen earthquakes and the interpretation of residual variability will be based primarily on this event-wise cross-validation evaluation.

**Concluding note**

We believe that adding this explicit grouped-by-event cross-validation evaluation will directly address the referee's methodological requirement for strict event-wise independence, strengthen the robustness of our reported predictive performance and variability analyses, and provide a sound basis for interpreting improvements relative to baseline models.