

Comments on “Earthquake Catalog and Continuous Waveforms from a Two-Week Distributed Acoustic Sensing experiment on Kefalonia Island, Greece” by Bocchini et al.

This paper analyses a period of high seismic activity in the Kefalonia region (Greece) during two weeks in 2024. The authors combine data from seismic stations and a 15 km-long telecommunications cable using DAS. They construct an enhanced earthquake catalogue with a factor of ~ 37 increase compared to the NOA catalogue. The DAS dataset is also made available, which is valuable for testing and benchmarking new methods.

The authors have done a large amount of work, and the paper addresses an interesting and relevant topic. In my opinion, the manuscript is suitable for publication in *Solid Earth* after the following comments are addressed.

General comment

I think the manuscript mixes different main goals, and this makes its overall focus unclear. In some parts, the paper is presented mainly as the release of a new DAS dataset for future benchmarking and method development. In other parts, the focus is on the methodological workflow, the integration of DAS and land seismic stations, and the construction and analysis of an enhanced earthquake catalogue. In my view, the methodological workflow and the integration of DAS and seismic station data are the central aspects of the paper, while the public release of the dataset is an added value.

I also think the manuscript goes too far in stating what future users may do with the dataset. It is useful to mention possible applications, but in several places the text moves from presenting the dataset to making broad claims about its future use.

Specific comments (in the order they appear in the manuscript)

- L15 and throughout the manuscript: please use distributed acoustic sensing in lowercase.
- L15: remove “on a telecommunications cable”, as this is repeated below.
- L22: the enhancement results in 2,780 events. I suggest also adding here the increase with respect to the official catalogue.
- L28-30: *“The aim is to provide a resource for researchers to test, develop, and benchmark DAS processing algorithms on tectonic earthquakes and to investigate the physical processes that drive complex seismic sequences”*. Please see my general comment above.
- L38: “..., and monitoring...”
- L40: “...large volumes of data that introduce...”
- L59: From *“The seismic sequence occurred...”* to *“...along with other natural events in the region”* does not read well.
- L58: *“Here we present a new dataset...”* and L62: *“In the following, we present a new earthquake catalog...”*, please see my general comment above.

- L68-69: *“The dataset released with this paper...”*, please see my general comment above. This idea is repeated again in L97, where it could be kept. Also, nothing is said here about the metadata being shared as well (for example, cable and channel coordinates).
- L84: The meaning of DAS is already given in the Introduction.
- L87-88: *“The DAS measurement period ranges from early July to late September 2024”*, but the study focuses only on two weeks (2-15 August 2024). Please explain why this time period was selected.
- L89: *“National Observatory of Athens (NOA)”*.
- Figure 1 caption: *“Grey triangles indicate public seismic stations from the HUSN not used in this study”*. Please explain why these stations were not used.
- L120-122: Please explain how the cable location refinement is performed, or provide a reference.
- L127: The sentence *“Automatic picking of P-wave arrivals along the fiber resulted from applying PhaseNet-DAS (Zhu et al., 2023)”* should be moved earlier (around L121), when P-wave arrivals are first introduced.
- L132-134: The sentence *“We observe that we can successfully record signals from teleseismic earthquakes, mostly on the offshore portion of the cable (Fig. S4).”* is distracting in this context and could be moved elsewhere.
- L134-137: *“Expected amplitude scaling with earthquake magnitude is also evident: for example, an amplitude ratio of ~230-240 between events of ML 0.6 and ML 3.0 located ~1.7 km apart (epicentral locations from NOA catalog) is consistent with the theoretical value of ~250 that is expected if the events were co-located (Fig. 2c).”*
 (i) It is not clear how the amplitude ratio of 230-240 is obtained from Fig. 2c. Please specify how amplitudes are measured (e.g., which channels or averaging procedure are used); (ii) Please clarify that the theoretical value (~250) is computed as $10^{(\Delta M)}$.
- L165: *“...; without additional...”*
- L176: *“...6,817 detections...”* This number is different from the one given in the abstract, because it includes false detections. Please make this distinction clear.
- L165-166: *“The detector is unable to distinguish between P- and S-waves, without additional processing this can lead to P- and S-waves from a given earthquake being classified as separate events.”* Please clarify what this additional processing is, or indicate clearly if it is described later in the manuscript.
- L192: Please clarify whether Ts-Tp is determined manually and how events are classified when both phases are not observable.
- L218: *“We compute the average...”*.
- L230-232: *“We note that we convert strain to strain-rates as one of the required pre-processing steps for PhaseNet-DAS.”* Please provide evidence or a reference for this statement, as this is not described as a required step in the original PhaseNet-DAS manuscript. The model is generally considered robust to input type. In addition, this statement seems inconsistent with L231-232 (*“We observe P and S arrivals of good quality despite PhaseNet-DAS being trained on raw strain rate data sampled at 100 Hz”*). If the model is robust, the need for this conversion should be clarified.

- L245-247: *“We revised the original velocity model by removing the second layer at 0.5 km depth ($V_p = 5.47$ km/s), maintaining lower velocities from 0 to 2 km depth (Table S1) to reflect the presence of unconsolidated sediments beneath the offshore cable segment.”* Please clarify how this modification is constrained. For example, is it based on borehole data, reflection/refraction studies, or other independent observations?
- L254-255: *“A total of 280 of the 284 earthquakes have semi-major error ellipse axis values and vertical errors < 5 km.”* This statement is not very informative, as Fig. S8 shows that most location errors are already well below this threshold.
- L265-267: *“We estimate local magnitude (ML) of located earthquakes using seismic station data following the empirical equation of Hutton and Boore (1987). We apply a bandpass filter to the waveforms in the band 2-30 Hz and calculate the maximum amplitude on the two horizontal components at each of the four seismic stations...”* Please clarify that the NOA catalogue computes local magnitudes using the Hutton and Boore (1987) relation. In addition, if this relation is used here, the amplitudes should be instrument-response corrected and converted to Wood-Anderson-equivalent amplitudes.
- L280-282: *“A comparison of earthquake magnitudes estimated here and common events in the revised NOA catalog shows a correlation coefficient of $R = 0.96$ and a near one-to-one relationship, suggesting a strong correlation (Fig. S9).”* I do not think the correlation coefficient is the most relevant metric here. A high R^2 value only shows that both magnitude estimates covary, but it does not demonstrate agreement. In this case, the important comparison is with the 1:1 line, and Fig. S9 shows some disagreement. Please quantify the residuals and discuss the deviations from the 1:1 relationship rather than emphasizing only the high correlation coefficient.
- L295-300: The reference to Fig. S10 here is confusing, because these figures refer to the CCC among template events, which are introduced only later (L314-319).
- L320-325: The choice of the CCC threshold of 0.52 needs a clearer justification. It is not obvious how this specific value is determined, and the precision of this threshold gives the impression of a well-defined criterion, while it seems to be based on an empirical relation. In addition, I do not think it is enough to say that other users can test different thresholds with the provided data. Since this threshold directly affects the results, its effect should be assessed within the manuscript itself.
- L335-342: The use of amplitude ratios to compute relative magnitudes implicitly assumes that absolute calibration is not required. While this approach is valid in principle, it relies on strong assumptions (similar path, radiation pattern, and frequency content), which may not hold for event pairs with non-negligible separation (CCC threshold = 0.52). This should be discussed more explicitly. In addition, the use of different frequency bands (5-20 Hz for DAS and 2-30 Hz for seismic stations) may affect amplitude ratios and should be justified.
- L345-347: the validation based on a high coefficient of determination ($R^2 = 0.92$) and a near 1:1 slope is not sufficient to demonstrate agreement. As discussed above, correlation reflects covariance but not accuracy. Please quantify the residuals between initial and relative ML estimates and discuss possible biases.

- L357: The reference to Fig. S11 here may be incorrect.
- L404-405: The interpretation of the strong amplification observed along the marine segment seems too limited. This behaviour may not only reflect shallow crustal structure, but also differences in DAS sensitivity to body and surface waves, mechanical coupling, cable construction, and installation conditions along the array, especially across the onshore-offshore transition.
- L406: Section 6.1 (Distorted waveforms). The description of waveform distortion is interesting. The authors note that identifying the exact cause is challenging and suggest cycle-skipping as a possible explanation. Phase unwrapping and the associated $\pm\pi$ ambiguity occur at the interrogator level, before any decimation. In this study, only decimated data (250 Hz) are available, which makes it difficult to verify directly whether true phase-cycle ambiguity occurred during acquisition. However, the observation that the distortion is largely reduced after bandpass filtering suggests that the observed step-like changes, while preserving the overall waveform shape, resemble a baseline shift and may therefore also be related to high-frequency processing artifacts or other amplitude-dependent instrumental effects. I think this possibility is worth examining more carefully.
- L423: Section 7 (Applications). I suggest renaming this section as Section 6.2 of the Discussion and adjusting the title (e.g., "*Potential applications of the dataset*"). In addition, in line with my general comment, I think some statements in this section are too strong. I suggest moderating these statements.