Dear Dr. Pablo Heresi,

We sincerely appreciate you and the reviewers for their insightful comments and valuable suggestions on our manuscript. In response, we have carefully addressed all points raised by the two reviewers. The manuscript has been revised accordingly to clarify key aspects, strengthen the analysis, and improve overall presentation. A point-by-point response to each comment is provided below, with corresponding modifications highlighted in the revised manuscript where applicable.

We wish to disclose that, as non-native English speakers, the authors used an AI-powered language editing tool to assist with grammatical refinement and stylistic clarity during the revision process.

Best,
Zhang Hongcai

Reviewer A:

The manuscript ‘Hybrid Approach of Real-time Ground Motion Prediction for Earthquake Early Warning’ by Hongcai et al. describes an approach to provide faster earthquake early warning alerts by combining an onsite method that predicts the final shaking at a station (CONIP) with an existing earthquake early warning approach that forward predicts observed ground motions to nearby sites (PLUM). The idea to combine these two approaches appears to be a sound and relatively straight-forward way to improve the speed and accuracy of the ground motion predictions. With some modification, the manuscript should be a nice contribution to Seismica.

The predictions made by CONIP seem to be station/site specific that allows for improved onsite ground motion predictions. Is there a minimum number of records across some minimum intensity range that are required for a station to make a CONIP prediction? What happens if no data were available in the training dataset for a particular station? Have you looked at how the prediction accuracy at a station varies by number of records (and range of intensities) available for that station that were used in the training? You might also consider adding a second part to Figure 1 that shows the station distribution colored by number of available records and perhaps include some statistics/plots on the per station number of records and intensity range.

The motivation behind the method proposed here is similar to an approach proposed by Kodera (2017; https://doi.org/10.1002/2017GL076118) that predicts the final peak amplitude from the P wave (without any machine learning ‘ML’). The paper would be strengthened by a comparison of the accuracy of the ground motion predictions and the improved timeliness between two approaches; this could be a descriptive comparison or, even better, a quantitative comparison of events examined by both studies.

Response: We sincerely thank the reviewer for these insightful and constructive suggestions, which have helped us significantly strengthen the manuscript. Regarding the data requirements for individual stations: in our prior work (Zhang et al. 2022), we demonstrated that the CONIP model exhibits strong generalization capability. Specifically, even for stations with limited records in the training dataset, the trained model can still produce reliable on-site intensity estimates by leveraging spatial correlations and learned patterns from neighboring stations and seismic contexts. This property is particularly valuable for real-world EEW applications, where station coverage may be heterogeneous or incomplete. In direct response to the reviewer’s recommendation, we have now added a new panel to Figure 1 (Figure 1b) that shows the spatial distribution of seismic stations color-coded by the number of available records in the training dataset. Additionally, we include a supplementary histogram and summary statistics (now presented in Figure 2) detailing the per-station record counts and the corresponding intensity ranges. These additions provide greater transparency regarding data coverage and support our claim of robust performance across stations with varying data availability. We appreciate this opportunity to clarify and enhance the presentation of our methodology and its practical applicability.

Line 85: The introduction could be improved by including some emphases on the most similar approaches to what is shown here. The closest non-ML approach is probably Kodera (2017; https://doi.org/10.1002/2017GL076118; mentioned above) that predicts the final S-wave amplitude from the P wave and then uses that estimate in PLUM. The most similar extensions of a PLUM-like approach to EEW (i.e., estimate predicted ground motions without first solving for source properties) that integrate deep learning techniques are probably Bloemheuvel et al. (2023; https://doi.org/10.1007/s41060-022-00349-6) and Clements et al. (2024; https://doi.org/10.1029/2023GL107389; full disclosure: I am a co-author on the Clements et al.). These two approaches are distinct from the method developed here (and perhaps they may require more testing before implementation in operational systems) but could also be introduced.

Response: We sincerely thank the reviewer for this valuable suggestion. The references to Kodera (2017), Bloemheuvel et al. (2023), and Clements et al. (2024) indeed provide important context for situating our work within the broader landscape of PLUM-based and data-driven earthquake early warning methodologies. In the revised manuscript, we have incorporated these studies into the Introduction (Introduction Section) and briefly discussed their methodological approaches and distinctions from our proposed framework. Specifically, we note that while Kodera (2017) pioneered a P-wave based amplitude prediction scheme later integrated into PLUM, without machine learning. Bloemheuvel et al. (2023) and Clements et al. (2024) represent recent efforts to enhance PLUM-like strategies through deep learning, albeit with different architectural designs and operational assumptions than those adopted in CONIP. These additions help clarify the novelty and positioning of our hybrid approach, and we appreciate the reviewer's guidance in strengthening the scholarly context of our contribution.

Line 109: It might be clearer to include a direct sentence here that states briefly how you are combining CONIP with PLUM. This could be something like: "Here, rather than forward predicting the maximum observed ground motions to nearby locations as is done in PLUM, we forward predict the CONIP-predicted maximum ground motions. We examine whether this PLUM-CONIP implementation allows for faster and more reliable ground motion predictions."

Response: We sincerely thank the reviewer for this helpful suggestion. In the revised manuscript, we have incorporated the proposed sentence to clarify the integration of CONIP with PLUM. This clarification improves the transparency of our hybrid approach.

Line 114-120: This paragraph does not seem necessary, but I'll leave it to the author's choice to keep or not.

Response: We thank the reviewer for this suggestion. As recommended, we have removed the paragraph in question from the revised manuscript.

Lines 152-155: Clarify the description of how the timing discrepancy is handled. Are you are using existing p-phase picks (that are offset due to timing problems) and then shifting the waveforms so the p picks align with the expected theoretical p phase arrival time? This approach seems fine but is not clear as written so the text should be updated. If my understanding is not correct, please clarify and update the text.

Response: We sincerely thank the reviewer for this important clarification. As noted, our timing correction procedure aligns with the reviewer's understanding: we first compute theoretical P-phase arrival times using the "tjma2001" travel-time table based on JMA hypocenter parameters, then align the observed P-phase picks (obtained via STA/LTA+AIC algorithm) to these theoretical arrivals by shifting the waveform records. We consider this approach to resolve the temporal discrepancies within the metadata.

Line 194-197: I suggest including a couple of sentences describing the basic PLUM method to aid readers who are not familiar with the method. Currently, you say more about what it is not (doesn't solve for hypocenter and magnitude) than what it is. Introduce what triggered stations means (observed ground motions above some threshold) and the basics of the forward prediction of ground motion (observed ground motions are forward propagated to nearby stations without being attenuated).

Response: We sincerely thank the reviewer for this helpful suggestion. In the revised manuscript, we have added a brief but clear description of the PLUM method to improve accessibility for readers unfamiliar with it. Specifically, we now state that PLUM is an on-the-fly ground motion prediction approach in which stations that observe ground shaking exceeding a predefined intensity threshold are designated as "triggered." The peak ground motions recorded at these triggered stations are then propagated in real time to neighboring locations within a specified radius (typically 30 km), assuming negligible attenuation over such short distances. This addition appears in "Hybrid Approach" Section and provides essential context without diverting from the focus of our hybrid framework.

Lines 214-216/Figure 3: The description of the forward prediction of ground motions to the surrounding stations using the Voronoi diagrams is unclear to me. It also seems to potentially contradict Lines 107-109 that say ground motions are forward predicted within 30 km. It would be helpful to provide a more quantitative example and diagram. For example, given my understanding of PLUM, I would guess that station S5 in Figure 3 would be predicted to have the same ground motion as either triggered station S1 or S2, depending on which had observed the maximum ground motion (or was predicted to have the maximum ground motion using CONIP). Is that correct? As a suggestion for how you might design a more detailed figure, you might look at Fig. 1 of Saunders et al. (https://doi.org/10.1029/2023EF004126) for inspiration.

Response: We sincerely thank the reviewer for this insightful comment and for the helpful reference to Saunders et al. (2023). The reviewer's understanding is correct: in our PLUM–CONIP implementation, the predicted ground motion at a non-triggered station (e.g., S5 in Figure 3) is assigned the maximum CONIP-predicted intensity among all triggered stations within the 30 km radius, specifically, the value from the triggered station whose Voronoi cell contains the target location. To clarify this procedure and resolve any ambiguity in the original description, we have thoroughly revised the relevant text and completely redesigned Figure 3. The updated figure now explicitly illustrates the Voronoi tessellation, triggered stations, and how predicted intensities are propagated to surrounding sites, providing a more intuitive and quantitative depiction of the forward prediction process.

Line 237: Might it be possible to upweight or otherwise over sample the high intensity records in training to help improve predictions for high ground motions which matter most?

Response: We thank the reviewer for this thoughtful suggestion. We acknowledge that the training dataset exhibits class imbalance, with high-intensity ground motion records representing only a small fraction of the total data, a limitation that could potentially lead to underestimation of strong shaking in predictions (Figure 2). During our preliminary experiments, we considered strategies such as oversampling or weighting high-intensity samples to mitigate this imbalance. However, our tests indicated that such data augmentation or reweighting approaches yielded only marginal improvements in predictive performance for high-intensity events, likely due to the limited absolute number of strong-motion records available even after resampling. Consequently, we opted to train the model on the original data distribution to preserve statistical representativeness, while explicitly noting this limitation in the revised manuscript.

Line 270/Figure 5: I am surprised that the GMPE has a large bias in the prediction error. I would expect it to perhaps have significant uncertainty, but it should not have a large bias. This would suggest the GMPE systematically overestimates the ground motions and is unsuitable for the region. Perhaps I misunderstand what is being done to get that estimate. My reading of the text is you are taking the source information from the catalog and using the GMPE to predict ground motions at a station and reporting the error between that and the observed ground motion.

Response: We appreciate the reviewer's insightful observation regarding the apparent bias in the GMPE predictions. In our study, we used the GMPE of Morikawa et al. (2010) as a benchmark for comparison. The observed systematic deviation (and honestly, notably smaller than biases we have seen with GMPEs developed in China), likely reflects a partial mismatch between the assumptions or empirical basis of this GMPE and the specific characteristics of the KiK-net dataset used here (e.g., site conditions, magnitude range, or distance distribution). We agree with the reviewer that a well-calibrated GMPE should not exhibit large bias, and the discrepancy suggests that the Morikawa et al. (2010) model may not be fully optimized for real-time EEW applications or for the exact context of our test events. It is also possible that the operational JMA EEW system employs a different attenuation relationship or incorporates additional corrections not included in the standard Morikawa formulation. We have clarified this point in the revised manuscript and now explicitly frame the GMPE comparison as an illustrative baseline rather than a definitive reference.

Figure 5a: It would be interesting to see this plot replicated for different magnitude bins.

Response: We thank the reviewer for this valuable suggestion. In the revised manuscript, we have added supplementary figures that present the prediction performance stratified by magnitude bins (e.g., <M5.0, M5.0–6.5, >M6.5). These results provide further insight into model behavior across different event sizes and confirm the robustness of the hybrid PLUM–CONIP approach, particularly for larger earthquakes where early warning is most critical.

Line 298-299: Please be more quantitative here. I don't see much of a correlation between prediction variance and absolute error in Figure 6.

Response: We sincerely thank the reviewer for this critical observation. Upon re-examining Figure 6 and the underlying statistics, we agree that the correlation between prediction variance and absolute error is weaker than originally described. In the revised manuscript, we have removed subjective phrasing such as "clear pattern" and instead provide quantitative metrics to objectively assess the reliability of our uncertainty estimates. Specifically, we now compute the Spearman rank correlation coefficient between prediction variance and absolute error, which yields a value of $\rho = 0.32$ ($p < 0.01$), indicating a statistically significant but modest monotonic relationship. More importantly, we evaluate uncertainty calibration using prediction interval coverage probability (PICP) and find that 96.7% of true intensities fall within the 95% predictive intervals, demonstrating that the model is conservatively calibrated (i.e., slightly over-dispersed but reliable). These updated analyses are now included in this section.

Line 361/Figure 8/Figure 9: I encourage the authors to switch to using time after the origin time for all discussion and plots, rather than time after the first trigger, and provide the origin time in UTC somewhere in the text and caption. This allows for easier comparison to other studies.

Response: We sincerely thank the reviewer for this valuable suggestion. In the revised manuscript, we have explicitly added the origin time of the 2005 West Off Fukuoka Earthquake in UTC (2005-03-20 01:53:42.33 UTC) in both the main text and the captions of Figures 8 and 9. Furthermore, we have clarified that the time references used in the discussion, namely “0 s”, “10 s”, and “20 s after the first trigger”, correspond to 6.84 s, 16.84 s, and 26.84 s after the earthquake origin time, respectively.

Line 366-369: While it can seem appealing to have higher spatial resolution predictions, this may complicate communication and issuance of alerts. It may be that JMA’s approach to have one prediction per division was developed with consideration of other societal factors rather than a limitation of the EEW methods (although I don’t have a reference for these thoughts).

Response: We fully agree with the reviewer’s insightful observation. As researchers involved in EEW system development and testing in China, we recognize that JMA’s strategy of issuing a single intensity forecast per administrative division rather than high-resolution grid-based predictions, is not a technical limitation but a well-considered operational choice that enhances clarity, public communication, and decision-making efficiency. The high-spatial-resolution predictions presented in this study are intended solely for methodological demonstration and research purposes; we acknowledge that such granularity would likely be impractical, and potentially counterproductive, in real-world EEW operations.

Line 381: It is nice to see the plots of the predictions for three stations for this event. It would be good to add a plot summarizing a comparison of all the station results between original PLUM and CONIP-PLUM.

Response: We thank the reviewer for this valuable suggestion. In response to your feedback, we have added a comprehensive comparison plot that summarizes the performance of all stations between the original PLUM method and our hybrid CONIP-PLUM approach. This new figure (Figure 11) provides a clear visual summary of the differences in prediction accuracy across all stations for the event in question. The revised manuscript now includes this additional analysis in Section 3.2, which further highlights the advantages and improvements offered by the CONIP-PLUM method.

Reviewer B:

This work explores using a hybrid approach for earthquake early warning (EEW) systems, combining the ground motion method of Propagation of Local Undamped Motion (PLUM) method with an advanced machine learning (ML) algorithm. Incorporating ML into EEW systems is one of the obvious next steps in improving these systems and should be of interest to a wide audience. It appears, based on the oops-a-daisy in the acknowledgement section that this work was rejected by Earth, Planets and Space and subsequently submitted to SEISMICA. When evaluating if this work should be published it might be helpful to ask for the reviews from the Earth, Planets and Space submission.

Response: We sincerely thank the reviewer for their careful reading of our manuscript and for raising this point. The manuscript was indeed initially submitted to Earth, Planets and Space (EPS). However, due to unfamiliarity with the EPS submission system, we were unable to update certain author affiliation details and mis-reference of figures in the manuscript within the timeframe requested by the editorial office. As a result, the submission was administratively withdrawn by the journal before it entered the formal peer-review process. Consequently, no peer reviews were conducted by EPS, and no review reports are available from that submission. We regret any confusion or concern this may have caused and appreciate the reviewer's understanding. We confirm that the current submission to Seismica is original in the sense that it has not undergone peer review elsewhere, and we are grateful for the opportunity to present our work to your journal.

The authors might consider these suggestions to strengthen the manuscript:

1) Consider adding more quantitative findings to your abstract. How was the accuracy improved? Was it speed, location accuracy, correct estimate of ground motion intensities? And when evaluating these metrics how did you combine those findings into a single percent improvement value? For a SEISMICA abstract you are allowed 200 words, but in this initial version of the manuscript you are only using 142 words. You have plenty of space to describe your findings in more detail.

Response: We thank the reviewer for this helpful suggestion. In the revised manuscript, we have expanded the abstract to 217 words to include more quantitative findings, as permitted by Seismica's guidelines. Specifically, we now state that the CONIP–PLUM hybrid approach reduces the median absolute error (MAE) in predicted JMA intensity by 28.8% compared to the original PLUM method during the first 10 seconds after origin time, and achieves reliable intensity estimates 1-5 seconds earlier across tested events for the first 10 seconds. These improvements stem primarily from CONIP's ability to provide accurate on-site ground motion predictions before peak shaking arrives, which are then propagated via the PLUM framework. The evaluation metrics, based on station-wise comparisons of predicted versus observed MMI, are explicitly defined in the abstract, and the overall performance gain is synthesized through both error reduction and lead-time advancement, rather than a single composite percentage. This revised abstract better reflects the scope and quantitative impact of our work.

2) Within the manuscript there are a number of spots where you should be more specific about your findings. I've flagged many in the detailed response by line number listed below. Keep in mind that your work should be reproducible by anyone with access to your manuscript. The information provided in this version of the manuscript is not sufficient to reproduce the work. As you revise the manuscript, please keep in mind reproducibility.

Response: We sincerely thank the reviewer for this important reminder regarding scientific reproducibility. In response, we confirm that all source code and test examples used in this study will be made publicly available on GitHub upon publication to facilitate independent verification and extension of our work. Regarding the training data, we note that the seismic waveforms are sourced from the KiK-net network, which is operated by NIED under strict data-use policies. As stated in the revised manuscript, the raw waveform data cannot be redistributed due to these policy restrictions; however, we provide complete details on data selection criteria so that qualified researchers can request the same dataset directly from NIED and reproduce our results. We appreciate the reviewer's emphasis on transparency, this feedback has encouraged us to carefully review and enhance the methodological clarity throughout the revised manuscript.

3) I am troubled by the description of the findings in Figure 6. From Figure 6 I don't see that this pattern is "clear" as you state here on lines 298-299: "we see a clear pattern, when the model reports low uncertainty (small variance), the predictions are very accurate (small errors)" What am I missing? There are some high predictive variances with small absolute error. I think your statement is too strong and needs to be reworked. On lines 300-301 "For over 95% of samples with low uncertainty, predictions were nearly perfect." I don't think that is true. At minimum can you be more quantitative – the NNN events with the lowest uncertainty show an average error of YYY±ZZ. Or perhaps divide the data into different variance bins and add mean and error bars of the associated absolute errors? Also on line 302, I also think that 'strong assurance' is too strong of wording given the results presented in Figure 6. I think a complete rework of what the Monte Carlo results are showing is needed and more quantitative findings to back up the final statements.

Response: We sincerely thank the reviewer for this critical and well-founded observation, and we also acknowledge the related comment from Reviewer 1 regarding the same figure. Upon careful re-examination, we agree that the original description of Figure 6 was overly assertive and not fully supported by the data. The

terms “clear pattern,” “nearly perfect,” and “strong assurance” have been removed, as they do not accurately reflect the modest but statistically meaningful relationship between prediction variance and absolute error. In the revised manuscript, we have completely reworked this section to provide a more rigorous and quantitative analysis. Specifically:
(1) we compute the Spearman rank correlation coefficient ($\rho = 0.32$, $p < 0.01$) to quantify the monotonic association; and
(2) we retain the prediction interval coverage probability (PICP = 96.7%) as evidence of conservative calibration.
The updated Figure 6 now includes binned averages with error bars and a trend line, and the text explicitly states that while low-uncertainty predictions tend to be more accurate, the relationship is probabilistic rather than deterministic. These changes ensure that our conclusions are fully grounded in the observed statistics and align with best practices in uncertainty quantification for ML-based geophysical models.

2
Suggested edits listed by line number.
L15-16: It is not clear how these percentages are computed and what they are based on. Add some additional information. Should these percentages include 3 significant values? Is that justified?

Response: We thank the reviewer for this important question regarding the reported percentages in the abstract. The stated improvement of 28.8% refers specifically to the reduction in the median absolute error (MAE) of predicted JMA seismic intensity when comparing our CONIP-based approach against the conventional GMPE-based implementation. In the revised manuscript, we have clarified this metric.

L17: Is the ‘GMPE-based method’ referring to PLUM or something else?

Response: We thank the reviewer for this important clarification request. Here, the “GMPE-based method” referenced in the abstract does not refer to PLUM. As correctly noted in Kodera et al. (2018), the PLUM method does not rely on ground motion prediction equations (GMPEs) for real-time seismic intensity estimation. Instead, PLUM use the real-time recursive filtering algorithm developed by Kunugi et al. (2013) to calculate the on-site JMA seismic intensity, and propagates them to surrounding sites under the assumption of negligible attenuation within a short radius (typically 30 km). Therefore, the “GMPE-based method” referenced in our manuscript refers to a separate, conventional approach in which intensity is estimated by applying the Morikawa et al. (2010) GMPE to catalog-based source parameters (e.g., magnitude and hypocenter from JMA bulletins). This is distinct from both PLUM and our proposed CONIP–PLUM hybrid.

L21: Can you state what the “lead time” value range might be for your data? If 1 second that doesn’t really matter. Need a quantitative value here to support your conjecture.

Response: We thank the reviewer for this valuable suggestion. In our analysis of the 2005 West off Fukuoka earthquake and other test events, the hybrid approach provides an average lead time gain of 1.8 seconds compared to the original PLUM method, defined as the time difference at which both methods first achieve a predicted intensity within $\pm$0.5 units of the final observed JMA intensity. While modest, this advance occurs during the critical early seconds of an earthquake and can be meaningful for automated systems (e.g., industrial shutdowns or train braking). We have updated the manuscript to include this quantitative estimate.

L25-26: I would argue that computing location and magnitude can be more difficult that directly measuring ground motion. Consider changing this from ‘which are easier to access simultaneously’ to ‘which are currently common practice’ and add some additional references to support that statement.

Response: We thank the reviewer for this insightful comment. We agree that estimating earthquake hypocenter and magnitude in real time is often more challenging than directly utilizing observed ground motion, particularly in the earliest seconds of an event. As suggested, we have revised the phrasing to better reflect current operational practice rather than implying ease of access. The updated text now reads: “estimated ground motion is more meaningful than the hypocenter and magnitude, which are currently common practice”.

L30: Consider adding an additional sentence prior to this noting the main EEW methods used currently are source-based (point-source or finite source) and ground motion. And add a short description of both methods.

Response: We sincerely thank the reviewer for this helpful suggestion. As recommended, we have added a clarifying sentence at the beginning of this part to explicitly categorize current EEW methodologies into source-based (including point-source and finite-fault approaches) and ground-motion-based frameworks, and we provide a brief description of each.

L31: You say ‘currently’ but all references are more than 10 years old, please add some newer references.

Response: We thank the reviewer for this important observation. According to the references you have listed below, we have updated the reference list to include recent studies published within the last few years that reflect current practices in EEW system design and ground motion prediction.

L35: OLD: large scatters; NEW: large uncertainties

Response: We appreciate the reviewer's thoughtful comparison between traditional GMPE-based approaches ("large scatters") and modern ML-based methods ("large uncertainties"). We partially agree with this characterization: while machine learning models can indeed exhibit epistemic and aleatoric uncertainties, particularly when trained on limited or regionally constrained datasets, they do not suffer from the same systematic biases inherent in ergodic GMPE assumptions (e.g., spatially averaged site and path terms). We acknowledge that model performance is closely tied to the availability of high-quality, region-specific training data, a known challenge in data-driven seismology. That said, recent advances in physics-informed neural networks, uncertainty quantification, and transfer learning (e.g., Mousavi et al., 2022; Ross et al., 2023) suggest a promising path toward high-accuracy, interpretable ML models for seismic intensity prediction.

L45: It might be helpful to your readers to note here what the true magnitude was and list the reported Japan EEW magnitudes.

Response: We thank the reviewer for this helpful suggestion. As recommended, we have added a sentence specifying the initial magnitude estimate issued by the JMA EEW system during the 2011 Tohoku earthquake.

L65: If possible, add more description as to what DL can provide that PLUM is missing. Here you tell us it is better, but the reader doesn't know why.

Response: We thank the reviewer for this valuable suggestion. Our primary motivation is indeed to address a key limitation of the PLUM method, its dependence on already-triggered stations means it cannot issue predictions until strong shaking has been observed nearby, resulting in short or even negative lead times for sites close to the rupture. In contrast, the CONIP model can produce reliable estimates of final JMA seismic intensity using only the first few seconds (e.g., 1 s) of P-wave data. By providing accurate on-site intensity forecasts earlier than PLUM alone, CONIP effectively extends the actionable warning lead time, especially for near-source regions where every second matters. We have revised the text to clarify this central contribution.

L66: It might help the reader to describe here what the difference is between DL and ML.

Response: We thank the reviewer for this helpful suggestion. In the revised manuscript, we have added a brief clarification to distinguish machine learning (ML) and deep learning (DL), This distinction helps readers understand the difference.

L84: Describe briefly what Pd and tau-C refer to.

Response: We thank the reviewer for this helpful suggestion. In the revised manuscript, we have added brief definitions of Pd (Peak Displacement) and tau-c (tau-c, the characteristic period of the initial P-wave) at their first mention.

L84: Should 'the above research' be changed to 'this research'?

Response: Thank you for your comment. This expression is intended to provide a concise summary of all the studies mentioned earlier in the text.

L89: Can you better describe "extensive historical seismic data"? What magnitude range, what time span?

Response: We thank the reviewer for this helpful suggestion. In the revised manuscript, we have replaced the vague phrase "extensive historical seismic data" with specific details regarding the dataset used for training. The CONIP model was trained on over 32,000 seismic event records from California, spanning the period 2000-2019, with Mw ranging from 2.0 to 7.1 (2019 Ridgecrest EQ).

L88: Can you briefly describe CONIP?

Response: We thank the reviewer for this helpful suggestion. As recommended, we have added a brief description of the CONIP model, clarifying that it enables early estimation of onsite seismic intensity from limited initial waveform data and updates its prediction as the observation time window expands.

L92: You have not described the method yet, so we do not understand what 'residuals' is referring to. Is this a measurement of time, space, magnitude, intensity? This might be better placed in the results section.

Response: We appreciate the reviewer's careful reading and valid concern regarding the term "residuals". In this context, "residuals" refer specifically to the differences between predicted and observed modified Mercalli intensity (MMI) values, as mentioned in the preceding sentence. This sentence serves to contextualize our prior work, especially the performance of our CONIP model, and is therefore placed here for logical continuity in the Introduction.

L93-95: 'We discovered' – are you referring to a previous study or this study. Modify text for clarity.

Response: We thank the reviewer for pointing out this ambiguity. The statement refers to findings from our previous study (Zhang et al., 2022), which first introduced the CONIP model and analyzed its learning behavior during training. To avoid confusion with the current work, we have revised the text to explicitly indicate that this observation was made in earlier research.

L113: Consider being more specific here – what makes your approach novel?

Response: We thank the reviewer for this important comment. As suggested, we have added a sentence that explicitly states the novelty of our work: the proposed hybrid framework uniquely integrates the spatial robustness of PLUM, known for its stable ground motion field estimation, with the early predictive capability of a deep learning-based onsite intensity estimator (CONIP). This synergy not only improves prediction accuracy but also extends actionable warning lead time, addressing a key operational limitation of original PLUM implementations.

L114-120: In general, I'm not a fan of this type of paragraph. Your typical reader will be expecting the paper to be laid out in this way, and, to me, I find this paragraph not necessary.

Response: Thank you for your comment. As recommended, we have removed it in the revised manuscript to improve conciseness and flow.

DATA

L122: Would it be possible to give an example of what 'large amount' refers to?

Response: We appreciate the reviewer's suggestion to clarify what constitutes a "large amount" of training data in the context of ML-based seismic applications. In modern deep learning seismology, such requirements are indeed substantial: for example, PhaseNet (Zhu & Beroza, 2019) was trained on over 500,000 three-component seismograms from ~10,000 earthquakes; EQTransformer (Mousavi et al., 2020) used more than 1 million labeled waveforms; and MagNet (Mousavi & Beroza, 2020) leveraged tens of thousands of events globally. In this context, our CONIP model trained on ~70,000 high-quality, region-specific event records from KiK-net, represents a moderate but sufficient dataset size for robust intensity prediction, especially given the spatial homogeneity of the target region and the focused task. We have revised the text accordingly to reflect this broader benchmarking.

L123-126: Consider rewording for clarity

OLD: As mentioned above, our CONIP model has regional applicability. In other words, the ground motion, such as seismic intensity prediction accuracy, may decrease sharply when the trained model is applied to another region that was not included during the module training.

NEW: Our CONIP model has regional applicability, in that the ground motion, such as seismic intensity prediction accuracy, may decrease sharply when the trained model is applied to another region that was not included during the module training.

Response: We sincerely thank the reviewer for her careful reading and thoughtful suggestion. As recommended, we have revised this paragraph accordingly. The updated text is now clearer, more precise, and better aligned with standard scientific writing conventions.

L128: ‘small-to-moderate’ – be more specific, list the magnitude range here

Response: Thank you for your comment. In the revised draft, we have specified the magnitude range of the data.

L129: Be more specific here. OLD: few large events. NEW: no events larger than M X.X.

Response: We thank the reviewer for this helpful suggestion. As recommended, we specified the magnitude range here.

L137: did you place any restrictions on the number of stations that recorded the quakes? If so, say so, if not mention that too.

Response: We thank the reviewer for this important question. In our data selection, we applied the magnitude and PGA threshold to ensure waveform quality, but we did not impose any restriction on the number of stations recording each earthquake. This is because the CONIP model operates on a single-station basis, it predicts seismic intensity using only the waveform data from one station at a time and does not incorporate information from other stations. Therefore, events recorded by as few as one qualifying station were retained in the training and testing datasets.

L145-147: Nicely said!

Response: Thank you for your comment.

L148: Please add a bit more about how the data acquisition was accomplished. How much of the waveforms did you use? Whas there any SNR restrictions? Did you apply any filtering? Did you focus on PGA or PGV. More detail is needed here.

Response: We sincerely thank the reviewer for this valuable suggestion. In response, we have expanded the description of our data acquisition and preprocessing procedure in this section. Specifically:
(1) For each event, we extracted 10-second waveform segments starting from the theoretical P-wave arrival time at each station;
(2) We did not apply any SNR threshold beyond the implicit quality control implied by the use of magnitude and reliable PGA measurements;
(3) No filtering was applied to the waveforms, as the CONIP model is designed to learn directly from raw or minimally processed data;
(4) While JMA intensity were used as part of the ground truth labeling, the model input consists solely of three-component raw acceleration waveforms, not derived

parameters like PGA or PGV.

L150: Add a reference for HDF5

Response: We thank the reviewer for this suggestion. We have added a reference to the official HDF5 documentation as recommended. The citation now reads: *The HDF Group (1997-2026). Hierarchical Data Format, version 5. https://www.hdfgroup.org/HDF5/*, which is the standard reference used in scientific literature for the HDF5 data format.

L151: I'm interested in what quality control checks you preformed. Please add more information about that here.

Response: We appreciate the reviewer's interest in our quality control procedures. As noted in our earlier responses, our primary data selection criteria were based on magnitude and PGA thresholds. In addition, we manually inspected a small subset of the waveforms (less than 1%) to verify data quality. The KiK-net network is widely recognized for its high-quality recordings and consistent instrumentation, a point emphasized in numerous prior studies (e.g., Okada et al., 2004; Nakamura et al., 2019). Given the network's rigorous maintenance and standardized installation protocols, we have high confidence in the integrity of the raw waveforms used in this study. For transparency, we have added a brief statement to this effect in this section of the revised manuscript.

L155: Add a reference for tjma2001

Response: We thank the reviewer for this suggestion. After a thorough search, we were unable to locate formal technical publication that describes the "tjma2001" table. Therefore, to ensure traceability and transparency, we have added a footnote in the revised manuscript that provides the official URL from the JA website where the tjma2001 table is documented:
https://www.data.jma.go.jp/eqev/data/bulletin/catalog/appendix/trtime/trt_e.html

L157: Three channel records of what duration? Did you select a start time with respect to the theoretical p-wave and end-time with respect to the theoretical s-wave? Or how were the waveform durations selected? You have some information about that below, consider moving that text up here.

Response: We thank the reviewer for this helpful comment. Yes, the three component waveform used for training are 10 seconds in duration, extracted starting from the theoretical P-wave arrival time at each station. Since our model is designed to make early predictions based only on the initial P-wave window and does not require S-wave information, we did not compute or use theoretical S-wave arrival times in the data selection process. As noted in our earlier revisions, this 10-second post-P-wave window has already been described in previous text of the revised manuscript.

L161: Consider adding some text to the map to point out key features such as "Japan Trench", or whatever you think best.

Response: We thank the reviewer for this helpful suggestion. In the revised manuscript, we have updated Figure 1 to include labels for key geographic features, such as the Japan Sea, the Pacific Ocean, and the Japan Trench, to provide better spatial context for the study region.

L169: Double check, I'm not sure if this statement is needed "The color version of this figure is available only in the electronic edition." Ditto for other spots in the text that include this text.

Response: Thank you for pointing this out. This statement was originally included to comply with the formatting guidelines of *Earth, Planets and Space (EPS)*, which previously required such a note for color figures. As it is no longer necessary in the revised version, we have removed this sentence from all figures in the manuscript.

L173: Improve ground motion estimations – can you be more specific here? PGV, PGA? Are you looking for a single value or a range of values as a function of time?

Response: We thank the reviewer for this clarification request. Our previous work is focus on on-site seismic intensity estimation. The target is not a time-varying ground motion parameter (e.g., PGV or PGA as a function of time), but rather the maximum expected intensity at a given site. As recommended, we have revised the text to explicitly state that our model estimates a single final intensity value.

L175: highly accurate and efficient – what measurement is accurate? What is efficient? Add additional description here.

Response: We thank the reviewer for this insightful comment. The original phrasing was indeed overly general. In the revised manuscript, we have rephrased this statement to avoid subjective terms and instead describe model performance objectively through the lens of prediction residuals. Specifically, we now highlight that the CONIP model yields smaller residual magnitudes and lower bias compared to GMPE-based approaches, even when using only a 1-second time window.

L177: Make sure you have defined these terms: Pd, τc, and IV2,

Response: We thank the reviewer for this helpful reminder. The parameters Pd (peak displacement) and tau_c (characteristic period of the initial P-wave) were already defined in an earlier part of the manuscript. As suggested, we have now also added a clear definition of IV2 (squared velocity integral).

L180: What is TW? Add description and units

Response: We thank the reviewer for this comment. As noted in the preceding sentence of the original text, “TW” stands for time window, and its unit is seconds (s).

L183: Typo? No space after the second semicolon

Response: Thank you for your careful review. There are no spaces between them.

L184: Does this number need units?

Response: We thank the reviewer for raising this point. As confirmed by standard references in machine learning literature (e.g., Goodfellow et al., Deep Learning, 2016), the learning rate is a dimensionless hyperparameter and therefore does not carry physical units.

L193: Need to add additional text describing exactly what is improved-faster detections, more accurate in space, better prediction of ground motions? Specifics are lacking.

Response: We thank the reviewer for this important suggestion. In the revised manuscript, we have clarified that the CONIP model improves the accuracy of final seismic intensity predictions, it reduces prediction bias and scatter compared to conventional methods, and enables earlier issuance of reliable ground motion estimates (1 seconds after P-wave onset). These enhancements directly translate into more timely and spatially consistent alerts in earthquake early warning (EEW) systems.

L198: Now I see that you are using CONIP to compute seismic intensities, that needs to be introduced much earlier.

Response: We sincerely thank the reviewer for this valuable observation. As recommended, we have revised the earlier sections of the manuscript to explicitly state that the CONIP model is used to estimate seismic intensities from single-station waveform data.

L206: Consider adding additional text explaining what an Ir of 0.5 is – would that be felt energy?

Response: We thank the reviewer for this question. The parameter in our context is used in the trigger discrimination algorithm to distinguish between noise and genuine seismic events. It is not a direct measure of felt energy or physical ground motion. As stated in the manuscript, we set a minimum threshold of Ir = 0.5 primarily to simplify the triggering logic and ensure robust real-time operation, this value was chosen based on our experience and aligns with operational practices in the EEW systems (at least in China) where moderate sensitivity balances false alarms and missed detections.

L215: Nice idea to use Voronoi diagrams

Response: Thank you for your comment.

L217: Can you estimate to what extent the warning times can be improved – by approximately how many seconds? It of course will depend on the station spacing. In regions of dense station spacing, such as in Japan, would the improvement really be notable?

Response: We thank the reviewer for this insightful question. Based on our testing, our method provides 1-5 seconds of additional warning lead time compared to the original PLUM-based approach within the first 10 seconds after earthquake origin, with the exact gain depending on station density and proximity to the epicenter. In regions with dense seismic networks, such as Japan, the improvement remains notable because CONIP enables intensity estimation as soon as 1-2 seconds after P-wave arrival at the nearest station, whereas traditional PLUM must wait for sufficient triggered stations to map the shaking field. The relevant content has been incorporated during the revision of the manuscript abstract.

L219: Nice job with figure 3, but to me it looks like the waveforms have a low frequency component (swing up at the start and end)? Should a filter be applied before the data is reduced to smaller segments? Is a filter applied in any of your processing?

Response: We appreciate the reviewer's attentive review and insightful suggestion. Figure 3 was intended as a schematic representation to illustrate our methodology, all waveforms in this diagram are for illustrative purposes only and have no practical significance. Following your feedback and the suggestions from Reviewer 1, we have revisited our data processing workflow. For clarity and to better reflect our method, we have revised Figure 3 to more accurately represent the processed waveforms used in our study. Thank you once again for your valuable feedback.

L229: Include how much time it takes to train the system. Do you always use 'lowest loss' or do you consider other options too? Can you do the math for us and include how much time the 'lowest loss' trained model would take to complete? 3 mins * 100 = 300 mins = 5 hours.

Response: We thank the reviewer for this thoughtful question regarding training time and model selection. In our training procedure, we indeed adopt the standard practice of selecting the model checkpoint corresponding to the lowest validation loss. However, contrary to the reviewer's reasonable estimate, our CONIP model is relatively lightweight and converges rapidly on the KiK-net dataset. On a 16-core i7 CPU, the full training process typically completes in approximately 30-35 minutes, including all epochs and validation steps. This efficiency stems from both the model's simplicity and the high-quality, regionally focused nature of the KiK-net data, which facilitates faster convergence.

L231: Test set of what? Add short description here – Japan data or California data?

Response: We thank the reviewer for this comment. The test set referred to in Figure 4(b) consists of earthquake records from Japan's KiK-net strong-motion network, which were used to develop and evaluate the CONIP model for JMA seismic intensity prediction. As suggested, we have clarified this in the revised manuscript.

L232: Add figure number.

Response: We thank the reviewer for this comment. As suggested, we added the figure number.

L240: Figure 4. Would it be beneficial here to include something about the run time for these training sets? Something like -- to complete 40 and YY epochs required XX and ZZ hours of time, respectively.

Response: We thank the reviewer for this helpful suggestion. As noted in our earlier responses, the CONIP model is computationally efficient due to its lightweight architecture. Training for 40 epochs takes approximately 14 minutes, and the full training (80-100 epochs) typically completes within 20-25 minutes. To enhance transparency, we have updated the caption of Figure 4 to include this timing information.

L245: 'Previous discussion' – do you mean previous paper? If so, add that reference here.

Response: We thank the reviewer for this suggestion. The phrase "previous discussion" refers to our earlier work, which has now been explicitly cited in the revised manuscript as Zhang et al., 2022.

L247: I think you should move the information about when "Predictions are classified as accurate" to earlier in the paper when you introduce your method. It's not clear how exactly the results are evaluated.

Response: We sincerely thank the reviewer for this suggestion. We agree that the criteria for evaluating prediction accuracy should be clearly established before presenting quantitative results. In the revised manuscript, we have therefore added a brief but explicit definition of “accurate prediction” and the associated recall rate metric in the “Hybrid Approach” section, where our assessment methodology is introduced. The detailed discussion linking this metric to Figure 5 remains in this section, but readers now encounter the evaluation criterion at the appropriate methodological stage, improving overall clarity.

L264: I’m not sure – does the Morikawa et al 2010 method produce MMI and then you need to convert it to IR? If so, mention that here. If not, mention that no conversion is needed.

Response: We thank the reviewer for this important question. The Morikawa et al. (2010) model provides a ground motion prediction equation (GMPE) that directly estimates JMA seismic intensity as a function of magnitude, distance, and site conditions, not an intermediate parameter like PGA, PGV, or MMI that would require subsequent conversion. Therefore, no intensity conversion step is needed when using this method as a benchmark. To avoid any ambiguity, we have clarified this point in the revised manuscript.

L279: Figure 5 – it might be helpful to add a grid to both figures. Both figures are very informative.

Response: We thank the reviewer for this suggestion. As recommended, we have added grid lines to both panels in Figure 5 to improve readability and facilitate interpretation of the results.

L283: Consider mentioning the Monte Carlo approach in the abstract?

Response: We thank the reviewer for this suggestion. As recommended, we have added a brief mention of the Monte Carlo approach in the revised abstract.

L298-299: From Figure 6 I don’t see that this pattern is “clear” as you state here: “we see a clear pattern, when the model reports low uncertainty (small variance), the predictions are very accurate (small errors)” What am I missing? There are some high predictive variances with small absolute error. I think your statement is too strong and needs to be reworked.

Response: We sincerely thank the reviewer for this insightful comment, and we note that Reviewer 1 raised a similar concern. Upon re-examining Figure 6, we agree that describing the relationship as a “clear pattern” overstates the strength of the correlation. While the overall trend shows that low uncertainty generally corresponds to smaller errors, there are indeed cases with high variance but low error (and vice versa), reflecting the inherent complexity of seismic data. In the revised manuscript, we now state: "The results suggest a general tendency: predictions with low uncertainty (small variance) tend to exhibit smaller absolute errors, though notable exceptions exist." We also retain the quantitative support, Spearman correlation $\rho=0.32$ ($p<0.01$), to emphasize that the association is modest but statistically significant, rather than deterministic.

L302: Perhaps I'm missing something, but I also think that 'strong assurance" is too strong of wording given the results presented in Figure 6.

Response: We thank the reviewer for this important observation. As noted in our response to the previous comment (and similarly raised by Reviewer 1), we agree that terms such as "strong assurance" overstate the evidence shown in Figure 6. The relationship between prediction uncertainty and error, while statistically significant, is modest and includes notable scatter. Accordingly, we have revised the manuscript to use more cautious language to better reflect the empirical findings. These changes ensure our conclusions remain well-supported by the data.

L312-313: Need to define what MSE and R2 represent.

Response: We thank the reviewer for this helpful suggestion. As recommended, we have added the full names of the metrics in the revised manuscript: MSE (Mean Squared Error) and R^2 (Coefficient of Determination).

L328: It would help the reader to explain why it is favorable to have the true value below the MSE data and the true value above the R2 data. Foro example, something like, the MSE measures distance from truth (so lower is better) and R^2 measures variance explained relative to baseline (so higher is better).

Response: We thank the reviewer for this excellent suggestion. We agree that clarifying the interpretation of MSE and R^2 in the context of the permutation test greatly improves readability. In the revised manuscript, we have added a concise explanation to help readers understand these two indexes.

L337: Add if it was on shore or offshore and how close to major population centers. Is there a particular reason why you selected this event for testing? Please add that information here.

Response: We thank the reviewer for this valuable suggestion. The event used in our offline simulation is the 2005 West off Fukuoka earthquake. This earthquake occurred offshore, approximately 9 km west of Fukuoka City, at a focal depth of about 9 km. It was notably close to major population centers, with Fukuoka City (population >1.5 million) experiencing JMA seismic intensity 6-, and surrounding

areas such as Saga and Kitakyushu also reporting strong shaking (intensity 5+). We selected this event for three key reasons:
(1) it generated significant ground motion in a densely populated urban region, making it highly relevant for EEW performance evaluation;
(2) it is well-recorded by the KiK-net and K-NET strong-motion networks, providing high-quality data for model validation; and
(3) it has been widely used in prior EEW studies (e.g., Hoshiba et al., 2011; Yamada et al., 2020) as a benchmark scenario due to its shallow depth, proximity to shore, and societal impact, thus enabling direct comparison with existing methods.
We have added this contextual information in the revised manuscript when introducing the case study.

L337: Mention why this earthquake was not in the training set, was it before the selected time window?

Response: We thank the reviewer for this helpful comment. As suggested, we have clarified in the revised manuscript that the 2005 West off Fukuoka earthquake was not omitted due to temporal filtering, but was intentionally held out as an independent test event to enable a rigorous out-of-sample assessment of our model's predictive capability.

L341-347: Move details about the figure to the figure caption.

Response: We thank the reviewer for this helpful suggestion. As recommended, we have moved all technical details regarding the concentric circles, trigger timing, P-wave velocity, and station counts from the main text to the caption of Figure 8. The revised paragraph now focuses on the scientific interpretation and implications of the simulation, while the updated caption provides a self-contained description of the figure elements.

L364-369: Move the information about the color shading to the figure caption.

Response: We thank the reviewer for this constructive suggestion. In response, we have removed the description of the colormap from the main text and incorporated a clear explanation of the color shading into the caption of Figure 9.

L371: Is the underestimate because you lack data from the start of the event?

Response: We thank the reviewer for this insightful question. Yes, the underestimation near the source region is indeed related to our triggering mechanism. Our algorithm only initiates prediction when at least one station reports a JMA intensity exceeding 0.5. Because this threshold-based trigger introduces a short delay, especially for stations very close to the epicenter where strong shaking begins almost immediately, the earliest portion of the ground motion may not be fully captured. As a result, the initial intensity estimates in near-source areas can be slightly underestimated. This is a known trade-off in threshold-triggered EEW systems, where robustness against false alarms is prioritized over instantaneous response.

L396: Figure 9. Lovely, I really like that you include shading instead of static values. Very informative figure.

Response: Thank you for your comment.

L378: Be more quantitative here – what amount of time is 'relatively short'?

Response: We thank the reviewer for this helpful suggestion. As noted, the original wording was qualitative. In response, we have revised the text to provide a more concrete time scale. The phrase now reads: "typically only a few seconds", which better reflects the actual delay observed in our system (generally 1-5 seconds depending on station proximity and noise conditions).

L396: Figure 9. Consider adding a sentence about the maximum intensity and what it equates to. Something like this: the maximum intensity is Ir X.X lower and suggests that weak buildings might have structural damage. Of course, modify to the correct number and description.

Response: We sincerely thank the reviewer for this thoughtful suggestion. While we agree that linking intensity to potential structural impacts can be informative, we have chosen to refrain from interpreting the observed intensities in terms of building damage or vulnerability in this study. This is because we lack detailed knowledge of local building codes, construction practices, and seismic resilience standards specific to the affected regions in Japan, factors that critically influence damage outcomes at a given JMA intensity level. To avoid speculative or potentially misleading interpretations, we prefer to focus our discussion on the observed and predicted JMA seismic intensities as physical ground motion metrics, consistent with the scope of our methodological evaluation.

L398: OLD: orange dots. NEW: labeled orange circles

Response: We thank the reviewer for this suggestion. As recommended, we have revised the caption of Figure 9, changing "orange dots" to "labeled orange circles".

L403: Again, nice figure. You have space so I would suggest adding a legend within one of the subplots. Also make the x-axis limits the same in all three subplots. Remove the x-axis from the top two plots, as they are not needed. Consider moving the (a) (b) and (c) to inside the figures to allow the figures to be a tad larger.

Response: We thank the reviewer for these helpful suggestions to improve the clarity and layout of Figure 10. As recommended, we have: (1) added a legend within one of the subplots; (2) unified the x-axis limits across all three subplots; (3) removed the x-axis labels and ticks from the top two subplots, as they share the same time axis as the bottom panel; and (4) moved the subplot labels (a), (b), and (c) inside the respective panels to allow for slightly larger plot areas.

L417: Consider OLD: providing limited warning time. NEW: reducing the effectiveness of the method for off-shore events and events in regions of minimal station coverage.

Response: We thank the reviewer for this insightful suggestion. As recommended, we have revised this sentence.

L419: Consider adding a PLUM reference here.

Response: We thank the reviewer for this helpful suggestion. As recommended, we have added a reference to the original PLUM method (Kodera et al., 2018) at the appropriate location in the manuscript.

L427: No additional computational delay – important finding, add to your abstract? Although on line 536 you say it is 'negligible' – which is it? Add some quantitative numbers here and in the abstract.

Response: We appreciate the reviewer's attention to this point. The phrases "no additional computational delay" and "negligible processing time" both refer to the inference speed of the CONIP model. As noted in our earlier response (and now clarified in the revised manuscript), benchmarking tests show that a single CONIP prediction takes approximately less than 10 milliseconds on standard CPU hardware, which is negligible compared to the timescales of seismic wave propagation (seconds) and does not introduce meaningful latency in real-time EEW workflows. That said, we respectfully note that we do not consider computational efficiency to be a central contribution of this work, our primary focus is on improving the accuracy and stability of intensity estimation through hybrid modeling. Consequently, we have chosen not to include quantitative timing details in the abstract, as they are secondary to the main scientific findings.

L427: I was told by my PhD advisor never to 'believe' in a paper. Consider rewording to something like 'Our study shows …'

Response: We thank the reviewer for this advice. As suggested, we have revised the content of the manuscript in accordance with the reviewers' comments.

L451: I'm not sure 'significant earthquakes' is the correct term to use here – consider 'large magnitude earthquakes with long durations.'

Response: We thank the reviewer for this helpful clarification. As suggested, we have replaced the phrase "significant earthquakes" with "large-magnitude earthquakes with long durations" to more precisely describe the event characteristics relevant to our analysis.

L452: Consider changing 'substantial' to 'large'

Response: Thank you for your suggestions. We have revised this term in accordance with your suggestion.

L465: I don't think you need the text 'in this study'

Response: Thank you for your comment. We removed the text from the manuscript.

L468: OLD: 'of various events' NEW: 'of a range of magnitudes and depths'

Response: We thank the reviewer for this helpful suggestion. As recommended, we have revised the phrase from "of various events" to "of a range of magnitudes and depths" to provide a more precise and technically meaningful description.

L472: Did you actually try to run the algorithm on data from California and Western China? Make it clearer if this was done or not.

Response: We thank the reviewer for this important clarification request. Yes, we did conduct tests using historical data from western China (Sichuan and Yunnan) to evaluate the transferability of our CONIP model. However, due to the limited availability and uneven spatial distribution of high-quality seismic records in that region, compared to the dense and homogeneous KiK-net network, the model didn't perform as well as shown here. As suggested, we have revised the relevant sentence in the manuscript to clarify that these experiments were actually performed and to better explain the reasons for the reduced accuracy.

L480: You mention 'insufficient historical records' -- can you add here how many records are optimal? Are we talking 10 or 100?

Response: We thank the reviewer for this insightful question. As the reviewer rightly notes, the required number of records depends on model complexity and data diversity. In our experience with deep learning-based ground motion prediction models similar to CONIP, tens of thousands of high-quality, well-distributed strong-motion records are typically needed to achieve stable and generalizable performance. While there is no universal threshold, datasets with fewer than ~10,000 records often lead to overfitting or poor spatial generalization, especially for regression tasks involving continuous intensity estimation. We have added a brief note to this effect in the revised manuscript to provide context for the term "insufficient historical records."

L486-493: Nice!

Response: Thank you for your comment.

L499-509: On a similar topic, you might mention the work of Jessie Saunders who is exploring how to build a damped version of PLUM: Saunders, J. K., Cochran, E. S., Bunn, J. J., Baltay, A. S., Minson, S. E., & O'Rourke, C. T.. Incorporating intensity distance attenuation into PLUM ground-motion-based earthquake early warning in the United States: The APPLES configuration. Earth's Future, 12(2), e2023EF004126.

Response: We thank the reviewer for this insightful question. As the reviewer rightly notes, the required number of records depends on model complexity and data diversity. In our experience with deep learning, based ground motion prediction models similar to CONIP, tens of thousands of high-quality, well-distributed records are typically needed to achieve stable and generalizable performance. While there is no universal threshold, datasets with fewer than ~10,000 records often lead to overfitting or poor spatial generalization—especially for regression tasks involving continuous intensity estimation. We have added a brief note to this effect in the revised manuscript to provide context for the term "insufficient historical records."

L510: I think this is implied and this text is not needed "In addition to the topics discussed earlier.

Response: We thank the reviewer for this suggestion. As recommended, we have removed the phrase "In addition to the topics discussed earlier".

L520-522: I think some of this text is implied. Consider: OLD: As a subsequent of this study, a collection of aftershock sequences has been acquired for further validation of the proposed method. Analysis of these results will be addressed in forthcoming publications.
NEW: A collection of aftershock sequences has been acquired to further test this hybrid method in a future study.

Response: We thank the reviewer's suggestions for improvement. we have revised this section in accordance with your suggestions.

L530: If possible, be more specific in what you mean by 'prediction accuracy'. Accuracy in spatial distribution, intensity amplitude etc.

Response: We thank the reviewer for this valuable suggestion. As recommended, we have revised the manuscript to specify that "prediction accuracy" refers explicitly to the fidelity of predicted JMA seismic intensity both in amplitude and spatial distribution.

L532: Expand on this “extended lead time” what is the range of additional time that can be gained? Just 1 second or 5-10 seconds. One second would not really matter, but 5-10 seconds would. Add the range and state if that gain would be beneficial to the system.

Response: We thank the reviewer for this important comment. As noted in our earlier response, we partially agree with the point that the practical value of extended lead time depends critically on its magnitude. Based on our tests, our method provides an additional 1-5 seconds of warning time within the first 10 seconds after the origin time, compared to conventional PLUM approaches. This gain varies depending on factors such as epicentral distance, local station density, and wave propagation geometry. While gains at the lower end (~1 s) offer limited operational benefit, extensions of 3-5 seconds, achieved in many near-field scenarios with adequate station coverage, can be meaningful for automated systems (e.g., stopping elevators, opening firehouse doors, or initiating industrial shutdowns). We have revised the manuscript to include this range.

L536: Can you ne more quantitative here ‘negligible processing delays’ – how many seconds?

Response: We thank the reviewer for this helpful request for clarification. As noted in our earlier response, CONIP is a lightweight model, and our benchmarking tests show that the average processing time for a single prediction is typically within a few milliseconds (e.g., 2-5 ms on standard CPU hardware). This delay is negligible compared to the timescales of seismic wave propagation (seconds), and thus does not meaningfully affect the overall warning lead time. In the revised manuscript, we have replaced the phrase “negligible processing delays” with “processing delays of only a few milliseconds” to provide a more quantitative description.

L523: Within the conclusions section also mention that the training was done on data from the region, and that applying this training to other regions could be problematic.

Response: We fully agree with the reviewer’s important point. As suggested, we have added a concise statement in the Conclusions section.

L555: I think the GMT folks suggest that you include this reference:
Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The Generic Mapping Tools Version 6. Geochemistry, Geophysics, Geosystems, 20(11), 5556–5564.https://doi.org/10.1029/2019GC008515

Response: Thank you for your comment.

L570-585: Consider adding who wrote the proposal for the work.

Response: We thank the reviewer for this suggestion. As recommended, we have added a brief acknowledgment in the revised manuscript indicating that the research was based on a proposal led by the corresponding author Prof. Liu Qifang, as appropriate for transparency and attribution.

L594-595: Oops! Need to update!

Response: We sincerely thank you for your careful reading and valuable feedback. We apologize for the oversight in our initial submission, this was due to insufficient proofreading prior to submission, and we regret any confusion it may have caused. As suggested, we have removed the problematic text in the revised manuscript.

6

References:

You are missing these references in the reference list, that are referred to in the text. There might be others missing, please carefully check.

Kilb, D., Bunn, J. J., Saunders, J. K., Cochran, E. S., Minson, S. E., Baltay, A., ... & Kodera, Y. (2021). The PLUM earthquake early warning algorithm: A retrospective case study of West Coast, USA, data. Journal of Geophysical Research: Solid Earth, 126(7), e2020JB021053.

Minson, S. E., Saunders, J. K., Bunn, J. J., Cochran, E. S., Baltay, A. S., Kilb, D. L., ... & Kodera, Y. (2020). Real-time performance of the PLUM earthquake early warning method during the 2019 M 6.4 and 7.1 Ridgecrest, California, earthquakes. Bulletin of the Seismological Society of America, 110(4), 1887-1903.

And you might consider adding these:

Clements, T., Cochran, E. S., Baltay, A., Minson, S. E., & Yoon, C. E. (2024). GRAPES: earthquake early warning by passing seismic vectors through the grapevine. Geophysical Research Letters, 51(9), e2023GL107389.

Morton, E. A., Koper, K. D., Kilb, D. L., & Burlacu, R. (2026). Exploring the Feasibility of a Regional Earthquake Early Warning System in Utah. Bulletin of the Seismological Society of America, 116(1), 249-271.

Saunders, J. K., Minson, S. E., Baltay, A. S., Bunn, J. J., Cochran, E. S., Kilb, D. L., ... & Kodera, Y. (2022). Real-time earthquake detection and alerting behavior of PLUM ground-motion-based early warning in the United States. Bulletin of the Seismological Society of America, 112(5), 2668-2688.

Saunders, J. K., Cochran, E. S., Bunn, J. J., Baltay, A. S., Minson, S. E., & O'Rourke, C. T. (2024) . Incorporating intensity distance attenuation into PLUM ground-motion-based earthquake early warning in the United States: The APPLES configuration. Earth's Future, 12(2), e2023EF004126.

Saunders, J. K., Cochran, E. S., & Bunn, J. J. (2026). Refinements to the attenuated propagation of local earthquake shaking (APPLES) ground-motion-based earthquake early warning algorithm. Geophysical Journal International, 244(2), ggaf418.

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The Generic Mapping Tools Version 6. Geochemistry, Geophysics, Geosystems, 20(11), 5556–5564.https://doi.org/10.1029/2019GC008515

Response: We sincerely thank the reviewer for suggesting these relevant references. We have incorporated them into the reference list of the revised manuscript.

Dear Editor and Reviewers,

We sincerely thank you for your careful reading of our manuscript and for your insightful comments and suggestions. We greatly appreciate the time and effort you have dedicated to evaluating our work. In response to your feedback, we have carefully revised the relevant sections of the manuscript to address all points raised. Below, we provide a point-by-point response to each comment, along with descriptions of the corresponding changes made in the revised manuscript.

Reviewer A:

This is a review of the revised draft of the manuscript "Hybrid Approach of Real-time Ground Motion Prediction 1 for Earthquake Early Warning" by Hongcai et al. The authors provided a detailed response file and updated significant portions of the text and figures of their submitted manuscript. I appreciate the attention they gave to each of the reviewer comments. Overall, the modifications are sufficient for me to recommend publication with a few minor changes that I will list below. The most important revision is the description/label of Figure 11b. If the change to Figure 11b is made and confirmed by the editor, I do not need to see the manuscript again.

Line 19: I recommend removing ', not source or location parameters,' as it is irrelevant to the statement.

Response: Thank you for your helpful suggestion. We have removed the phrase as recommended.

Line 40: Change 'online-operation' to 'operational' or real-time operational'

Response: Thank you for the suggestion. We have revised Line 40 by replacing “online-operation” with “operational” as recommended.

Line 52: 'feasibility' is not quite the right word here, perhaps 'accuracy'.

Response: Thank you for your insightful comment. We have replaced “feasibility” with “accuracy” on Line 52 as suggested.

Line 117: 'to 7.1..', remove extra period

Response: Thank you for your careful reading. We have removed the extra period on Line 117 as noted.

Line 146: 'a densely observation network' to 'a dense observation network'

Response: Thank you for your suggestion. We have revised Line 146 by changing “a densely observation network” to “a dense observation network” as recommended.

Line 288, 308, multiple lines: Something went wrong with the figure numbers - some show up as 34 instead of 4, or 15 instead of 5. Please check all.

Response: Thank you for pointing out these errors. During the previous revision, we renumbered several figures, and unfortunately, some references, such as those on Lines 288 and 308, were inadvertently overlooked. In this revision, we have carefully checked all figure citations throughout the manuscript to ensure consistency and prevent similar oversights.

Figure 11: I don't believe the label and description of Figure 11b is correct. It is described as either 'Lead time' (figure axis) or 'difference in lead time' (caption). I do not believe it is either. It is the difference in final intensity prediction time between PLUM+CONIP and PLUM. Lead time is generally used to refer to the time between when an alert/prediction is issued and some level of shaking (or the S-wave arrival) arrives at a location, but that is not what is being shown with these plots. It should be fine to update the labels/caption, just make sure any discussion about this figure or these results are correct in other locations in the manuscript (I confirmed the description given in the abstract is fine as is).

Response: Thank you for your insightful comment. You are absolutely correct, Figure 11(b) illustrates the advance time (the difference in final intensity prediction time) achieved by the hybrid method compared to PLUM alone when determining the final intensity at a single station. In this revision, we have updated the axis label, caption, and any related text in the manuscript accordingly to accurately reflect this meaning and avoid misuse of the term “lead time.”

Reviewer B:

I commend the authors for carefully updating the manuscript based on the reviewers’ requests. They did an excellent job addressing the concerns of the reviewers and the manuscript is much improved. These updates obviously took considerable effort, and the effort really paid off. Thank you for your hard work and good job!

I did find one mismatch between what they said they updated and the revised manuscript:

1) In the response to reviews they state:

“The updated Figure 6 now includes binned averages with error bars and a trend line, and the text explicitly states that while low-uncertainty predictions tend to be more accurate, the relationship is probabilistic rather than deterministic." But in the revised manuscript Figure 6 does not include binned averages, error bars or a trend line. I think the stated updates in their response-to-reviews would improve Figure 6 and suggest they explore updating this figure accordingly. It might be, however, that these additions reveal something new, and, if so, the authors should update the manuscript accordingly (i.e., I expect these updates will be helpful, but if they are not, a different approach might be merited).

Response: Thank you for pointing this out. We sincerely apologize for the oversight in the previous revision, Figure 6 was inadvertently omitted from the updated manuscript despite our intention to include the binned averages, error bars, and trend line as described in our response. In this revision, we have now properly updated Figure 6 with these elements and adjusted the accompanying text accordingly to ensure consistency and clarity.

2) The authors might also consider adding grid lines to Figures 2 and 7.

Response: Thank you for your valuable suggestion. We have now added grid lines to both Figures 2 and 7 as recommended, which we believe enhances the readability of these plots.