The authors present an analysis of a 6-year glider observational dataset in the Cretan Sea. The region is highly relevant to Eastern Mediterranean variability, and several changes described over the study period appear consistent with recent findings and with what is being observed more broadly in the area. However, the manuscript in its current form lacks a clear narrative and prioritization of objectives, and many conclusions are insufficiently supported by quantitative results and methodological detail. For these reasons, I recommend rejection in its current form, but encourage re-submission after substantial restructuring and additional analysis, which in my view go beyond a single major-revision cycle. I hope the authors will view these comments as guidance to strengthen the scientific robustness and clarity of the work.

We thank the reviewer for the thorough reading of our manuscript and for the constructive comments and suggestions that will help improve our work. We have addressed all points raised and will revise the manuscript accordingly. Our responses are provided below, with reviewer comments in black and our replies following each point in blue italics.

Major comments

1. **Clarify the main scope and re-balance the manuscript accordingly**

At present, it is difficult to identify a single central research question (or a small set of clearly prioritized objectives). The abstract places strong emphasis on the winter 2022 convection/mixing event, but the manuscript does not consistently follow through with event-focused analysis beyond MLD and selected transects. If winter 2022 is intended as a main focus, it should be supported with additional diagnostics (see comment 5). If not, the abstract and discussion should be revised to better reflect the broader goals (seasonal/interannual variability, anomalies vs climatology, water-mass evolution, etc.).

We thank the reviewer for this important observation. We agree that the manuscript lacked a clearly prioritised structure, and we will revise the abstract, introduction, and section accordingly. As the dataset used in this manuscript originates from a sustained monitoring program rather than targeted experiment, the study is inherently descriptive in scope; however, this does not preclude a clear hierarchy of objectives. In the revised manuscript we articulate two explicit and prioritised objectives, as follows.

Objective 1 (Sections 3 and 4.3): To document the hydrographic state and seasonal-to-interannual variability of the Cretan Sea between 2017 and 2023. This includes the characterisation of water masses, the identification of the occurrence and depth of winter intermediate-water formation, the detection of exceptional events (e.g. winter 2022), and the temporal evolution of Transitional Mediterranean Water (TMW) within the basin. This analysis provides observational evidence on the status and evolution of the key characteristics of the area over the study period and addresses the following questions: What is the typical hydrographic structure across seasons and years? Are there any exceptional events recorded? Did the Cretan Sea exhibit the ability to produce deep waters, in addition to the intermediate waters formed annually? Was there any TMW inflow during this period, and how was the resident TMW transformed within the Cretan Basin over time?

Objective 2 (Section 4.1-4.2): To quantify variability in the intermediate and deep layers over the glider observation period, as well as deviations relative to historical conditions. The objective is prioritised as follows. First, we quantify variations in temperature and salinity in the intermediate (200 - 600 m) and deep (600 - 1000 m) layers and assess their statistical robustness, placing the Cretan Sea signal in the *broader context of Mediterranean and global ocean change. This component will be further supported in the revised manuscript according to your suggestion (see comment 6). Second, we identify and quantify deviations from the recent climatology (2000 - 2015) based on Iona et al. (2018) (see comment 7).*

Winter 2022 is not the primary focus of the study. It is presented as a notable extreme event captured within the observational record, contextualised against the background seasonal variability documented in Objective 1. The abstract and discussion will be revised to make this clear, removing language that could be read as event-centric. Full ERA-5-based atmospheric diagnostics for winter 2022 are beyond the scope of the present paper and will be addressed in a dedicated future study focused on that event.

2. Too many figures relative to the amount of quantified results

There are many figures (some could be merged), yet relatively few results are explicitly quantified in the text. In several places the figure captions/results text remain descriptive, while key numbers (magnitudes of changes, anomalies by layer, interannual differences, uncertainties) are not summarized clearly. I recommend reducing redundancy, merging related panels, and adding concise quantitative summaries (tables or compact metrics) to strengthen the Results section.

We agree with this assessment and will act on it accordingly. In the revised manuscript: (i) Figs. 6 and 5 will be merged into a single composite figure; (ii) Figs. 8 and 9 (winter 2022 transects and density difference fields) will be consolidated into one figure; (iii) Figs. 14 and 15 (TMW T-S evolution and core property time series) will be merged (iv) same for Figs. 16 and 17. For each remaining figure, the corresponding results paragraph will be strengthened with explicit quantitative statements: layer-mean values, anomaly magnitudes by depth, inter-mission differences with standard deviations, and trend statistics were applicable.

The inclusion of satellite ADT/geostrophic velocity is potentially valuable to interpret mesoscale control on the transects, but in the current manuscript it remains largely qualitative. Please either (i) reduce/relocate ADT panels (in a potential Appendix) if only illustrative, or (ii) leverage ADT quantitatively (eddy polarity/strength indices per mission; distance to eddy center; relation between ADT anomalies and isopycnal heave / T–S anomalies; simple eddy tracking) to demonstrate how mesoscale circulation explains the observed hydrographic variability.

We agree. The ADT panels currently embedded within Fig. 6 (panels d, one per season) will be moved to a new Appendix A, serving as illustrative reference material. The main text will retain only a brief qualitative reference to the mesoscale context for each representative transect, noting the dominant eddy polarity from the ADT field. A quantitative mesoscale analysis linking ADT anomalies to isopycnal heave and T-S variability is an interesting direction that we plan to develop in a dedicated follow-up study.

3. Spatial definitions and data-selection methodology need to be explicit

Several analyses refer to “east/central/west” regions, specific profiles, or climatological comparisons, but the manuscript does not clearly define:

- the spatial extent and boundaries of the “east/central/west” regions,
- the exact locations used for “representative” profiles and how they are selected,

- the spatial/temporal filters used to extract/aggregate data for each analysis.

These definitions are critical for reproducibility and interpretation, and should be stated explicitly (ideally with a map and/or a table of coordinates/criteria).

We agree that these definitions were insufficiently presented and will add them to the revised manuscript. The key definitions are as follows. The glider transect extends from 23.66 deg E (western end) to 26.15 deg E (eastern end) approximately parallel to and north of the Cretan coastline. For the purposes of spatial description within a transect, we divide the transect into three sectors: western (23.66 deg - 24.5 deg E), central (24.5 deg - 25.5 deg E), and eastern (25.5 deg - 26.15 deg E). These boundaries will be marked explicitly on Fig. 2 in the revised manuscript.

“Representative” profiles and transects in figures (e.g., Figs. 6, 8, 9, 14, 16, 17) are selected as the single east-to-west transect for each mission that falls within the defined time window (see transect_time_windows.csv available through Zenodo under comment 6). For trend calculations (Section 4.1), mission-averaged layer means are computed from that one E -> W or W ->W transect, binned horizontally at 4 km resolution. The E1M3A fixed buoy is located at 35.726 deg N, 25.130 deg E, near the central sector of the transect. All of the definitions will be included in a revised Methods’ section (Section 2) with a supplementary table of coordinates and selection criteria.

4. Structure: Results vs Discussion are currently mixed

The Results section contains substantial interpretation and comparison that would fit better in the Discussion, while the Results themselves often lack quantitative reporting. I suggest a clearer separation:

- Results: quantitative statements tied directly to figures/tables (values, anomalies, slopes, uncertainties).
- Discussion: interpretation, mechanisms, comparison to literature, broader context.

We agree and will restructure the manuscript accordingly. The following material will be moved from Results to Discussion: (i) the mechanistic interpretation of the anticyclone in the autumn 2017 transect (currently in Section 3, Fig. 6 description); (ii) the attribution of intermediate-water formation to winter atmospheric forcing and its comparison to Velaoras et al. (2014) (currently in Section 3.1); (iii) the comparison of glider-derived trends to other Mediterranean studies and the global ocean (currently in Section 4.1); (iv) all speculative statements regarding post-2022 TMW and its Eastern Mediterranean implications (currently in Section 4.3). The Results section will be rewritten to report only quantitative values tied directly to figures and tables, without interpretation. The Discussion will synthesise these results in the context of regional and basin-scale dynamics and prior literature.

5. Attribution and context for the 2022 strong convection/mixing event

If winter 2022 is a major highlight, it would benefit from stronger supporting evidence on the atmospheric forcing and buoyancy loss. For example, ERA5-based diagnostics over the event window (turbulent heat fluxes, wind stress, buoyancy flux components, evaporation–precipitation, air–sea temperature differences) could substantiate the proposed driver and help distinguish local vs basin-scale forcing.

As noted in our response to Comment 1, winter 2022 is treated as a significant extreme event within the observational record, but it is not the primary focus of the study. Full ERA-5 atmospheric diagnostics (turbulent heat fluxes, buoyancy flux components, air-sea temperature differences) are therefore beyond the scope of the present paper. However, we will retain and strengthen the existing atmospheric context already in the manuscript, i.e., the -3 deg C March anomaly over Greece documented by the AMS State of the Climate 2022 report and the supporting MLD evidence from Teruzzi et al. (2024) and Potiris et al (2024) across adjacent areas. This combination of independent observational evidence from the glider, CTD casts at E1M3A, and regional model/satellite-derived MLD fields will provide sufficient qualitative context for the event within the revised scope of the paper. A dedicated study with full atmospheric forcing analysis is planned as follow-up work.

6. Trend analysis: statistical methodology and significance need more detail

Given the short record (6 years), limited number of missions, and uneven seasonal sampling, I am cautious about strong statements regarding trends without detailed statistical justification. The manuscript should specify:

- the statistical method used (regression model and assumptions),
- how seasonality is handled (or why it cannot be removed),
- which parameters are used to assess significance (e.g., p-values, confidence intervals, r),
- the reported values of these parameters,
- whether trends remain robust under sensitivity tests (e.g., excluding winter 2022, restricting to consistent seasons, or restricting to consistent depth coverage, which is at present variable).

If possible, additional confirmation should be attempted via independent sources (e.g., regional in-situ data, Argo where available, and/or CMEMS reanalysis) over the same region and time span.

Finally, the E1M3A buoy time series is not described in enough methodological detail (sampling frequency, temporal coverage, QC/processing), and the manuscript does not explicitly discuss any discrepancies between buoy- and glider-derived trend estimates. Please add these details and a quantitative buoy–glider trend comparison over the overlapping period.

We thank you for the valuable suggestions and guidance in strengthening the support of the derived statistical trends. We have addressed all points to provide a more comprehensive statistical justification.

Temporal trends were estimated using ordinary least squares (OLS) linear regression of the form $y(t) = \beta_0 + \beta_1 * t$*, where t is decimal year and* β_1 *is the rate of change per year. The model assumes linearity, independence of residuals, and constant variance of errors (von Storch and Zwiers, 1999, p. 150). In our specific case, these assumptions are partially supported by the data structure: because the input values are mission-averaged layer means (one value per deployment, with missions separated by weeks to months), temporal autocorrelation at the mission scale is expected to be limited. The restriction of the analysis to the intermediate and deep layers (below 200 m), where seasonal variability is strongly attenuated as shown in Fig. 10, further reduces the risk of a seasonal cycle contaminating the residuals.*

Because the dataset spans only six years and missions are not evenly distributed across seasons (winter and autumn are overrepresented; see Fig. 3), a formal seasonal decomposition cannot be applied without artefacts. Instead, we restrict trend estimation to the 200 - 600 m and 600 - 1000 m layers, where Fig.10 demonstrates that the seasonal signal is strongly damped relative to the surface. This choice is further supported by the sensitivity tests described below, as suggested, which show that the trends are not driven by any single season.

The six-year trends reported in the manuscript are calculated using a linear regression method of the mean values of each variable, in the intermediate (200–600 m), and deep (600–1000 m) layers. Mean values are calculated considering the in-situ values (without any interpolation) of the profiles collected in one glider transect (and not the whole mission) as represented in Figure 2, between longitudes [23.66 26.15]. The reported trends have been assessed significant through the corresponding p-values in the 95% confidence bounds.

Intermediate layer (200–600 m):

Temperature trend: 0.059 deg C/yr, SE = 0.0187, 95% CI: [0.018,0.100] deg C/yr, p=0.0081, R^2 = 0.46

*Salinity trend: 0.0205 psu/yr, SE = 0.0044, 95% CI: [0.0109, 0.0300] psu/yr, p = 5.4*10^-4, R^2=0.65*

Deep layer (600–1000 m):

Temperature trend: 0.075 deg C/yr, SE = 0.0190, 95% CI: [0.034, 0.117] deg C/yr, p = 0.0019, R^2 = 0.57

Salinity trend: 0.0215 psu/ yr^{-1}*, SE = 0.0040, 95% CI: [0.0128, 0.0302] psu/yr, p = 1.7*10^-4, R^2 = 0.71.*

The manuscript will be revised to replace all preliminary trend values (including those in Table 2 and Section 4.1 of the manuscript) with the updated values reported in order to enhance the robustness of the trend analysis and incorporate the necessary sensitivity tests.

Copernicus Reanalysis data cannot be used to assess trends derived from the glider observations. The primary reasons are methodological: (i) the Med-MFC reanalysis combines model dynamics with data assimilation, so its layer means are not directly comparable to in-situ transect averages; (ii) preliminary comparisons revealed systematic and mission-dependent offsets in the intermediate layers that we could not explain. A dedicated model–glider intercomparison is planned as future work.

To support the glider-based results with regional in situ observations, we instead rely on the E1M3A buoy time series, which provides the most suitable temporal coverage compared to other available observations in the area, such as Argo floats. We will include all necessary methodological detail and explicitly discuss the estimation of trends from both the glider and buoy datasets.

To ensure reproducibility, the analysis was performed using MATLAB code, which is available through zenodo: https://doi.org/10.5281/zenodo.19347841 .

REVISED TREND ANALYSIS:

To better quantify the statistical robustness of the six-year trends, additional methodological details and sensitivity analyses are provided. For each depth layer (200–600 m and 600–1000 m), mission-averaged means were computed using one east-to-west transect per deployment, thereby ensuring equal weighting of missions irrespective of the number of profiles acquired. As mentioned above, linear trends were estimated using ordinary least squares (OLS) regression of the form y(t)=beta_0 + beta_1* t, over 2017–2023.

Given the relatively short duration (N = 14 missions over 2017–2023) and uneven seasonal sampling (winter and autumn more frequently sampled than spring and summer), the estimated trends should be interpreted as multi-year tendencies rather than long-term climatic trends.

Glider full dataset (2017–2023):

Intermediate layer (200–600 m)

Temperature trend: 0.059 deg C/yr, SE = 0.0187, 95% CI: [0.018, 0.100] deg C/yr, p = 0.0081, R^2 =0.46

Salinity trend: 0.0205 psu/yr, SE = 0.0044, 95% CI: [0.0109, 0.0300] psu/yr, p = 5.4 *10^-4, R^2 = 0.65

Deep layer (600–1000 m)

Temperature trend: 0.075 deg C/yr, SE = 0.0190, 95% CI: [0.034, 0.117] deg C/yr, p = 0.0019, R^2 = 0.57

Salinity trend: 0.0215 psu/ yr^{-1}, SE = 0.0040, 95% CI: [0.0128, 0.0302] psu/yr, p = 1.7*10^-4, R^2 = 0.71

In both layers, the 95% confidence intervals for temperature and salinity trends do not include zero, supporting statistical significance. To evaluate robustness, the following sensitivity tests were undertaken:

1. Excluding winter 2022 (N = 13 missions): temperature and salinity trends remain significant in both layers. Intermediate layer: T slope = 0.054 deg C/yr (p = 0.021), S slope = 0.0189 psu/yr (p = 0.0017). Deep layer: T slope = 0.081 deg C/yr (p = 0.0020), S slope = 0.0228 psu/yr (p = 2.2*10^-5). This confirms that the observed signal is not an artefact of the anomalous 2022 convective event.
2. Restricting to the missions reaching ~1000 m (this is 8 of 14 missions): the deep-layer (600–1000 m) trends are significant and stronger than in the full dataset (temperature: slope = 0.097 deg C/yr, p = 3.1 x 10^-5, salinity: slope = 0.0256 psu/yr, p = 2.7 * 10^-5) demonstrating that uneven maximum depth coverage does not artificially generate the observed signal. The intermediate-layer trend is weaker and non-significant in this subsample (T: p = 0.34; S: p = 0.095), which is expected given the reduced N = 8 and the fact that 200 - 600 m are available for all 14 missions in the full dataset.
3. Season-restricted subsets: winter-only (N = 5 missions) and autumn-only (N = 4 missions) subsets both yield positive slopes in temperature and salinity across both layers (e.g. winter-only T 200 - 600 M:0.043 deg C/yr, autumn only: .051 deg C/yr), consistent in sign and order of magnitude with the full dataset result. However, they generally lose statistical significance (p = 0.10 - 0.14 for temperature; p = 0.06 - 0.12 for salinity) due to the small sample size (N = 4–5 missions), reflecting limited statistical power rather than reversal of the signal.

Buoy data description:

As an additional check on the OLS results, we applied iteratively reweighted least squares (IRLS) robust regression (bisquare weighting) to the full 14-mission dataset. The robust slopes are fully consistent with

OLS: T 200 - 600 m = 0.060 deg C/yr, T 600 - 1000 m = 0.072 deg C/yr, S 200 - 600 m = 0.215 psu/yr, 600 - 1000 m = 0.0206 psu/yr. All missions received high weights (mean weight >= 0.92, minimum weight >= 0.49, confirming the absence of leverage points that would distort the OLS estimates. The close agreement between OLS and robust slopes further supports the robustness of the reported trends.

The E1M3A mooring record spans 28 May 2007–26 Nov 2024 at 35.7263 deg N, 25.1307 deg E, with a median sampling interval of ~3 hours at fixed nominal depths of 250, 400, 600, and 1000m. After quality control (QC flags 1–2 accepted), we computed monthly means to reduce short-term autocorrelation and to provide a temporal resolution comparable to glider mission-scale averages. The resulting monthly time series contains 211 months (May 2007–Nov 2024). The overlapping period with the glider missions is Nov 2017–Nov 2023 (73 months, 14 matched glider missions). Layer means for the 200 - 600 m and 600 - 1000 m layers were derived by averaging at buoy values from the depth levels falling within each layer.

Buoy trends over the full record (all years):

Over the full monthly buoy record (2007–2024), the intermediate layer shows statistically significant warming and salinification, while the deep layer shows weaker/insignificant changes:

Layer 200–600 m (N=174 months): Temperature: +0.0246 deg C/yr (95% CI 0.0179–0.0313, p≈1.6×10^(-11), Salinity: +0.00919 psu/yr (95% CI 0.00708–0.01130, p≈4.9×10^(-15)

Layer 600–1000 m: Temperature: +0.00813 deg C/yr (p=0.039; low explained variance, R^2≈0.03), Salinity: not significant (slope ≈ −0.00089 psu/yr, p=0.545)

These results indicate that, in the long-term buoy context, the strongest and most robust multi-year tendency below 200 m is intermediate-layer warming and salinification, whereas deep-layer trends are much weaker and not consistently significant.

Quantitative buoy–glider comparison over the overlapping period:

To support glider trend findings, buoy trends were recomputed over the overlapping period (Nov 2017–Nov 2023; 73 months) using the same monthly-mean approach. Overlap-period buoy trends show:

- *Layer 200–600 m (N=71): Temperature: +0.0234 deg C/yr (p=0.059; marginal), Salinity: +0.0278 psu/yr (p≈2×10^(-6); significant)*
- *Layer 600–1000 m (N=66): Temperature: +0.0788 deg C/yr (p≈1.1×10^(-8); significant), Salinity: +0.0216 psu/yr (p=0.0033; significant)*

Both platforms agree in sign for all four variable-layer combinations during 2017 - 2023. The slope magnitudes are comparable, particularly for the deep layer: buoy and glider both estimate deep-layer temperature warming of ~0.079 and ~0.075 deg C/yr respectively, and deep salinity increase of ~0.022 and 0.022 psu/yr (ratio 1.0). In the intermediate layer, the buoy temperature trend (~0.023 deg C/yr, p = 0.059) is marginally non-significant over the overlap period, whereas the glider estimate (0.059 deg C/yr, p = 0.008) is significant. This discrepancy in magnitude likely reflects the difference in spatial footprint between the glider transect and the single fixed-point buoy at 35.7263 deg N, 25.1307 deg E, as well as the different temporal sampling (monthly buoy means vs mission-averaged glider values). The fact that *both platforms agree strongly in the deep layer, and in the sign and order of magnitude in the intermediate layer, lends confidence to the reported trends. The difference between buoy trends over the full record (2007 - 2024) versus the overlap window (2017 - 2023) further illustrate the sensitivity of short-window linear fits to multi-year regional anomalies. Hence the 2017–2023 trends should be interpreted as multi-year tendencies rather than long-term climatic rates.*

von Storch, H. and Zwiers, F.W. (1999). Statistical Analysis in Climate Research. Cambridge University Press.

7. External references / climatology description should be more concrete

The manuscript refers to an external climatology and uses it as a reference. Please describe more clearly:

- which dataset(s) underpin the climatology,
- how it was constructed (at least a short methodological summary),
- how the comparison is made (same spatial box? same season? interpolation/averaging method?).

If additional historical observations are available (SeaDataNet/CMEMS/ship CTDs/Argo), they could be used to provide context beyond the 6-year record and to corroborate the reported changes.

We will expand the climatology description in the revised manuscript. The climatology used is the Mediterranean Sea Hydrographic Atlas of Iona et al. (2018). It was constructed by applying a variational inverse method (DIVA – Data Interpolating Variational Analysis) to historical in-situ temperature and salinity profiles compiled from multiple databases (SeaDataNet, World Ocean Database, MEDAR/MEDATLAS), covering the period 1950 - 2015. We use the most recent seasonal climatology period available (2000 - 2015) and extract the winter season fields for the Cretan spatial domain. The climatology is provided on a regular grid with standard depth levels at 5, 10, 30, 75, 200, 400, 700, and 1000 m. Our comparison is performed as follows: for each winter glider mission (2018 - 2023), mission-mean profiles of temperature and salinity are computed by averaging all profiles within the transect longitude range (23.66 deg - 26.15 deg E); these are then interpolated onto the climatology depth levels; and the anomaly at each level is computed as the glider mission mean minus the climatological winter mean for the same depth. The resulting anomaly profiles are then averaged across six winter missions to produce the mean deviation shown in Fig. 13. This comparison approach, its spatial domain, and the climatology source will all be stated explicitly in the revised Methods' section. Regarding additional historical data: Argo float coverage in the Cretan Sea is sparse for the study period, and shipborne CTD casts are available only at discrete times and locations (Fig. 2). These data were not sufficient to construct an independent multi-year time series comparable in resolution to the glider record. We will note this limitation explicitly in the revised Discussion.

8. Presentation and clarity

Several figures are difficult to read (small labels, low contrast, dark text on dark backgrounds). Increasing font sizes and improving contrast would substantially improve accessibility.

Consider merging figures and reducing the total number by combining related information into fewer, more information-dense plots.

We thank the reviewer for these suggestions and will implement them fully. All figures will be regenerated with a minimum font size of 10 pt for axis labels and tick marks, and 11 pt for panel titles and legends, ensuring legibility at typical journal column widths. Colorbar labels and in-figure annotations will be revised to use high-contrast colors (avoiding dark text on dark backgrounds, particularly in the deep-layer panels of Figs. 6,8, and 12). Regarding figure merging: as noted in our response to Comment 2, Fig. 6 will be reduced and combined with 5, the content of Fig. 8 will be streamlined and consolidated with Figs. 7 and 9, Figs. 14 and 15 will be combined and Figs. 16 and 17 will be merged reducing the total figure count from 17 to 12.

Summary recommendation

The dataset and topic are valuable, but the manuscript requires substantial restructuring, clearer definitions and methods, stronger quantitative reporting in the Results, and more robust support for trend statements (including statistical significance and/or external validation). If the authors choose to resubmit a substantially reworked manuscript, the work could become a solid and publishable contribution to Eastern Mediterranean variability studies.

We hope that the revised manuscript will meet the reviewer’s expectations.