Review 1

The role of the physical circulation, especially fronts and eddies, in controlling phytoplankton abundance and community structure is one of wide interest, making overlap with ecology, biodiversity, remote sensing, non-linear dynamics and biological-physical interactions. This paper therefore is likely to be widely read. It takes a variety of methods for locating fronts and uses them to highlight front related anomalies, which themselves vary across a region with strong large scale biogeochemical gradients (the Mediterranean). It is clearly written and provides both a good technical basis and, based on the results, some interesting lines of future study for application to other ocean regions. I should say at the outset that I do not have detailed knowledge of the methods used but, while I think the main results are well argued and clearly presented, I feel that there needs to be more discussion about assumptions and limitations before I can recommend publication.

We thank the reviewer for their careful reading of our manuscript and for their constructive and helpful comments. We have considered each remark and revised the manuscript accordingly. Below, we provide a point-by-point response to all comments and indicate the corresponding changes made to the manuscript.

R1.1 : It is a strength of the paper that the authors deliberately use multiple methods to detect and quantify frontal activity, acknowledging that there is still no consensus on the optimal one to use, should such an optimal one exist. While I understand that the paper should not go into too much detail describing those methods, I think that it is important that the method by which the high-resolution salinity field was constructed needs to be described and discussed. I was surprised that it was possible to re-construct a salinity field at the high resolution used but, as I said, this is not my field so I looked into the cited papers. They in turn refer to Olmedo E. et al. Improving SMOS Sea Surface Salinity in the Western Mediterranean Sea through Multivariate and Multifractal Analysis. Remote Sensing. 2018; 10(3):485. https://doi.org/10.3390/rs10030485. Section 3.4 of that paper suggests that a linear relationship between salinity and temperature is used to increase the resolution of the product (their Equation 8). There is also no apparent validation of the performance of the product at the high resolution scales in that paper. Particularly if salinity is to be combined with temperature to create density, as in the paper under review, this is a major assumption. Two questions arise from this. First, if salinity is being reconstructed using temperature at small scales, how well is the component of salinity independent of temperature (which leads to temperature not being perfectly correlated with density) reconstructed? Second, to what extent does the frontal map using density differ from that using temperature not because of small-scale salinity but because of differences between the two different sources of high-resolution temperature data that are used and which are unlikely to match perfectly at these small scales? I am not going to suggest that this needs to somehow be corrected but I do think it is something that needs to be acknowledged and the possible consequences for the robustness of the results to be discussed.

We agree with the reviewer that information on the SSD product was missing in the original submission. We have thus briefly introduced some relevant details in the revised text. In fact, the 4DMED product does not use the linear T–S approach of Olmedo et al. (2018); it uses the multivariate optimal interpolation of Droghei et al. (2016), as re-adapted and described in

Sammartino et al. (2022). In that algorithm, SST is an auxiliary covariate in the multivariate optimal interpolation. It is not a direct proxy for SSS/SSD and does not impose a deterministic T–S/D relation. It modulates the interpolation weights of salinity/density observations according to thermal similarity, thereby exploiting the physical T–S covariance to improve water-mass distribution coherence and effective resolution

At scales finer than the effective resolution of salinity observations, part of the SSS/SSD structure is indeed inferred from SST. Temperature and salinity are therefore not fully independent at those scales (which is physically-justified in the cited literature). There is still totally independent salinity/density information at the scales resolved by SMOS, SMAP and in situ data, and salinity/density can vary along isotherms, but the high-resolution texture is partly inherited from SST. This approach has been demonstrated to significantly improve SSS/SSD reconstruction accuracy (see cited literature). We clarified also that Sammartino et al. (2022) include spatial spectral analysis and high-resolution validation against totally independent thermosalinograph data. Consequently, while the large- and mesoscale salinity variability is constrained by independent salinity observations, the finest-scale structures in the reconstructed density field are not fully independent of SST.

This point is particularly relevant for our analysis because the HI diagnostic is specifically designed to detect spatial heterogeneity at relatively small scales. We therefore agree that differences between SST-derived and density-derived HI cannot be interpreted straightforwardly as the contribution of small-scale salinity structures independent of temperature. The density-based HI incorporates salinity information, but its fine-scale spatial structure is partly constrained by SST.

We have clarified this limitation in the revised manuscript and have consequently adopted a more cautious interpretation of the comparison between SST- and density-derived fronts. In particular, we do not use this comparison to quantify the specific contribution of small-scale salinity gradients to frontal structures. Rather, the density product provides a complementary diagnostic that incorporates the available salinity information, while acknowledging that the finest spatial scales are not fully independent from SST.

We have modified the data section :

“*Density fields are available for the entire Mediterranean basin, on a regular grid with a spatial resolution of 1/24°, and a time interval of 1 day, from 1 January 2016 to 1 August 2022, and are part of the 4DMed-Sea project. These fields were in fact obtained through a multivariate optimal interpolation algorithm specifically designed to combine satellite UHR SST data, sea surface salinity images from various satellite sources and in situ measurements (Sammartino et al., 2022). Input satellite-derived density images are obtained by remapping both UHR SST and satellite salinity measurements over the final grid and applying the equations of state for seawater density. Classical optimal interpolation provides gap-free fields by averaging observations with weights that depend on spatial and temporal covariance. The multivariate algorithm used in Sammartino et al. (2022), originally introduced by Droghei et al. (2016), includes an additional decorrelation factor that increases/decreases observations’ weights based on their thermal differences with respect to the interpolation point. In practice, density observations are weighted more if they lie on the same isothermal as the target point. This strategy exploits the physical covariance*

between temperature and salinity/density and forces interpolated patterns to reproduce more consistent water-mass distributions at the small scales than classical algorithms, effectively increasing the resolution of the interpolated surface fields. Independent density information remains at scales resolved by the satellite and in situ data, but the high-resolution texture is inherited from SST. Full details on the algorithm, including spatial spectral analysis and high-resolution validation based on thermosalinograph data, can be found in Sammartino et al. (2022).”

We have added to the discussion section :

“*An important limitation of the density-based frontal diagnostic is that, although the 4DMED product incorporates independent satellite and in situ salinity observations, its finest-scale spatial structure is constrained by high-resolution SST through the multivariate interpolation. Density-derived HI therefore cannot be interpreted as a fully independent measure of small-scale salinity-driven frontal variability.*”

Reference : Droghei, R., Nardelli, B. B., & Santoleri, R. (2016). Combining in situ and satellite observations to retrieve salinity and density at the ocean surface. Journal of Atmospheric and Oceanic Technology, 33(6), 1211-1223.

There are other places where more detail is needed too. While they are more minor, they are still significant for those who wish to apply the approach themselves:

R1.2 : The evidence and metric used to determine that a 7x7 pixel window was the optimal (lines 170-171) are not given

We agree that the explanation for the choice of the HI window size was not sufficiently detailed in the original manuscript. The window size was initially assessed through a sensitivity analysis using 3 × 3, 5 × 5 and 7 × 7 pixel windows (Figure R1.2.1). Note also that in our initial study, we considered both weak (5 < HI < 10) and strong fronts (HI > 10) categories. The sensitivity analysis showed that the window size mainly affects the detection of weak fronts. The mean fraction of valid pixels classified as weak fronts increased from 21.51 ± 9.51% for the 3 × 3 window, to 26.44 ± 5.83% and 35.13 ± 5.46% for the 5 × 5 and 7 × 7 windows, respectively (Figure R1.2.2 and the Table R1.1.1). This indicates that the two smaller windows (3 × 3 and 5 × 5) tend to detect less of the spatial heterogeneity associated with weak frontal structures.

However, the fraction of valid pixels classified as strong fronts (HI > 10), which is the only front category considered in the manuscript, remained much more coherent: 4.18 ± 0.23%, 3.92 ± 0.15% and 3.74 ± 0.10% for the 3 × 3, 5 × 5 and 7 × 7 windows, respectively. In addition, the comparison between the 5 × 5 and 7 × 7 windows showed that the fraction of strong front pixels differed by only −0.18 ± 0.11 percentage points on average over 365 days analysed (Figure R1.2.2 and the Table R1.1.1).

Rather than considering the 7 × 7 window as an optimal window, our choice was initially made in the context of our broader analysis, in which we considered both weak and strong fronts. We then restricted the analysis presented in the manuscript to strong fronts, whose identification is less affected by the choice of window size. We have clarified these details in the revised manuscript, and removed the term “optimal” in the revised manuscript.

"*The 7 × 7 pixel window was selected following a sensitivity analysis performed with 3 × 3, 5 × 5 and 7 × 7 pixel windows. We applied this analysis for two categories of SST heterogeneity: weak fronts (5 < HI < 10) and strong fronts (HI > 10). The window size affected the detection of weaker SST heterogeneity, with the fraction of pixels classified as weak fronts increasing with window size. In contrast, the fraction classified as strong fronts, which are the fronts considered in this study, remained relatively stable across the three window sizes. We finally decided to focus only on the strong front category, hereafter referred to as front, and to keep the 7 × 7 window, while verifying that the classification of strong fronts was not substantially affected by the window size.*"

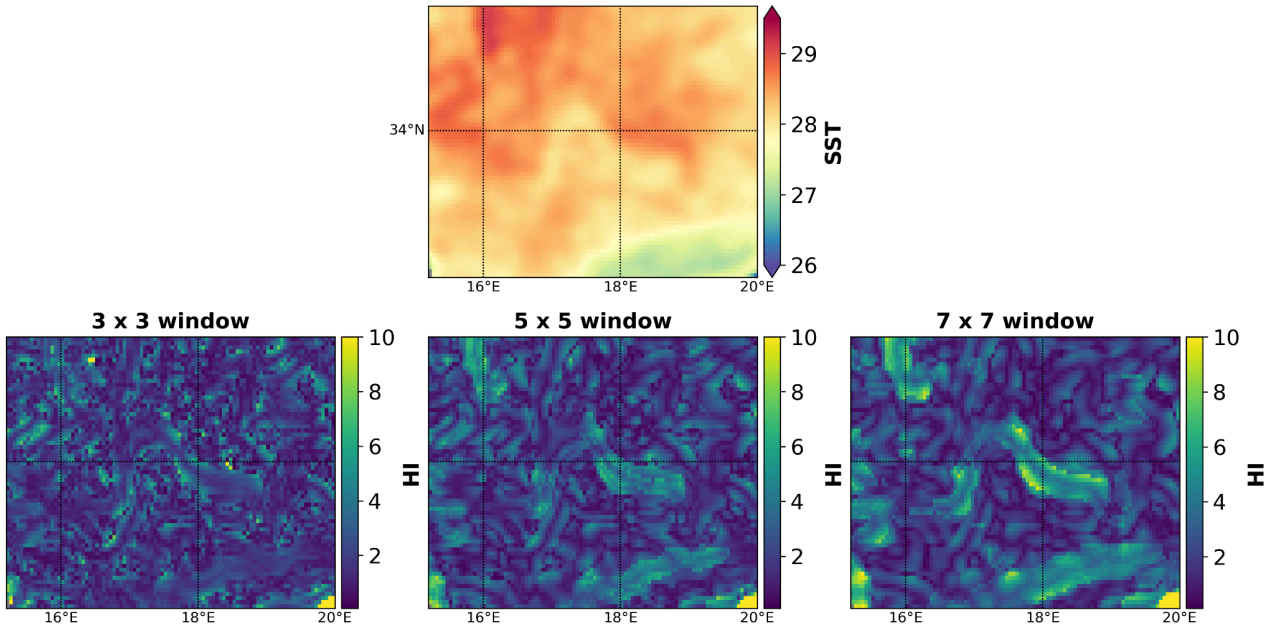


Figure R1.2.1 : Snapshots of 15 July 2012 : SST fields (top), HI with 3 x 3 window (left), 5 x 5 window (center), 7 x 7 window (right)

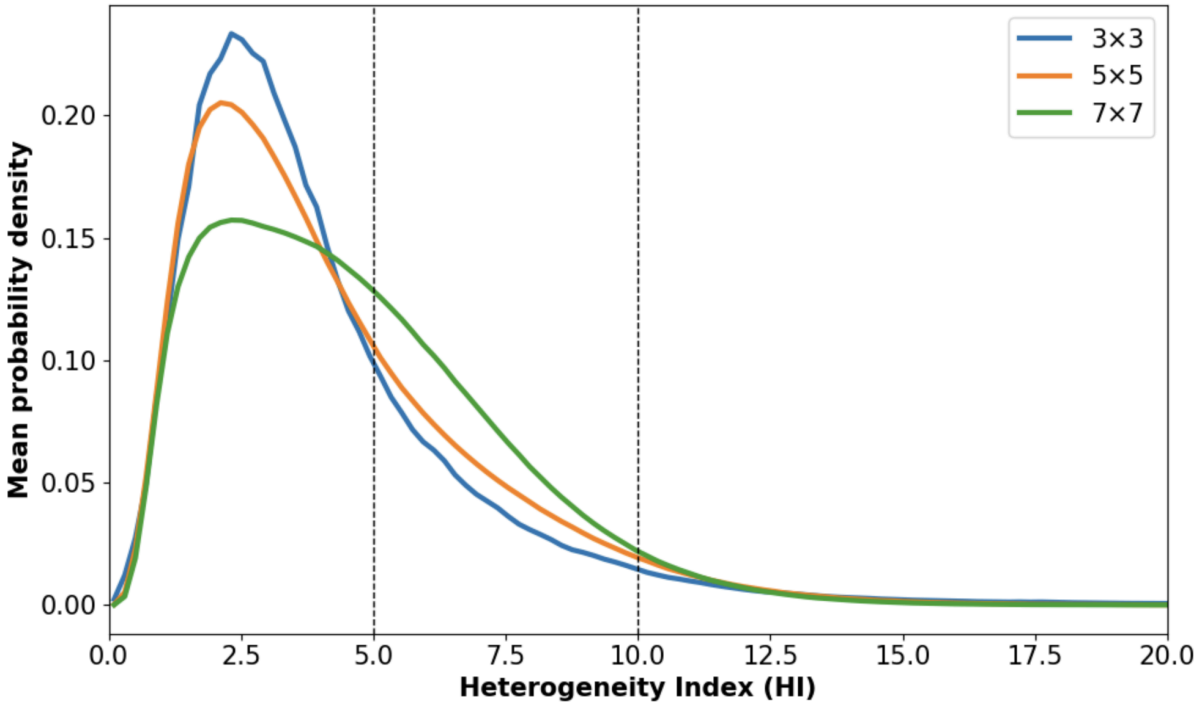


Figure R1.2.2 : Sensitivity of the Probability Density Function of HI pixels for the year 2012 over the entire Mediterranean Sea, for each window size. The blue line is for the 3 x 3 pixel window, the orange line is for the 5 x 5 pixel window and the green line is for the 7 x 7 pixel window. The vertical dashed lines indicate the thresholds used to distinguish weak (5 < HI < 10) and strong (HI > 10) fronts.

Window size	Weak fronts (%)	Strong fronts (%)
3x3	21.51 ± 9.51	4.18 ± 0.23
5x5	26.44 ± 5.83	3.92 ± 0.15
7x7	35.13 ± 5.46	3.74 ± 0.10

Table R1.1.1 : Sensitivity of HI-based front classification to the size of the window. Values represent the mean percentage ± standard deviation of valid pixels classified as weak (5 < HI < 10) and strong (HI > 10) fronts over a year.

R1.3 : The alternative weightings for the HI used by Liu and Levine (2016) are not mentioned or discussed (lines 173-174)

We thank the reviewer for pointing this out. Our study adopts the HI formulation proposed by Haëck et al. (2023), who explicitly describe and discuss the methodological differences with Liu and Levine (2016), including the weighting coefficients and the computation of bimodality. As our objective was not to compare alternative HI formulations, we have therefore modified

the Methods section to focus on the formulation actually used in our study and removed the reference to Liu and Levine (2016) from this section.

However, we consider the comparison with Liu and Levine (2016) relevant when interpreting our results in the broader context of previous studies. We have retained this reference in the Discussion, where we compare the Chl-a enhancements associated with SST-based fronts across studies and acknowledge differences in HI formulation, spatial scales, regions and data used and that may contribute to the differences observed. This comparison is intended to contextualize our findings rather than to assess the relative performance of the two HI formulations.

"*Our results are consistent with previous studies reporting enhanced Chl-a associated with frontal dynamics \citep{Liu2016, Haeck2023}. The magnitude of this enhancement, however, differs among studies. \citet{Haeck2023} reported an increase of more than 60\% in the vicinity of the Gulf Stream and \citet{Liu2016} found an increase of 38 \% on average, with values up to 83 \% within the North Pacific subtropical gyre. The lower Chl-a enhancements observed in our study (4-35 \%) may arise from both methodological and environmental factors. Regarding the HI methodology, we adopted the formulation of \citet{Haeck2023}, which differs from that of \citet{Liu2016} in the weighting of the HI components and the computation of bimodality. For example, \citet{Liu2016} used coefficients based on the annual maxima of the components, whereas Haëck et al. (2023) used their annual standard deviations to reduce the influence of extreme values on the weighting. These methodological differences may influence the frontal structures identified and the associated Chl-a excess values. More discussion about their technical details can be found in \citet{Haeck2023}. In addition, we used a larger window (7 x 7 pixels) to compute HI compared to \citet{Liu2016}, and we may underestimate the effect of smaller fronts on Chl-a. Finally, the North Pacific and the North Atlantic Subtropical Gyre and the Mediterranean Sea differ substantially in their physical and biogeochemical environments, which may also contribute to differences in the magnitude of the biological response to fronts.*"

R1.4 : It was not evidenced why 300 samples was the optimal number to use to assess Chl-a differences (lines 203-205)

We agree with the reviewer that the original manuscript did not provide enough details for the choice of 300 samples. In the revised manuscript, we further explain the sensitivity of the Brunner–Munzel test to the random selection of the 300 observations. For each region and season, 300 observations were randomly taken from both the frontal and non frontal Chl-a distributions, and the procedure was repeated 100 times. The resulting sensitivity analysis is summarized in Table R1.4.1 The first three columns report the region, season, and the mean Chl-a excess over fronts with its standard deviation. The fourth column reports the median p-value obtained from the 100 random realizations. The last column gives the fraction of realizations for which the Brunner–Munzel test yielded a significant difference between frontal and non frontal Chl-a distributions ($p < 0.05$). Thus, a fraction close to 100% indicates that the statistical significance is largely insensitive to the random selection of 300 observations, whereas lower fractions are associated with a greater sensitivity.

Compared with the results reported in the original version, only two cases changed, one from significant to non-significant (winter in the Alg-Tyr basin), and one from non-signicant to

significant (autumn in the Ion-Lev basin). In both cases, the lower fraction of significant realizations (27% and 66%, respectively), show that the results obtained previously resulted from the random realization. Overall, most regions and seasons show consistent significance across the 100 realizations, with significant fractions larger than 70%. These results show that our main conclusions are generally robust to the random selection of observations. Therefore, while 300 samples should not be considered as an optimal sample size, this sensitivity analysis demonstrates that 300 samples provides a sufficiently consistent basis for comparing the frontal and non frontal Chl-a distributions.

Region	Season	Excess ± STD (%)	p-value	Fraction p < 0.05
Alboran	Winter	3.9 ± 23.3	0.003	88%
	Spring	33.8 ± 32.2	<0.001	100%
	Summer	29.9 ± 21.4	<0.001	100%
	Autumn	21.9 ± 20.9	<0.001	98%
Gulf of Lion	Winter	NaN	0.182	28%
	Spring	4.5 ± 28.0	<0.001	89%
	Summer	14.2 ± 11.2	<0.001	100%
	Autumn	12.2 ± 9.6	<0.001	99%
Algerian - Tyrrhenian	Winter	NaN	0.226	27%
	Spring	16.6 ± 14.1	0.016	75%
	Summer	21.8 ± 9.7	<0.001	100%
	Autumn	13.0 ± 7.4	<0.001	100%
Ionian - Levantine	Winter	8.2 ± 7.8	<0.001	94%
	Spring	15.8 ± 9.6	<0.001	100%
	Summer	18.3 ± 8.9	<0.001	100%
	Autumn	11.0 ± 8.3	0.014	66%

Table R1.4.1 : Sensitivity of the Brunner-Munzel test to random sampling of 300 samples from frontal and non-frontal Chl-a distributions. The colour scale indicates the intensity of the Chl-a excess, with green associated to positive excesses and pink to negative excesses. The p-value column represents the median p-value obtained from the 100 random realizations. The Fraction p column gives the fraction of realizations for which the Brunner–Munzel test yielded a significant difference between frontal and non frontal Chl-a distributions (p < 0.05).

"*To assess the robustness of this choice, we repeated the random sampling procedure 100 times, drawing 300 observations from both frontal and background distributions for each region and season. The fraction of realizations yielding p < 0.05 was used to evaluate the sensitivity of the statistical significance to the random sampling. Most cases showed a high*

fraction of significant realizations, indicating that the main results were generally robust to the random selection of 300 observations. "

R1.5 : It is not made clear if, and if so how, the different datasets were put onto the same resolution grid. SST is given as 1/20 deg, SSS as 1/24 deg and FSLE as 1/25 deg.

We agree that the methods used to homogenize the spatial grids were not sufficiently described in the original manuscript. The HI was calculated from the SST and 4DMED-SEA density fields on their respective native grids and the resulting HI fields were subsequently interpolated onto the 4 km Chl-a grid. The AVISO and 4DMED FSLE fields, as well as the regionalization data based on Roy et al. (2021) and the PFT data, were also interpolated onto the Chl-a grid. Therefore, the Chl-a grid was used as the common analysis grid, while the HI calculations were performed at the original resolution of the SST and 4DMED density datasets. This procedure has now been explicitly described in the revised manuscript.

"*HI was calculated directly from the SST and density fields on their respective grids, before the resulting HI fields were interpolated onto the 4 km Chl-a grid. Similarly, the AVISO and 4DMED FSLE fields, as well as the regionalization data based on Roy et al. (2021) and the PFT data, were interpolated onto the Chl-a grid prior to the statistical analyses. Thus, the Chl-a grid was used as the common analysis grid. All variables were regridded using linear spatial interpolation onto the Chl-a grid.*"

R1.6 : Another area of discussion needed is the uncertainty on the results that quantify the anomalies of Chl-a and community structure associated with fronts. This applies for both (HI and FSLE) methods of front detection. Some of the differences are small (a few percent) but where is the threshold to regard a signal as significant? I accept that calculating this might not be straightforward but it requires at least an acknowledgement and ideally a suggestion for how future work might address this.

In the present study, statistical significance is assessed by comparing Chl-a distributions over frontal and non-frontal conditions using the Brunner–Munzel test, with $p < 0.05$. However, this statistical test does not provide an uncertainty estimate for the magnitude of the Chl-a or community structure anomaly. We have added to the Discussion :

"*While the statistical tests identify significant differences of Chl-a distribution over frontal and non-frontal conditions, statistical significance should not be interpreted as evidence that all observed anomalies are ecologically meaningful. In particular, small anomalies should be interpreted cautiously, as their uncertainty and ecological relevance cannot be fully assessed from the present analysis.*"

R1.7 : For the HI approach, there is already an aspect that can be better described. The justification for and sensitivity of results to the choice of threshold in Heterogeneity Index to decide whether a pixel is in a front are not given (lines 177-178) and should be provided.

We used HI > 10 to identify frontal pixels following the approach of Haëck et al. (2023), who defined HI < 5 as background (non frontal conditions), 5 < HI < 10 as weak fronts, and HI > 10 as strong fronts. In their study, the threshold of 10 was supported by the distribution of HI and was used to distinguish regions characterized by stronger SST heterogeneity. Similarly, we analyzed the distribution of HI and the distribution of Chl-a following HI over 1998-2021,

and we present here an example for the Alboran region (Figure R1.7.1). Most of HI values (in red) are below HI = 5 and their density decreases when HI increases, suggesting that fronts with large SST heterogeneity are less frequent. Also, Chl-a (in blue) increases with HI, especially after HI = 10. (We also investigated the bump of Chl-a occurring when HI < 3, and we think that it is largely explained by large-scale blooms occurring in spring).

In our initial analysis, we considered both weak and strong fronts in order to assess how varied the Chl-a response with frontal intensity (Figure R1.7.2). The stronger Chl-a excesses were associated with strong fronts (in plain line) compared with weak fronts (in dashed line), confirming that the response varies with frontal strength. In some cases, the Chl-a excess associated with weak fronts were equal to the background values. We finally decided to retain a single category of fronts (the strong fronts hereafter called fronts) as making the distinction between these two categories did not provide additional information that would change the interpretation of the results. We did the same approach for the FSLE-based fronts. Nevertheless, we have added these two figures to the Appendix section, to further illustrate and discuss the sensitivity of the Chl-a response to HI.

Hence, rather than considering HI > 10 as an optimal threshold determined through a sensitivity analysis, we use it as a consistent criterion to focus our study on the strongest frontal structures and their associated biological response. This choice is also consistent with the approach of Haëck et al. (2023), and with Tzortzis et al. (2021), who showed that some weak fronts may primarily act as dynamical barriers without necessarily being associated with a detectable biological enhancement.

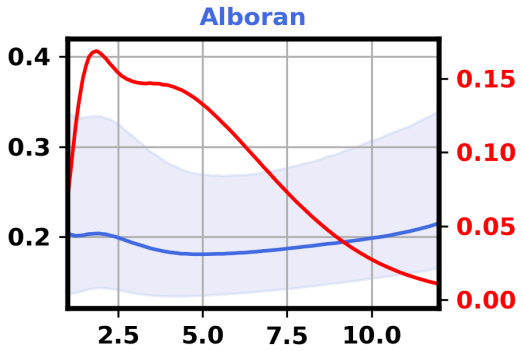


Figure R1.7.1: PDF of Chl-a following HI in the Alboran region between 1998-2021. PDF of Chl-a is in blue, and PDF of HI is in red.

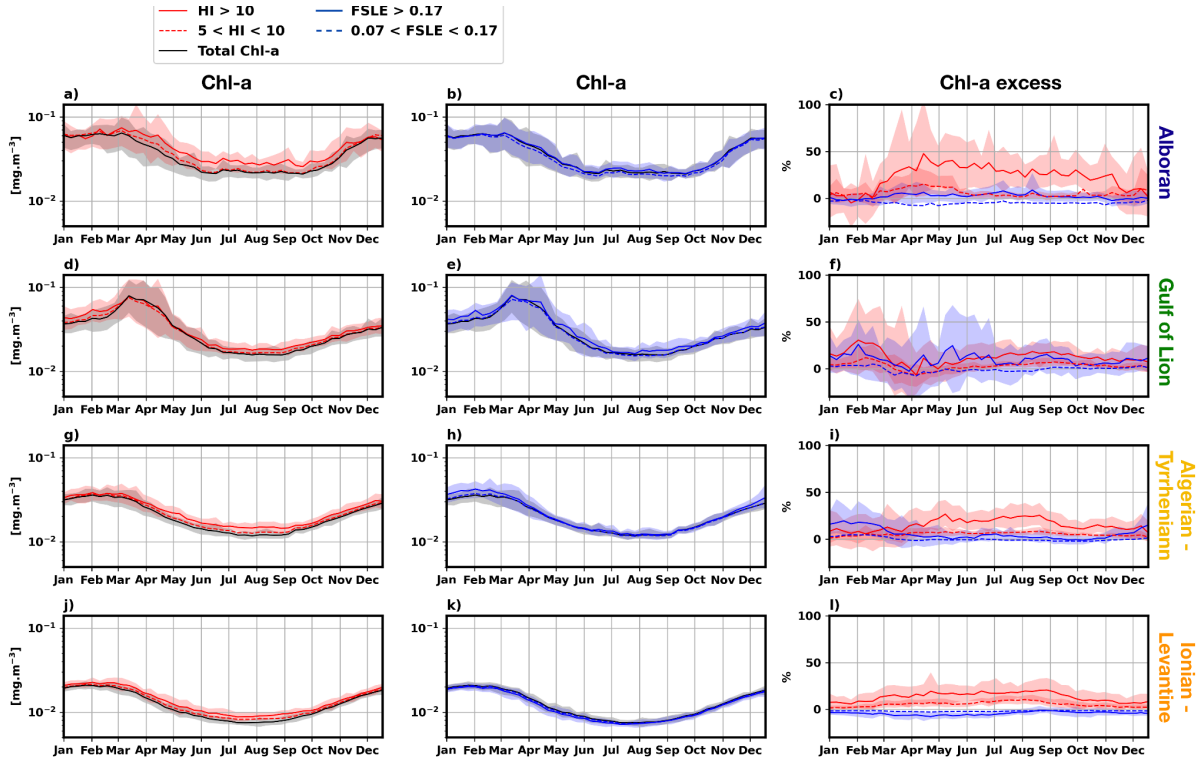


Figure R1.7.2 : Seasonal variability of Chl-a median values and excess of Chl-a averaged over the period 1998–2021, for the different regions, over weak (dashed line) and strong fronts (plain line) detected from the SST and FSLE field.

I also have a few suggestions for the discussion.

R1.8 : Might it be possible in future to distinguish whether the magnitude of anomalies associated with a front changes as a result of the front's strength or whether they are just related to the fraction of a window that is occupied by fronts?

We have added this to the discussion section :

"*Initially, we used the distinction between weak and strong fronts to assess how the Chl-a response varied with frontal intensity. The comparison suggests that the magnitude of the Chl-a excess varies with frontal strength, with larger excesses over strong fronts. However, because the HI is calculated over a finite spatial window, we cannot exclude a contribution from differences in the fraction of the window occupied by frontal pixels. Disentangling these two effects would require a dedicated analysis and represents an interesting avenue for future work.*"

R1.9 : How is the agreement of the results with those from two previous papers (lines 296-301) affected by the fact that those two papers used different parameter values when calculating HI?

We addressed this question with the remark **R1.3** and we have clarified to the Discussion section**:**

“Our results are consistent with previous studies reporting enhanced Chl-a associated with frontal dynamics \citep{Liu2016, Haeck2023}. The magnitude of this enhancement, however, differs among studies. \citet{Haeck2023} reported an increase of more than 60\% in the vicinity of the Gulf Stream and \citet{Liu2016} found an increase of 38 \% on average, with values up to 83 \% within the North Pacific subtropical gyre. The lower Chl-a enhancements observed in our study (4-35 \%) may arise from both methodological and environmental factors. Regarding the HI methodology, we adopted the formulation of \citet{Haeck2023}, which differs from that of \citet{Liu2016} in the weighting of the HI components and the computation of bimodality. For example, \citet{Liu2016} used coefficients based on the annual maxima of the components, whereas Haëck et al. (2023) used their annual standard deviations to reduce the influence of extreme values on the weighting. These methodological differences may influence the frontal structures identified and the associated Chl-a excess values. More discussion about their technical details can be found in \citet{Haeck2023}. In addition, we used a larger window (7 x 7 pixels) to compute HI compared to \citet{Liu2016}, and we may underestimate the effect of smaller fronts on Chl-a. Finally, the North Pacific and the North Atlantic Subtropical Gyre and the Mediterranean Sea differ substantially in their physical and biogeochemical environments, which may also contribute to differences in the magnitude of the biological response to fronts.”

R1.10 : Can the authors provide anything beyond the current short statement (lines 321-322) of stratification, nutricline depth and depth of front circulation as possible causes for the signal they see? For example, does subsurface silicate vary across the Mediterranean in a way consistent with the signal seen with diatoms?

The observed enhancement of diatoms at fronts is consistent with a mechanism involving the enhancement of surface nutrients, but our observations do not allow us to identify which nutrient is primarily responsible.

R1.11 : Can the authors offer any suggestions for what the next steps should be beyond calling for longer time series? How can we start to separate strength of front, lifetime of front, underlying nutrient concentrations etc to have a causal understanding of what we see?

We agree that the mechanisms underlying the observed regional and seasonal differences deserve further discussion. The enhancement of diatoms at fronts is consistent with enhanced nutrient supply, but the satellite observations used here are restricted to the surface and do not provide information on the vertical structure of the fronts, nutricline depth, or vertical nutrient fluxes. We therefore cannot determine whether the observed response is controlled by silicate, nitrate or other nutrient limitations, nor can we directly relate it to the depth of frontal circulation. We have clarified this limitation in the revised Discussion and avoided attributing the observed patterns to a specific nutrient mechanism.

“Moving from the statistical relationships identified here toward a causal understanding will require combining observations with process-resolving numerical models. Satellite observations provide the spatial and temporal coverage needed to characterize the surface biological response to fronts, but they do not resolve the subsurface processes that ultimately control nutrient supply and phytoplankton growth. A promising next step would therefore be to use numerical models, evaluated against the satellite relationships documented here, to explicitly resolve the three-dimensional frontal circulation, vertical

nutrient fluxes and phytoplankton response. Such an approach would make it possible to separate the respective roles of front intensity and lifetime, background nutrient conditions and nutricline depth, and to test the mechanisms responsible for the regional and seasonal patterns identified in this study."

There are also a few minor presentation issues:

R1.12 :The decision to have one result using HI and one using FSLE on each of Figures 3 and 4 rather than both HI ones on one and both FSLE ones on the other seems counter-intuitive.

We thank the reviewer for this remark, we therefore changed the organization of our figures. Here an example for the SST-based and density based fronts :

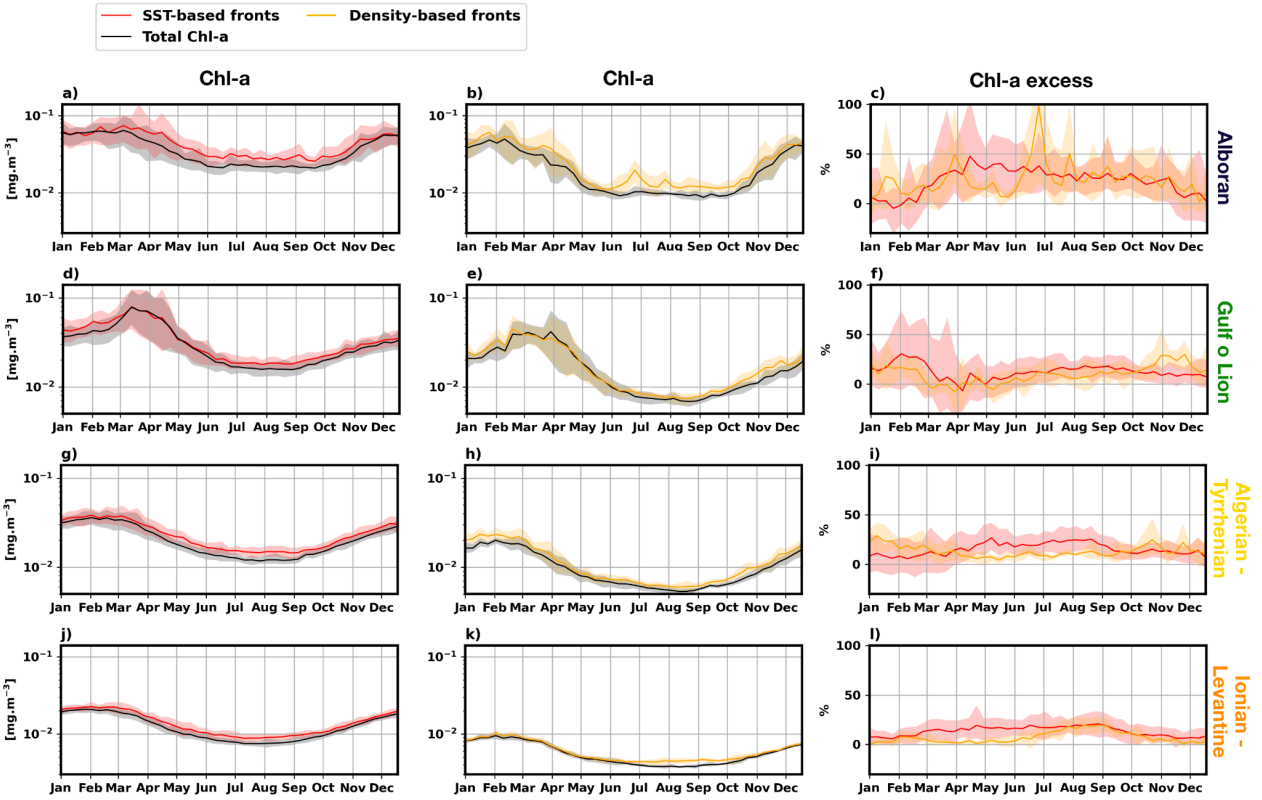


Figure R1.7.2 : Seasonal variability of Chl-a median values and excess of Chl-a, for the different regions, over fronts detected from the SST field (in red) and density fields (in orange). Chl-a and Chl-a excess are averaged over 1998-2021 for the SST-based fronts and over 2016-2021 for the density-based fronts.

R1.13 : I think a reader would appreciate an example image showing how the different methods identify fronts in a region.

We thank the reviewer and we therefore added a new figure :

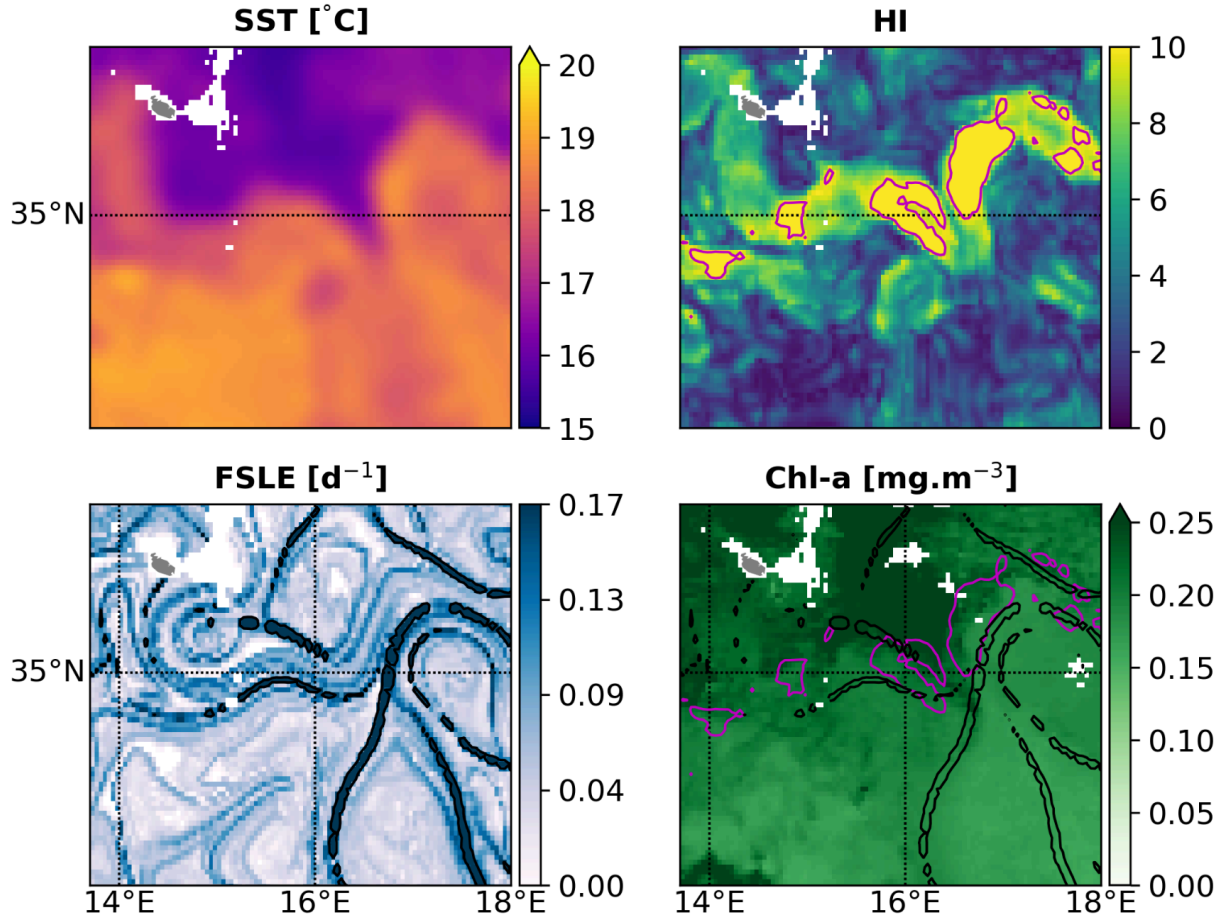


Figure R1.13: Snapshots of SST, HI, FSLE and Chl-a on 19 January 2011. The line contour in magenta corresponds to the HI values of 10, and the line contour in dark corresponds to the FSLE values of 0.17 d^{-1}.

R1.14 : Line 161 – "implies to compute"

Review 2

General Comment

This study quantifies changes in surface phytoplankton biomass and community composition over fronts in the Mediterranean Sea. The authors find that biomass is consistently enhanced over fronts and diatoms are favored, but with regional and seasonal variability such that the strongest effects were found in wintertime and oligotrophic regions. Although focusing on a different region, these results and SST-frontal identification methods closely mirror those of the North Atlantic region presented in previous studies by the authors (Lévy et al. 2025 and Haëck et al., 2023). From my perspective, the most interesting and novel component of this study is comparing whether phytoplankton have a stronger response to SST-derived, density-derived, or FSLE-derived fronts. Unfortunately, this is the component of

the paper for which I have the most methodological concerns. I believe the authors would produce a more convincing analysis if they calculated FSLE independently rather than relying on existing datasets so that they could conduct controlled tests, changing only one factor at a time (e.g., input velocity grid resolution, initial Lagrangian particle separation, final particle separation). I have many other concerns regarding the interpretation of FSLE, listed in the comments attached. I believe that this study has potential, but the manuscript needs considerable improvements. Therefore, I cannot recommend this pre-print for publication primarily due to its incremental novelty from the authors' previous work, insufficient methodological descriptions (Major Comment 1), and major methodological concerns (Major Comment 2). The manuscript would also benefit from English language editing, and to that end I provide many recommendations in the Technical Comments attached.

We thank the reviewer for their careful reading of our manuscript and for their constructive and helpful comments. We have considered each remark and revised the manuscript accordingly. Below, we provide a point-by-point response to all comments and indicate the corresponding changes made to the manuscript.

Major Comments

Methodological details are unclear or neglected in many instances:

R2.1 : Lines 92-93: Can the authors clarify if the spatial interpolation of the SST over cloud-covered regions is part of the product, or if they did it themselves? If the interpolation was done by the authors, what percentage of the pixels were interpolated?

We thank the reviewer for pointing out this ambiguity. The spatial interpolation over cloud-covered areas is part of the SST product processing and was not performed by us. We have clarified this point in the revised manuscript.

" *The SST product provides spatially interpolated fields over areas affected by cloud cover, yielding a spatially complete daily product.*"

R2.2 : Lines 100-101: Can the authors clarify if they are applying the "multivariate optimal interpolation algorithm", or if this was done previously by Sammartino et al.? Either way, an explanation is needed to describe what this means. It is also unclear why the authors describe a depth-resolved machine learning algorithm to derive density, if they only end up using surface data from satellite and in situ data.

The multivariate optimal interpolation algorithm is part of the 4DMED-Sea product and previously done by Sammartino et al. (2022). It follows the approach described by Droghei et al. (2016) and adapted for the Mediterranean Sea by Sammartino et al. (2022). The multivariate optimal interpolation is a statistical method that combines observations from different sources to produce a spatially continuous field while accounting for their spatial covariance and associated uncertainties. For the density product, it combines satellite-derived observation and in-situ measurements of sea surface salinity, and high resolution SST data.

We also agree that the description of the depth-resolved density reconstruction was unnecessary and potentially confusing, since only the surface density field is used in our analysis. We have therefore removed this information.

"*Density fields are available for the entire Mediterranean basin, on a regular grid with a spatial resolution of 1/24°, and a time interval of 1 day, from 1 January 2016 to 1 August 2022, and are part of the 4DMed-Sea project. These fields were in fact obtained through a multivariate optimal interpolation algorithm specifically designed to combine satellite UHR SST data, sea surface salinity images from various satellite sources and in situ measurements (Sammartino et al., 2022). Input satellite-derived density images are obtained by remapping both UHR SST and satellite salinity measurements over the final grid and applying the equations of state for seawater density. Classical optimal interpolation provides gap-free fields by averaging observations with weights that depend on spatial and temporal covariance. The multivariate algorithm used in Sammartino et al. (2022), originally introduced by Droghei et al. (2016), includes an additional decorrelation factor that increases/decreases observations' weights based on their thermal differences with respect to the interpolation point. In practice, density observations are weighted more if they lie on the same isothermal as the target point. This strategy exploits the physical covariance between temperature and salinity/density and forces interpolated patterns to reproduce more consistent water-mass distributions at the small scales than classical algorithms, effectively increasing the resolution of the interpolated surface fields. Independent density information remains at scales resolved by the satellite and in situ data, but the high-resolution texture is inherited from SST. Full details on the algorithm, including spatial spectral analysis and high-resolution validation based on thermosalinograph data, can be found in Sammartino et al. (2022).*"

R2.3 : Lines 104-112: I recommend providing an equation for FSLE in the paper.

The FSLE equation has now been added to the revised manuscript.
"*\mathrm{FSLE} = \frac{1}{\tau}\ln\left(\frac{\delta_f}{\delta_0}\right)*"

R2.4 : Additionally, these methods are hard to follow when the text switches back and forth with some numbers provided as decimals and others as fractions. If I understand correctly, the input velocity fields are coarser for the AVISO product (1/4 degree) compared to the 4DMED-SEA product (1/24 degree), and the choices for the Lagrangian particle initializations are also different. For AVISO, the particles are seeded at 1/50 (0.02) degree and FSLE is calculated as the time at which the particles spread 30 times the initialization (0.6 degree). The 4DMED-SEA product initializes at 1/24 degree, and calculates FSLE at the time at which particles spread 30 times the initialization (30/24 degree). Wouldn't this result in the detection of entirely different sized fronts, such that the ones in 4DMED-SEA are much larger? Do both datasets use the same interpolation schemes and timestepping methods? Is 4DMED-SEA also running particle trajectories backward-in-time? These choices all could affect the conclusion of which front identification method best highlights the fronts that result in a biological response.

We have taken this comment into account and have revised the way in which the two products are presented. We agree that the different velocity fields, spatial resolutions, particle-separation parameters, and trajectory configurations are important when interpreting

the comparison. We have therefore revised the Discussion to discuss these methodological differences and reinforce the idea that we do not look for the best product at detecting biologically relevant fronts, but rather to test the robustness of the front–biology relationship when using different dynamical diagnostics.
Both products use backward-in-time FSLEs, but they are derived from different velocity fields and use different Lagrangian configurations. For the AVISO product, the input geostrophic velocity field has a spatial resolution of 1/4° and the resulting FSLE field is provided at 1/25°. Particle trajectories are integrated backward-in-time using a fourth-order Runge–Kutta scheme, with bilinear interpolation in space, with a 6 hour integration time step. For the 4DMED-FSLE product, the input velocity field has a spatial resolution of 1/24°. The trajectories are also integrated backward-in-time using a fourth-order Runge–Kutta scheme, a bilinear spatiotemporal interpolation, a 1 hour time step. The two products use the same separation factor, $\delta f/\delta 0 = 30$, but different absolute separation scales. For AVISO, $\delta 0 = 1/50°$ (= 0.02°) and $\delta f = 0.6°$, whereas for 4DMED-FSLE, $\delta 0 = 1/24°$ (0.0417°) and $\delta f = 30/24° = 1.25°$.

Hence, the 4DMED-SEA velocity field resolves a broader range of smaller-scale dynamical variability before the Lagrangian diagnostic is applied. In contrast, the AVISO-FSLE uses a coarser velocity field input, with the small-scale structure of the resulting FSLE field emerging from the choice of the initial particle separation. The two products reach fine-scale FSLE structures through different pathways: 4DMED starts from a higher-resolution velocity field that already represents smaller-scale variability, whereas the AVISO product derives fine-scale Lagrangian structures from a coarser velocity field through the trajectory-based FSLE configuration. We added in the FSLE identification method section :

“We used *different FSLE products to test the robustness of the relation between fronts and phytoplankton response. They differ in the resolution of their velocity fields input and in their Lagrangian configuration. The AVISO product is derived from a 1/4° geostrophic velocity field, but uses a relatively small initial particle separation (δ0 = 1/50°), allowing to capture subgrid dynamics and the representation of finer filaments. In contrast, the 4DMED-FSLE is based on a higher-resolution velocity field (1/24°), which explicitly represents a broader range of smaller-scale dynamical variability before the Lagrangian diagnostic is applied. Hence, the two products reach fine-scale FSLE structures through different pathways: 4DMED-FSLE starts from a higher-resolution velocity field that already represents smaller-scale variability, whereas the AVISO product derives fine-scale Lagrangian structures from a coarser velocity field through the trajectory-based FSLE configuration. These differences provide a useful test of whether the observed front–phytoplankton relation persists when the dynamical diagnostic is changed.* ”

R2.5 : Lines 117-123: Can the authors comment on if the phytoplankton algorithm was validated in the Mediterranean Sea? Does it work effectively in coastal, oligotrophic, and eutrophic systems? I find it curious that the derived community structures look very similar between the oligotrophic and eutrophic regions (Fig 1c).

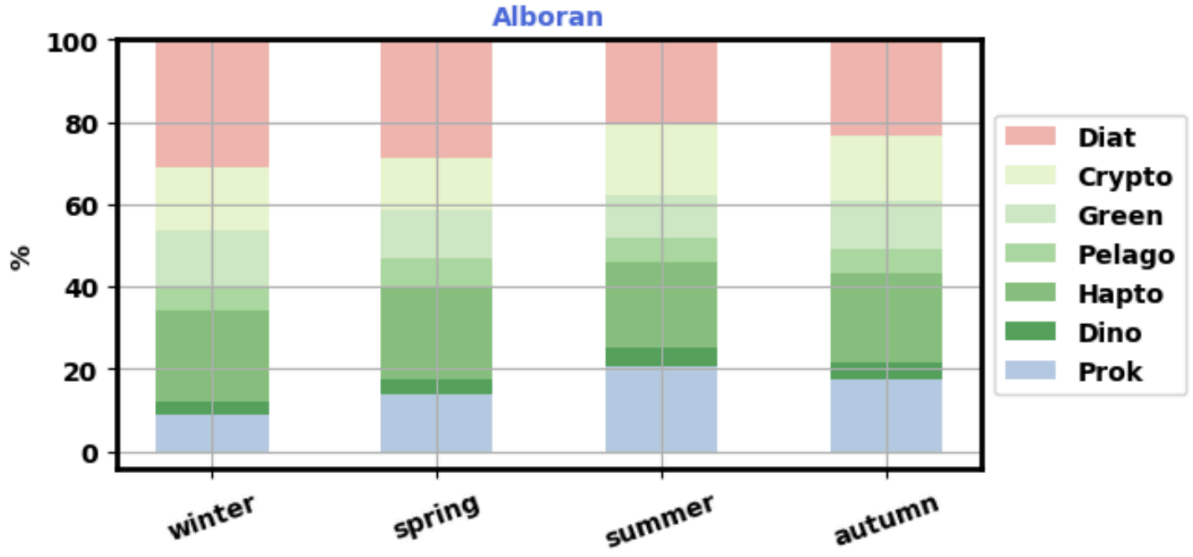
We thank the reviewer for raising this point. The phytoplankton pigment retrieval used in our study is based on the Self-Organizing Map (SOM) algorithm developed by El Hourany et al. (2019a). It was trained using a large database of in situ HPLC pigment measurements

collocated with satellite observations and showed robust performance in retrieving the concentrations of the different diagnostic pigments at the global scale. This approach was then adapted for the Mediterranean Sea by El Hourany et al. (2019b). The Mediterranean SOM-PFT was evaluated against independent in situ HPLC observations collected throughout the basin. The cross-validation of the pigment retrieval yielded R^2 values ranging from 0.62 to 0.85, with an average RMSE of 0.015 mg m^{-3}. The derived PFTs were also evaluated against in situ pigment-based classifications, with for example agreement for diatoms (94.6%) and Synechococcus (75.5%). The temporal variability of the reconstructed PFTs was also evaluated against the long-term BOUSSOLE time series in the Ligurian Sea. This reference has been added in the revised manuscript.

El Hourany et al. (2019b) also showed that the method capture the ecological gradient of the basin, including the predominance of picophytoplankton in oligotrophic waters and the occurrence of diatoms and other larger phytoplankton groups in more productive and coastal environments.
Regarding the relatively similar community structures between regions in Figure 1c, we believe this mainly reflects the fact that the figure represents the mean PFT composition over the full study period. The Mediterranean PFT study shows pronounced seasonal and regional variability. These temporal changes can be smoothed when averaged over the full study period (Figure R2.6.1).

We modified in the Data section :
"*The phytoplankton pigment retrieval used is based on the Self-Organizing Map (SOM) algorithm developed by \citet{ElHourany2019} for the global-scale, and then adapted and evaluated specifically for the Mediterranean Sea by \citet{ElHourany2019b}.*"

Reference : El Hourany, R., Abboud-Abi Saab, M., Faour, G., Mejia, C., Crépon, M., & Thiria, S. (2019). Phytoplankton diversity in the Mediterranean Sea from satellite data using self-organizing maps. *Journal of Geophysical Research: Oceans*, 124, 5827–5843. https://doi.org/10.1029/2019JC015131



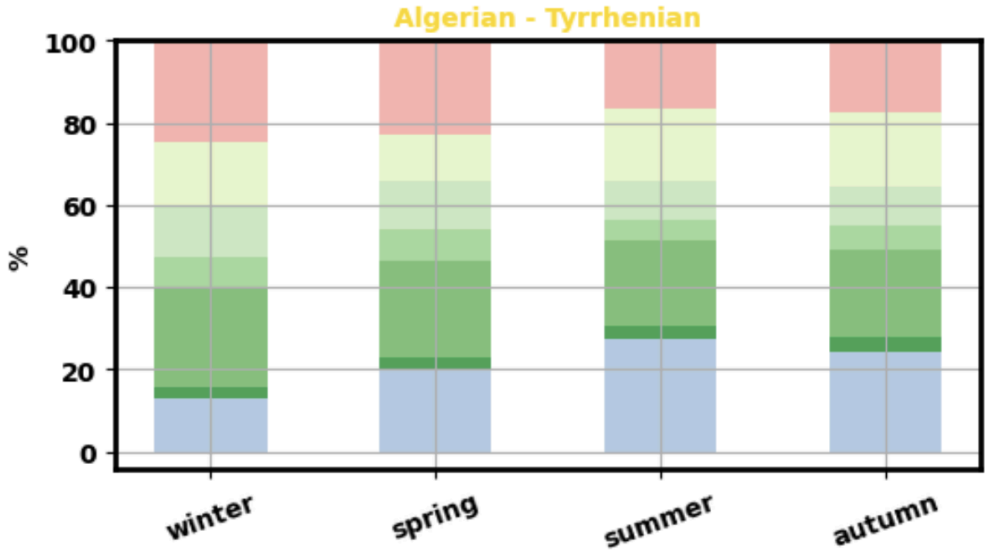


Figure R2.6.1 : Seasonal average of phytoplankton assemblages from pigment-derived phytoplankton groups data for the Alboran and the Algerian-Tyrrhenian regions.

R2.6 : Lines 134-137: How was mixed layer depth calculated? How was the clustering "simplified"?

Regarding the MLD, El Hourany et al. (2021) used the global monthly mixed layer depth (MLD) climatology of Holte et al. (2017). It is based on approximately 1.5 million Argo temperature profiles collected between January 2000 and April 2018. The MLD was determined from these profiles using the method of Holte and Talley (2009). Monthly MLD values were then averaged within 1° × 1° boxes over the 2003–2020 period to obtain the climatological MLD used in the bioregionalization.

Concerning the clustering, El Hourany et al. (2021) identified seven bioregions (C1–C7, Figure 2.6.1). For our study, we simplified the original bioregionalization to obtain four large-scale regions adapted for our regional analysis. First, C7 was excluded because it represents a heterogeneous class largely associated with coastal waters and shallow areas, which are strongly influenced by processes such as river discharge and sediment resuspension. We also excluded the Adriatic and Aegean Seas because these regions exhibit particularly complex and heterogeneous physical and biogeochemical characteristics. This is also apparent from the original regionalization, in which the Adriatic alone is divided among several bioregions (C1, C2, C5, and C7). After these exclusions, four main bioregions remained and were used as the basis for our study: C2, C3, C5, and C6. We then adjusted the boundaries of these classes to obtain four geographically coherent regions corresponding to the Alboran, Gulf of Lion, Algerian–Tyrrhenian and Ionian–Levantine regions. These adjustments were deliberately limited to merging or reassignment of neighboring areas rather than redefining the underlying physical/biogeochemical classification. For example, the Rhodes Gyre, which is identified as C5 in the original regionalization, was included in our Ionian–Levantine region (region C6) to maintain a coherent regional framework for our analysis. We have revised the manuscript to describe this procedure and to make clear that the four regions are a simplified geographical grouping

of the original El Hourany et al. (2021) bioregionalization, rather than the result of a new clustering analysis.

"*We then simplified the original seven-bioregion classification to define four larger and geographically coherent regions. We did not perform a new clustering, but rather regrouped and adjusted the original bioregions according to their geographical coherence and the objectives of this study. The bioregion C7 was excluded because coastal phytoplankton dynamics are strongly influenced by river runoff and continental-shelf processes. We also excluded the Adriatic and Aegean Seas because of their complex and heterogeneous physical and biogeochemical characteristics, which are reflected in the presence of several distinct bioregions in these areas. The remaining bioregions were regrouped and, if necessary, their boundaries were adjusted to delineate four large regions: the Alboran (light blue), the Gulf of Lion (green), the Algerian-Tyrrhenian (yellow), and the Ionian-Levantine (orange) regions. For example, the area of the Rhodes Gyre, which belongs to C5 in the original classification, was included in our Ionian–Levantine region to maintain geographical coherence within this large-scale regionalization. Finally, coastal pixels with a water depth shallower than 100 m were excluded from the analysis.*"

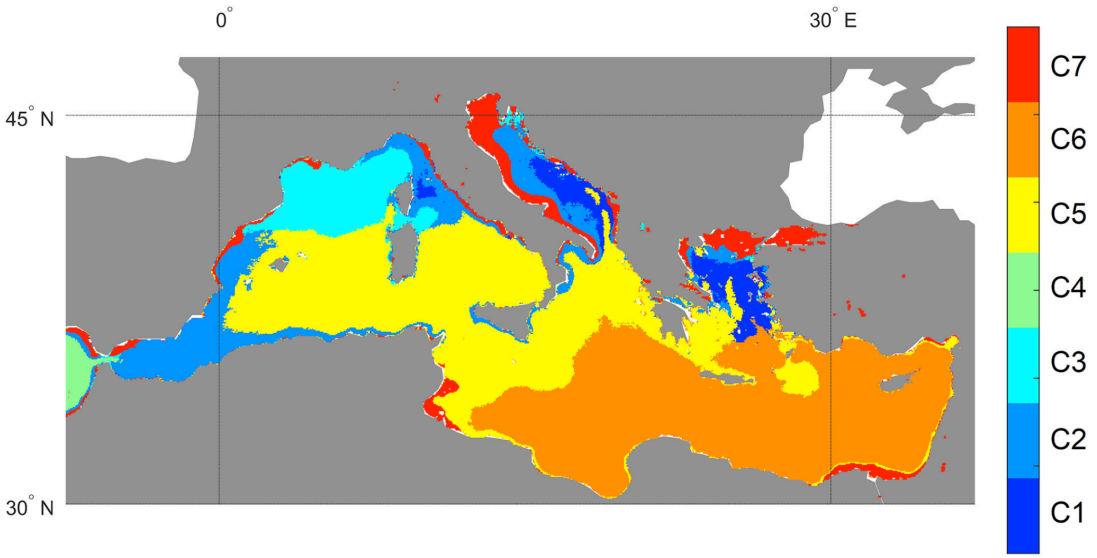


Figure R2.6.1 : Original regionalization from El Hourany et al. (2021).

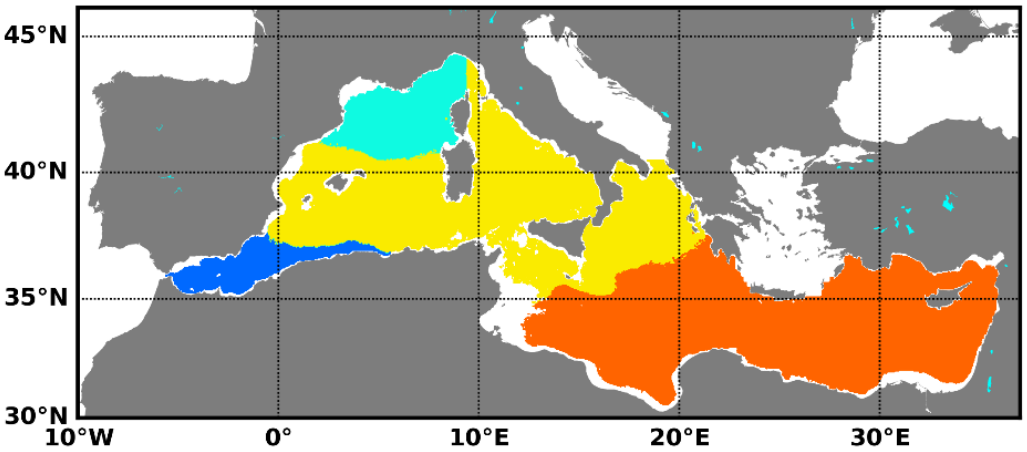


Figure R2.6.2 : Final regionalization used in our paper.

R2.7 : Lines 157-158: Can the authors clarify how “all fronts are projected onto a common grid”?

We agree that the steps used to homogenize the spatial grids were not sufficiently described in the original manuscript. The HI was calculated from the SST and 4DMED density fields on their respective native grids and the resulting HI fields were subsequently interpolated onto the 4 km Chl-a grid. The AVISO and 4DMED FSLE fields, as well as the regionalization data based on Roy et al. (2021) and the PFT data, were also interpolated onto the Chl-a grid. Therefore, the Chl-a grid was used as the common analysis grid, while the HI calculations themselves were performed at the native resolution of the underlying SST and 4DMED density datasets. This procedure has now been explicitly described in the revised manuscript.

“*HI was calculated directly from the SST and density fields on their respective native grids, before the resulting HI fields were interpolated onto the 4-km Chl-a grid. Similarly, the AVISO and 4DMED FSLE fields, as well as the regionalization data based on Roy et al. (2021) and the PFT data, were interpolated onto the Chl-a grid prior to the statistical analyses. Thus, the Chl-a grid was used as the common analysis grid. All variables were regridded using linear spatial interpolation based on their latitude and longitude coordinates onto the Chl-a grid. No temporal interpolation was applied.*”

R2.8 : Line 170-171: “the 7x7 pixels window was the most optimized choice for our study”; what was the logic to draw this conclusion?

This point has also been made by the reviewer at the comment **R1.2**. Please refer to the response we provided.

R2.9 : Line 174: What does the “L2 norm of the difference” mean here ?

Following Haëck et al. (2023), the bimodality consists of comparing the observed SST histogram with a Gaussian fit of the same histogram. More specifically, for each SST bin (i), we calculate the difference between the histogram value (H_i) and the corresponding value of the Gaussian fit (G_i), and compute the L2 norm as (B = (\sum_i (H_i - G_i)^2) ^{1/2}). Thus, (B) provides a measure of how the SST distribution differs from a Gaussian distribution, with larger values indicating stronger non-Gaussianity and bimodality. We have clarified this definition in the revised manuscript.

"*The bimodality (B) is defined as the L2 norm of the difference between the SST histogram and a Gaussian fit of the histogram. For each SST bin (i), we compute the difference between the histogram value (H_i) and the corresponding value of the fitted Gaussian (G_i), and define (B) as (B = (\sum_i (H_i-G_i)^2)^{1/2}). Thus, (B) quantifies how the SST distribution differs from a Gaussian distribution, with larger values indicating stronger non-Gaussianity and, in particular, bimodality.*"

R2.10 : Line 177-178 & 197-200: At first, the authors claim that HI index > 10 is used to delineate the fronts from the background circulations, but then give further explanation later. I think these sections should be combined into one coherent thought and provide a quantitative assessment of the Chl-a over the strong and weak fronts ("close to" is vague). Can one standalone pixel be considered a front by the HI definition? What is the minimum size of the fronts?

We thank the reviewer for this comment. We have combined the different elements into one coherent thought in the revised manuscript.
Concerning the strong and weak HI-based fronts discussion, please refer to the response made to the comment **R1.7** of the reviewer 1.

Concerning the standalone pixels and spatial extent of the detected fronts remarks, the HI algorithm is a pixel-based diagnostic. Following Haëck et al. (2023), HI is calculated for each pixel from the SST heterogeneity within a 7 × 7 pixel window centered on that pixel. Thus, the method identifies pixels located in areas of enhanced SST heterogeneity rather than delineating individual frontal structures with a prescribed minimum area or width. Consequently, an isolated pixel can satisfy the HI > 10 criterion, although this should be interpreted as a pixel located over a strong frontal signal rather than as a complete front by itself. The 7 × 7 window defines the spatial scale over which the SST heterogeneity is evaluated, rather than imposing a minimum size on the resulting front. This point has been added in the method part about the limits of the method.

"*A pixel satisfying the HI > 10 criterion indicates that the pixel is located within a region characterized by enhanced SST heterogeneity, but it does not necessarily represent an entire coherent frontal structure by itself. In particular, the method does not impose a minimum spatial extent or width for a front. Consequently, isolated strong-HI pixels can occur without forming a spatially continuous front. This should be kept in mind when interpreting the frontal Chl-a signal, which represents the response of pixels located over strong frontal signatures rather than the response of individually delineated frontal structures.*

R2.11 : Lines 180-184: Line 180: If using backward-in-time trajectories, doesn't the FSLE identify convergence, not stretching? I suggest clarifying if the previous studies cited used

this FSLE threshold in the same domain as this study. What are the units of this threshold (1/time)? What sensitivity tests were conducted? What was the logic to choose an "optimal value"? As with the HI, can one standalone pixel of high FSLE be considered a front? If so, this may be very sensitive to numerical instability (for example, see Reijnders et al. https://doi.org/10.1029/2025MS005470)

We have clarified the description of the FSLE diagnostic and the choice for the threshold used in the analysis.

We used a threshold of 0.17 d^{−1}, which is within the range of values previously used to identify strong Lagrangian structures in the Mediterranean Sea. For example, d'Ovidio et al. (2004) reported Mediterranean FSLE values of approximately 0.1–0.6 d^{−1}, while De Monte et al. (2012) used 0.1 d^{−1} to identify fronts. Recently, Sudre et al. (2023) used region-specific FSLE thresholds in the Mediterranean Sea, with values ranging from approximately 0.09 to 0.25 d^{−1}. Our value of 0.17 d^{−1} therefore falls within the range of thresholds previously applied in the Mediterranean. We agree that describing this threshold as an "optimal value" would be too strong, and we have removed this word. More details are provided for the Major comment **R2.15** (below)**.**

We also did a sensitivity analysis, described in detail below in response to the next comment **R2.15**, using thresholds of 0.10, 0.17, and 0.25 d^{-1} over 10 years. This analysis was designed to assess the robustness of frontal occurrence and of the associated biological response to the threshold choice.

As for the HI analysis, we initially distinguished between weak and strong FSLE fronts, using a lower threshold of 0.07 d^{−1} for weak fronts, and compared their Chl-a distributions with the total Chl-a (Figure R2.15.2). We finally decided to retain a single category of fronts (the strong fronts hereafter called fronts) as making the distinction between these two categories did not provide additional information that would change the interpretation of the results. More details are provided for the Major comment **R2.15** (below)**.** We have added Figure R2.15.2 to the Appendix section, to further illustrate and discuss the sensitivity of the Chl-a response to the HI and FSLE values.

Finally, as for the previous point about HI, we agree that a single high-FSLE pixel should not be interpreted as an entire frontal structure. Strong FSLE values are generally located along filaments that delineate attracting Lagrangian structures, where isolated high-FSLE pixels may also occur. Pixels exceeding the threshold should be interpreted as locations associated with fronts signatures, and quantify the biological response statistically over these pixels rather than considering each pixel as an individual front.

"*FSLEs identify regions of strong convergence and are widely used to detect frontal structures and transport barriers. We used a threshold of 0.17 d^{−1} to identify pixels lying into frontal structures. Previous Mediterranean studies used FSLE thresholds within a similar range. For example, d'Ovidio et al. (2004) reported FSLE values of approximately 0.1–0.6 d^{−1} while De Monte et al. (2012) used a threshold of 0.1 d^{−1}. Recently, Sudre et al. (2023) used region-specific thresholds ranging approximately from 0.09 to 0.25 d^{−1}. Thus, our threshold falls within the range of thresholds previously applied in the Mediterranean Sea. Note that our initial approach was to distinguish between weak and*

strong FSLE fronts, with a lower threshold of 0.07 0.1 d^{−1} for weaker fronts. We finally decided to retain only one category but the results for weaker fronts are presented in the Appendix section."

R2.12 : Lines 193-194: How is "total chl-a" defined? Regional average? Seasonal average? Whole domain?

By "total Chl-a", we refer to the Chl-a without applying any front classification, and considering all valid pixels within each region. For each day between 1998 and 2021, we computed the spatial mean Chl-a over all valid pixels of the corresponding region, independently of whether pixels were classified as fronts. We then averaged these daily regional values over the entire 1998–2021 period. The resulting seasonal signal therefore represents the average seasonal variability of Chl-a per region over the 24-year period.

"*Total Chl-a refers to all valid Chl-a pixels without applying any front classification. For each day between 1998 and 2021, we calculated the spatial mean Chl-a over each region and then averaged the daily values over the full study period to obtain a multi-year daily climatology. The resulting seasonal cycle therefore represents the mean seasonal variability of regional total Chl-a over 1998–2021 or 2016-2021.*"

R2.13 : Line 205: How big are the sample pools from which the draw of 300 comes from? How was 300 chosen? A bootstrapping analysis would strengthen the statistical analysis here (e.g., redrawing 300 samples many times and computing a confidence interval for the statistic).

We thank the reviewer for this comment. A resume of the size of each pool has been summarized by the table R2.13, with the column N_front. The size of the pool depends on the size of the regions and the cloud coverage.

Concerning the justification about the choice of 300 and more information about the statistical test, please refer to the response that we provided to the **R1.4** remark of the reviewer 1.

Region	Season	N_front	Excess ± STD (%)	p-value	Fraction p < 0.05
Alboran	Winter	96342	3.9 ± 23.3	0.003	88%
	Spring	207402	33.8 ± 32.2	<0.001	100%
	Summer	838261	29.9 ± 21.4	<0.001	100%
	Autumn	786532	21.9 ± 20.9	<0.001	98%
Gulf of Lion	Winter	243157	NaN	0.182	28%
	Spring	198467	4.5 ± 28.0	<0.001	89%
	Summer	591460	14.2 ± 11.2	<0.001	100%
	Autumn	953037	12.2 ± 9.6	<0.001	99%
Algerian - Tyrrhenian	Winter	513736	NaN	0.226	27%
	Spring	676410	16.6 ± 14.1	0.016	75%
	Summer	1118904	21.8 ± 9.7	<0.001	100%
	Autumn	907946	13.0 ± 7.4	<0.001	100%
Ionian - Levantine	Winter	850055	8.2 ± 7.8	<0.001	94%
	Spring	500723	15.8 ± 9.6	<0.001	100%
	Summer	967880	18.3 ± 8.9	<0.001	100%
	Autumn	1129789	11.0 ± 8.3	0.014	66%

Table R2.13 : Extension of the Table R1.4.1, Sensitivity of the Brunner-Munzel test to random sampling of 300 samples from frontal and non-frontal Chl-a distributions. The colour scale indicates the intensity of the Chl-a excess, with green associated to positive excesses and pink to negative excesses. The N-frontal column indicates the total number of valid pixels available over frontal conditions. The p-value column represents the median p-value obtained from the 100 random realizations. The Fraction p column gives the fraction of realizations for which the Brunner–Munzel test yielded a significant difference between frontal and non frontal Chl-a distributions (p < 0.05).

R2.14 : Line 211-213: How was the phytoplankton group anomaly computed? Same as chl-a? Is the median used for chl-a, but mean for the groups ?

The phytoplankton pigment-derived data provide the relative proportion (%) of each phytoplankton group rather than a concentration. For each day, we calculated the regional spatial mean proportion of each group over pixels identified as frontal conditions and compared it with the corresponding spatial mean over the regional domain (without any front classification) to compute the frontal anomaly. Hence, we computed the spatial median for Chl-a and the spatial mean for phytoplankton pigment-derived data.

"*We also investigated the local impact of fronts on phytoplankton community structure. For each day, we calculated the regional spatial mean proportion of each phytoplankton group over pixels identified as frontal conditions and compared it with the corresponding spatial mean over the regional domain (without any front classification) to compute the frontal anomaly. The frontal anomaly for each group was then calculated as the difference between its mean proportion within fronts and its mean proportion over the region.* "

R2.15 : Major comment 2 : The way the FSLE is applied to define fronts from a single threshold is concerning to me especially given that the density of occurrence is 100% in parts of the Alboran region, and generally much higher than the HI methods throughout the entire domain (Figure 2d). As mentioned in my earlier comment, I believe that the AVISO and 4DMED datasets are detecting differently sized fronts. Using a single threshold for defining fronts from the FSLE throughout the entire domain and for both datasets is also not sufficiently justified beyond citing previous papers and claiming a sensitivity test was conducted without showing those results. This is especially important given that one of the major conclusions of this paper is that the HI method is superior to the FSLE method in identifying fronts with biological reactions. Such a claim goes against previous studies of the region that suggest using both together (Sudre et al. 2023: https://www.nature.com/articles/s41597-023-02615-z; d'Ovidio et al. 2009: https://doi.org/10.1016/j.dsr.2008.07.014 (a highly relevant study that is not cited or discussed in this paper)). The authors speculate that the FSLE is not correlated with the biological response because of the "effective resolution" (Line 256, 305), without much support (no citations, e.g., Yang 2025: https://link.springer.com/article/10.1007/s41064-024-00318-7).

We thank the reviewer for raising this important point. We agree that two distinct issues need to be clarified: (i) the sensitivity of the results to the FSLE threshold used to define fronts, and (ii) the comparison between the FSLE and HI diagnostics. We will repeat some details that have been already addressed in the previous responses, but we believe it is important to present the full process coherently in light of the relevant reviewer's comment.

(i) Our analysis initially distinguished between weak FSLE fronts (0.07 < FSLE < 0.17 d^{-1}) and strong FSLE fronts (FSLE > 0.17 d^{-1}). The threshold values were selected based on the range of FSLE values reported in previous studies of Mediterranean Sea, as well as the probability density functions (PDF) of Chl-a regarding FSLE values (Figure R2.15.1). D'Ovidio et al. (2004) reported FSLE values of approximately 0.1–0.6 d^{-1}, while Sudre et al. (2023) used region-specific thresholds ranging from 0.119 d^{-1} in the Western Mediterranean to 0.254 d^{-1} in the Alboran Sea. Concerning the PDF, most of FSLE values are below 0.07 d^{-1}) and their density decreases when FSLE increases, suggesting that stronger FSLE-based fronts are less frequent ; (2) the response of Chl-a varies with FSLE depending on the region. Chl-a increases in the Alboran region and its larger values occur with stronger fronts, Chl-a also increases in fronts in the Gulf of Lion but the strong fronts density is quite small, in the Algerian-Tyrrhenian Chl-a seems to be less influenced by FSLE and in the Ionian-Levantine region Chl-a decreases with FSLE. We also investigated the response of the Chl-a, presented on Figure R2.15.2 and Tables R2.15.1 and R2.15.2. Chl-a excess associated with weak FSLE structures was small and mostly not significant compared to the stronger fronts. We therefore focused our study on the stronger fronts,

defined by FSLE > 0.17 d^{-1}, as making the distinction between these two categories did not provide additional information that would change the interpretation of the results.

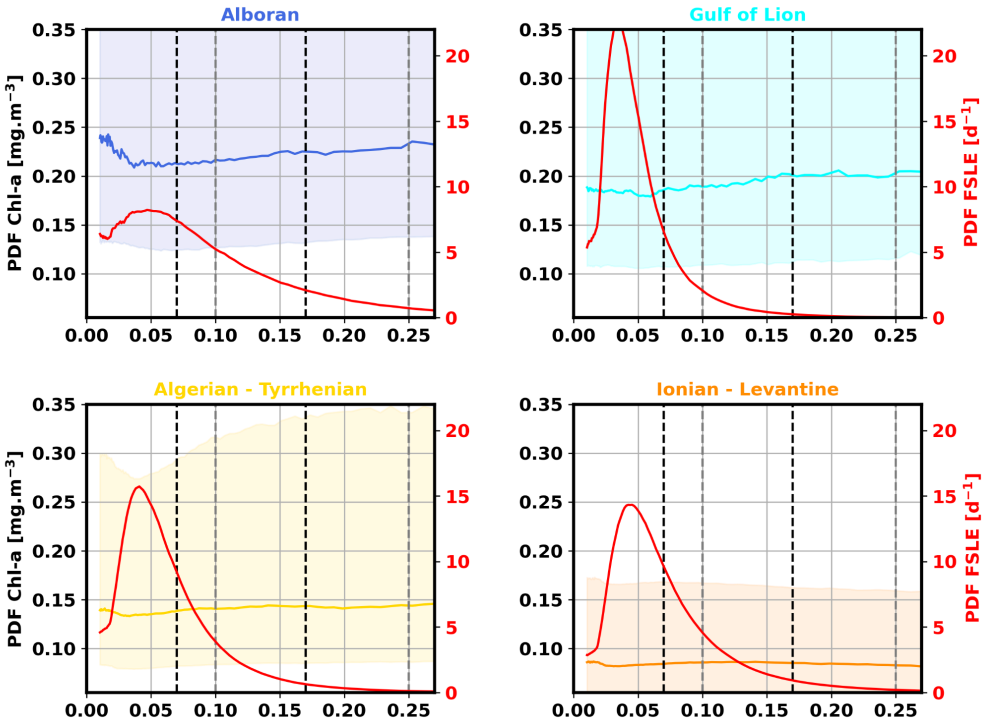


Figure R2.15.1 : PDF of Chl-a following FSLE in between 1998-2021. PDF of Chl-a is in colors, and PDF of FSLE is in red.

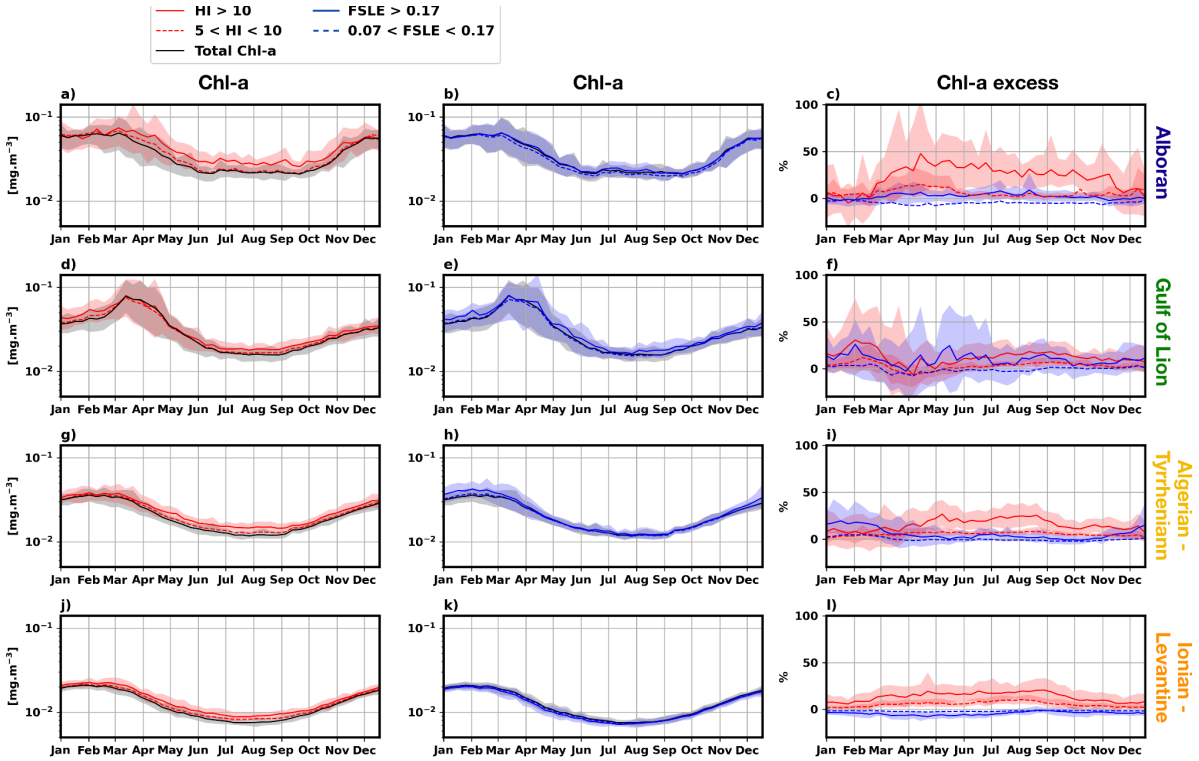


Figure R2.15.2 : Seasonal variability of Chl-a median values and excess of Chl-a averaged over the period 1998–2021, for each region (per row), over weak (dashed line) and strong fronts (plain line). Total Chl-a are shown in black, values computed over fronts from SST fields in red and fronts from AVISO-FSLE in blue. The plain lines represent the climatological mean of 8-day distributions, and the envelopes mark the 10\% and 90\% percentile across the entire time series.

	Region	Season	Excess ± STD (%)	p-value	Fraction p < 0.05
0	Alboran	Winter	NaN	0.125	27%
1		Spring	-6.4 ± -6.4	0.004	84%
2		Summer	-4.8 ± -4.8	<0.001	100%
3		Autumn	-5.3 ± -5.3	0.007	82%
4	Gulf of Lion	Winter	NaN	0.476	6%
5		Spring	NaN	0.434	6%
6		Summer	NaN	0.442	3%
7		Autumn	NaN	0.406	10%
8	Algerian - Tyrrhenian	Winter	NaN	0.403	5%
9		Spring	NaN	0.350	8%
10		Summer	NaN	0.254	21%
11		Autumn	NaN	0.330	8%
12	Ionian - Levantine	Winter	NaN	0.254	16%
13		Spring	NaN	0.440	10%
14		Summer	NaN	0.205	22%
15		Autumn	NaN	0.379	11%

Table R2.15.1 : Sensitivity of the Brunner-Munzel test to random sampling of 300 samples from weak frontal and non-frontal Chl-a distributions. Weak fronts are defined as 0.07 < FSLE < 0.17 d^{-1}. The colour scale indicates the intensity of the Chl-a excess, with green associated to positive excesses and pink to negative excesses. The p-value column represents the median p-value obtained from the 100 random realizations. The Fraction p column gives the fraction of realizations for which the Brunner–Munzel test yielded a significant difference between frontal and non frontal Chl-a distributions (p < 0.05).

Region	Season	Excess ± STD (%)	p-value	Fraction p < 0.05
Alboran	Winter	NaN	0.287	12%
	Spring	NaN	0.227	20%
	Summer	NaN	0.477	3%
	Autumn	NaN	0.134	38%
Gulf of Lion	Winter	13.3 ± 13.3	0.036	57%
	Spring	NaN	0.201	34%
	Summer	10.3 ± 10.3	0.006	70%
	Autumn	NaN	0.149	30%
Algerian - Tyrrhenian	Winter	15.1 ± 15.1	0.002	84%
	Spring	NaN	0.500	5%
	Summer	NaN	0.469	5%
	Autumn	NaN	0.627	1%
Ionian - Levantine	Winter	-3.8 ± -3.8	0.027	61%
	Spring	NaN	0.064	47%
	Summer	-3.9 ± -3.9	0.013	69%
	Autumn	NaN	0.092	41%

Table R2.15.2 : Sensitivity of the Brunner-Munzel test to random sampling of 300 samples from strong frontal and non-frontal Chl-a distributions. Strong fronts are defined as FSLE > 0.17 d^{-1}. The colour scale indicates the intensity of the Chl-a excess, with green associated to positive excesses and pink to negative excesses. The p-value column represents the median p-value obtained from the 100 random realizations. The Fraction p column gives the fraction of realizations for which the Brunner–Munzel test yielded a significant difference between frontal and non frontal Chl-a distributions (p < 0.05).

Hence, our choice was intended to focus on intense Lagrangian structures while maintaining a common threshold across all regions. We did sensitivity tests with the thresholds 0.10, 0.17, and 0.25 d^{-1}. We selected a large range according to the region-specific thresholds in Sudre et al. (2023) ranging from 0.119 d^{-1} in the Western Mediterranean to 0.254 d^{-1} in the Alboran Sea. We examined both the fraction of the domain classified as fronts and the associated Chl-a response. We present here the Chl-a results for the Alboran and the Ionian-Levantine regions (Figure R2.15.3 Tables R2.15.2, R2.15.3 and R2.15.4).

In the Alboran region, the fraction of the regions was ranging from 24.7% for 0.10 d^{-1}, 9.6% for 0.17 d^{-1} to 3.4% for 0.25 d^{-1}. In the Ionian-Levantine region, the fraction of the regions was ranging from 17.7 % for 0.10 d^{-1}, 4.2% for 0.17 d^{-1} and to 1.06% for 0.25 d^{-1}. For all, the seasonal variabilities were preserved, with higher frontal occurrence in summer and autumn and lower occurrence in winter and spring. The seasonal pattern of

Chl-a remained also consistent across the tested thresholds. The lower threshold, 0.10 d^{-1}, identifies a larger fraction of the domain and includes weaker fronts structures, resulting in a weaker Chl-a excess with weaker statistical significance. In contrast, the larger thresholds select fewer, stronger structures and with larger Chl-a excess and a stronger statistical significance. We therefore retained 0.17 d(^{-1}) as an intermediate threshold within the range of region-specific values reported by Sudre et al. (2023) for the four regions considered in our study, without selecting either the most inclusive or the most restrictive threshold. The sensitivity analysis has been added to the Appendix, and its implications are discussed in the Discussion section. We have also revised the comparison between the HI- and FSLE-based results to avoid comparing the values of Chl-a excesses obtained using the two diagnostics, as these values are directly influenced by the choice of HI and FSLE thresholds. This point is discussed further below.

We would like to note that the differences in showing excess values according to the statistical significance between Table R2.15.2 and the Table in the original manuscript result from the random realization. For example, in the original manuscript, the Chl-a excess in winter in the Alboran region was statistically significant, whereas it is no longer significant in the revised manuscript. This difference is consistent with the low fraction of significant p-values (12%), indicating that only a small proportion of the repeated random samplings yielded a statistically significant result. This example highlights the importance of reporting the fraction of significant p-values in the revised manuscript

"*We performed a sensitivity analysis using three FSLE thresholds: 0.10, 0.17, and 0.25 d(^{-1}) and we examined the response of Chl-a. These values span a range consistent with the region-specific thresholds reported by Sudre et al. (2023), which vary from 0.119 d(^{-1}) in the western Mediterranean to 0.254 d(^{-1}) in the Alboran Sea. The lower threshold (0.10 d(^{-1})) identified a larger fraction of each domain, including weaker structures, and was associated with smaller Chl-a excess and weaker statistical significance. Inversely, the higher threshold (0.25 d(^{-1})) selected fewer, more intense structures, associated with larger Chl-a excesses and stronger statistical significance. Hence, 0.17 d(^{-1}) has been retained as an intermediate threshold within the range of region-specific values for the four regions considered in our study, avoiding both the most inclusive and the most restrictive thresholds. However, using a single threshold across all regions limits our ability to account for regional differences in frontal dynamics and in Chl-a responses. Future studies could build on the regionalisation approach proposed by Sudre et al. (2023) by defining region-specific FSLE thresholds. Such an approach could help better account for regional variability in frontal activities and refine the assessment of the biological responses."*

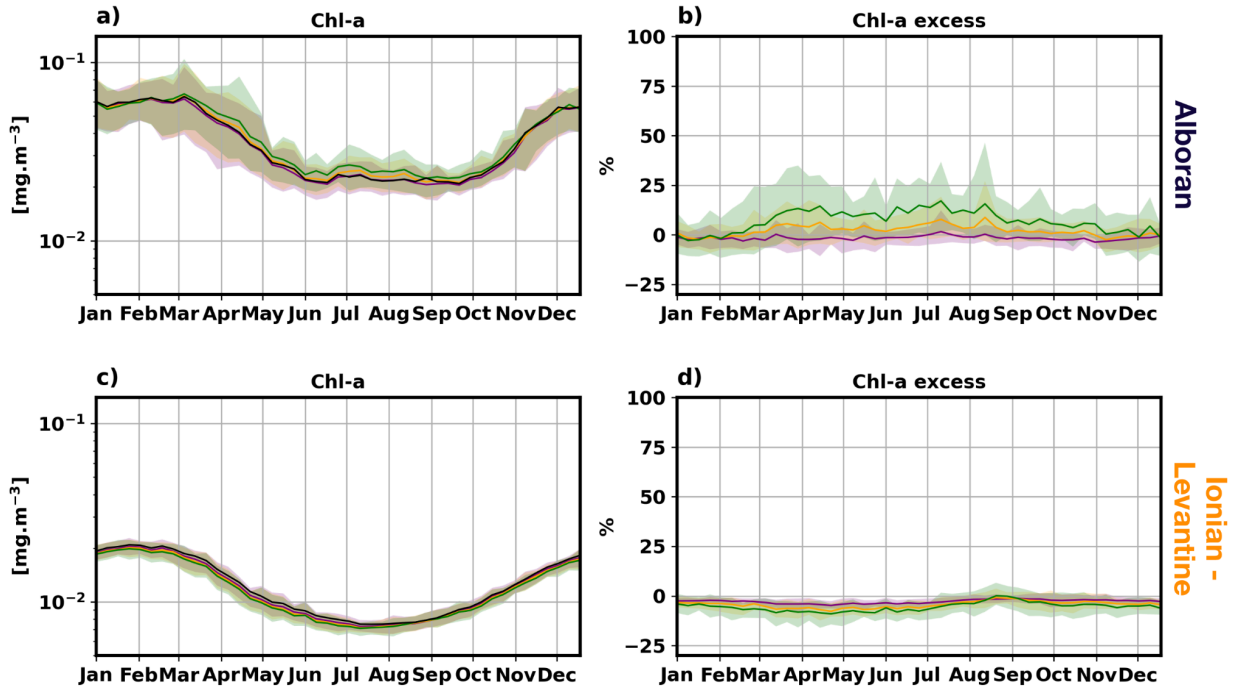


Figure R2.15.3 : Sensitivity analysis of the FSLE threshold. Seasonal variability of Chl-a median values and Chl-a excess averaged over the period 1998–2021, for the Alboran region and the Ionian-Levantine region. Chl-a associated with the FSLE-based front of 0.1 d^{-1} is in purple, 0.17 d^{-1} in yellow and 0.25 d^{-1} in green. Total Chl-a seasonal average is in black line.

Region	Season	Excess ± STD (%)	p-value	Fraction p < 0.05
Alboran	Winter	NaN	0.212	15%
	Spring	-1.8 ± -1.8	0.040	55%
	Summer	-0.7 ± -0.7	0.046	53%
	Autumn	-2.3 ± -2.3	0.044	54%
Gulf of Lion	Winter	NaN	0.319	12%
	Spring	NaN	0.385	4%
	Summer	NaN	0.063	43%
	Autumn	NaN	0.135	31%
Algerian - Tyrrhenian	Winter	NaN	0.138	35%
	Spring	NaN	0.456	4%
	Summer	NaN	0.425	9%
	Autumn	NaN	0.499	3%
Ionian - Levantine	Winter	NaN	0.120	34%
	Spring	NaN	0.303	26%
	Summer	NaN	0.083	37%
	Autumn	NaN	0.309	20%

Table R2.15.3 : Sensitivity of the Brunner-Munzel test to random sampling of 300 samples from strong frontal and non-frontal Chl-a distributions. Strong fronts are defined as FSLE > 0.10 d^{-1}. The colour scale indicates the intensity of the Chl-a excess, with green associated to positive excesses and pink to negative excesses. The p-value column represents the median p-value obtained from the 100 random realizations. The Fraction p column gives the fraction of realizations for which the Brunner–Munzel test yielded a significant difference between frontal and non frontal Chl-a distributions (p < 0.05).

Region	Season	Excess ± STD (%)	p-value	Fraction p < 0.05
Alboran	Winter	NaN	0.208	16%
	Spring	NaN	0.518	3%
	Summer	11.9 ± 11.9	0.006	76%
	Autumn	NaN	0.283	16%
Gulf of Lion	Winter	NaN	0.086	37%
	Spring	22.2 ± 22.2	0.041	57%
	Summer	21.5 ± 21.5	0.010	74%
	Autumn	NaN	0.531	7%
Algerian - Tyrrhenian	Winter	21.7 ± 21.7	<0.001	100%
	Spring	NaN	0.422	15%
	Summer	NaN	0.425	9%
	Autumn	NaN	0.513	4%
Ionian - Levantine	Winter	-5.1 ± -5.1	0.004	81%
	Spring	-8.0 ± -8.0	0.011	75%
	Summer	-4.5 ± -4.5	0.002	84%
	Autumn	-4.1 ± -4.1	0.010	71%

Table R2.15.4 : Sensitivity of the Brunner-Munzel test to random sampling of 300 samples from strong frontal and non-frontal Chl-a distributions. Strong fronts are defined as FSLE > 0.25 d^{-1}. The colour scale indicates the intensity of the Chl-a excess, with green associated to positive excesses and pink to negative excesses. The p-value column represents the median p-value obtained from the 100 random realizations. The Fraction p column gives the fraction of realizations for which the Brunner–Munzel test yielded a significant difference between frontal and non frontal Chl-a distributions (p < 0.05).

(ii) We have also clarified how the comparison between HI and FSLE is presented. We reinforce the idea that we do not look for “the best product“ at detecting biologically relevant fronts, but rather to test the robustness of the Chl-a response to fronts when using different approaches and different products. Our results show that the biological signal associated with HI fronts is consistently associated with a Chl-a enhancement, with more statistically significant differences than those observed for FSLE-based fronts. Besides, we have adapted our discussion to avoid directly comparing the magnitude of Chl-a excesses between the two methods. These estimates are sensitive to methodological choices, including the thresholds used to identify HI- and FSLE-based fronts. Instead, we do a greater emphasis on the patterns and consistency of the Chl-a response across diagnostics.

We have also revised our discussion of the “effective resolution” of FSLEs. Rather than attributing the weaker biological response directly to effective resolution, we now discuss

that the pixel-by-pixel spatial colocalisation used in our analysis may not be equally suited to the two approaches. For example, the SST-based HI is calculated from SST fields at a resolution of 1/20°, which is relatively close to that of the Chl-a product. This facilitates the spatial correspondence between the detected SST-based structures and the biological signal. By contrast, FSLE ridges represent Lagrangian structures whose spatial extent and position may not be fully captured by a direct pixel-by-pixel comparison with Chl-a fields. Moreover, our sensitivity analysis shows that lowering the FSLE threshold, thereby identifying a larger fraction of the domain, does not necessarily lead to a larger Chl-a excess. This suggests that the magnitude of the biological signal depends not only on the diagnostic used, but also on how the structures are spatially compared with the Chl-a field.

For these reasons, we clarified in the introduction section our reason of using different approaches :
"*These approaches provide complementary views of frontal dynamics, but differ in their spatial resolution, physical interpretation and sensitivity. SST-based methods can exploit high-resolution satellite imagery but are affected by cloud cover and identify thermal structures that do not necessarily correspond to dynamically active density fronts. FSLEs provide a Lagrangian perspective on surface transport, but their effective resolution depends on the velocity field input and may limit the accuracy with which frontal positions are identified \citep[e.g.,][]{Lehahn2007}. Hence, previous studies have highlighted the value of combining complementary front-detection approaches (d'Ovidio et al., 2009; Sudre et al., 2023), which motivated us to assess whether the inferred biological responses are robust across front detection methods*"

And to the discussion section :
"*The Chl-a response differed between the HI and FSLE diagnostics, with HI-based fronts consistently associated with Chl-a enhancement, whereas the signal associated with FSLE-defined fronts was generally weaker and less statistically significant. This pattern was observed across different products of the two diagnostics, including HI calculated from SST and surface density fields and two FSLE products based on different velocity fields. However, these differences should not be interpreted as evidence that one diagnostic is superior to the other, as HI and FSLE capture different characteristics of frontal dynamics. Methodological factors may contribute to the weaker Chl-a signal associated with FSLE-based fronts. The pixel-by-pixel colocalisation used in this study may not equally capture the biological signatures associated with the two diagnostics. For example, the SST-based HI is calculated from daily satellite fields at a resolution of 1/20°, relatively close to that of the Chl-a product, which facilitates the spatial correspondence between detected SST-based structures and biological observations. In contrast, FSLE filaments represent Lagrangian structures whose spatial extent and position may not correspond directly to local Chl-a maxima at the pixel scale. Finally, phytoplankton responses to frontal dynamics may not be instantaneous (Gangrade and Franks, 2023; Gangrade and Mangolte, 2024), and therefore the systematic co-localisation of fronts and of their biological response, as assumed here and in previous studies (e.g., Haëck et al., 2023; Lévy et al., 2025), may underestimate the impact of fronts when the biological signal is advected away from the front."*

Minor Comments

R2.16 : The title is misleading. This study specifically focuses on regions within the Mediterranean Sea, whereas the phrase “over oceanic regions” suggests a systematic study of a wider domain. Additionally, while I believe the authors are using the phrase “trophic” in reference to eutrophic versus oligotrophic, I believe that word is more commonly associated with the food web (trophic transfer), which may lead readers to believe that multiple trophic levels (e.g., zooplankton) are studied.

We thank the reviewer for this comment, now the title is :
“*Phytoplankton community responses to oceanic fronts across Mediterranean biogeochemical regimes, seasons, and frontal diagnostics”*

R2.17 : The authors state that their focus is on “fine scale fronts” (Line 59), without specifically defining what they mean by “fine scale”. For example, what is the specific length scale of interest? In Line 33, the authors state “Their spatial scale spans a large range, from a few hundred meters to hundreds of kilometers in length.” Is this entire range considered fine scale in this study? If not, I suggest the authors clearly articulate why their focus is on “fine scale” fronts, rather than all fronts.

The second paragraph in the Introduction section has been modified to define “fine scale” :
“*They are ubiquitous features observed everywhere in the ocean, with quite variable characteristics. Here, we focused on fine-scale fronts, structures ranging from a few hundred meters to hundreds of kilometers in length*.”

R2.18 :. It would be useful for readers to somewhere (maybe in the intro) provide context & citations regarding the traits of gleaners and opportunists, e.g., “larger phytoplankton” (Line 326) that are “fast-growing” (line 334) refers to diatoms, and “oligotrophic-adapted organisms” (line 328) refers to the prokaryotes.
We added this to the second paragraph in the introduction section :
“*Beyond changes in phytoplankton biomass, fronts can also lead to contrasting responses among phytoplankton groups, reflecting differences in their ecological traits and nutrient resources. Smaller phytoplankton are adapted to efficiently exploit low nutrient concentrations, whereas larger phytoplankton, such as diatoms, can rapidly benefit from nutrient pulses owing to their high potential growth rates \citep{Mangolte2022}.*”

And this to the discussion section :
“*These results are consistent with previous studies showing that larger phytoplankton, such diatoms, can benefit from nutrient pulses owing to their high potential growth rates \citep{Levy2015,Liu2021,Mangolte2022,Mangolte2023,Levy2025}, whereas smaller phytoplankton, such as prokaryotes, are often better adapted to persist under oligotrophic conditions \citep{Mangolte2022}*”

R2.19 : I do not understand how the last paragraph (Lines 364-369) is relevant or logical, since frontal occurrence (Xing et al. 2024) and chl-a (Yang et al. 2023) in the Mediterranean Sea have not been projected to change.

As previous remarks have also been made on this part, it has been removed.

R2.20 : The software used to conduct the study was not provided by the authors. Technical Corrections

It has been added.
“ *The software used to conduct this study is Python.* “

We want to thank the reviewer for all the next minor comments (**R2.21 - R2.48)**, and we will address them directly in the revised manuscript.

R2.21 : It is assumed that readers are familiar with the currents within the sub-regions shown in Figure 1. I suggest providing a figure or subplot that shows the currents and gyres referenced throughout the paper.

R2.22 : Incorrect usage of “groups” versus “group” in many cases: Suggest changing “phytoplankton groups data” to “phytoplankton groups” or “phytoplankton group data” (e.g., Lines 7, 271, Fig 1 caption); “phytoplankton groups mosaic” to “phytoplankton group mosaic” (Line 154); “groups” to “group” (Line 211, 212, 213, 275)
R2.23 : Citations needed to support the sentence in Lines 22-23 & the sentence in Line 33
R2.24 : Line 27: the word “constitute” is not appropriate here
R2.25 : The sentence in Lines 36-37 is incoherent; suggest re-writing
R2.26 : Inappropriate use of the word “besides”, suggest removing it in Lines 40, 127, 282
R2.27 : Line 45: “lagrangian” should be capitalized
R2.28 : Line 54: Suggest changing “Advances in observational and detectable methods has significantly improved…” to “Advances in observation and detection methods have significantly improved…”
R2.29 : “fine scale” and “fine-scale” are used interchangeably, suggest remaining consistent throughout
R2.30 : The content in Lines 54 – 74 is strangely organized, and the flow of thoughts is difficult to follow:
R2.31 : The sentence in Line 57 feels out of place here, and repetitive with the content in Lines 68-71
R2.32 : Suggest moving the sentences in Line 61-66 to the end of the paragraph
R2.33 : Line 72: I believe “lying” should be “relying”
R2.34 : The sentence in Lines 97-99 is incoherent and a run-on sentence; suggest re-writing
R2.35 : Line 122: I believe “their” should be “they”
R2.36 : Fadida et al. 2026 (https://doi.org/10.5194/os-22-329-2026) is a regional-specific and relevant study to cite within lines 129-131
R2.37 : Line 145: “green region” appears to be cyan in color; recommend using the regional name rather than the color in this description for the reader
R2.38 : When discussing the “other eukaryote” groups, it would be helpful for the readers to remind that this means “non-diatom eukaryotes” or “diatom-excluded eukaryotes” (e.g., Lines 152, 273)
R2.39 : Recommend changing “implies to compute” to “computes” (Line 161) and “to assume” to “assumes” (Line 162).

R2.40 : Recommend removing the sentence in Lines 163-164 because they are confusing in this spot, and repetitive with sentence at Line 172. The sentence in line 164-165 is also repetitive with the second sentence of the same paragraph.
R2.41 : Line 167: I believe “28 km square window” should be “28 km x 28km square window”
R2.42 : Line 197: Recommend changing “make a difference” to “differentiated”
R2.43 : Line 220: Recommend changing “fronts with SST” to “fronts detected with SST”
R2.44 : Line 226: “bassin” typo
R2.45 : Line 229: “associated to finer currents” to “associated with finer currents”
R2.46 : Figure 3 & 4 captions: “ans” typo
R2.47 : Line 247: “fronts effect” to “frontal effect”
R2.48 : Line 299: “more than 60%” of chl-a? “front scale” to “frontal scale”.