Response to François Ribalet (Reviewer 1)

We thank the reviewer for their constructive comments and appreciate the time dedicated to reviewing our manuscript. We address each comment below, and the corresponding revisions in the manuscript are highlighted in yellow.

General comments:

This manuscript links two decades of thermohaline profiles to phytoplankton assemblage shifts at a single Mediterranean coastal station. The dataset is impressive, the analytical approach is interesting, and the writing is clear. However, I have a few concerns, mainly that the authors often use a causal language that doesn't match what the data show, and several analytical decisions lack justifications, which I explain below:

My first concern is that main conclusion covers a decade with no biological data. The abstract states that conditions favoring larger nanoeukaryotes have grown less frequent since 2004. The PicoNano time series begins in 2014. Ten years of the claimed shift were never observed. The paper shows that thermohaline structures changed after 2004. It shows that different structures associate with different assemblages after 2014. From these two facts it infers a third: that assemblages probably changed between 2004 and 2014. My issue is the language in the abstract and discussion. It reads as observed fact rather than inference. The causal chain: wind shift, then mixing, then smaller nanoeukaryotes, then reduced copepod lipids, then sardine decline. Each link rests on correlation across separate time series. None are modeled together. The chain's foundation is a single Wilcoxon test at p = 0.031, uncorrected, out of four tests run.

We agree that this distinction was not sufficiently clear in the original manuscript. Our intention was not to present these ecological changes as directly observed between 2004 and 2014, but rather as an inference based on combining (i) the long-term changes in thermohaline structures observed since 2004 and (ii) the associations between thermohaline structures and PicoNano assemblages established from 2014 onwards. We have therefore revised the Abstract and the Discussion to consistently distinguish observations from inferences.

We also acknowledge that the Wilcoxon test mentioned by the reviewer does not remain statistically significant after correction for multiple testing. This point is addressed in more detail in our response to minor comment No. 4. More generally, we have revised formulations that could be interpreted as implying causal or directly observed relationships. The proposed ecological cascade is now explicitly presented as a plausible interpretation of the combined observations, rather than as a demonstrated sequence of causal events.

The abstract was modified as follow:

(l. 15-17 in revised manuscript)"*We show that highly homogeneous water columns are associated with different PicoNano assemblages compared to mixed but less homogeneous water columns. Overall, these results suggest that thermohaline conditions associated with larger nanoeukaryotes have become less frequent since 2004.*"

The discussion was modified as follow:

(l. 567-568 in revised manuscript*) "These observations coincided with a shift in the dominant wind direction within the north-western sector, from 350° to 340° in 2004 (Appendix F1).*

(l.572-573 in revised manuscript*) "Between 2008 and 2018 the recurrent predominance of winds from 330° was also consistent with the increased occurrence of well-mixed waters.* "

(l. 585 - 587 in revised manuscript) "*Interestingly, Garcia et al. (2024) independently reported a decrease in the lipid content in zooplankton 300-500 µm size fraction, mainly composed by copepods, since 2009. On the one hand, a phytoplankton-based diet is known to be associated to a higher lipid content in copepods (Lee et al., 2006).*"

(l. 589-594 in revised manuscript*) "One possible hypothesis is that the suspected increased prevalence of small eukaryotic nanophytoplankton in winter may have reduced the contribution of larger nanophytoplankton to copepod diets. This could have contributed to the observed decrease in their lipid content. The hypothesis of a shift from autotrophic prey to heterotrophic and mixotrophic prey has already been suggested by Garcia et al. (2024). If this hypothesis is correct, such changes could also have affected the quality of the favourite preys ingested by S. pilchardus which are copepods... "*

Second, the clustering pipeline rests on undocumented choices. The classification of 1,619 TS profiles into nine vertical structures is the analytical backbone of the paper, yet several decisions that directly determine the result are presented without validation. First, the authors report that BIC selected 9 clusters using a VVV model, but BIC has a well-known tendency to favor a larger number of clusters as sample size grows, and it is not guaranteed to identify ecologically meaningful groupings. No sensitivity analysis is provided showing how the biological or temporal results change under 7 or 11 clusters. Second, the clustering is performed on the first 4 MFPCA scores, justified by a "plateau" in explained variance, but the proportion of total variance actually retained by 4 components is never reported. This information is needed to assess whether the low-dimensional representation is adequate. Third, the uncertainty threshold of 0.25 (posterior probability > 0.75) is cited to Scrucca et al. (2023) as a general guideline, not validated for this specific dataset or application. With ~25% of profiles unclassified, the "NC" category is large enough to affect temporal trend estimates; the sensitivity of GAM results to this threshold needs to be tested.

The objective of the clustering was not to directly identify ecologically meaningful groups, but to identify recurring thermohaline structures from TS profiles. The ecological relevance of these structures was evaluated independently by comparing their associated nutrient concentrations and PicoNano assemblages. Therefore, although we agree that the BIC criterion does not guarantee ecologically meaningful clusters, this was not its intended purpose in our study.

From this perspective, overestimating the number of clusters is less problematic than underestimating it. Clusters with similar hydrological and biological characteristics can subsequently be merged, as illustrated by clusters 6 and 9, which were both identified as well-mixed waters and exhibited similar nutrient concentrations and PicoNano assemblages. In contrast, selecting too few clusters could merge distinct thermohaline structures associated with different PicoNano assemblages, thereby masking biologically relevant differences. We acknowledge that alternative partitions are possible and that a finer subdivision of some hydrological conditions may exist. Nevertheless, the nine-cluster solution provided a meaningful representation of the recurring thermohaline structures while retaining sufficient biological observations within each group for subsequent ecological comparisons.

Regarding the MFPCA representation, we agree that the proportion of variance explained by the first four components should have been reported. This information has now been added to the manuscript to justify the adequacy of the low-dimensional representation: *(l. 235-236)* in revised manuscript*)* "*The first four principal component scores were retained, accounting for 93.66% of the total variance.*"

Regarding the uncertainty threshold, we agree that it may influence the proportions of the different clusters and consequently the GAM estimates. Following your recommendation, we performed a sensitivity analysis using alternative uncertainty thresholds. The results are presented below and were added in appendix. The sensitivity analysis showed that the estimated temporal patterns and the associated statistical significance of the GAM smooth terms were robust across the range of uncertainty thresholds tested. The only exception was the trend for stratified waters, whose statistical significance varied with the selected threshold.

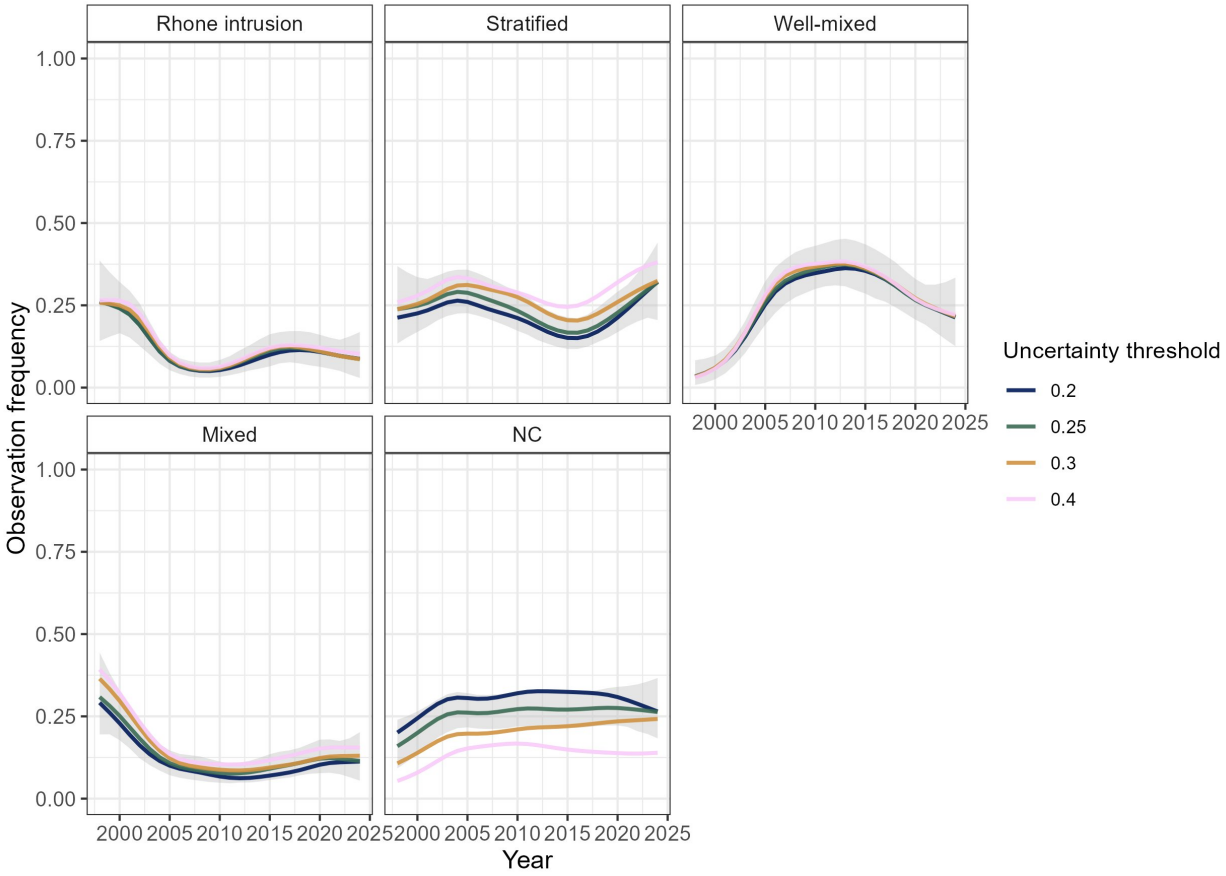

Figure 1: Evolution of the annual observation frequency of aggregated vertical structures obtained from model-based clustering using different uncertainty thresholds. Solid curves represent the multinomial Generalized Additive Models (GAMs) fitted to the observed annual frequencies of each vertical structure for each uncertainty threshold. For each vertical structure, the grey ribbon represents the 95% pointwise confidence interval estimated by bootstrapping for the uncertainty threshold of 0.25 used in the study.

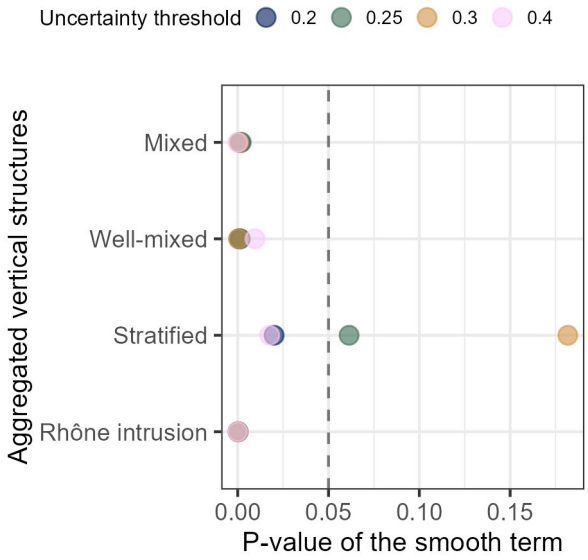

Figure 2: P-values of the χ^2 tests for the smooth terms of each aggregated vertical structure obtained from model-based clustering using different uncertainty thresholds.

Finally, the Prochlorococcus signal may be an instrument artifact rather than ecology. The paper acknowledges, briefly, that surface Prochlorococcus counts likely miss high-light-adapted ecotypes. These cells lose pigment under strong light and drop below detection. This matters because the paper's central summer-autumn pattern is exactly what that artifact would produce. Stratified oligotrophic water brings high light and photo-inhibition, pushing cells below detection. Mixed water brings low-light-adapted cells to the surface, where the instrument sees them. The paper treats the resulting abundance pattern as ecology, but no evidence is given to rule out instrument bias as the source. The authors should run a sensitivity analysis that removes Prochlorococcus from the analysis and report whether the assemblage separation holds.

We acknowledge that the instrument may miss high-light-adapted cells and that this could explain the observed patterns, as already discussed in the manuscript (l. 486-489): "*Because high-light–adapted ecotypes dominate Mediterranean surface waters in summer and autumn (Mella-Flores et al., 2011), the apparent decline of Prochlorococcus in more oligotrophic conditions may correspond to the increase of undetected high-light–adapted cells with a very low amount of photosynthetic pigments, under the limit of sensitivity of flow cytometry.*" As stated previously (l. 486), "*our data likely reflect only the low-light–adapted fraction of the population.*" We also specify later (l. 544-545) that "*our results suggested that the presence of low-light-adapted Prochlorococcus and picoeukaryotes at the surface in summer appeared to be mainly attributed to vertical mixing, consistent with their low tolerance to high irradiance (Zubkov et al., 2000).*" Accordingly, our conclusions regarding Prochlorococcus apply only to the detectable low-light-adapted fraction and should not be extrapolated to the total Prochlorococcus population. We added the following sentence for clarity: (l 497-499 in revised manuscript) "*Accordingly, the patterns described here for Prochlorococcus should be interpreted as applying to the detectable low-light-adapted fraction detectable by flow cytometry rather than to the total Prochlorococcus population present in situ.*"

Regarding the proposed sensitivity analysis, removing Prochlorococcus from the assemblage analysis would not address this limitation, as the objective of the study is to characterize the observed PicoNano assemblages measured by flow cytometry rather than to estimate the

abundance of individual taxa. Excluding one group would fundamentally alter the assemblage definition without resolving the underlying detection limit of the instrument.

Minor comments:

1. *Wind direction is circular (angular) data. Applying a regression tree (CART) to degrees treats the 30° difference between 320° and 350° as equivalent to any other 30° difference, while the 40° gap between 350° and 10° becomes 340°. Please discuss or justify the use of linear CART on circular wind direction data.*

   The regression tree was applied only to wind directions within the northwestern quadrant (270°–360°), rather than to the full 0°–360° circular domain. We focused exclusively on this sector because northwesterly winds are the dominant winds during well-mixed and mixed conditions (Fig. E2) and correspond to the well-known influence of the Mistral on water-column mixing (Millot, 1979). Within this restricted angular range, the circular discontinuity at 0°/360° is not encountered, making the use of a standard CART regression appropriate. Although this restriction was already described in the Methods section (l. 297: *"For each year between 1998 and 2024, the 95th percentile of wind intensity was calculated for each wind direction of the northwestern sector (270°–360°)...*"), we omitted to mention it in the legend of Figure F1. This has now been corrected.

   (l 731-732 in revised manuscript) "*Piecewise constant regression (Classification and Regression Trees, CART (Ripley, 2025), blue line) of the mean annual dominant wind direction (black dots) in the northwestern quadrant (270°–360°) at the Marignane Météo-France station.* "

2. *Please describe the gating strategy and how did you manage consistency of gate boundaries across operators and years, over a multi-decade series.*

   The flow cytometry analyses and gating were performed as part of the SNO SOMLIT monitoring programme and independently of the present study. The gating strategy used across the SOMLIT time series comes from a well-established scheme (Vaulot, Marie, Partensky, Simon et al., going back to the early 1990s): combining light-scattering parameters and natural photosynthetic pigment induced fluorescence (chlorophyll, phycoerythrin) allows the optical resolution of the various cytometric groups differing in size (forward scatter related) and pigment content (different red and orange fluorescence intensities). Data acquisition was triggered on red fluorescence to get rid of non-fluorescent particles, and subpopulations interactively defined by regions and gates identified from the combination of the cytometry variables (parameters). Synechococcus is discriminated by its orange fluorescence from phycoerythrin on an orange vs red fluorescence cytogram, while Prochlorococcus, smaller and less fluorescent, is distinguished from picoeukaryotes on the forward scatter vs red fluorescence bivariate distribution.

   Consistency across the monitoring series is supported by the standardized and centralized analytical procedure of the SOMLIT network. Samples collected at the different SOMLIT stations are processed according to a common protocol before flow cytometry analysis. Pico- and nanoplankton flow cytometry analyses are centralized at

the BioPIC platform (Observatoire Océanologique de Banyuls, France) which is the reference platform for pico- and nanoplankton analyses across the SOMLIT monitoring network. That centralization is itself a major argument for consistency as the samples were processed through a common reference laboratory with a shared instrument lineage and a shared expertise validated, by two independent French experts in flow cytometry (Dominique Marie from the Roscoff Biological marine station and Gérald Grégori from the MIO). For the 2014–2024 period corresponding to the data used in our study, the analyses were performed by the same operators, further limiting inter-operator variability.

Since fixation artifacts are a classic source of unwanted inter-year variability independent of gating itself, the SOMLIT network decided to use a single validated protocol for its 12 marine stations: the protocol for flow cytometry analysis of phytoplankton cultures and natural samples (Marie et al. 1999, improved in 2014) built on earlier work including testing effects of freezing and fixation on cell recovery for Prochlorococcus, Synechococcus, picoeukaryote, and nanoeukaryote cytometric groups, with earlier groundwork from Troussellier, Courties & Zettelmaier (1995) on flow cytometric analysis of coastal lagoon bacterioplankton and picophytoplankton, specifically addressing fixation and storage effects.

The following sentence was added to the revised manuscript: (l 146-48 in revised manuscript) ”*Phytoplankton samples collected within the SNO SOMLIT were analysed at the BioPIC platform (Observatoire Océanologique de Banyuls, France). Flow cytometry analyses followed the protocols described in Marie et al. (1999, 2014), and the gating strategy followed the scheme described in Marie et al. (2000).* ”

Marie, D., Partensky, F., Vaulot, D., and Brussaard, C.: Enumeration of Phytoplankton, Bacteria, and Viruses in Marine Samples, Current Protocols in Cytometry, 10, 11.11.1–11.11.15, https://doi.org/10.1002/0471142956.cy1111s10, 1999.

Marie, D., Simon, N., Guillou, L., Partensky, F., and Vaulot, D.: Flow Cytometry Analysis of Marine Picoplankton, in: In Living Color, edited by Diamond, R. A. and Demaggio, S., pp. 421–454, Springer, https://doi.org/10.1007/978-3-642-57049-0_34, 2000.

Marie, D., Rigaut-Jalabert, F., and Vaulot, D.: An improved protocol for flow cytometry analysis of phytoplankton cultures and natural samples, Cytometry Part A, 85, 962–968, https://doi.org/https://doi.org/10.1002/cyto.a.22517, 2014.

Troussellier M, Courties C, Zettelmaier S. 1995. Flow cytometric analysis of coastal lagoon bacterioplankton and picophytoplankton: Fixation and storage effects. Est Coast Shelf Sci 40:621-633.

3. *Zero-inflated counts (RedPicoProk, OraNano) enter a DDA that assumes normality. Please describe how zeros where handled.*

DDA was used solely as a descriptive exploratory method and can be viewed as a particular form of PCA that identifies the axes maximizing the variance between predefined groups. Consequently, the identification of the discriminant axes does not require the assumption of multivariate normality. No specific treatment was applied to zero counts. Because the normality assumption was not satisfied, no statistical inference was performed (i.e., no tests of discriminant axis significance or MANOVA). The

relevance of the discriminant axes was assessed solely based on the proportion of discriminant inertia explained.

4. *Four Wilcoxon tests are conducted comparing wind intensity between well-mixed and mixed conditions at 320°, 330°, 340°, and 350°. Did the author correct for multiple comparisons? The p = 0.031 result for 330°, the only result highlighted, corresponds to a false discovery rate that is not controlled. With four tests at α = 0.05, the familywise error rate is ~18.5%.*

   As mentioned in the Methods section (l. 292): "*Because the objective was to identify potentially relevant wind directions within the north-west sector rather than to test a single global hypothesis, the analysis was considered exploratory and no correction for multiple comparisons was applied. P-values should therefore be interpreted as indicative of potential direction-specific patterns rather than as confirmatory statistical evidence.*"

   Accordingly, the result obtained at 330° (p = 0.031) should not be considered as evidence of a statistically significant effect, but rather as an exploratory indication of a possible difference in wind intensity between well-mixed and mixed conditions. This is why we referred to it as "a potential tendency" in the Discussion (*"a potential tendency was observed for well-mixed conditions to occur under relatively stronger winds from 330° compared with mixed conditions*" (l. 568)). The objective of this analysis was not to provide definitive evidence of a wind-direction effect, but to identify potential patterns that may warrant further investigation.

   Nevertheless, we acknowledge that the lack of statistical significance after correction for multiple comparisons should be emphasized. We therefore performed an additional sensitivity analysis using Hommel's correction, after which no comparison remained significant. This information has now been added to the Results section: (l. 433-434 in revised manuscript) "*However, after correction for multiple comparisons using Hommel's method, this difference was no longer statistically significant.*"

5. *The mean-normalized nutrient index (sum of concentrations each divided by its temporal mean) gives equal weight to each nutrient regardless of its ecological role or dynamic range, so a nutrient with high absolute concentration and low variance will dominate the index. Please justify this choice.*

   We respectfully disagree with this point. Because each nutrient concentration is divided by its own temporal mean, differences in absolute concentration are removed and each normalized variable has a temporal mean of one. Consequently, a nutrient with a high absolute concentration but low temporal variability does not dominate the index; its normalized values remain close to one. Instead, the temporal variability of the index is mainly driven by nutrients with larger relative fluctuations. Our objective was precisely to construct a dimensionless index reflecting the overall relative availability of dissolved nutrients rather than their absolute concentrations.

6. *The COAST-HF profiler operates only from April to November, creating a systematic seasonal gap in the high-frequency portion of the combined dataset. Please discuss whether this seasonal truncation introduce a potential source of bias in the cluster distribution or temporal analysis.*

   This is an important point. However, neither the seasonal distribution of clusters nor the temporal trend analyses were based on the combined dataset. Both were performed

exclusively using the SOMLIT dataset, which provides year-round sampling and is therefore unaffected by the seasonal gap in the COAST-HF profiler. We acknowledge that this was not sufficiently explicit in the original manuscript and have clarified this in the Methods section relative to the GAM. (l. 311-312 in revised manuscript) "*As for the comparisons, the model was fitted only to TS profiles from the SOMLIT dataset, which were sampled consistently throughout the year.*"

7. *The paper identifies three structures associated with freshwater intrusions but is unable to classify Northern Current intrusions at all. The Discussion acknowledges this limitation but does not address its consequences for the completeness of the temporal trend analysis or the NC category composition.*

   Northern Current intrusions may co-occur with the hydrological structures identified from the TS profiles. Our classification is based on the vertical thermohaline structure of the water column rather than on the origin of the water masses. Consequently, a Northern Current intrusion does not necessarily define a distinct hydrological structure. For example, a water column classified as well mixed remains well mixed regardless of whether the water originates from the Northern Current or not. The main limitation therefore concerns situations where a Northern Current intrusion modifies the thermohaline properties sufficiently to increase the classification uncertainty, potentially leading to assignment to the NC category. However, the proportion of unclassified profiles remained stable throughout the study period, suggesting that this source of uncertainty did not vary over time and is therefore unlikely to bias the temporal trend analysis. We compared the dates of documented Northern Current intrusions with our sampling dates and found very little overlap, which also prevented us from characterizing these events as a separate category. Consequently, although Northern Current intrusions may influence the thermohaline characteristics of individual profiles, we do not expect them to substantially affect the long-term trends reported here.

8. *The link between small nanoeukaryotes and reduced copepod lipid content rests on a separate study, different organism, different years. The two series are never formally compared, and confounders (e.g., temperature effects on copepod lipid metabolism, prey switching) are not considered.*

   Our intention was not to formally compare our observations with the time series reported by Garcia et al. (2024) or to infer a causal relationship between the two studies. Rather, this paragraph was intended as a discussion of a possible ecological mechanism linking our observations with previously reported changes in the ecosystem. We agree that the original wording may have overstated this interpretation. As mentioned above, we have therefore revised the paragraph to make its speculative nature more explicit.