Response to Reviewers

Dear Reviewers,

We appreciate the time and effort that you have dedicated and are grateful for your insightful comments which have improved the manuscript. We have incorporated the constructive suggestions, and have highlighted the changes within the manuscript and marked them in blue color. Here is a point-by-point response to your comments and concerns.

Reviewer 1

The authors have addressed most of my comments in the revision. However, I am still concerned about the model's performance.

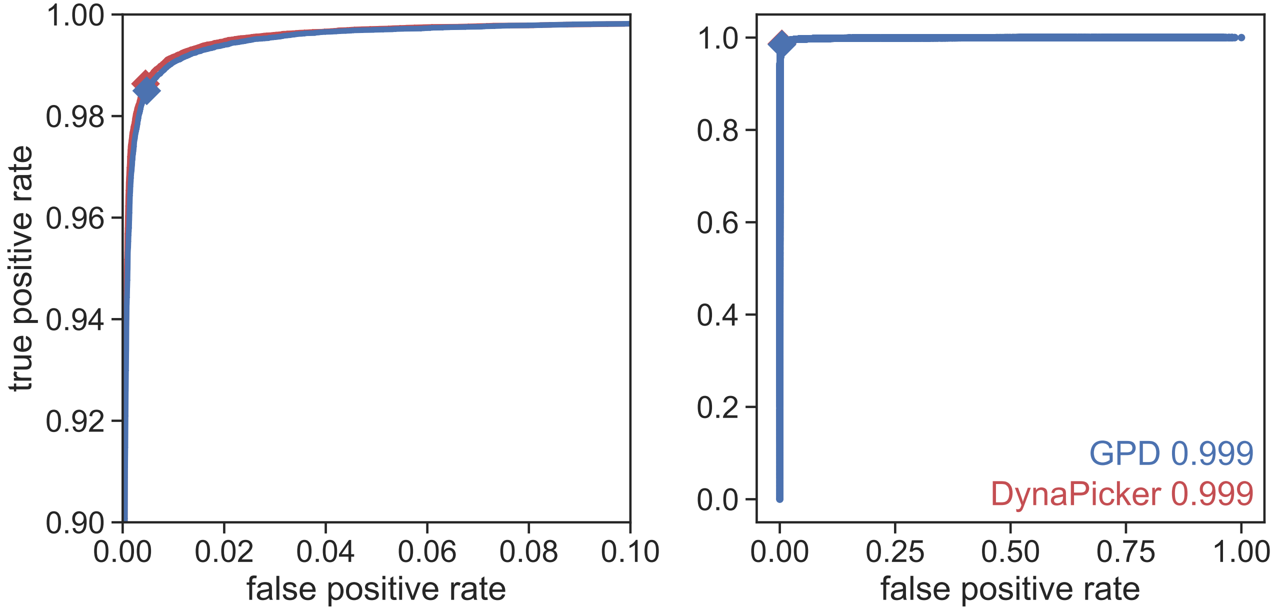
I am unclear why the authors do not follow Figure 1 from "Which Picker Fits My Data?" to compare the model performance. The plot of the ROC curve is a very clear way to compare different models. My concern is that the model performance is actually worse.

The figure provided in the response file is also not very clear. Which dataset do the authors use to produce the plot? Please also plot the curves of models like GPD, PhaseNet, EqT, etc. in the same plot. If all models have a high true positive rate on your dataset, please zoom in on the range between like 0.9-1 for a clear comparison.

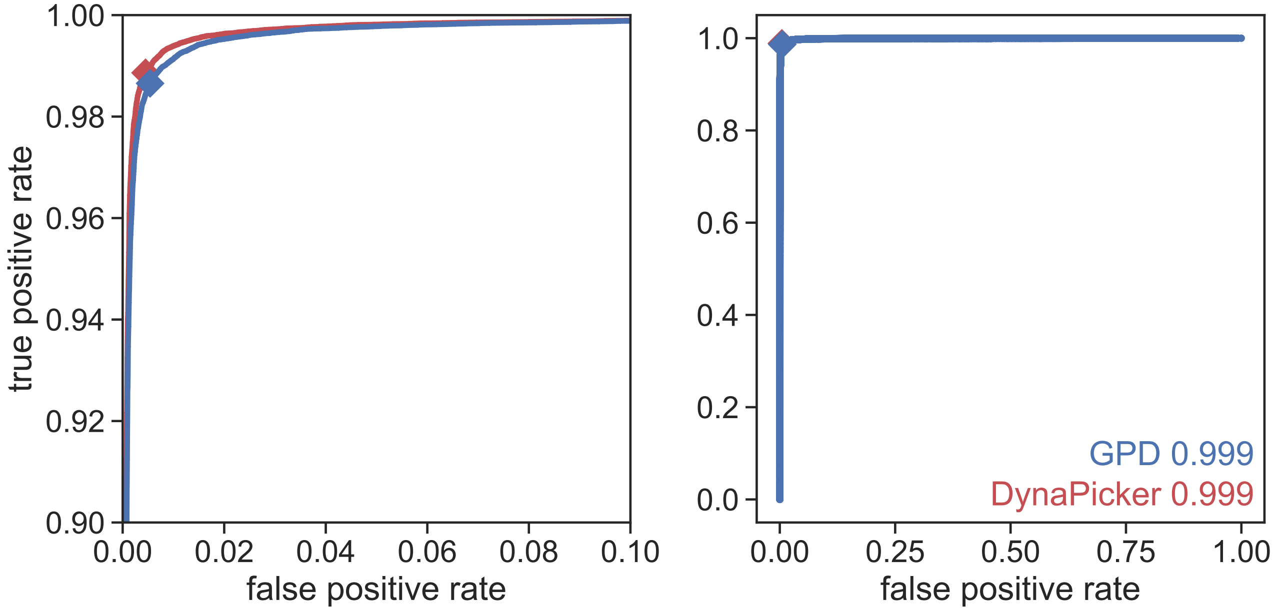
Response: We thank the reviewer for the constructive suggestion. In accordance with the Figure 1 presented in Münchmeyer et al.'s study titled "Which Picker Fits My Data?", we have generated receiver operating characteristic (ROC) curves for the P/S phases of the DynaPicker and GPD models using the SCEDC dataset, as depicted in Figure 1. This figure includes both a zoomed-in view within the range of 0.9-1 and the complete ROC curves. From Figure 1, we can observe that in comparison to GPD, the DynaPicker exhibits (a) a high true positive rate and (b) a slightly superior F1-score for P/S phase identification. Additionally, it is noteworthy that both the PhaseNET and EqT models are trained on the data with known ground truth arrival time for the P/S phases. However, since the SCEDC dataset lacks ground truth phase arrival times, we refrain from presenting ROC curves for these models in this context.

Reviewer 2

The manuscript proposes a new deep-learning picker that leverages dynamic convolutional neural networks for detecting and picking seismic phases from windowed or continuous waveform data. The authors then combined the previously published



(a) ROC curve of P-phase.



(b) ROC curve of S-phase.

Figure 1. Receiver operating characteristics for detection results from in-domain experiments for P-phase. The right subfigure allows us to assess the full curves of the performance evaluated on the SCEDC dataset. The left subfig shows zoomed-in parts of the upper left corner, with the zoom level to allow distinguishing the different models. Models were selected to maximize the Area Under Curve (AUC) score. Numbers in the corners indicate the test AUC scores. Markers indicate the point with the configuration associated with the highest F1 score.

CREIME model for magnitude estimations of waveform windows that have high P-wave probabilities. The authors have evaluated the performance of their picker and their combined workflow on open-source seismic datasets and aftershocks following the Turkey earthquake. The technical part of the manuscript is overall solid. The authors also corrected some previously raised issues. I think this manuscript is worthy of documentation at SE. I have only minor concerns regarding the motivation behind the proposed method and its application to the Kahramanmaras aftershock sequence. Below are my detailed comments:

The authors stated that ‘most of the prevalent CNN-based models perform inference using static convolution kernels, which may limit their representation power, efficiency, and ability for interpretation.’ However, the reported improvements are not that significant (from 98% to 99%). Can the authors elaborate on how these limitations and differences would affect the application of these models?

Response: In the context of phase classification, the DynaPicker model did not show a superior performance improvement. However, its application to the task of phase arrival time picking yielded noteworthy results. Through extensive testing on the continuous data sourced from multiple datasets, DynaPicker demonstrated an enhanced ability to detect earthquake events, showcasing robust and reliable performance in this specific application. Additionally, the DynaPicker performs better in phase classification when confronted with noisy data (as shown in the plots including the previous response).

Hence, this capability is indicative of the DynaPicker's resilience in capturing intricate features and patterns. In contrast, models relying on static convolution kernels exhibit limitations and differences that hinder their ability to capture the nuanced delineation of features and patterns.

2. The manuscript covers the Kahramanmaras aftershock too briefly. The entire article only has Section 5.5 on this topic. The authors may consider weakening the emphasis on the application of the Kahramanmaras aftershock in the title, as this part only occupies a small portion of the paper. Moreover, it lacks the phase association and event location using P and S phases from multiple seismic stations. Since the focus is on earthquake monitoring, earthquake location information is crucial. Besides, Section 6.2, it's not 'the live data of the Turkey earthquake'. You are using downloaded continuous waveforms, which is different from the live data stream.

Response: We express our gratitude for the reviewer's valuable suggestions.

- The application of the Kahramanmaras aftershock constitutes a minor portion of the paper, we have consequently revised the manuscript title.
- The significance of phase association and event location utilizing P and S phases from multiple seismic stations play a key role in the practical application of earthquake monitoring. We intend to dip into these aspects in our upcoming work.
- Within the manuscript, we have substituted the term "the live data of the Turkey earthquake" with "the continuous waveforms of the Turkey earthquake" in Section 6.2.