



Supplement of

Three decades of marine bioregions evolution through analysis of phytoplankton phenology

Wilhem Riom et al.

Correspondence to: Wilhem Riom (wilhem.riom@imev-mer.fr)

The copyright of individual parts of the supplement might differ from the article licence.

Section S1: method flowchart

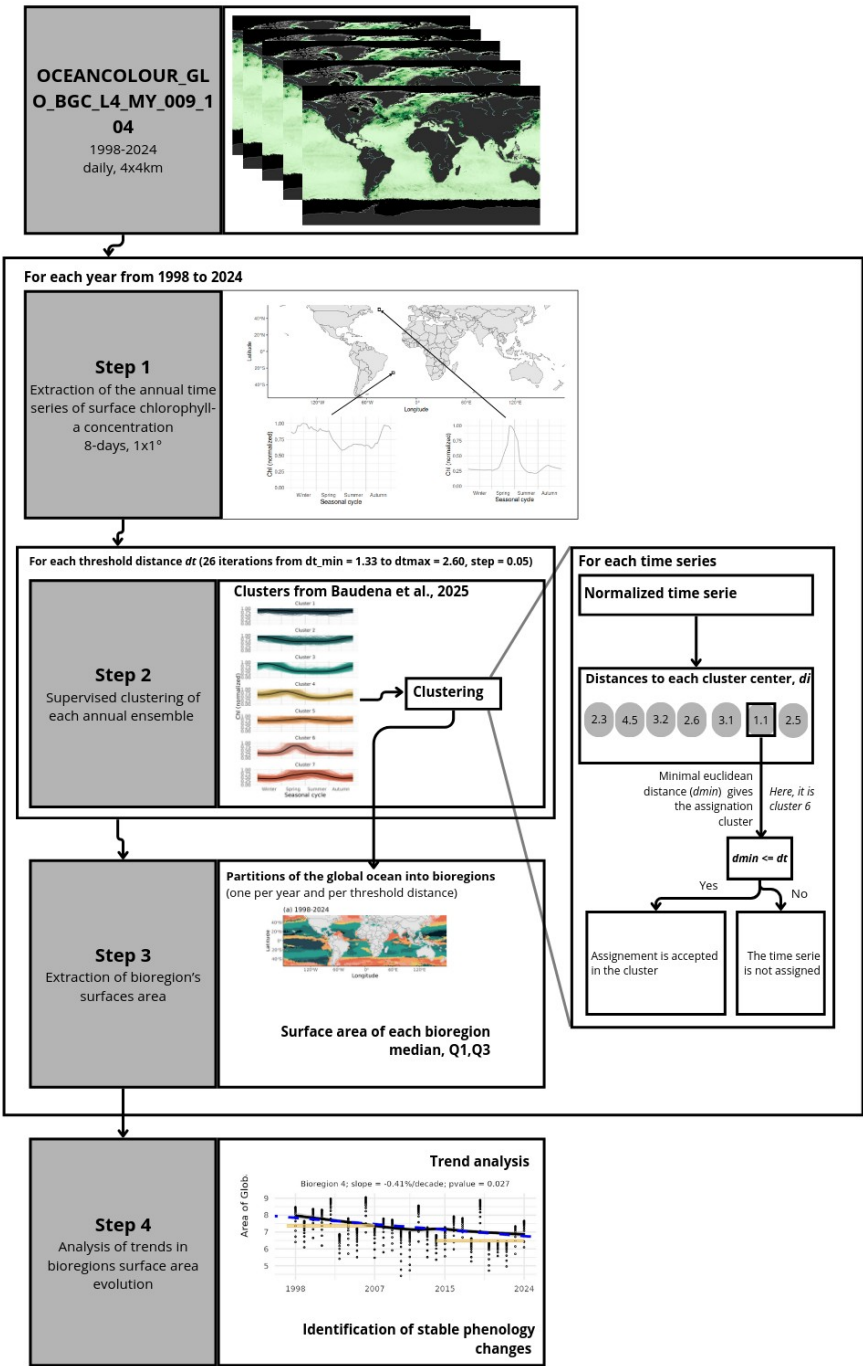


Figure S1: Flowchart of the method (adapted from Mayot *et al.*, 2016).

Step 1 : Extraction of the annual time series of surface chlorophyll-a concentration at 8 days interval, over a $1 \times 1^\circ$ grid.

The following steps are inspired from supplementary section “S.1 Calculation of the reference climatology” from Baudena *et al.*, 2025 and describe the method used to derive annual ensembles composed of times series of surface chlorophyll-a concentration for each year **Y** between 1998 and 2024.

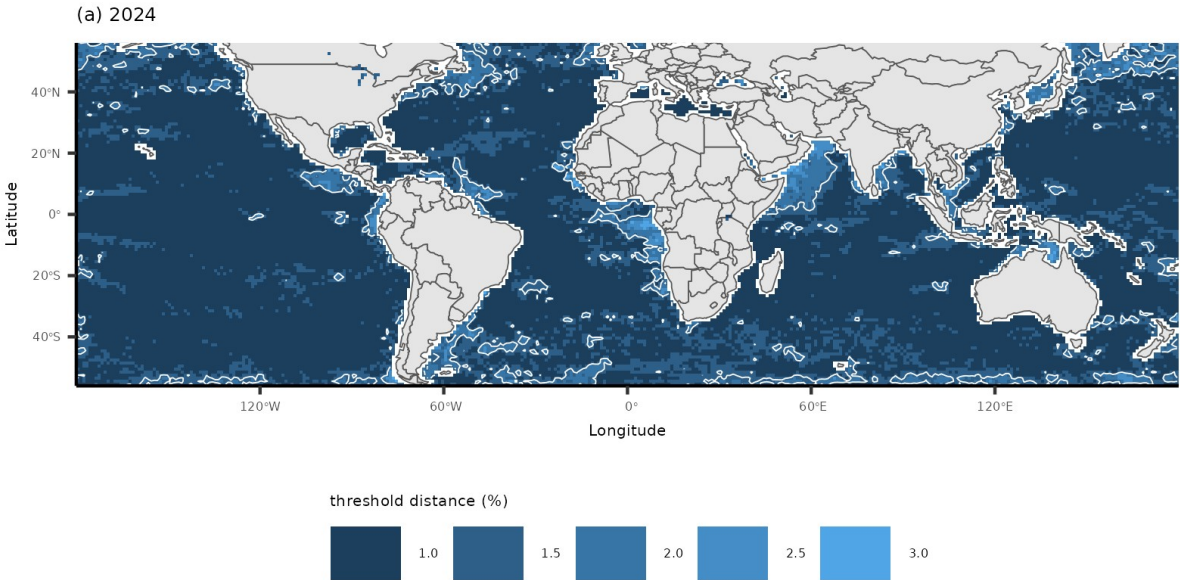
1. We consider the chlorophyll-a field of the 1st January of year **Y**. Then, we exclude all grid cells within 1° (~ 111 km) of the coast.
2. We reduce the spatial resolution from the nominal resolution (0.04°) to 1° . For this purpose, we build a grid of 1° resolution covering the global ocean. For each grid cell, we take the median of all the chlorophyll-a values within the grid cell (~ 625 values per cell). If, in that grid cell, less than 50% of the values are available (due e.g. to cloud coverage), we flag that cell as not available (NA).
3. The temporal resolution of the time series is eight days. Thus, step 1 and 2 are repeated for all the days between the 1st and the 8th January of year **Y**. Hence, for a given grid cell, we obtain 8 chlorophyll-a values. We consider the median of these values as representative of the chlorophyll-a concentration in that grid cell for the period between the 1-8 January of year **Y**. Values flagged as NA are excluded from the median. If all eight values in a grid cell over an eight-day period are marked as NA, the chlorophyll-a value in that grid cell (for the 1-8 January **Y** period) is marked as NA. This is repeated for all the grid cells.

Step 3 is repeated for each of the time slices composing a climatological year (each time slice is composed of 8 days: 1-8 January; 9-16 January, etc.)

4. For each grid cell, if the annual time series has less than 75% of available values, we exclude that grid cell from the analysis. Similarly, if the annual time series shows temporal gaps of more than 4 weeks, we discard that grid cell. Otherwise, we remove any existing temporal gaps using a monotonic splice interpolation (*pchip* interpolate function in the python package *scipy*). This type of interpolation is chosen because it performs a non-linear interpolation and, at the same time, does not create artificial peaks in the time series.
5. Each time series is smoothed using a running mean over 40 days (i.e. 5 time steps).
6. We shift the annual time series located in the Southern hemi-sphere of 6 months.
7. Finally, we normalise each time series by dividing it by its annual maximum value.

Section S2: Method sensitivity to the threshold distance

This section presents the rationale behind the usage of a set of 26 discrete threshold distances to quantify the uncertainty of the supervised clustering method. In Baudena *et al.* (2025), the unsupervised clustering of the reference ocean color climatology, assigned all annual time series in one of seven clusters. Each cluster possesses a most dissimilar element, characterized by a maximal euclidean distance to the cluster centroid (mean element). These seven distances range from ~1.33 (in cluster 1) to ~2.60 (in cluster 7). The supervised clustering technique, used in this study, needs a maximal threshold distance to prevent assignment of too dissimilar time series into one of the seven clusters. We empirically based the range of acceptance threshold from the minimal to maximal cluster classification distance from Baudena *et al.* (2025) with a regular step of 0.05. This step was chosen as small enough to discretize the cumulative distribution functions and account for their variations. For each individual year, we obtain twenty six different estimates of the bioregions. As an illustration, Fig S2a shows the classification distance for the year 2024. Larger distances (dissimilarities) are found in coastal regions and in domains of reduced satellite data availability (high latitudes, Arabian Sea, western African coast, Modi *et al.*, 2022). Fig. S2b shows the amount of classified time series for each annual ensemble in respect of the threshold distance. The smallest classification distance ensures correct assignment of 88% of the time series while the largest ensures close to 100 %.



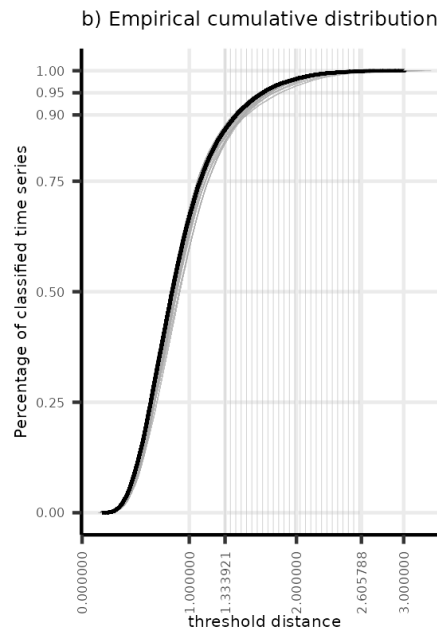


Figure S2: a) Spatial distribution of the clustering distance for the year 2024. The white contour indicates the smallest threshold distance considered in the supervised clustering. b) Empirical cumulative distribution of the amount of assigned time series in respect of the threshold distance for all annual ensembles (grey lines) and for the year 2024 (black line). The selected discrete threshold distances are highlighted by grey vertical lines ranging from 1.33 to 2.60.

Cited Literature

Baudena, A., Riom, W., Taillandier, V., Mayot, N., Mignot, A., and D'Ortenzio, F.: Comparing satellite and BGC-Argo chlorophyll estimation: A phenological study, *Remote Sensing of Environment*, 326, 114743, <https://doi.org/10.1016/j.rse.2025.114743>, 2025.

Mayot, N., D'Ortenzio, F., Ribera d'Alcalà, M., Lavigne, H., and Claustre, H.: Interannual variability of the Mediterranean trophic regimes from ocean color satellites, *Biogeosciences*, 13, 1901–1917, <https://doi.org/10.5194/bg-13-1901-2016>, 2016.

Modi, A., Roxy, M. K., and Ghosh, S.: Gap-filling of ocean color over the tropical Indian Ocean using Monte-Carlo method, *Sci Rep*, 12, 18395, <https://doi.org/10.1038/s41598-022-22087-2>, 2022.

