Dear Editor,

We wish to thank you and the reviewers for the useful comments and suggestions that helped us in improving the manuscript. We have carefully addressed each comment from both reviewers enhancing the quality of the manuscript, particularly in the discussion of results and the implications of ORION for future studies and applications. The revised manuscript includes an improved discussion section, clarifications in the introduction and methodology, and a more detailed description of the datasets, as the reviewers suggested.

Each answer and change is reported below and marked in bold in the annotated manuscript. The few deleted sentences from the original are shown with strikethrough. A clean version of the revised manuscript is also provided.

Thank you for considering our manuscript for publication in Seismica.

Kind regards,

Emanuele Bozzi & co-authors

Reviewer 1

1.1 Emphasis on near real-time. "The emphasis on the "near real-time" feature in the title does not seem fully aligned with the manuscript content. Most sections focus on the methodology and application examples, while relatively little discussion is devoted to computational performance or processing latency.... it would **strengthen the manuscript** to include **additional discussion on computational efficiency** and **integration within operational seismic monitoring timelines**. For example, **how** do the authors envision **ORION** to **fit within the workflow of a standard seismic monitoring center in terms of timeliness**?"

We thank the reviewer for this observation. We have added a paragraph in the Discussion clarifying how ORION can fit within an operational seismic monitoring workflow. We here report the paragraph:

NEW DISCUSSION PAR. 1 (lines 262-270 of the annotated manuscript)

ORION has the potential to be integrated into standard operational seismic monitoring workflows, given its two-stage structure. Specifically, the spatial clustering step is performed only once per deployment and therefore implies no additional computational cost after this operation. In contrast, the waveform-attribute–based channel selection requires typically less than two minutes per event on an average laptop. This performance places ORION, even *in its current form, within a near–real-time processing window compatible with routine post-trigger analysis in seismic monitoring centres. Furthermore, to reduce the computational cost and make it more suitable for near-real time applications, ORION could be converted into more efficient programming languages exploiting GPUs parallelization and installed at interrogator level. This would enable the selector to operate on triggers with computational times on the order of less than one minute.*

1.2 Beyond earthquake location? "From the title, ORION appears to be presented as a general framework for DAS channel selection. However, the manuscript primarily focuses on earthquake location performance. **It would be valuable to briefly discuss potential applications beyond location**, such as **earthquake mechanism determination** and other array seismology applications."

Also 1.3

Thanks for these valuable observations. We indeed focused on earthquake location to show a result of the application of the channel selector. Nevertheless, we are currently exploring the impact of ORION on other source parameters and on event magnitude. A tool that identifies the relative (to each section, to guarantee spatial and azimuthal coverage) best channels should have an impact (increasing accuracy) on source parameter estimation with DAS, especially for near real time estimations. Moreover, the idea behind spatial clustering can be easily transferred to e.g., beamforming, strengthening the individual section estimations. These two suggestions for the discussion have been added to the manuscript and surely help the discussion of its implications beyond what's reported in the manuscript.

NEW DISCUSSION PAR. 2. (lines 279-284 of the annotated manuscript)

Although we demonstrated the effects of the proposed channel selection strategy on event location for a seismic sequence, such an approach is expected to positively impact other source parameter estimations as well. Higher-quality traces should, in principle, enhance waveform-fitting–based methods for source parameter estimation (Dziewonski et al., 1981; Tarantola, 2005). Finally, clustering the cable into sections with similar geometrical properties may benefit array-processing techniques, such as beamforming, given their assumptions about local spatial coherence of the seismic wavefronts (Rost and Thomas, 2002).

1.3 Notes on EEW. "A **short discussion** on the **potential relevance of ORION for earthquake early warning** systems would be beneficial. Given the emphasis on "near real-time" operation, readers like me may naturally wonder whether the method could be adapted for real-time EEW workflows. Even a brief perspective on feasibility, challenges, or limitations in this context would strengthen the manuscript."

We thank the reviewer again for the important comments which help broaden the implications of the study and the discussion section. We note that in its current form, ORION's channel selection step, requiring one to two minutes per event, is too slow for strict EEW applications, where latency targets are typically on the order of seconds. However, the one-time nature of the spatial clustering step means that the geometric pre-processing is already complete before any event occurs. A "computationally reduced" version of ORION, relying only on pre-computed static quality metrics (e.g., long-term average SNR or pre-event RMS noise levels), that identifies "superchannels", rather than event-by-event waveform evaluation, could in principle operate within EEW-compatible timescales.

I here report the new paragraph regarding these comments in the Discussion section:

NEW DISCUSSION PAR. 3 (lines 284-290 of the annotated manuscript)

Moreover, some of the principles proposed here can be translated into a more computationally efficient version of the algorithm for earthquake early warning (EEW) applications or any other application requiring a latency of the order of seconds. Assuming that spatial clustering has already been performed, long-period channel attributes could be used to identify a subset (e.g., few tens of channels) of quasi-stable channels (Biondi et al., 2025), supported by a final selection based on current waveform attributes. As a result, the number of these channels would be comparable to the station density typically employed in operational EEW systems.

1.4 Customization and prior information. "Can **ORION incorporate customized attributes** defined by users? Researchers may want to add their own for different purposes."

See also 1.5

Yes, in principle every waveform/channel attribute can be used (if quantitative), as the core of the selection is a score method on spatial clusters. I added this idea in the Discussion section:

NEW DISCUSSION PAR. 4 (lines 255-261 of the annotated manuscript)

We propose a selection based on three representative waveform attributes: SNR, local waveform coherence, and pre-event noisiness of the channel. However the same framework is extendable, without altering the basic code architecture, to *any other quantifiable channel or waveform attribute. Given that the code is open source, users can tailor the attribute-based selection to their specific application, incorporating alternative or additional features that may be better suited to a given dataset. Moreover, prior information about the cable, such as proxies for ground coupling, can potentially contribute to the selection of the channels as stable or quasi-stable attributes.*

1.5 Customization and prior information. "Is there a **mechanism to integrate prior information about the cable**, such as a **'coupling index'** or other quality metrics?"

See also 1.4.

Yes, as a level-zero step, one can introduce a stable parameter, which contributes to the final scoring. We added a sentence regarding this opportunity in the Discussion section (please see NEW DISCUSSION PAR. 4).

Minor comments

1.6 — "Abstract and Line 58: Please provide the full name of ORION at its first mention."

Done. Thanks.

1.7 — "**Line 43**: Consider **adding references** to support the statement regarding 'common practice'."

I've added references to articles that subsample the cable at fixed arbitrary steps (e.g., every gauge length).

The updated sentence is at lines : 48-49 of the annotated manuscript

1.8 — "Line 61-67: As **this section lists the key features of ORION, more concise wording would improve clarity**. It may also be preferable to **avoid qualitative descriptors** such as 'robust' or 'high-quality' unless they are quantitatively supported."

We have revised lines 61–67 to replace qualitative descriptors with more precise and targeted language. Specifically, "robust to varying types of cable geometry" has been rephrased to "applicable to cable geometries ranging from rectilinear to highly curvilinear configurations" The phrase "high-quality channels" has been replaced with "quality metrics (SNR, local coherence, and pre-event noise levels)". Moreover we have added subtitles to each point, which sum up the key requirements that guided ORION's development.

We here report the revised paragraph (lines 65-74 of the annotated manuscript):

ORION is designed to meet the following requirements:

• Robusteness: i.e., applicable to cable geometries ranging from rectilinear to highly curvilinear configurations,

• Distribution: i.e., ensuring sufficient spatial and azimuthal coverage of selected channels along the cable

• Tunability: i.e., allowing users to adjust the relative weighting of waveform attributes and spatial coverage, enabling them to focus on quality metrics (signal-to-noise ratio, local coherence, and pre-event noise levels) or spatial/azimuthal coverage,

• Scalability: i.e., supporting the retrieval of a defined number of channels, e.g., for a successive integration of DAS data with existing conventional networks,

• Streaming: i.e., operating in near real-time to spatially reduce DAS data.

1.9 — "Line 105, 117-118: The **automatic clustering step** appears to be fundamental to the workflow. While the Supplement provides a helpful demonstration, the **main text** would **benefit from a more quantitative description of this step**. For example, **explicit definition of the time window, mathematical formulation of the clustering criteria, description of tunable parameters etc.**"

We agree that the description of the spatial clustering step needed an expansion. The revised section now includes: (a) the definition of the local azimuth window used to compute local cable orientation; (b) the mathematical criterion for cluster membership in the modified DBSCAN algorithm, and (c) the key tunable parameters.

NEW SPATIAL CLUSTERING SECTION (lines 126-144 of the annotated manuscript)

DAS arrays often exhibit strong azimuthal variations over short distances. Combining channels with different azimuths can degrade signal coherence during stacking due to incompatible axial sensitivities, and signals may undergo severe degradation near abrupt changes in fiber direction. We address these issues with a modified version of the DBSCAN algorithm (Ester et al., 1996) that enforces spatial contiguity while clustering channels considering the local azimuth as clustering metrics (Figure SUPPL2). The local azimuth at each channel is estimated over a sliding window of N_w contiguous channels, converted to vector components to handle the circular nature of azimuth. A *channel is added to an existing cluster if its Euclidean distance in the standardized feature space, comprising cumulative along-fiber distance and the two azimuth components, does not exceed a certain value ε (estimated automatically via the k-distance heuristic, Satopää et al., 2011). The method requires a minimum of N_min contiguous channels to form a valid cluster and excludes channels with abrupt azimuthal changes. When few unclustered channels occur inside an otherwise quasi-rectilinear section, a new cluster is initiated to maintain spatial continuity. This design permits smooth azimuthal variations within a cluster, which are less problematic for stacking than sharp changes. To prevent excessively long rectilinear or gently curved sections from dominating the distribution of selected channels, clusters exceeding a maximum number of channels L_max are further subdivided into shorter segments. This procedure ensures balanced spatial coverage, especially for rectilinear cables or borehole deployments.*

1.10 — "Line 205-208: It would be great to **add more information**, for example, the **sampling rate of the data tested**, and the **volume of the data processed** each time for reference."

See also 2.4

We've added two supplementary figures that summarize the principal features of each dataset (Figures Suppl6 and Suppl7). We interpreted the "volume of data processed" as the original number of channels of each deployment and thus we also provide the reduction factor obtained with ORION.

1.11 — "Figure 3: **Are all plotted waveforms auto-selected, or only those marked with triangles?** The **legend entry** 'auto-selected channels on DAS data' (yellow) is slightly ambiguous. Consider adding a **separate label** for manually selected channels if applicable."

All plotted waveforms are auto-selected by ORION. The triangles mark a further manual subset chosen for display purposes, illustrating that the reduced channel count produced by ORION is small enough to allow visual inspection and subjective evaluation. We clarified the figure caption and legend accordingly (Caption of Figure 3 in the annotated manuscript).

(a-f) A further manual subset of selected channels (marked with grey triangles) is highlighted below each panel to illustrate that the reduced channel count provided by ORION allows straightforward visual inspection.

1.12 — "**Figure 4**: **There appears to be a gap in selection frequency in the middle of the marine section**. Given that the attribute scores appear high and the geometry stable, could the authors clarify the reason for this?"

This occurs because the channel selection is currently performed within the 10th–90th percentile range of the quality scores. This is a preventive measure to avoid outliers, such as locally coherent noisy transients. While this approach may exclude channels with the highest overall scores (i.e., very-good), we have found it to be a good compromise in all the previous tests we conducted. Since the goal is to develop a tool that provides good-quality channels distributed along the entire cable length, not the overall best channels, this currently represents the best compromise we have identified. This is already mentioned section 3.2., but I clarified it and further mentioned in the figure:

Section 3.2 (lines 159-162 in the annotated manuscript):

Finally, we prevent selecting localized and energetic coherent noise bursts, possibly misclassified as high-quality channels, by discarding quality scores outside the 10th–90th percentile range. Although this may omit some of the highest-scoring channels, it provides a more robust selection, especially for near-real-time applications.

Figure 4 caption in the annotated manuscript:

Notably, the slightly lower frequency of selected channels in the middle of the marine section reflects the overall highest SNR, coherence, and RMS noise scores in that region, which are often excluded by the 10th–90th percentile thresholds used in ORION to ensure robustness to noise bursts.

1.13 — "Additionally, **please consider adding a description of the 'light blue' color** directly in the figure."

Done. I've put the "marine section" label for clarification.

1.14 — "Section 5.3: It would be interesting to see **future evaluations** of **integrated network** performance using **selected DAS channels**. A brief **forward-looking statement** could further strengthen this section."

That's part of a project that is being finalized these months and we'll be part of a separate manuscript, which will implement ORION-selected channels into integrated DAS + seismometer event location. Nevertheless, I inserted a sentence in section 5.3 on this topic:

NEW 5.3 PAR. 5 (lines 228-230 of the annotated manuscript)

The selected channels are furthermore well-suited for integration with data from conventional seismometers, an approach currently being developed within an ongoing project aimed at joint DAS–seismometer earthquake location.

****

Reviewer 2

2.1 — "1. Motivation: The **authors** could **consider broadening the motivation for spatial downsampling** by highlighting the **unique challenges** associated with **DAS deployments in remote and harsh environments**, such as remote volcanoes, glaciers, and future planetary applications, for example, on the Moon or Mars. In these settings, **severely limited data transmission bandwidth** makes **in situ channel selectio**n not merely a practical convenience, but an **operational necessity** for project or mission success."

We thank the reviewer for the interesting points which have been used to broaden the introduction section.

I here report the introduction paragraph (lines 40-44 of the annotated manuscript):

This is further reinforced in remote and harsh environments, such as volcanic flanks (Nishimura et al., 2021), glaciers (Klaasen et al., 2023; Walter et al., 2020; Hudson et al., 2021), or prospective planetary deployments (Zhai et al., 2024), where limited data transmission bandwidth makes in-situ channel selection not just a practical convenience but an operational necessity.

2.2 — "2. Storage strategy: The manuscript explains that spatial clustering is performed once, while channel selection is waveform dependent and therefore event specific. Could the authors **clarify how this is envisioned in practice**? Is **ORION** intended to **permanently reduce stored data volume** by discarding non-selected channels, or is it **primarily a real-time analysis tool** while the full dataset remains archived? Since DAS provides high-resolution spatial information, **understanding whether the non-selected data are preserved or removed would help clarify** the operational strategy."

ORION, in our view, is intended to support rapid analysis with DAS, without recurring to a blind selection based exclusively on channel quality (possibly lacking azimuthal coverage) or fixed interval selection. A possible implementation can consider selecting channels for the near real time data analysis and keeping the full dataset for successive analysis. Right now ORION outputs an .mseed file ready to use for seismological analysis containing the selected channels, thus not the entire original dataset. The user is kept to decide whether to keep the original data or just the selected channels. We clarified this point at the beginning of the methodology section:

NEW METHODS PAR. 1 (lines 108-113 of the annotated manuscript)

Selected channels are intended to support near real-time analysis without recurring to selection based exclusively on channel quality (possibly lacking azimuthal coverage) or fixed-interval decimation. In this context, non-selected channels are not discarded and can be preserved for successive offline analysis, supported by algorithms that perform efficient low-loss compression for long-term experiments (Seguí et al., 2025).

2.3 — "**Many DAS applications**, such as ambient noise interferometry or structural health monitoring, **rely on continuous data** rather than discrete events. How would the **authors recommend** applying **ORION** to these **'event-less' studies**? Would a **static selection based on long-term average noise levels be appropriate**, or would the **downsampling significantly degrade** the **spatial resolution required for noise correlation?**"

See also 1.3 and 1.4.

We thank the reviewer for raising this interesting point. ORION currently operates on detected events, but in principle can also be applied to continuous data. Extending the framework to periods without detections is an interesting direction for future development; in such cases, waveform statistics computed over a rolling window (e.g., the preceding 24 hours) could replace per-event scoring to guide channel selection based on background noise content. Regarding noise correlation, which is another potential application of DAS technology, we believe that spatial downsampling as performed by ORION, retaining good-quality channels with adequate azimuthal coverage, could support more efficient noise correlation analyses without sacrificing significant spatial resolution. Moreover, since the user can tune the minimum number of channels per cluster and the maximum subsection length, the spatial resolution of the selected subset can be adjusted to meet the requirements of a given application.

A new clarifying section has been added to the Discussion section to describe the possible adaptation of ORION to continuous recordings:

NEW DISCUSSION PAR. 6 (lines 292-295 of the annotated manuscript)

ORION currently operates on detected events, but can in principle be applied to continuous data. For instance, waveform statistics computed over a rolling window (e.g., the preceding 24 hours or longer) could replace per-event scoring to guide channel selection. If a fixed channel geometry is required, noise statistics from a pre-operative period could instead be used to define a static set of representative channels for each spatial cluster.

2.4 — "**Given the diverse environments tested**, including submarine, volcanic, and urban settings, were the relative contributions of SNR, coherence, and pre-event RMS similar across sites, or **did they require case-specific tuning**? Adding a **brief table or comment** on the stability of these parameters would be helpful for future users."

Thanks for highlighting this point. The relative contribution of SNR, coherence and pre-event RMS is stable across the different tests, with the exception of the SAFE example, where we tested a different combination for an individual event (1-1-0.3). Overall, when a priori information on specific requirements (e.g., SNR and coherence are a better proxy for channel quality, compared to noise RMS) of the datasets is lacking, we suggest keeping the relative contribution at 1-1-1, to avoid unnecessarily downweighting one waveform attribute. We agree that a table that summarizes the parameters adopted in the tests and application to real data is missing. Therefore, we have added Figure 6 and 7 in the supplementary file, which also contains the principal features of the data.

2.5 — "Regarding **Supplementary Figure 5a**, channels **1783-1923 appear to have very low SNR**. While the manuscript mentions that suboptimal channels are sometimes included to maintain azimuthal coverage, these specific traces look almost purely noise-dominated. **Could the authors explain if there is a 'minimum quality floor' below which a channel is rejected** even if it is needed for geometry?"

In the kefalonia application, we had the necessity to cover the marine section with lots of subsections, given its gentle geometry and thus the presence of a very long single spatial cluster, compared to the complex (lots of clusters) onland sections. As a consequence, sometimes, lower quality channels are selected, forced by the necessity to cover as much as possible to cable, even when that section is overall noisy. Moreover, in the Kefalonia example, we kept the final percentile of acceptance very-low, at 1st percentile. Therefore, we have privileged (in relative speaking) spatial coverage over absolute quality of the waverforms. This necessity is further supported by the event location test, given the fact that the cable is short and the seismicity is outside the network. We adopt this cautious approach to avoid discarding data that, although noisy, may still carry useful information. Nevertheless, we demonstrated that the overall quality of the waveforms is better than a conventional fixed interval subsampling (e. g., Fig Suppl5).