

Reply to Rev1

All responses to the reviewers' comments, as well as the corresponding updates introduced in the manuscript, are highlighted in red to facilitate quick and easy tracking

This study presents the work carried out by deep-sea benthic observatories located at key sites in the Mediterranean region, including data acquisition, quality assurance and quality checks, and data banking. These autonomous or cabled multidisciplinary systems are equipped with physical sensors, such as current meters and CTDs, as well as biogeochemical sensors, primarily optical turbidity sensors. The methodology involved applying and standardising rigorous quality assurance and quality control protocols for the different types of instruments, as well as filling data gaps. These datasets help remedy the critical shortage of deep-ocean observations below 2,000 metres and improve assessments of high- and low-frequency variability in the thermohaline and hydrodynamic conditions of these sites. The authors present some results, including the detection of deep-water warming trends at all sites, and a notable change in the direction of dominant currents in the Ionian Sea over a decade.

Overall, the manuscript is well presented and illustrated. The datasets are easily accessible via the provided links. However, I have a few comments and suggestions.

We thank the reviewer for the positive evaluation of our manuscript and for the accurate summary of the study. We are also grateful for the reviewer's comments and suggestions provided below, which have helped us improve the clarity and quality of the manuscript. Our detailed responses are reported point by point

In the introduction, the authors rightly point out that long-term observations in the deep regions of the Mediterranean Sea are rare. They should also mention the various types of operational deep observatories that complement their own. Notably, these include the HYDROCHANGES network (Schroeder et al., 2013; <https://ciesm.org/our-science/hydrochanges/>), which has continuously monitored hydrological and hydrodynamic variability at various locations in the Mediterranean since 2002. This includes deep sites in the Adriatic and northwestern basins, as well as key locations such as the Corsican, Balearic, Sicilian and Gibraltar straits. Further results concerning the long-term evolution of hydrological conditions in the deep basins of the Mediterranean can be found in Chiggiato et al. (2023). It would be interesting to include a perspective on the convergence of measurements and the complementarity of the different datasets in the introduction or conclusion.

We thank the reviewer for this suggestion. The Introduction has been revised accordingly [lines 58–63 and 67–68] to improve clarity and provide additional context, and relevant references have been added.

Regarding the post-processing of current meter data, I was interested in the results in Figure 3, which compare current measurements obtained by an ADCP and a punctual current meter. According to Table 1, the acoustic punctual current meter (Falmouth 3D-

ACM) operates at 2 Hz, while the ADCP (RDI WH 300 kHz) operates at 2.8×10^{-4} Hz (1 sample/hour), and the comparison is made using hourly data. The hourly average of 7200 measurements from the punctual current meter is then compared to the hourly measurement from the ADCP. Is the hourly ADCP measurement the result of a single ping, given that the standard deviation for a single ping is at least 2 cm/s? Or is it the result of averaging an ensemble of pings, in which case the standard deviation would be lower? Information on the acquisition modes specific to certain instruments and the standard deviation associated with the measurements for each instrument is important for making comparisons. Additionally, it would be preferable for this figure to plot the 1:1 line rather than the least-square regression line.

The Falmouth 3D-ACM current meter recorded velocities at 2 Hz. For comparison with the ADCP record, the time series was smoothed using a 1-hour moving average and sampled at hourly intervals to obtain an hourly estimate of the flow. The RDI Workhorse 300 kHz ADCP was configured with 100 pings per ensemble and a ping interval of 6 s, resulting in ensemble durations of approximately 10 min. Each reported velocity profile therefore represents an internal average of multiple pings, and the hourly ADCP measurements used in this study correspond to these ensemble-averaged profiles. In accordance with reviewer's suggestion, Figure 3 has been revised to display the 1:1 line as well as the least-squares regression line. The text has been updated accordingly [lines 229-239].

Figure 4 illustrates the variation in data quality across different sensors and sites, categorising data as good (flag = 1), interesting/suspicious (flag = 3), or missing/removed (flag = 9). Firstly, the flag labels should be consistent between the figures and the legend. Furthermore, it is surprising that most of the data (between 50% and 90%) for the different sensors is considered interesting/suspicious. Could the authors provide details and explain why this percentage is so high after the data has undergone a quality check? Are these percentages associated with the raw or final processed data?

We thank the reviewer for this comment. The original figure contained an error in the color reference of the quality flags, which were inadvertently inverted. This has been corrected in the revised figure, where the labels are now properly aligned. As a result, most of the data correspond to the good quality flag, while the suspicious and rejected categories represent only a small fraction of the dataset. The figure and caption have been corrected accordingly in the revised manuscript [lines 293-294].

Figure 6 shows the quality control performed on a time series of current intensity, as measured by an ADCP, as well as the two-level time series. Why does the Hovmöller diagram not start in mid-December 2003, as it does for the time series? This would enable us to observe the filtering of the initial data, which appears very noisy in the upper part of the water column sampled by the ADCP. The depth and altitude relative to the bottom of the measurements should also be indicated systematically in the figure titles and legend for the time series to help guide the reader.

We originally displayed the Hovmöller diagram over a shorter time window to allow a clearer comparison between the raw and quality-controlled data, focusing on a

representative month where the effect of the filtering could be easily visualized. However, following the reviewer's suggestion, the Hovmöller diagram has now been extended to cover the full time series, aligned with the temporal range shown in the time series panels below. This modification allows the reader to observe the initial noisy period in December 2003 and the effect of the quality control on those data. In addition, the labels have been revised to report the absolute depth of the measurements, improving readability and making the information on sensor position clearer in the figure titles and legends. Caption has been updated accordingly [lines 388-390].

Figure 8 shows time series for which missing data has been filled using a method that combines singular spectral analysis and optimal interpolation. It may be helpful to review the definitions of the K1 (lunisolar diurnal) and M2 (principal lunar semidiurnal) tide components, and the origin of the inertial frequency f . The sentences "As shown in Figure 8, the warming signal in the reconstructed time series confirms the robustness of the spectral estimate" (lines 443–446) are unclear to me. Could you please explain your reasoning more clearly?

We agree that the original wording could be misleading. The sentence was intended to indicate that the spectral estimates derived from the reconstructed time series are supported by the 95% confidence limits, within which the main spectral features are contained. We have therefore rephrased this part of the manuscript [lines 454–470] to clarify this point and improve the overall explanation.

Finally, it would be interesting if the authors could provide an example of the simultaneous variation of different variables, such as temperature, salinity, current, and turbidity / oxygen. This could be over a short period to illustrate an event for which high-frequency measurements are useful, or over a longer period to illustrate seasonal or longer-term variability.

We appreciate the reviewer's comment. Figure 7 is intended to summarize the variability observed in the temperature time series at the four sites, together with the hodographs relating measured velocities to water-mass densities, in order to highlight the main dynamical features of the observed variability. The figure was carefully designed to provide a comprehensive overview within a single panel, while avoiding the inclusion of additional plots that would largely repeat the same information and unnecessarily increase the length of the manuscript. We believe that this approach helps maintain clarity and conciseness while still conveying the key dynamical aspects of the dataset.