Marine Environmental Research 219 (2026) 108101

Contents lists available at ScienceDirect

Marine Environmental Research

journal homepage: www.elsevier.com/locate/marenvrev

A novel three-step biologically informed ocean partitioning strategy: insights from toxigenic phytoplankton in a coastal upwelling system

M.J. Lima [a,*], I. Caballero [b], A.B. Barbosa [a]

[a] Centre for Marine and Environmental Research (CIMA), Aquatic Research Network (ARNET), Universidade do Algarve, Campus de Gambelas, Faro, 8005-139, Portugal

[b] Institute of Marine Sciences of Andalusia (ICMAN), Spanish National Research Council (CSIC), Campus Universitario Río San Pedro s/n, Puerto Real, 11519, Spain

ARTICLE INFO

Keywords:
Novel partitioning strategy
Group-specific
Harmful algal blooms
Environmental variability
Remote sensing
Empirical-statistical models

ABSTRACT

Ocean partitions are often based on readily accessible variables, such as abiotic factors and chlorophyll-a concentration, but provide limited insight into biological patterns. This study developed a three-step partitioning strategy prioritizing environmental factors that best described the variability patterns of harmful algal bloom (HAB)-forming taxa off SW Iberia. These included the producers of amnesic shellfish poisoning (ASP), diarrhetic shellfish poisoning (DSP), and paralytic shellfish poisoning (PSP). First, dimensionality reduction and unsupervised classification were applied to three environmental datasets, derived from remote sensing and model outputs, covering a 19-year period. Second, different empirical-statistical models were used to determine which datasets best explained the abundance of HAB-producers, available for an 8-year period in different classified coastal production areas. Finally, the best datasets were used to derive partitions prioritizing the variability of different HAB groups, at a pixel level. The first classifications identified up to 12 regions, with four to five located in the coastal-slope domain, with a variable configuration depending on the dataset. The best predictor datasets and models identified five regions (two inner-shelf, two outer-shelf/slope, and one transitional coastal-ocean region), representative of HAB groups. No clear distinctive partitions were identified for different groups, namely for ASP- and DSP-producers, likely due to the combined influence of upwelling and freshwater discharges, along with submarine topographic features. Our partitioning strategy can be applied to other marine systems and taxonomic groups. Future improvements, including more complete environmental and biological datasets, could enhance the value of biologically informed environmental partitions as proxies for species abundance.

1. Introduction

Marine systems are highly dynamic, being influenced by ocean circulation patterns and atmosphere-ocean-land interactions, and their variability ultimately impacts marine food web structure and function. This natural heterogeneity is especially pronounced in coastal ecosystems, which are highly productive, resourceful, and subject to increasing anthropogenic pressures (Boyd et al., 2016; Ayyam et al., 2019). Thus, representative sampling of the global ocean, regular and at increasingly finer spatial scales, remains a major challenge. Ocean surface partitioning into provinces, i.e., coherent, homogeneous regions with similar properties and defined boundaries, is widely accepted (Reygondeau and Dunn, 2019; Marchese et al., 2022). The delineation of provinces is useful to extrapolate sparse *in situ* data, test hypothesis, assess environmental variability and trends, improve bio-optical algorithms and global model performance (see Bock et al., 2022 and references therein), and support region-specific ocean management (see review by Krug et al., 2017a).

Over the past decades, several partitions of the marine environment at different scales have been established based on a set of abiotic and biotic variables (see review by Krug et al., 2017a). Due to their far more readily and increased availability compared to biological data, the classical approach is to consider mostly physico-chemical (abiotic) features that are assumed to be representative of the ocean biogeochemical properties. These include sea surface temperature (SST), salinity, mixed layer depth (MLD), euphotic zone depth, wind components, surface current velocity, turbidity, bathymetry, photosynthetically available radiation, and nutrient concentrations (e.g. Oliver et al., 2004; Gregr

* Corresponding author.
E-mail addresses: mjlima@ualg.pt (M.J. Lima), isabel.caballero@csic.es (I. Caballero), abarbosa@ualg.pt (A.B. Barbosa).

https://doi.org/10.1016/j.marenvres.2026.108101
Received 26 January 2026; Received in revised form 20 April 2026; Accepted 4 May 2026
Available online 5 May 2026

and Bodtker, 2007; Oliver and Irwin, 2008; Fendereski et al., 2014; Krug et al., 2018a; Suominen et al., 2021; Bolaños-Durán et al., 2025). However, the complexity of biological processes cannot be fully captured by classifications based solely on abiotic properties (e.g. Zhao et al., 2025).

Phytoplankton, as the foundation of most marine food webs, have been frequently used to partition the ocean surface, considering chlorophyll-a concentration (Chl-a; used as a proxy for phytoplankton biomass; e.g. Marchese et al., 2022; D'Ortenzio and Ribera d'Alcalà, 2009; Krug et al., 2017b; Ferreira et al., 2019; Arena et al., 2025), phytoplankton assemblage composition (e.g. Taylor et al., 2011; Hofmann Elizondo et al., 2021; Konik et al., 2024; Vargas-Yáñez et al., 2024), and phenology (e.g. Konik et al., 2024; Mayot et al., 2016; Krug et al., 2018b; Ferreira et al., 2021; Kheireddine et al., 2021; Pramlall et al., 2024). However, most partitions based on phytoplankton-related variables (Chl-a) do not necessarily reflect the spatial-temporal variability patterns of the abundance of specific functional groups of phytoplankton, with different functional traits (Weithoff and Beisner, 2019), including harmful algal bloom (HAB)-forming taxa. HABs have multiple negative impacts on marine ecosystems, human health, and socio-economic activities (e.g. Gobler and Sunda, 2012; Shumway et al., 2018; Trainer et al., 2020). Thus, HAB prediction, based on modelling approaches and solid monitoring efforts, has become a priority for coastal management.

The region off southwestern Iberian Peninsula (SWIP) is a very heterogeneous marine domain, located at the transition between North-East Atlantic subtropical and temperate waters, and particularly vulnerable to climate change (Krug et al., 2017b). SWIP is influenced by coastal upwelling (Relvas et al., 2007) and riverine inputs (Prieto et al., 2009; Caballero et al., 2014), and frequent HAB events have been reported for coastal areas extensively used for resource extractive activities (see Lima et al., 2022). Most global partitioning schemes identified two regions in the SWIP area (Sherman, 1991, 1994; Longhurst, 2007; Aquarone et al., 2008; Spalding et al., 2007, 2012; Fay and McKinley, 2014), whereas regional partition studies based on Chl-a variability (Krug et al., 2017b; Ferreira et al., 2019; Navarro and Ruiz, 2006), phytoplankton phenological indices (Krug et al., 2018b; Ferreira et al., 2021) or a combination of Chl-a and abiotic variables (Krug et al., 2018a; Caballero et al., 2014; Muñoz et al., 2015), identified up to 10 regions, two to four of which were included in coastal domains. Yet, these partitions do not necessarily represent the variability patterns of HAB groups, a critical issue for coastal ecosystems.

In this context, this study aimed to partition the coastal areas off SWIP using a novel approach that prioritizes environmental factors that best describe spatial-temporal variability patterns of the abundance of the most common toxigenic phytoplankton taxa associated to the most frequent HAB events in the area (see Lima et al., 2022 and references therein). The specific study objectives were to: (1) partition the coastal domain off SWIP into environmentally homogeneous regions, based on three different datasets comprising Chl-a only, Chl-a combined with other satellite-derived variables, and the integration of modelled and satellite-derived variables, during a *ca.* 19-year period (July 2002 – December 2021); (2) assess the value of these datasets as predictors of the abundance of different HAB groups for specific coastal regions, using empirical-statistical models of varying complexity; and (3) generate more biologically-informed coastal partitions using the best-performing datasets and model types for three potentially toxigenic HAB groups. These included diatoms and dinoflagellates, representing groups with contrasting functional traits and niche preferences (Weithoff and Beisner, 2019). Based on previous studies (Krug et al., 2017b, 2018a; Ferreira et al., 2019, 2021; Caballero et al., 2014; Lima et al., 2022), we anticipated a separation between the western and southern SWIP coasts, and differences between the partitions oriented for distinct phytoplankton groups.

2. Material and methods

2.1. Study area

The region off SWIP (35-39°N, 12-5°W; see Fig. 1) forms the northernmost part of the Iberian Canary Eastern Boundary Upwelling System, one of the most biologically productive systems in global oceans (Carr and Kearns, 2003). SWIP includes both zonally (south) and meridionally (southwest) oriented coastlines which are influenced by different oceanographic regimes (Relvas et al., 2007; García-Lafuente et al., 2006). The south sector of the study area includes a semi-sheltered basin, the Gulf of Cádiz (GoC), connected with the Mediterranean Sea through the Strait of Gibraltar, as well as variable nutrient-rich freshwater inputs from rivers, including Guadiana, Piedras, Tinto-Odiel, and Guadalquivir (Krug et al., 2017b; Caballero et al., 2014), and submarine groundwater discharges (Hugman et al., 2015; Piló et al., 2018). There are also other small confined coastal systems that drain directly into the Portuguese south (Arade estuary, Alvor and Ria Formosa coastal lagoons) and southwest coasts (Mira estuary, Odeceixe, Aljezur and Bordeira streams (Cabral et al., 2012; Cardoso et al., 2012).

The study area is affected by mesoscale and sub-mesoscale features (e.g., upwelling filaments, fronts, cyclonic and anticyclonic eddies, coastal countercurrents), often shaped by topographic structures including capes (e.g., Cape São Vicente, CSV, Cape Santa Maria, CSM, Cape Trafalgar) and submarine canyons (Relvas et al., 2007; García-Lafuente et al., 2006; Peliz et al., 2004, 2007, 2009; García Lafuente and Ruiz, 2007). Along the SWIP southwestern coast, summer northerly winds promote upwelling-favourable conditions, shifting to downwelling-favourable conditions during autumn (Fiúza et al., 1982; Relvas and Barton, 2002). Along the SWIP southern coast, westerly winds are associated with upwelling-favourable conditions, less intense compared to the west coast, and upwelling relaxation is associated with the development of a nearshore, warm coastal countercurrent flowing

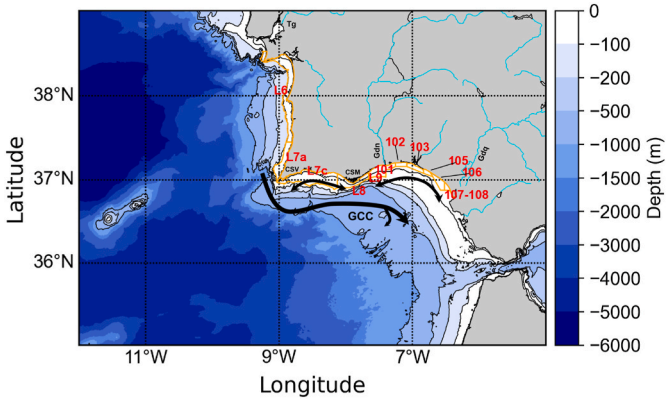

Fig. 1. Study area (whole domain), off southwestern Iberia, with information on bathymetry and major river systems. CSV and CSM depict the location of prominent topographic features, Cape São Vicente and Cape Santa Maria, respectively. Tg, Gdn and Gdq indicate the discharge points of Tagus, Guadiana and Guadalquivir river systems, respectively. Black arrows represent the main circulation patterns, as defined by de Oliveira Júnior et al. (2022). The thicker arrow identifies an equatorward slope current (Gulf of Cádiz Current, GCC). Black lines represent the coastline, and the 20 m, 100 m, 200 m, 500 m and 1000 m isobathymetric contours. The orange polygons delimit classified shellfish coastal production areas (L6, L7a, L7c, L8, L9, 101, 102, 103, 105, 106, 107-108), labelled in red. Limits of these areas were retrieved from the Portuguese Institute of the Sea and Atmosphere (https://www.ipma.pt/en/bivalves/docs/) and Consejería de Sostenibilidad y Medio Ambiente de la Junta de Andalucía (https://portalrediam.cica.es/descargas/index.php/s/descargas?dir=/Inf_archivo/08_AMBITOS_INTERES_AMBIENTAL/02_LITORAL_MARINO/04_SOCIOECONOMIA/ZonasProduccionMoluscos). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

westwards (Relvas et al., 2007; García-Lafuente et al., 2006; Garel et al., 2016). In addition, the Gulf of Cadiz current, flowing southeast (see Fig. 1), typically stronger in the summer, can also extend the upwelling influence into the southern shelf (García-Lafuente et al., 2006; Garel et al., 2016; Sánchez and Relvas, 2003; Criado-Aldeanueva et al., 2006).

2.2. Environmental data

A set of environmental variables were acquired for the period 2002-2021, for the whole study area (Fig. 1 and Table S1). Due to the limited availability of *in situ* data, satellite remote sensing (wind speed and direction; SST; photosynthetically available radiation, PAR; absolute dynamic topography, ADT; and Chl-a), and model-derived products (salinity; MLD; nutrient concentrations; and geostrophic currents), for surface and near-surface (0.49 m), respectively, and at different spatial-temporal resolutions, were used. All mapped variables were regridded to the same spatial resolution, 4-km, and to two temporal resolutions, monthly and monthly climatologies (monthly averages for 2002-2021).

2.2.1. Upper ocean physical and chemical variables

Daily level 3 satellite-derived SST data, with a spatial resolution of 4 km, were retrieved from MODerate resolution Imaging Spectroradiometer, on-board Aqua satellite (MODISA), available at NASA's Ocean-color database (https://oceancolor.gsfc.nasa.gov/, accessed January 15, 2023). SST data were limited to nighttime passes, to avoid diurnal solar heating effects (Robinson, 2010), and high-quality data (quality flag 0). Daily level-3 mean surface PAR at 4 km spatial resolution, from MODISA, were also retrieved from the same database. Standard level-3 flags were applied for quality control (LAND, HISOLZEN, NAVFAIL, FILTER, HIGLINT), and quality-failed pixels were automatically masked and considered invalid. Daily ADT data, at 0.25° spatial resolution, were acquired from the 'sea level gridded data from satellite observations for the global ocean from 1993 to present' dataset, version vDT2021, provided by Copernicus Climate Change Service (C3S; Copernicus Climate Change Service, Climate Data Store, 2018). This product is derived from a multi-mission altimeter dataset, which combines altimetry measurements from a 2-satellite constellation: a reference mission (T/P, Jason-1, Jason-2, Jason-3), and a complementary mission (ERS-1/2, Envisat, SARAL/Altika, Sentinel-3A). This ADT data corresponds to the sea surface height above the geoid and is estimated as the sum of the sea level anomaly and the mean dynamic topography for a 20-year reference period (1993-2012).

Daily sea surface zonal (U) and meridional (V) wind fields, at 0.25° spatial resolution, were obtained from the Blended Seawinds version 2.0 dataset (BSW), available at NCEI-NOAA (NOAA, 2020). BSW is based on multiple scatterometers, standardized across platforms, hence allowing a high quality and more complete temporal and spatial coverage of ocean wind vectors (Zhang et al., 2006).

Daily salinity and MLD data, at 0.083° spatial resolution, were obtained from the Atlantic -Iberian Biscay Irish (IBI) - Ocean Physics Reanalysis database (Product ID: IBI_MULTIYEAR_PHY_005_002), provided by Copernicus Marine Service (CMS; https://marine.copernicus.eu/, accessed March 3, 2023). MLD is estimated as the ocean depth at which the water potential density has increased by 0.01 kg m^{-3} relative to the near-surface value at 10 m depth level, and values were limited by local bathymetry. For MLD, product quality was assessed by comparing it with the mixed layer depth climatology, for the period 1993-2016, provided by (de Boyer Montégut et al., 2004). The product reproduced the main meridional gradient (deeper MLD in the north), although an overestimation was detected in the northwestern part of the IBI domain (see quality information document available at https://documentation.marine.copernicus.eu/QUID/CMEMS-IBI-QUID-005-002.pdf; E.U. Copernicus Marine Service Information, 2023a). For salinity, the product estimations were compared to the World Ocean Atlas 2023 climatology, with higher salinities detected along the northern European coasts (see quality information document available at https://documentation.marine.copernicus.eu/QUID/CMEMS-IBI-QUID-005-002.pdf; E.U. Copernicus Marine Service Information, 2023a).

Daily zonal (V) and meridional (U) components of surface seawater velocity, at 0.25° spatial resolution, were acquired from CMS's Global Total (COPERNICUS-GLOBCURRENT), Ekman and Geostrophic currents at the surface and 15m dataset (ID: MULTI-OBS_GLO_PHY_MYNRT_015_003). Velocity fields are obtained by combining CMS satellite-derived geostrophic surface currents and modelled Ekman currents. Validation of surface currents, against the YoMaHa database (Lebedev et al., 2007), showed that zonal velocity is more accurate than the meridional velocity. Moreover, no global bias was detected, though increased uncertainties were observed near the equator (see quality information document available at https://documentation.marine.copernicus.eu/QUID/CMEMS-MOB-QUID-015-003.pdf; E.U. Copernicus Marine Service Information, 2023b). Daily data of the concentration of nitrate (NO_3^-), ammonium (NH_4^+), phosphate (PO_4^{3-}), silicon (Si) and iron (Fe), at 0.083° spatial resolution, were also retrieved from CMS's Atlantic-Iberian Biscay Irish-Ocean BioGeoChemistry Non-Assimilative Hindcast dataset (ID: IBI_MULTIYEAR_BGC_005_003). Model-derived nutrient data are in good agreement with the Biogeochemical Argo observations, but tend to overestimate low concentrations and underestimate high concentrations (see quality information document available at https://documentation.marine.copernicus.eu/QUID/CMEMS-IBI-QUID-005-003.pdf; E.U. Copernicus Marine Service Information, 2023c).

2.3. Phytoplankton data

2.3.1. Chlorophyll-a concentration

Daily level 3 satellite-derived Chl-a data, at 4 km spatial resolution, used as a proxy for total phytoplankton biomass, was retrieved for the whole study area from the European Space Agency (ESA)'s Ocean Colour Climate Change Initiative dataset (version 6.0; Sathyendranath et al., 2023). This merged product is created by shifting the wavelengths of five ocean color sensors: Sea-viewing Wide Field of View Sensor, MODISA, Medium Resolution Imaging Spectrometer, Visible Infrared Imaging Radiometer Suite, Sentinel 3A and 3B Ocean and Land Colour Instrument data to match MERIS bands, applying bias correction, and computing per pixel uncertainty estimates (Copernicus Climate Change Service, 2019). The atmospheric correction with POLYMER is performed independently for each sensor, and Chl-a is estimated using a blended combination of OCI, OCI2, OC2, and OCx algorithms (the updated OC3/OC4 band ratio algorithm; O'Reilly and Werdell, 2019), which attempts to weight the outputs of the best-performing algorithms based on the water types present, thereby improving performance in nearshore Case-2 waters (Copernicus Climate Change Service, 2019).

2.3.2. Abundance of toxigenic phytoplankton taxa

Abundance of toxigenic phytoplankton taxa associated with different human syndromes (Amnesic Shellfish Poisoning, ASP; Diarrhetic Shellfish Poisoning, DSP; Paralytic Shellfish Poisoning, PSP), in Portuguese and Spanish coastal waters, was obtained from the Portuguese Institute of the Sea and Atmosphere (IPMA; https://www.ipma.pt/pt/index.html/, accessed February 15, 2023) and Consejería de Agricultura, Pesca, Agua y Desarrollo Rural de la Junta de Andalucía (https://www.juntadeandalucia.es/agriculturaypesca/moluzonasprodu/, accessed February 23, 2023) public databases. Several nearshore sampling stations within classified shellfish coastal production areas (L6, L7a, L7c, L8, L9, 101, 102, 103, 105, 106, 11, 107 and 108; see Fig. 1 and S1) were considered (from north to south, and west to east): Sesimbra, Galapos, Comporta, Aljezur, Porto de Mós, Praia Dona Ana, Falésia, Praia de Faro, Culatra, Monte Gordo, Isla Canela, Barra del Terron, Punta Umbria, Mazagon, Matalascañas, Zona Marítima de Doñana, Doñana Norte, and Doñana Sur. The coastal area 11 corresponded to the aggregation of areas 107 and 108 prior to mid-2018, after which they were treated separately. In situ field sampling undertaken within the national monitoring programs

was not concurrent among sites. Sampling frequency was approximately weekly, and intensified during periods of historically high abundances of HAB-forming species and/or concentrations of phytotoxins in bivalve molluscs, or after occurrences above alert or threshold levels (see IPMA, 2013). Toxigenic taxa included ASP-producers (*Pseudo-nitzschia delicatissima* group, <3 μm width, and *P. seriata* group, >3 μm width); DSP-producers (*Dinophysis* spp., *Prorocentrum* spp. Except *P. minimum*, and *Phalacroma* sp.); and PSP-producers (*Gymnodinium* spp., including *G. catenatum*, and *Alexandrium* spp.). The coastal areas 107 and 108, along the Spanish coast, were aggregated, as HAB abundances were statistically not significantly different ($p > 0.05$; data not shown). Additionally, abundances were binned for each month to reduce computational demands. Phytoplankton abundance was quantified using the Utermöhl method (see Utermöhl, 1958), which involves settling a Lugol's iodine-preserved water sample, followed by analysis under inverted microscopy and identification based on morphological traits and taxonomic reference literature. Since this approach often does not allow identification at the species level, and the Portuguese public database does not provide species-specific information, abundance data were grouped according to the associated human syndrome, following (Lima et al., 2022). Abundances reported in the public databases as below the limit of detection (LOD) or as "ND" (undetected) were treated as zero. Data with other descriptions, including "NDi" (unavailable), "NR" (unfulfilled) or "NQ" (unquantified) were excluded.

Temporal and spatial variability of each HAB-forming group within individual coastal production areas, was examined using the Seasonal-Trend decomposition of time series based on Loess (STL), a non-parametric regression analysis, which allows the separation of the long-term trend from the seasonal (intra-annual) and remainder components (Kwok et al., 2023). STL decomposition has several advantages, including the capability of identifying a seasonal component that changes over time, the responsiveness to nonlinear trends, and the robustness in the presence of outliers and missing data. Although this method was initially proposed to accommodate missing observations, it becomes ineffective when gaps are extensive or occur consecutively (Hyndman and Athanasopoulos, 2018). Most coastal production areas had a percentage of missing observations less than 5%, but others (L8, 101, 103, 105, 106, 107 and 108 areas) contained extensive gaps (>60%) and were, thus, excluded from time series analysis. STL also requires the specification of a list of parameters. Following the original recommendations (Cleveland et al., 1990), the parameter controlling seasonality cycle length was set to 12. For the smoothing parameters, an iterative process was applied by varying the length of the seasonal and trend smoothing windows between 7 and 49. Then, the best combination of parameters was selected for each coastal production area and group (see Table S2), based on the lowest in-sample root-mean-squared error (RMSE) relative to a benchmark model with a periodic behaviour (seasonal = 999, trend = 25; see Cristina et al., 2016 for further details).

2.4. Three-step classification procedure of coastal domains that prioritize environmental variables best describing HAB variability

The workflow steps of the proposed biologically informed partitioning strategy are summarized in Fig. 2. Firstly, a dissimilarity analysis was applied to select the list of independent partitioning variables. The resulting variables were organized into three datasets, according to biological relevance and data source uncertainty. The first dataset comprised only Chl-a, as a proxy of phytoplankton biomass, i.e., with higher biological relevance. A second dataset incorporated other satellite-derived variables, in addition to Chl-a, specifically physical variables (SST, PAR, ADT and wind speed and direction), thereby also integrating potential environmental drivers of phytoplankton. The third dataset combined modelled and satellite-derived variables (hereafter referred to as mixed dataset), to maximize the availability and diversity of information, due to the lack of *in situ* observations for the study area. A three-step combined strategy, including data reduction and unsupervised classification, a suite of modelling techniques, and the reapplication of the unsupervised classification considering the best selected datasets for each HAB-forming group, at a pixel-level, was then used to partition the coastal areas off SWIP into regions that prioritize environmental variables that best describe the variability of key toxigenic phytoplankton taxa.

Following the classification procedure, differences in the environmental properties that best described variability for different HAB groups between coastal regions, were analysed using the Kruskal-Wallis test, a one-way analysis of variance on ranks, followed by multiple pairwise comparisons using the Dunn's test (Dunn, 1961). All statistical tests were considered at a 0.05 significance level. Graphical representations, pre-processing analyses and model development were performed using Python version 3.12.3 (https://www.python.org/, accessed January 15, 2025), and specific modules such as xarray, pandas, numpy, netCDF4, basemap, shapefile, scipy, sklearn, scikit-posthocs, statsmodels, seaborn and matplotlib. The following sections describe thoroughly each workflow step.

2.4.1. Pre-processing analysis: identification of partitioning variables

Spearman rank correlation coefficient (r_s) was used to evaluate the strength of monotonic relationships between the 16 potential partitioning variables, and the correlation distance (1- r_s) was then used to build a dissimilarity hierarchical cluster tree. Two groups of redundant, strongly correlated variables (dissimilarity below 0.25) were identified: NO_3^- and NH_4^+ concentrations, and SST and PO_4^{3-} concentration (data not shown). A single variable was selected from each group, NO_3^- concentration and SST, resulting in a dataset with 14 partitioning variables (SST, salinity, MLD, PAR, surface zonal and meridional components of wind, wind speed, zonal and meridional components of surface seawater velocity, ADT, NO_3^-, SiO_4^{4-}, Fe and Chl-a).

2.4.2. First step: data reduction and classification of the whole study area

The monthly climatologies (period: 2002-2021) of each of the three environmental datasets were used to train a self-organizing map (SOM; Kohonen, 1990), with the rows representing different pixels and the columns representing time steps (i.e., monthly climatologies of the variables included in each dataset). SOM is an artificial neural network technique that produces a low-dimensional (typically two-dimensional), discretized representation of the input data referred to as a neuron map (Kohonen, 2013). SOM partitions the initial data set so that each neuron on the map represents a specific pattern characterized by a synthetic multidimensional vector known as the reference vector (prototype vector or weight vector; El Hourany et al., 2019). The reference vector may be perceived as the position of the neuron in the data space, containing the same number of variables of the input dataset (in this case, 14). In contrast to other artificial neural networks, this type of representation preserves the topology and the neighbourhood relationships of the input space, which means that two neighbouring neurons on the map will be represented by two close reference vectors (according to the Euclidean distance) in the space of the input data, and conversely, two close observations in the input layer will be projected onto two close neurons of the map (Leloup et al., 2007). A hexagonal structure for the map was selected as it minimizes directional bias, resulting in a more uniform and accurate data representation (Kohonen, 2001; Liu et al., 2006). Batch learning was used to ensure that all observations are shown to the neurons prior to updating their values (Demuth et al., 2014). The neuron with the closest reference vector to the input data vector was found using the Euclidean distance, ignoring the dimensions corresponding to missing data. The number of neurons on the map must be predefined by the user and is dependent upon the level of detail desired in further analysis. Different neuron map sizes (ranging from 2 to 100 neurons) were tested, and the total error of each map size, estimated as the sum of quantization and topographic errors, was used to determine the optimal map size. The quantization error represents the average distance between each data vector and their best match unit (BMU), and

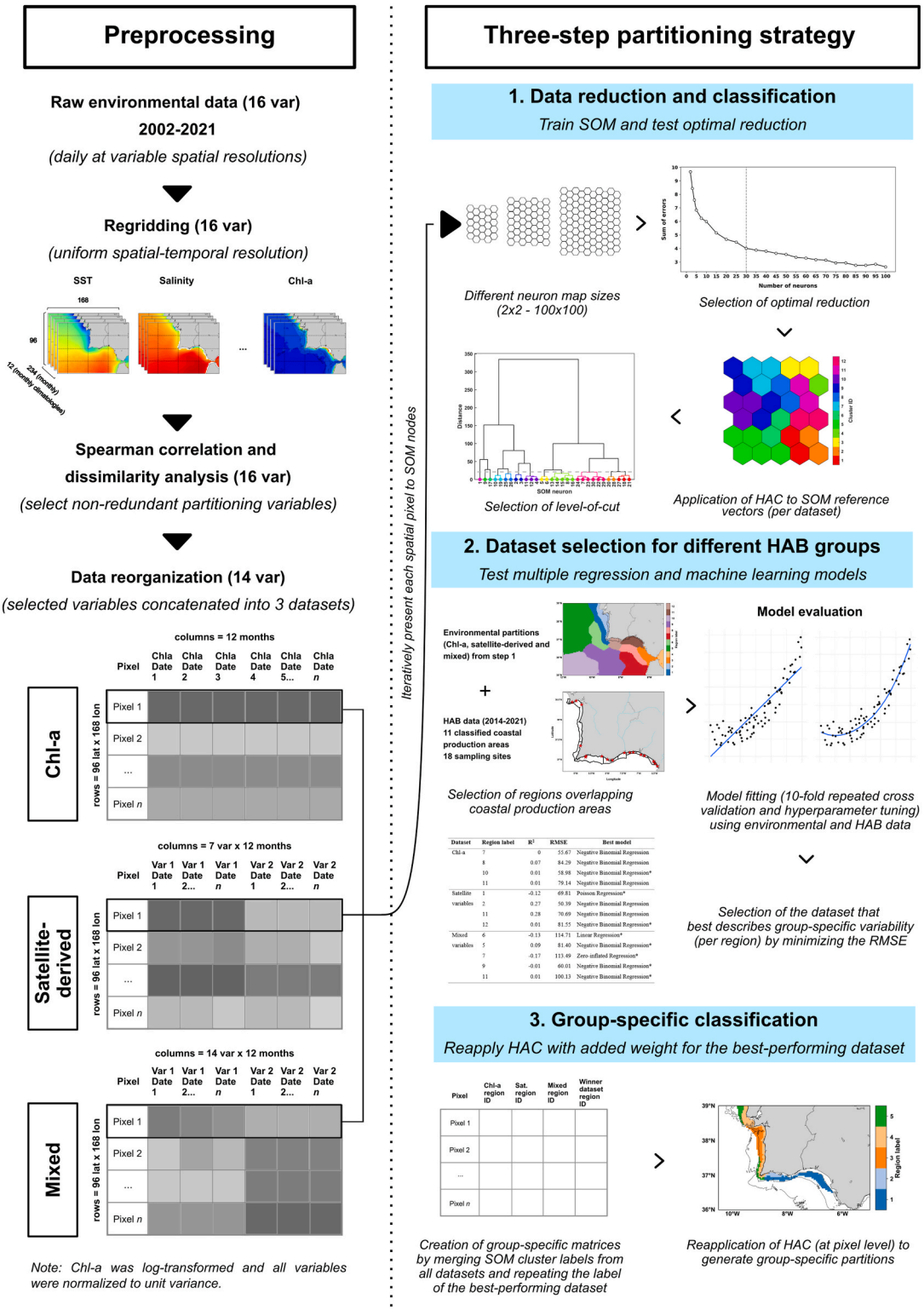

Fig. 2. Flow diagram representing all the stages of the proposed partitioning strategy that prioritizes environmental factors that best describe the variability patterns of the abundance of harmful algal bloom (HAB)-forming taxa off SW Iberia. Target taxa included producers of amnesic shellfish poisoning (ASP), diarrhetic shellfish poisoning (DSP), and paralytic shellfish poisoning (PSP). The preprocessing stage (left panel) involved regridding the raw environmental data to a uniform spatial resolution and two temporal resolutions (monthly and monthly climatologies), followed by Spearman's correlation and dissimilarity analysis to select non-redundant partitioning variables, and data organization into three datasets: Chl-a, satellite-derived and mixed (modelled and satellite-derived variables). The three-step partitioning strategy (right panel) included the following steps: (1) data reduction and unsupervised classification using self-organizing maps (SOM) and hierarchical agglomerative clustering (HAC) to define the initial environmental partitions from each dataset; (2) dataset selection for each HAB group using multiple empirical-statistical models, applied to regions overlapping classified coastal production areas; and (3) group-specific classification to generate the final (biologically informed) environmental partitions, by reapplying HAC, at the pixel level, with added weight for the datasets that best described the variability patterns of the corresponding HAB group. See text for further details.

the topographic error corresponds to the percentage of data vectors for which two BMUs are not neighbouring map units (Vesanto et al., 2000). The optimal number of neurons was defined as the first map size of three consecutive reductions in the total error less than 5% (Fendereski et al., 2014). After several iterations of the SOM algorithm applied to each dataset, considering different neuron map sizes, a map with 30 neurons was selected to proceed with the clustering analysis. The SOM Toolbox for Matlab was used with the monthly climatological value, for all datasets, composed of variables normalized to unit variance. Prior to normalization, Chl-a data were log-transformed (Campbell, 1995) to minimize the effect of very high values (particularly in coastal areas) and better represent the full range of variability, and to facilitate the identification of the shape of seasonal cycles by SOM (e.g. Marchese et al., 2022; El Hourany et al., 2021). Training length ranged from 50 to 200 epochs, and the neighbourhood radius from 0.05 to 1000, over four successive batch training runs.

The resulting SOM was then partitioned into regions with similar environmental conditions using the Hierarchical Agglomerative Clustering (HAC) analysis. HAC is an unsupervised learning technique (Wilks, 2011) that represents hierarchically the underlying relationship among objects through a dendrogram (Pramlall et al., 2024; Karna and Gibert, 2022). The dendrogram is formed by recursively merging objects that are closer in a n-dimensional space, using a division that minimizes differences between objects of a given cluster based on Euclidean distance, while maximizing differences between objects of different clusters, based on Ward's linkage, respectively. In contrast to other clustering approaches (e.g., K-means), HAC analysis does not require the number of clusters to be defined a priori. In fact, the number of clusters emerges from the clustering process itself and is often determined by visual inspection of the dendrogram, by identifying the largest horizontal gap among the major branches (Karna and Gibert, 2022). The level-of-cut varied between 4 and 12 clusters (regions), and the optimal level was determined by a combined assessment of branch length analysis and physical interpretability, based on reported oceanographic patterns for the study area (Relvas et al., 2007). The dendrogram structure was analysed to identify potential cut levels, and the level that provided greater discrimination in coastal areas while retaining physical meaning was selected.

2.4.3. *Second step: identification of datasets prioritizing environmental variables that best describe variability of HAB-forming groups in coastal areas*

Due to the limited availability of quantitative information required to develop fully mechanistic-numerical models (McGillicuddy et al., 2005; McGillicuddy et al., 2011; Ruiz-Villarreal et al., 2016), HAB prediction is mostly based on conventional empirical-statistical modelling strategies, in isolation or combined into hybrid models (see Lima et al., 2022 and references therein). These include different regression methods (e.g. Blum et al., 2006; Lane et al., 2009; Anderson et al., 2009; Sarma et al., 2023), generalized linear models (e.g. Anderson et al., 2010; Cusack et al., 2015; Zhou et al., 2021; Díaz et al., 2022), and generalized additive models (e.g. Lima et al., 2022; Torres Palenzuela et al., 2019; Detoni et al., 2024). Machine learning models have been increasingly used due to their capacity to deal with complex, often non-linear, noisy datasets, and outperforming traditional approaches (e.g., see reviews by Cruz et al., 2021; Park et al., 2024, and references therein; Díaz et al., 2025). In this context, a set of models of varying complexity, including Linear Regression, Polynomial Regression, Poisson Regression, Compound Poisson-Gamma Regression, Negative Binomial Regression, Zero-Inflated Regression, and Random Forests, was applied to analyse which of the three datasets (and associated predictors) best explained the variability of different HAB-forming groups in the regions, identified during the previous step, overlapping with the classified coastal production areas (Fig. 1 and S1). The response variables were the abundance of phytoplankton associated with ASP, DSP and PSP human syndromes, using both non-transformed and log-transformed values. Log-transformation was applied to minimize the influence of extreme values, approximate the distribution to normality and ensure model convergence. Abundance values were calculated as weighted means, according with the proportional contribution of each classified coastal production area within each region. Non-redundant predictor variables, identified using Spearman's rank correlation and subsequent dissimilarity analysis, included ocean physical (SST, Salinity, PAR, ADT, MLD, geostrophic current velocity), chemical (nutrient concentrations), and biological variables (Chl-a), and meteorological variables (wind direction and magnitude), corresponding to each region. Both predictor and response variables were aggregated to a monthly resolution.

Simple linear regression is typically used to describe the covariation between a continuous response variable and continuous predictor variables. However, it assumes that the relationship between variables is linear and that the residuals have constant variance (homoscedasticity) across all values of the predictors (Helsel et al., 2020), which is rarely suitable for phytoplankton abundance data. Poisson regression is designed for modelling count data (such as phytoplankton cell counts), although the majority of the datasets violate the basic assumption that the mean of the underlying distribution is equal to the variance (e.g. Lima et al., 2022). The Negative Binomial model provides a more flexible generalization of Poisson regression that explicitly accounts for overdispersion (i.e., when the variance is larger than the mean ;Cameron and Trivedi, 1998; Lloyd-Smith, 2007; Hilbe, 2011). However, when data also contains an excess of null observations, which is common in HAB monitoring, the Negative Binomial model might be inadequate. In such cases, two common alternatives are the compound Poisson-Gamma (or Tweedie) models and the Zero-Inflated models. Tweedie models account for zero outcomes as a natural consequence of the underlying compound Poisson process, with the probability of zeros determined by the expected value of the response. Zero-inflated models incorporate an additional zero-generating mechanism by combining a logistic component, which captures and explicitly models the excess zeros not attributable to the count distribution (referred to as structural zeros), with a count component that models the distribution of nonzero outcomes as well as zeros that arise probabilistically within the count process itself (referred to as sample zeros). This added flexibility comes at the cost of extra parameters and increased complexity, thus Tweedie models are generally preferred to zero-inflated models (Lambert, 1992; Dunn and Smyth, 1996; Zuur and Ieno, 2021). In this study, zero-inflated models could not be fitted to ASP- and DSP-producers due to the relatively reduced number of null observations. Random Forest models are an ensemble machine learning algorithm that has been successfully employed to HAB prediction (e.g. Valbi et al., 2019; Derot et al., 2020; Ottaviani et al., 2020). These models combine multiple decision trees, and for each tree, only a subsample of the input data is used for training, and a random selection of predictors is available for splitting at each decision node. The final output is then generated by averaging the results of all trees making up the forest (Breiman, 2001; Liaw and Wiener, 2002). The combination of the tree structure and the bootstrapping method, which uses sampling with replacement to make the selection procedure completely random (without prior assumptions), enables the random forest model to be robust against overfitting, provide unbiased internal estimates of the generalization error, and show high predictive performance (Breiman, 2001; Liaw and Wiener, 2002; Thomas et al., 2018).

The optimal parametric configuration for each combination of model features, HAB group, and method was selected by a hyperparameter optimization. This was based on a grid search approach, i.e., models using different parameters controlling the learning process were trained and evaluated, considering a 10-fold repeated cross-validation scheme, and the best models were selected based on cross-validated optimization metrics (Bergstra et al., 2011). The different sets of training parameters considered are summarized in Tables S3–S5. Models were evaluated by calculating the R^2 and the Root Mean Squared Error (RMSE). The former

quantifies the proportion of variance in the observed data explained by the model, with values ranging from 0 to 1, while the latter indicates the mean error between the predicted and measured values. In contrast to R^2, RMSE is scale-dependent on the target variable, and thus is reported in its original units. When considering log-transformed response variables, model performance metrics were computed using back-transformed predictions to ensure proper model comparison. These metrics have been widely applied in HAB prediction studies due to their interpretability and ability to capture different aspects of model performance (Mermer et al., 2025). The dataset that minimized the RMSE when predicting group-specific abundances was selected as the most suitable for explaining monthly variability patterns. Thus, for each HAB group, the best-performing environmental dataset (Chl-a, satellite-derived or mixed) was identified based on the model that performed best at predicting the abundance of the corresponding phytoplankton group, in regions that spatially overlapped with classified coastal shellfish production areas off SW Iberia.

2.4.4. Third step: classification of coastal areas prioritizing environmental variables that best describe the variability of specific HAB groups

After selecting the datasets that prioritized environmental variables that best described the variability of specific HAB groups, a fourth column was added to each group-specific dataset containing three columns with the original SOM cluster labels from the Chl-a, satellite-derived and mixed datasets, resulting from the first step of our partitioning strategy. The additional fourth column repeated the cluster label of the best-performing dataset, to attribute more weight in the classification process to the environmental variables that best described group-specific variability, thus eliminating the need to reuse the original datasets. HAC analysis was then reapplied to these three data matrices, at a pixel level. The optimal number of regions was determined by selecting the partition showing greater discrimination (i.e., higher number of regions) in the SWIP coastal areas.

3. Results

Basic statistical information on abiotic environmental variables and Chl-a, over the whole spatial domain, is summarized in Table S6. The

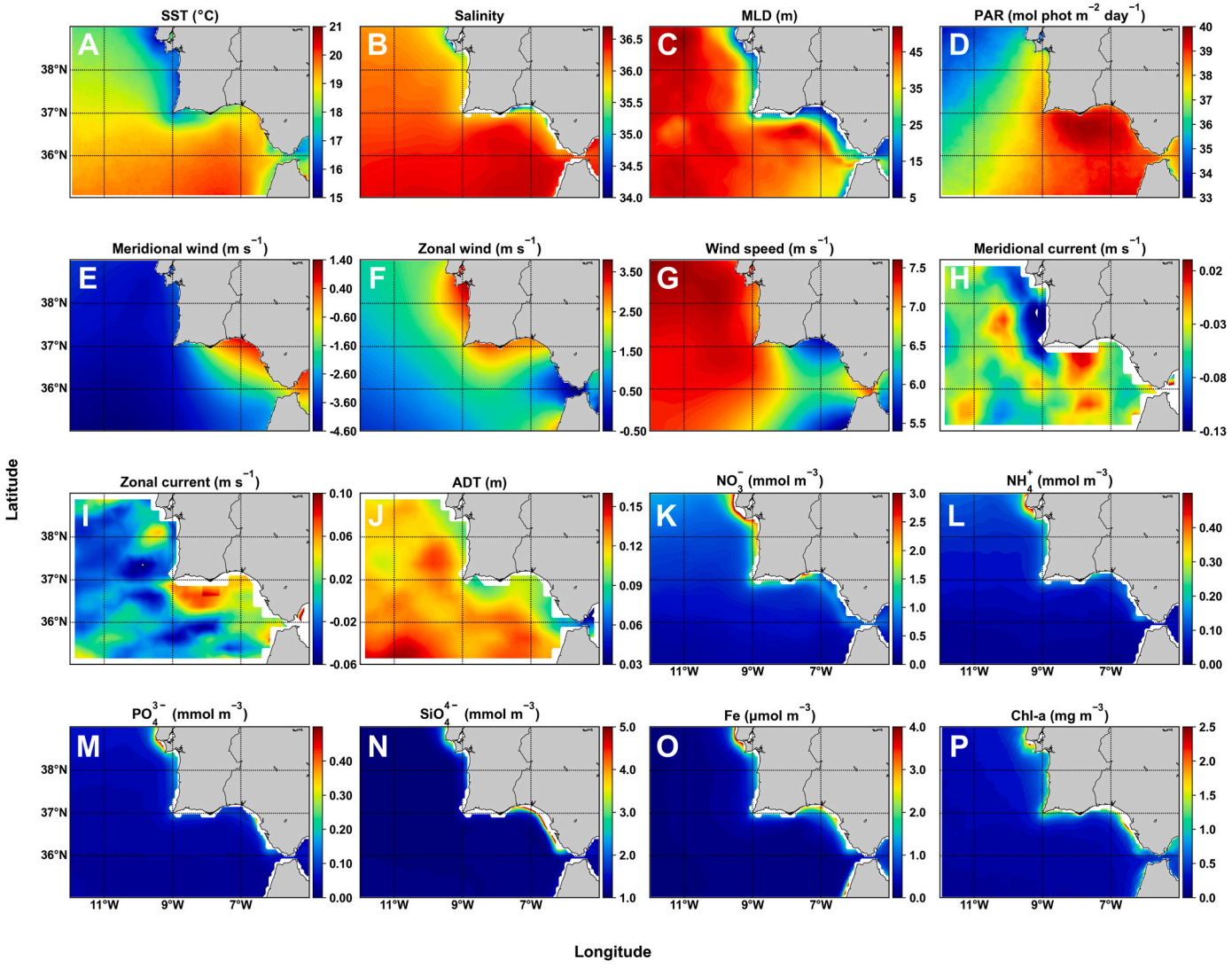

Fig. 3. Distribution of mean values at surface of (A) sea surface temperature (SST); (B) salinity; (C) mixed layer depth (MLD); (D) incident photosynthetically active radiation (PAR); (E) meridional and (F) zonal wind speed; (G) wind intensity; (H) meridional and (I) zonal geostrophic current speed; (J) absolute dynamic topography (ADT); (K) subsurface nitrate concentration (NO3-); (L) ammonium concentration (NH4+); (M) phosphate concentration (PO43-); (N) silicate concentration (SiO44-); (O) iron concentration (Fe); and (P) chlorophyll-a concentration (Chl-a), during the period between July 2002 and December 2021, off SW Iberia.

spatial distribution of environmental properties off SW Iberia revealed higher mean SST in southern domains (Fig. 3A), and higher salinity in the semi-sheltered GoC basin, while lower salinities occurred near major river mouths (e.g., Tagus, Guadiana and Guadalquivir; Fig. 3B; see Fig. 1 for location of river discharge points). Lower MLD (Fig. 3C) and weaker winds (Fig. 3G), coincident with higher PAR (Fig. 3D), were also observed in the GoC. Northerly winds ($V < 0\ m\ s^{-1}$; Fig. 3F) prevailed along the west coast, associated with equatorward currents (Fig. 3I). Westerlies ($U > 0\ m\ s^{-1}$) and easterlies (Fig. 3E) prevailed in the GoC and the Strait of Gibraltar, respectively, with the former associated with eastward currents (Fig. 3H). Strong coastal-offshore gradients were also detected, with increased ADT (Fig. 3J) and decreased nutrient concentrations (Fig. 3K–O) and Chl-a (Fig. 3P) towards offshore waters.

3.1. Variability of specific HAB-forming taxa in classified coastal production areas

The abundance of HAB groups varied during the 8-year period (2014-2021), including both seasonal and longer-period components, with marked differences observed across coastal production areas and groups (Fig. 4). Overall, for the seasonal component, ASP-producers exhibited higher variability in L7a, and DSP-producers in L7c, with the latter showing similar patterns across production areas. ASP-producers, showed bimodal annual cycles, with peaks occurring during spring and summer (L6, L7c) or autumn (L9). Yet, in L7a, a quasi-unimodal cycle with a summer peak was detected (Fig. 4A). DSP-producers in L6 and L7a showed bimodal annual cycles until 2016, with spring and summer peaks, but changed to unimodal cycles with a greater spring peak in the latest years (Fig. 4C). In L7c, a progression from bimodal to a unimodal cycle with a late-summer peak was observed. DSP-producers in L9 and 102 areas exhibited consistent unimodal annual cycles with a summer peak until 2017, transitioning to bimodal cycles with spring and summer/autumn peaks. PSP-producers, in all coastal production areas, except L9, exhibited a shift from bimodal (spring and summer/autumn peaks) to unimodal annual cycles, the latter with peak values from early to late spring (Fig. 4E). However, the results for the latter group should be interpreted with caution due to the large number of null observations, that likely introduced discrepancies into the decomposition of both seasonal and longer-period variability components.

Concerning the longer-period component, an increasing trend in the abundance of ASP- and DSP-producers was observed along the west coast (L6, L7a) until 2019, followed by a stabilization or decline (Fig. 4B–D). Over the south coast, ASP increased Copernicus Climate Change Service, Climate Data Store, 2018 onwards (L7c, L9; Fig. 4B), and DSP-producers showed no clear trend (L9, 102; Fig. 4D). PSP-producers showed a decreasing trend in all coastal production areas until late 2017, stronger for the southern areas (L7c, L9), followed by a stabilization (Fig. 4F).

3.2. Partitioning of the whole study area

The unsupervised classification of the whole study area identified 11 regions using the Chl-a dataset (Fig. 5A), and 12 regions using the satellite-derived and mixed datasets (Fig. 5B–C), with variable configurations across datasets. Five regions over the coastal domain were identified by the mixed dataset (Fig. 5C), and four by the other datasets (Fig. 5A–B). A coastal region covering the west coast, turning around

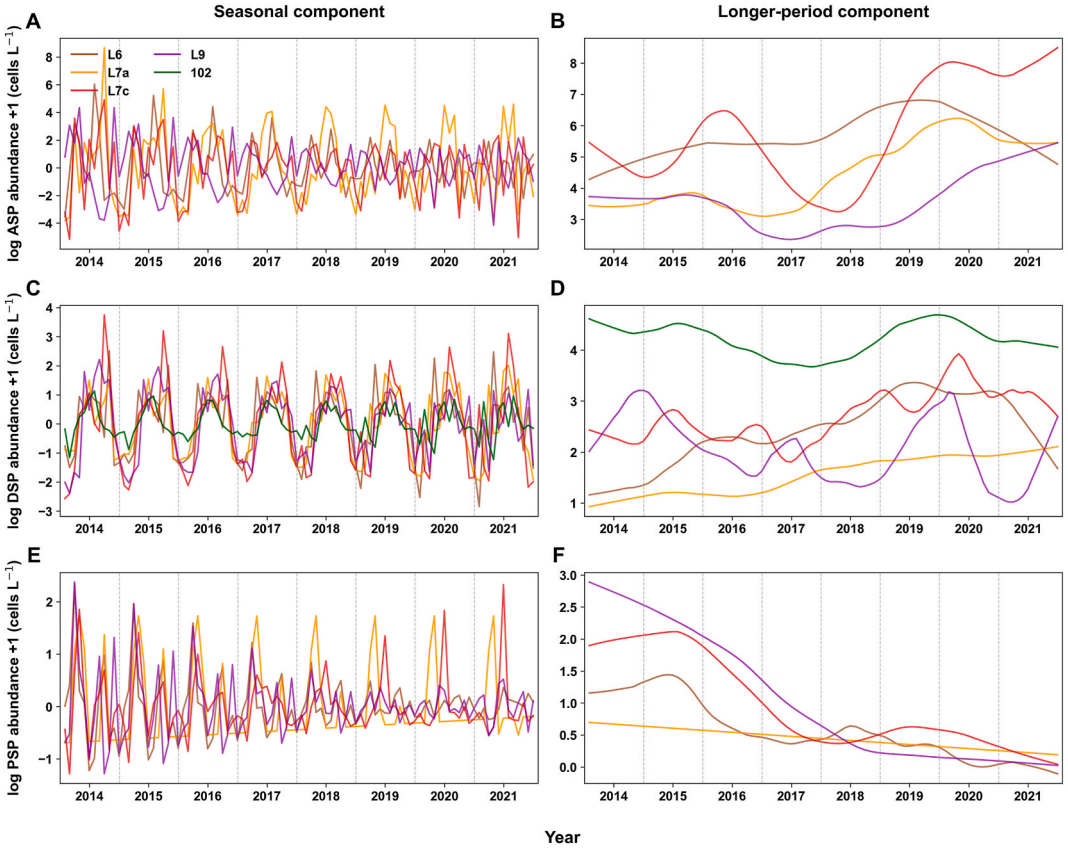

Fig. 4. Temporal variability of the abundance of toxigenic phytoplankton taxa, during the 2014-2021 period, including the seasonal (left column; panels A, C and E) and longer-period components (right column; panels B, D and F). Taxa included ASP (Amnesic Shellfish Poisoning), DSP (Diarrhetic Shellfish Poisoning) and PSP (Paralytic Shellfish Poisoning)-producers, in different classified coastal shellfish production areas off SW Iberia (L6, L7a, L7c, L9 and 102). Note differences in scale between harmful algal bloom (HAB) groups. Similarity between colours indicates proximity of production areas (see Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

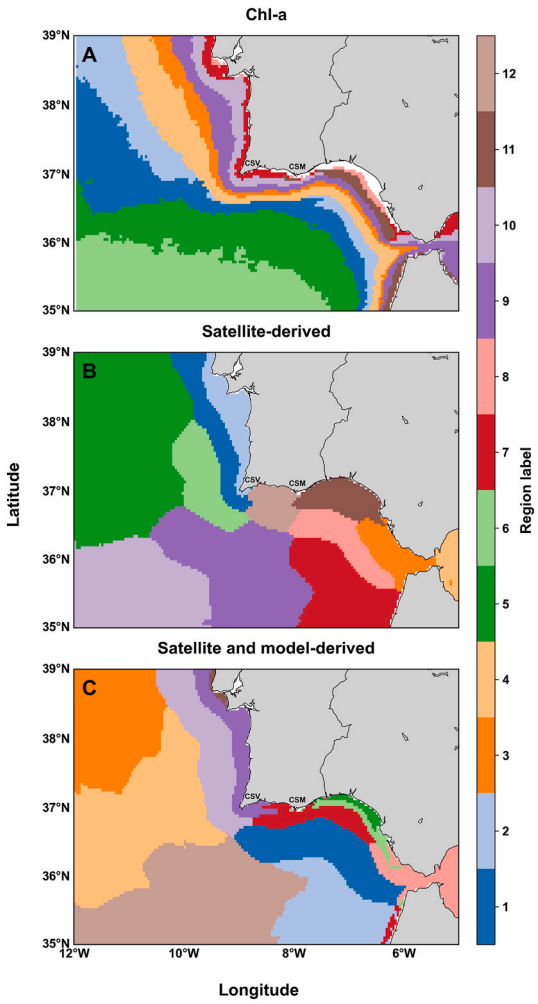

Fig. 5. Regions with coherent, co-varying environmental variability patterns off SW Iberia (period: 2002-2021), established based on the application of self-organizing maps and clustering classification methods to different datasets: (A) chlorophyll-a (Chl-a); (B) satellite-derived variables (including Chl-a); and (C) satellite- and model-derived variables. CSV and CSM depict the location of prominent topographic features, Cape São Vicente and Cape Santa Maria, respectively. Region labels are not comparable across datasets, i.e., colours do not correspond to the same regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

CSV and extending eastwards, was identified by all datasets. A clear separation between the western and eastern sectors of CSM was detected by the satellite-derived dataset (Fig. 5B). Two distinct regions over the shelf and slope, east of CSM, and an additional region influenced by freshwater (note the distinct small areas near the Tagus and Guadiana river mouths; see Fig. 1 for location of river discharge points), were also detected by the Chl-a and mixed datasets (see region 8 in Fig. 5A and region 11 in Fig. 5C).

3.3. Identification of the best datasets prioritizing environmental variables that best describe variability of HAB-forming groups

Distinct empirical-statistical models were applied to identify the best relationships between the potential environmental predictors included in the three datasets, and the abundance of HAB groups for each region overlapping different classified coastal production areas (Tables S3–S5). Overall, simpler algorithms (e.g., regression models) performed better than more complex ones (e.g., random forests; see Table 1 and Tables S3–S5), and ca. 50% of the models exhibited improved performance with log-transformed data. Negative binomial regression models showed the highest performance, whereas linear and Poisson regression models had the worst predictive capacity, with some exceptions for PSP-producers (Tables S3–S5). The best-performing models were obtained for ASP-producers, with 46% of the models achieving R^2 values above or equal to 0.30, suggesting moderate model predictive skills. Models for DSP and PSP-producers showed lower predictive capabilities, with average R^2 values of 0.05 and 0.02, respectively. Low R^2 values (and high errors) generally suggested that model performance was limited, even for the best models (Table 1).

For ASP-producers, the best dataset and model results, were obtained for the west and south coasts, with R^2 values between 0.45 and 0.55, for satellite and mixed datasets (see Table 1 and Fig. 5). For DSP group, the best dataset and model was associated with inner and outer shelf areas in the western sector of CSM, with an R^2 value of 0.56 for the satellite dataset. For PSP-producers, the best dataset and models were obtained for the western coast and the inner and outer shelf areas in the eastern sector of CSM, showing R^2 values around 0.27, for the satellite dataset (see Table 1 and Fig. 5).

3.4. Partitioning of coastal waters off SWIP prioritizing environmental variables that best describe variability of different HAB-forming groups

Basic statistical information on the environmental variables used to partition the coastal domain off SWIP, for different HAB groups and coastal regions, is summarized in Tables S7–S9. Our three-step partitioning strategy identified five regions predominantly within the coastal domain off SWIP, for all HAB groups (Fig. 6): two inner shelf areas in the southern and western sectors (depth <200 m; regions 2 and 4); two outer shelf-slope regions, extending up to the 1000 m isobath near the Tagus river mouth (regions 1 and 3 on the south and west coasts, respectively); and one transitional region between outer-shelf and oceanic waters off the west coast (region 5). Each coastal region showed a relatively small spatial coverage, accounting for 6-30% of the total coastal-slope domain (see Table S10). For the two inner shelf regions, higher nitrate concentrations and lower SST were detected on the west coast, whereas higher silicate and iron concentrations were observed on the south coast ($p < 0.05$; see Tables S7–S9 and associated Kruskal-Wallis results). The partitions prioritizing the variability in ASP- and DSP-producers displayed similar configuration patterns (Fig. 6A and B). However, the partition prioritizing the variability of PSP-producers showed a relative increase in the area covered by region 1, and a decrease in the area of region 2, both located on the south coast (Fig. 6C).

4. Discussion

A novel approach combining SOM, HAC and model fitting, was used to develop a more biologically representative environmental partition of the coastal domain off SW Iberia. The use of 14 independent partitioning variables, covering a 19-year period, identified a total of 11 or 12 regions in the whole study area, depending on the dataset, with 4 to 5 located in the coastal-slope domain. Despite the similarity in number, region configurations were different across datasets, then highlighting the need to identify which dataset best represented the environmental conditions for different HAB groups. Simpler empirical-statistical models outperformed more complex models and revealed that the satellite-derived and mixed datasets were more representative of the environmental conditions best describing HAB variability. The use of this information, at a pixel level, identified five environmentally coherent regions potentially representative of the variability of each HAB group. However, their value as proxies for HAB distribution patterns is constrained by the predictive skills of the best models and available datasets.

Table 1
Summary of the best-performing models used to predict the abundance of toxigenic phytoplankton groups responsible for ASP (Amnesic Shellfish Poisoning), DSP (Diarrhetic Shellfish Poisoning) and PSP (Paralytic Shellfish Poisoning) human syndromes, during the period 2014-2021, for coastal regions that spatially overlap with coastal shellfish production areas in the area of SW Iberia, considering distinct predictor datasets (chlorophyll-a, Chl-a, satellite-derived and mixed datasets). Information on the best-fitted model and the corresponding coefficient of determination (R^2) and root-mean-squared error (RMSE) is provided for each combination of harmful algal bloom (HAB) type, dataset and region. Region labels are not comparable across datasets. Asterisks indicate models where the response variable (group-specific abundance) was log-transformed. The total number of samples used for model fitting was 96. See Fig. 5 for region location considering different datasets, and Tables S3–S5 for information on model hyperparameters. Information on region's location relative to the coastline orientation and, over the south coast, relative to Cape Santa Maria (CSM), is provided after the region label (only for ASP-producers). C: coast; Sl: slope; W: west; S: south; wCSM: western sector of CSM; eCSM: eastern sector of CSM.

HAB taxa	Dataset	Region label	R^2	RMSE	Best model
ASP	Chl-a	7 (C)	0.17	18476.97	Negative Binomial Regression
		8 (C-eCSM)	0.01	16314.18	Negative Binomial Regression
		10 (Sl)	0.02	13117.81	Negative Binomial Regression
		11 (Sl-eCSM)	0.004	18345.94	Negative Binomial Regression
	Satellite variables	1 (Sl-W)	0.30	25587.14	Negative Binomial Regression
		2 (C-W)	0.45	12080.88	Negative Binomial Regression
		11 (C,Sl-eCSM)	0.05	14465.32	Negative Binomial Regression*
		12 (C,Sl-wCSM)	0.34	34117.80	Negative Binomial Regression
	Mixed variables	6 (C,Sl-eCSM)	0.14	18249.12	Negative Binomial Regression
		5 (C-eCSM)	0.36	22982.22	Negative Binomial Regression
		7 (C,Sl-S)	0.39	26080.22	Negative Binomial Regression
		9 (C-W)	0.55	18020.83	Negative Binomial Regression
		11 (C)	0	15042.61	Negative Binomial Regression*
DSP	Chl-a	7	0.001	545.10	Negative Binomial Regression*
		8	0.01	3076.34	Negative Binomial Regression
		10	0.01	775.61	Negative Binomial Regression*
		11	0.04	952.04	Negative Binomial Regression
	Satellite variables	1	−0.01	141.57	Negative Binomial Regression*
		2	0	908.28	Negative Binomial Regression*
		11	0.03	922.30	Negative Binomial Regression*
		12	0.56	1139.04	Negative Binomial Regression
	Mixed variables	6	0.02	217.36	Negative Binomial Regression*
		5	−0.01	1535.19	Negative Binomial Regression*
		7	0.01	559.18	Negative Binomial Regression*
		9	0	618.15	Negative Binomial Regression*
		11	0.001	1567.05	Negative Binomial Regression*
PSP	Chl-a	7	0	55.67	Negative Binomial Regression
		8	0.07	84.29	Negative Binomial Regression
		10	0.01	58.98	Negative Binomial Regression*
		11	0.01	79.14	Negative Binomial Regression
	Satellite variables	1	−0.12	69.81	Poisson Regression*
		2	0.27	50.39	Negative Binomial Regression
		11	0.28	70.69	Negative Binomial Regression
		12	0.01	81.55	Negative Binomial Regression*
	Mixed variables	6	−0.13	114.71	Linear Regression*
		5	0.09	81.40	Negative Binomial Regression*
		7	−0.17	113.49	Zero-inflated Regression*
		9	−0.01	60.01	Negative Binomial Regression*
		11	0.01	100.13	Negative Binomial Regression*

4.1. Partitioning the area off SWIP: towards more biologically meaningful partitions

Overall, our unsupervised partitions of the whole SWIP area revealed an inner-outer shelf gradient, and differences between meridionally and zonally oriented coastal areas. These gradients were previously captured by both macroscale (Hoepffner and Dowell, 2005) and mesoscale (Krug et al., 2017b, 2018a, 2018b; Ferreira et al., 2019, 2021) unsupervised partitions covering the study area. At a macroscale level, several studies delineated two to four regions off the SWIP area, with coastal and oceanic regions separated by distinct boundaries (Sherman, 1994; Longhurst, 2007; Spalding et al., 2007, 2012; Fay and McKinley, 2014). Other studies at the mesoscale level (see Table 2), namely those focusing the Gulf of Cádiz, also identified marked differences between coastal and oceanic regions (Navarro and Ruiz, 2006; Muñoz et al., 2015).

Our biologically informed partitioning strategy identified five distinct regions for each HAB group: two inner shelf, two outer shelf-slope, and one transitional region between coastal and oceanic waters off the west coast (see Fig. 6). This clear separation between inner and outer shelf areas was only achieved by (Krug et al., 2017b) for the eastern sector of CSM. However, in contrast to previous studies (see Table 2), our partitions did not separate regions primarily influenced by coastal upwelling and those influenced by freshwater inputs. As previously hypothesized, different regions were identified in the western and southern SWIP coasts. The existence of two inner-shelf regions, on the west (region 4) and south (region 2) coasts, probably resulted from the occurrence of stronger upwelling activity on the west coast, responsible for higher nitrate concentrations and lower SST, in respect to the south coast (Krug et al., 2017b; Ferreira et al., 2021; Relvas et al., 2007). Interestingly, the extension of the western inner-shelf region around CSV, into the southwest coast, corroborated the influence of upwelling in the northwestern sector of Gulf of Cádiz. The eastward-advected upwelled water masses may either be retained in a cyclonic eddy near CSM or continue to flow eastward and join the Huelva Front (García-Lafuente et al., 2006; Sánchez and Relvas, 2003; Criado-Aldeanueva et al., 2009). The expansion of the western outer-shelf

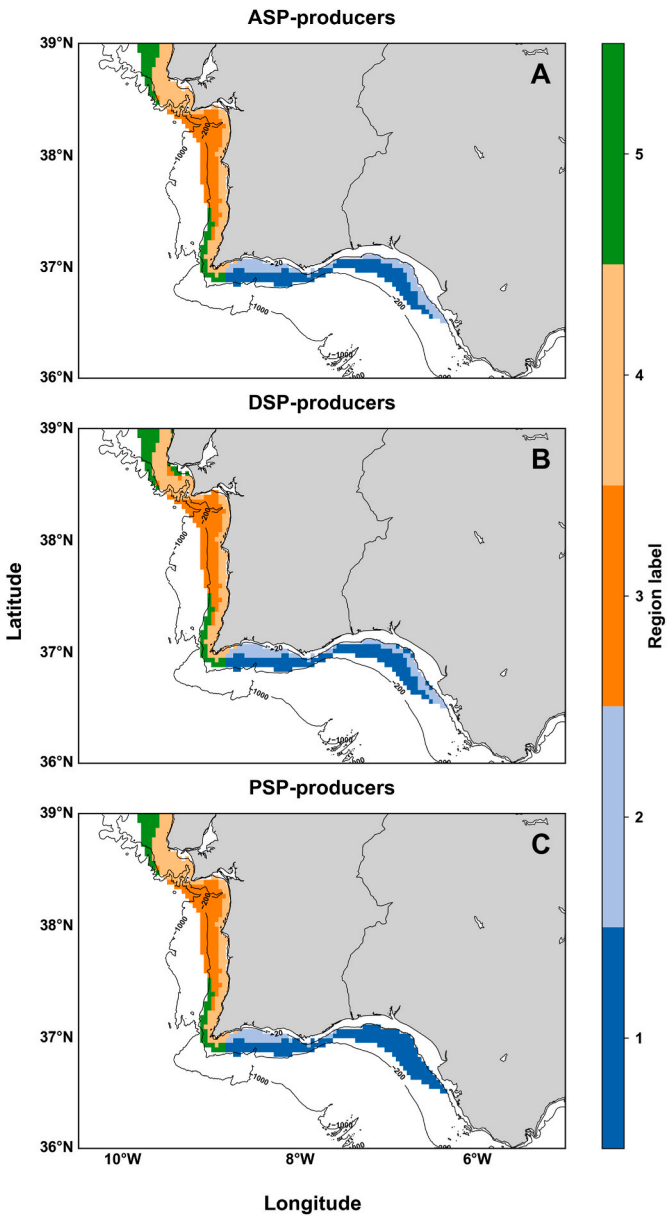

Fig. 6. Final partitions prioritizing environmental variables that best describe variability of the most frequently reported toxigenic phytoplankton taxa, over the coastal-slope domains off SW Iberia, for the period 2002-2021: (A) Amnesic Shellfish Poisoning (ASP)-producers, (B) Diarrhetic Shellfish Poisoning (DSP)-producers, and (C) Paralytic Shellfish Poisoning (PSP)-producers. Black lines represent 20 m, 200 m and 1000 m isobathymetric contours.

region up to the 1000 m isobath (region 3) is probably a reflection of intensified cross-shore processes associated with upwelling filaments, as reported in previous studies (Relvas et al., 2007; García Lafuente and Ruiz, 2007). The transitional outer shelf-ocean region on the west coast (region 5) probably indicated the influence of submarine topographic features, such as the Nazaré, Cascais, Lisboa/Setúbal and São Vicente canyons (see 1000 m isobathymetric contours, Fig. 6; Allin et al., 2016), that promote more oceanic conditions closer to the coast. The offshore extension of the southern outer-shelf region (region 1), limited mostly within the 200 m isobath, revealed the dominance of alongshore currents (García-Lafuente et al., 2006; Garel et al., 2016; Criado-Aldeanueva et al., 2009). Differences between inner and outer-shelf regions mostly over the southern coast, may also be explained by the influence of major river discharges (Guadiana and Guadalquivir), low-flow streams, lagoons, and small estuaries, as well as local submarine groundwater discharges (Hugman et al., 2015), and atmospheric wet deposition (Thompson et al., 2015; Zou et al., 2000). These nutrient sources may explain higher silicate concentration in the south coast, while the additional contribution from Saharan dust deposition (Escudero et al., 2005) may be responsible for the higher iron concentrations (see Tables S7–S9).

Although Chl-a alone was not identified as the best predictor dataset for any HAB group, our coastal-slope environmental regions showed good agreement with previous Chl-a or phenology-based partitioning studies (Krug et al., 2017b; Krug et al., 2018b; Ferreira et al., 2021; see Table 2 for a detailed comparison with previous studies), though expanding the detail to specific HAB groups. The identification of the mixed dataset, that includes nutrients, geostrophic currents and salinity in addition to satellite-derived variables, as the best predictor dataset for ASP-producers, probably reflected the specific requirements and functional traits of r-strategist opportunistic diatoms (Weithoff and Beisner, 2019; Marañón, 2015). The satellite-derived dataset was the best, yet with much lower predictive capacity, at predicting the abundance of both DSP- and PSP-producers, which include dinoflagellates that are less dependent on nutrient availability due to specific life traits, such as mixotrophy, motility and allelopathy (Smayda and Trainer, 2010; Glibert and Burford, 2017).

Nevertheless, the anticipated delineation of distinct coastal regions for each HAB group was not clearly detected. The similarity between the environmental partitions (regions) prioritizing ASP- and DSP-producers may represent the combined influence of coastal upwelling and riverine discharges. Yet, the best set of predictors, do not necessarily have the same influence (positive or negative) on HAB taxa, and can even show contrasting influences. For example, upwelling events or their environmental signatures (e.g., lower SST) can both have positive and negative effects on *Pseudo-nitzschia* blooms, due to nutrient enhancement and advective losses, respectively, for Iberian coastal waters (Torres Palenzuela et al., 2019; Moita, 2001; Silva et al., 2009; Palma et al., 2010; Vidal et al., 2017; Moita et al., 2016), and other coastal upwelling systems (see reviews by Trainer et al., 2010; Pitcher et al., 2017). River discharge or its proxies (salinity) have also been identified as predictors of *Pseudo-nitzschia* spp., showing overall negative effects, in the southern Iberian coastal waters (Lima et al., 2022), and other freshwater-influenced systems (Lane et al., 2009; Anderson et al., 2009, 2010, 2011; Liefer et al., 2009; Guallar et al., 2016). In turn, *Dinophysis* spp. Appear to be favoured by weak to moderate upwelling events, but also by upwelling relaxation and downwelling events, which promote the contact with their prey (*Mesodinium rubrum*) in northwestern Iberian shelf waters (Moita et al., 2006, 2016; Escalera et al., 2010; Díaz et al., 2016; Velasco-Senovilla et al., 2023), and southwest Portuguese coastal waters (Loureiro et al., 2005, 2011). Similarly to *Pseudo-nitzschia*, a negative relationship between *Dinophysis* and high river discharges is often described, likely reflecting increased advective losses (Moita et al., 2016). Yet, low to medium river discharges may also have a positive effect on *Dinophysis* sp., as reported for the south Portuguese coast (Lima et al., 2022) and other freshwater-influenced systems (e.g. Marzidovšek et al., 2024) associated with increased nutrient availability during periods of intensified haline-driven stratification (e.g. Moita et al., 2016). The configuration of the environmental partitions (regions) prioritizing PSP-producers suggested a reduced contribution from riverine discharges, particularly along the south coast, which contrasts with previous reports of positive relationships between river discharge and *Gymnodinium catenatum*, the main PSP-producing species in the study area (Lima et al., 2022). However, given the limited availability of non-zero observations for this HAB group, these results should be interpreted with caution.

4.2. Advantages and limitations of the novel partitioning strategy

Our three-step partitioning strategy has several advantages, but there are generic and study-specific limitations that need to be

Table 2
Overview of previous mesoscale and regional partitioning studies including the coastal-slope domain off SW Iberia, with information on the spatial and temporal coverage, partitioning variables, number of regions identified within our study area, similarities and differences relative to the regions identified in the present study, and associated reference, organized chronologically. SST: sea surface temperature; PAR: incident photosynthetically active radiation; Chl-a: chlorophyll-a concentration; MLD: mixed layer depth; Z_{eu}: euphotic depth; Sal: salinity; W^3: turbulent mixing index; U: zonal component of wind; V: meridional component of wind; PO_4: phosphate concentration; Fe: iron concentration; Lat: Latitude; Long: Longitude; CSV: Cape São Vicente; CSM: Cape Santa Maria; WC: west coast; SC: south coast; Gdn: Guadiana; Gdq: Guadalquivir; WSlp: western slope; SSlp: southern slope; NOff: north offshore; CoUp: coastal upwelling; CoBa: coastal river basins; CoMa: coastal continental margin; OcN: oceanic north; OcSW: oceanic southwest.

Spatial domain	Temporal coverage	Variables	N° regions covering our study area	Similarities	Differences	Reference
European Seas	1999-2000	SST, PAR, Chl-a	3 (coast-oceanic)	One region (class 8) detected near CSV and between Guadiana and Guadalquivir river mouths, and one transitional region between coastal and oceanic waters (class 4), mainly observed during spring.	No clear distinction between inner and outer shelf regions; dynamic season-specific regions.	Hoepffner and Dowell (2005)
Gulf of Cádiz	1998-2002	Chl-a	3 (coast)	One region (zone 2), similar to our region 4, under strong upwelling influence near CSV; one region (zone 3) partially covering our region 1, close to CSM; and one region (zone 4) that may partially cover our region 2, under greater influence of major river discharges.	Differentiation between upwelling and river-influenced areas; no clear distinction between inner and outer shelf regions; did not consider the west coast.	Navarro and Ruiz (2006)
Gulf of Cádiz & Alboran Sea	2005-2009	SST, Chl-a	3 (coast), 1 (slope)	One region (zone 1) partially covering our region 4, influenced by upwelling near CSV; a coastal region (zone 4) that coincides with our regions 1 and 2; and a region between the Guadalquivir river mouth and Cádiz (zone 5) mainly impacted by riverine discharges.	Differentiation between upwelling and river-influenced areas; additional slope region detected east of CSM (zone 3); no clear distinction between inner and outer shelf regions; did not consider the west coast.	Muñoz et al. (2015)
Southwest Iberia	1997-2012	Chl-a	4 (coast), 2 (slope), 1 (oceanic)	One west coast region (region WC), covering our regions 3 and 4, and one on the south coast (region SC), covering our regions 1 and 2, mainly influenced by upwelling associated with northerly and westerly winds, respectively. Two coastal regions near Guadiana and Guadalquivir river basins (Gdn, Gdq), partially overlapping the eastern sector of our region 2. Two western and southern slope regions (WSlp, SSlp), partially covering our regions 3 and 1, respectively; and one oceanic region (NOff) covering part of our transitional region 5.	Differentiation between upwelling and river-influenced areas, and between inner and outer-shelf areas on the eastern sector of CSM.	Krug et al. (2017b)
Southwest Iberia	2002-2011	MLD, SST, PAR, Z_{eu}, Sal, W^3, U, V, PO_4, Fe, Lat and Long	2 (coast), 2 (slope)	Two coastal regions, one confined to narrow areas on the west coast and near the Guadalquivir river mouth (EP3), and other extending the western and southern coasts (EP9), broadly covering all our regions.	Two slope regions over the western and southern coasts (EP11, EP6), with distinct offshore extensions; dynamic season-specific regions.	Krug et al. (2018a)
Southwest Iberia	1997-2015	Chl-a-derived phenological indices	2 (coast), 1 (slope)	A coastal region located on the west coast, extending eastwards from CSV (Upwelling-influenced), corresponding to our regions 3 and 4; and a slope region (Coastal-Slope), overlapping our regions 1 and 2.	One coastal region mostly located east of CSM next to major river basins (River-influenced), coastal wetlands and lagoons, and small estuarine systems.	Krug et al. (2018b)
Western Iberia	1998-2016	Chl-a	2 (coast), 1 (coast-slope), 1 (oceanic)	A coast-slope region (region C), which partially overlaps our regions 1 and 3; and a mostly oceanic region (region B), partially covering our regions 1 and 5.	Two coastal regions influenced by the Guadalquivir river discharge (region D), and smaller estuarine systems and rivers, likely also capturing the signal of coastal upwelling (region F). No clear distinction between inner and outer shelf regions.	Ferreira et al. (2019)
Western Iberia	1998-2018	Chl-a-derived phenological indices	2 (coast), 1 (slope), 2 (oceanic)	One coastal region concentrated near Sagres and scattered along the west coast, under stronger upwelling influence (CoUp), partially covering our region 4; one coastal region close to large rivers (Tagus, Guadiana and Guadalquivir), coastal lagoons and smaller river systems (CoBA), that partially overlaps our region 2; one slope region mostly in the west coast, transitioning from coastal to oceanic waters (CoMa), that might represent our region 5; and two regions (OcN and OcSW) that partially overlap our regions 1 and 2.	No clear distinction between inner and outer shelf regions.	Ferreira et al. (2021)

addressed. The region configurations, obtained in the first step, limited the application of the following steps to regions that overlapped with classified coastal production areas. The lack of *in situ* environmental data collected concurrently with HAB sampling, that created the need to use model-derived data, as well as the absence of other potential environmental drivers and/or predictors of HAB-groups, also limited the output of the first unsupervised classification. Thus, the relatively low spatial (4 km) and temporal (monthly) resolution of the environmental data, combined with issues associated with the HAB monitoring programs, likely explained the low (DSP- and PSP-producers) to moderate (ASP-producers) model predictive skills of even the best models. The combination of the abundance of different HAB taxa, into broad categories related with human syndromes, may also explain the limited model performance, since even species from the same genus can exhibit different niche preferences (e.g., *D. acuta* and *D. acuminata*; Moita et al., 2016; Escalera et al., 2010; Vale and Sampayo, 2003; Reguera et al., 2012; Díaz et al., 2019a; Batifoulier et al., 2013). In addition, the reduced number of sampling stations within each classified coastal production area, located nearshore (beach monitoring; see Anderson et al., 2009), and the high frequency of null abundances, namely for PSP group, limited model selection and precluded the use of independent training and test datasets (e.g. Bouckaert et al., 2004), which ultimately constrained model predictive skills. The simplified weight attribution approach used for the best-performing dataset in the third step of our partitioning strategy, combined with the low taxonomic, spatial and temporal resolutions, may have prevented the identification of distinct coastal regions for each group. Furthermore, extrapolating the best model results based on nearshore HAB data collected during an 8-year period, to the whole study period and the coastal regions overlapping with the classified coastal production areas, implicitly assumes that the statistical relationships between HAB abundances and environmental predictors remained consistent over time and space. These assumptions may not accurately reflect dynamic interactions between HAB groups and environmental conditions. However, the exploration of specific relationships between the abundance of HAB-forming groups and different predictors was beyond the scope of this study.

Despite these limitations, our partitioning approach allowed the delineation of biologically relevant regions, that partially captured the environmental variability associated with different toxigenic phytoplankton groups. While simplifying spatial heterogeneity, this strategy adds biological representativeness, providing a framework to design more cost-effective HAB sampling strategies that minimize oversampling in regions with similar environmental conditions, while reducing the risk of undersampling heterogeneous areas, and promoting a sustainable use of resources. Although this partitioning strategy is more attractive for areas with limited biological data, future studies would benefit from more complete environmental and biological datasets, with finer spatial and temporal coverage, increased taxonomic resolution, and inclusion of environmental drivers and/or predictors of target biological groups. For phytoplankton dedicated partitions, these should integrate both local (e.g., upwelling metrics, river discharges) and non-local predictors (e.g., offshore nutricline depth, climate indices; Kenitz et al., 2023). Specifically, for the DSP-producers *Dinophysis* spp., the additional inclusion of *M. rubrum* should be considered (Harred and Campbell, 2014; Díaz et al., 2019b). Higher-resolution datasets could enhance model performance and stability, and support independent model evaluation. The exploration of alternative weight attribution approaches in the third step of the proposed partitioning strategy could potentially support the identification of distinct partitions between groups, although this would require more comprehensive datasets. Furthermore, this approach can be transferred to other biological groups, namely using species occurrence data from global repositories (e.g., Ocean Biodiversity Information System, Global Biodiversity Information Facility), associated with distinct model types, which could support group-specific forecasting systems and ecosystem management in complex marine domains. Hyperspectral remote sensing missions, such as NASA's Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission, will provide extended, high-resolution phytoplankton functional type products that could be integrated into our partitioning strategy, offering a promising avenue to overcome the scarcity of *in situ* biological data (Bracher et al., 2026). All the abovementioned suggestions could refine the outputs of the three steps of our partitioning strategy, namely improving model predictive skills and the value of biologically informed environmental partitions as proxies for species abundance.

5. Conclusions

This study developed a novel three-step partitioning strategy of the coastal areas off SWIP, that prioritizes environmental factors that best describe variability patterns of different HAB taxa, at the pixel-level. The first unsupervised classification identified up to 12 regions over the whole study area, with 4 to 5 located in the coastal-slope domain. However, the configuration of these regions varied depending on the predictor datasets used, highlighting the need to select the set of variables that best represented environmental conditions specifically associated with different HAB groups. The best predictor datasets (satellite-derived and mixed datasets) and empirical-statistical models (simpler outperformed complex, machine-learning approaches), with low to moderate predictive skills, identified five spatially coherent regions, including two inner shelf, two outer shelf-slope and one transitional coastal-ocean region. Contrary to our hypothesis, no clear distinctive partitions were identified for different HAB groups, namely for ASP- and DSP-producers. Our biologically informed partitions apparently reflected the combined influence of upwelling patterns and riverine discharges, as well as the presence of submarine topographic features (canyons). This novel partitioning strategy simplifies the environmental heterogeneity, while adding specific biological significance to the partitions, thus expanding its value relative to previous abiotic-based partitioning exercises. Future studies would benefit from more complete datasets, for both environmental predictors and biological variables, and alternative weight attribution approaches (third step in the proposed strategy). Although our study focused on specific toxigenic phytoplankton groups off SWIP, this methodology is transferable to other marine systems and taxonomic groups, namely for areas with limited biological information, supporting effective ecosystem monitoring and management strategies, targeting specific organisms.

Funding statement

This work was supported by the FCT project UIDP/00350/2025 (CIMA; https://doi.org/10.54499/UID/00350/2025), LA/P/0069/2020 (ARNET; https://doi.org/10.54499/LA/P/0069/2020), and a PhD scholarship (2022.13748.BD; https://doi.org/10.54499/2022.13748.BD).

CRediT authorship contribution statement

M.J. Lima: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft. **I. Caballero:** Supervision, Writing – review & editing. **A.B. Barbosa:** Conceptualization, Formal analysis, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to express their gratitude to Instituto Português do

Mar e da Atmosfera (IPMA) and Consejería de Agricultura, Pesca, Agua y Desarrollo Rural de la Junta de Andalucía (abundance of toxigenic phytoplankton), European Space Agency (ESA)'s OC-CCI group (chlorophyll-a), NASA's Oceancolor database (sea surface temperature, and photosynthetically available radiation data), Copernicus Marine Service (salinity, mixed layer depth, currents, and nutrients data), NOAA's National Centres for Environmental Information (wind data), and Copernicus Climate Change Service (absolute dynamic topography) for providing free, regular, and high-quality products to the scientific community. The authors would also like to thank Dr Paulo Martel (University of Algarve) and Dr Roy El Hourany (University of the Littoral Côte d'Opale) for their insightful discussions and technical support in the implementation of self-organizing maps, and Prof. Wandrille Duchmin (Swiss Institute of Bioinformatics) for his expertise and support with model testing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marenvres.2026.108101.

Data availability

Data will be made available on request.

References

Allin, J.R., Hunt, J.E., Talling, P.J., Clare, M.A., Pope, E., Masson, D.G., 2016. Different frequencies and triggers of canyon filling and flushing events in Nazaré Canyon, offshore Portugal. Mar. Geol. 371, 89–105. https://doi.org/10.1016/j.margeo.2015.11.005.

Anderson, C.R., Kudela, R.M., Benitez-Nelson, C., Sekula-Wood, E., Burrell, C.T., Chao, Y., Langlois, G., Goodman, J., Siegel, D.A., 2011. Detecting toxic diatom blooms from ocean color and a regional ocean model. Geophys. Res. Lett. 38. https://doi.org/10.1029/2010GL045858.

Anderson, C.R., Sapiano, M.R.P., Prasad, M.B.K., Long, W., Tango, P.J., Brown, C.W., Murtugudde, R., 2010. Predicting potentially toxigenic *Pseudo-nitzschia* blooms in the Chesapeake Bay. J. Mar. Syst. 83, 127–140. https://doi.org/10.1016/j.jmarsys.2010.04.003.

Anderson, C.R., Siegel, D.A., Kudela, R.M., Brzezinski, M.A., 2009. Empirical models of toxigenic *Pseudo-nitzschia* blooms: potential use as a remote detection tool in the Santa Barbara Channel. Harmful Algae 8, 478–492. https://doi.org/10.1016/j.hal.2008.10.005.

Aquarone, M.C., Adams, S., Valdés, L., 2008. XIII-40 iberian coastal LME. In: Sherman, K., Hempel, G. (Eds.), The UNEP Large Marine Ecosystem Report: a Perspective on Changing Conditions in Lmes of the World's Regional Seas. United Nations Environment Programme, Nairobi, Kenya, pp. 553–561.

Arena, M., Pratolongo, P., Loisel, H., Tran, M.D., Jorge, D.S.F., Delgado, A.L., 2025. Full-resolution Sentinel-3 satellite observations of phytoplankton phenology in optically complex waters of the Northern patagonian shelf. Estuaries Coasts 48, 138. https://doi.org/10.1007/s12237-025-01572-7.

Ayyam, V., Palanivel, S., Chandrakasan, S., 2019. Coastal ecosystems and services. In: Ayyam, V., Palanivel, S., Chandrakasan, S. (Eds.), Coastal Ecosystems of the Tropics - Adaptive Management. Springer, Singapore, pp. 21–47. https://doi.org/10.1007/978-981-13-8926-9_2.

Batifoulier, F., Lazure, P., Velo-Suarez, L., Maurer, D., Bonneton, P., Charria, G., Dupuy, C., Gentien, P., 2013. Distribution of *Dinophysis* species in the Bay of Biscay and possible transport pathways to Arcachon Bay. J. Mar. Syst. 109–110. https://doi.org/10.1016/j.jmarsys.2011.12.007. S273–S283.

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.. In: https://proceedings.neurips.cc/paper_files/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html (accessed January 21, 2026).

Blum, I., Rao, D.S., Pan, Y., Swaminathan, S., Adams, N., 2006. Development of statistical models for prediction of the neurotoxin domoic acid levels in the pennate diatom *Pseudo-nitzschia multiseries* utilizing data from cultures and natural blooms. In: Algal Cultures, Analogues of Blooms and Applications. Science Publishers Inc., Enfield, New Hampshire, USA, pp. 891–916.

Bock, N., Cornec, M., Claustre, H., Duhamel, S., 2022. Biogeographical classification of the global ocean from BGC-argo floats. Glob. Biogeochem. Cycles 36. https://doi.org/10.1029/2021GB007233 e2021GB007233.

Bolaños-Durán, E., Muñiz-Castillo, A.I., Valenzuela-Quiñonez, F., Paz-García, D.A., 2025. Classifying the seascape of eastern tropical pacific based on physicochemical variables. Mar. Environ. Res. 210, 107294. https://doi.org/10.1016/j.marenvres.2025.107294.

Bouckaert, R.R., Frank, E., 2004. Evaluating the replicability of significance tests for comparing learning algorithms. In: Dai, H., Srikant, R., Zhang, C. (Eds.), Advances in Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, pp. 3–12. https://doi.org/10.1007/978-3-540-24775-3_3.

Boyd, P.W., Cornwall, C.E., Davison, A., Doney, S.C., Fourquez, M., Hurd, C.L., Lima, I.D., McMinn, A., 2016. Biological responses to environmental heterogeneity under future ocean conditions. Glob. Change Biol. 22, 2633–2650. https://doi.org/10.1111/gcb.13287.

Bracher, A., Dogliotti, A., Werdell, J., 2026. A Scientific Roadmap of Aquatic Hyperspectral Remote Sensing: Overview of Status, Challenges and Future Perspectives. International Ocean Colour Coordinating Group, Dartmouth, Canada.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Caballero, I., Morris, E., Prieto, L., Navarro, G., 2014. The influence of the Guadalquivir river on spatio-temporal variability in the pelagic ecosystem of the eastern Gulf of Cádiz. Mediterr. Mar. Sci. 15, 721–738. https://doi.org/10.12681/mms.844.

Cabral, H.N., Fonseca, V.F., Gamito, R., Gonçalves, C.I., Costa, J.L., Erzini, K., Gonçalves, J., Martins, J., Leite, L., Andrade, J.P., Ramos, S., Bordalo, A., Amorim, E., Neto, J.M., Marques, J.C., Rebelo, J.E., Silva, C., Castro, N., Almeida, P.R., Domingos, I., Gordo, L.S., Costa, M.J., 2012. Ecological quality assessment of transitional waters based on fish assemblages in Portuguese estuaries: the estuarine fish assessment index (EFAI). Ecol. Indic. 19, 144–153. https://doi.org/10.1016/j.ecolind.2011.08.005.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.

Campbell, J.W., 1995. The lognormal distribution as a model for bio-optical variability in the sea. J. Geophys. Res. Oceans 100, 13237–13254. https://doi.org/10.1029/95JC00458.

Cardoso, I., Cancela da Fonseca, L., Cabral, H.N., 2012. Ecological quality assessment of small estuaries from the Portuguese coast based on benthic macroinvertebrate assemblages indices. Mar. Pollut. Bull. 64, 1136–1142. https://doi.org/10.1016/j.marpolbul.2012.03.030.

Carr, M.-E., Kearns, E.J., 2003. Production regimes in four eastern boundary current systems. Deep Sea Res. Part II Top. Stud. Oceanogr. 50, 3199–3221. https://doi.org/10.1016/j.dsr2.2003.07.015.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. STL: a seasonal-trend decomposition procedure based on loess. J. Off. Stat. 6, 3–73.

Ocean colour daily data from 1997 to present derived from satellite observations. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019. https://doi.org/10.24381/cds.f85b319d. https://cds.climate.copernicus.eu/datasets/satellite-ocean-colour?tab=overview. (Accessed 30 April 2023).

Sea level gridded data from satellite observations for the global ocean from 1993 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.4c328c78 (accessed on 25 Mar 2023).

Criado-Aldeanueva, F., García-Lafuente, J., Navarro, G., Ruiz, J., 2009. Seasonal and interannual variability of the surface circulation in the eastern Gulf of Cadiz (SW Iberia). J. Geophys. Res. Oceans 114. https://doi.org/10.1029/2008JC005069.

Criado-Aldeanueva, F., García-Lafuente, J., Vargas, J.M., Del Río, J., Vázquez, A., Reul, A., Sánchez, A., 2006. Distribution and circulation of water masses in the Gulf of Cadiz from in situ observations. Deep Sea Res. Part II Top. Stud. Oceanogr. 53, 1144–1160. https://doi.org/10.1016/j.dsr2.2006.04.012.

Cristina, S., Cordeiro, C., Lavender, S., Costa Goela, P., Icely, J., Newton, A., 2016. MERIS phytoplankton time series products from the SW iberian peninsula (Sagres) using seasonal-trend decomposition based on loess. Remote Sens. 8, 449. https://doi.org/10.3390/rs8060449.

Cruz, R.C., Reis Costa, P., Vinga, S., Krippahl, L., Lopes, M.B., 2021. A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination. J. Mar. Sci. Eng. 9, 283. https://doi.org/10.3390/jmse9030283.

Cusack, C., Mouriño, H., Moita, M.T., Silke, J., 2015. Modelling *Pseudo-nitzschia* events off southwest Ireland. J. Sea Res. 105, 30–41. https://doi.org/10.1016/j.seares.2015.06.012.

D'Ortenzio, F., Ribera d'Alcalà, M., 2009. On the trophic regimes of the Mediterranean Sea: a satellite analysis. Biogeosciences 6, 139–148. https://doi.org/10.5194/bg-6-139-2009.

Díaz, P.A., Gormaz, R., Aguayo, P., Pérez-Santos, I., Saldías, G.S., Figueroa, R.I., Fernández, P.A., Álvarez, G., Rodríguez-Villegas, C., Schwerter, C., Cassis, D., Vera, R., Conca, C., 2025. Unveiling the 2017 *Karenia* bloom in NW Chilean Patagonia by integrating remote sensing and field data. Microorganisms 13, 2440. https://doi.org/10.3390/microorganisms13112440.

Díaz, P.A., Molinet, C., Seguel, M., Niklitschek, E.J., Díaz, M., Álvarez, G., Pérez-Santos, I., Varela, D., Guzmán, L., Rodríguez-Villegas, C., Figueroa, R.I., 2022. Modelling the spatial and temporal dynamics of paralytic shellfish toxins (PST) at different scales: implications for research and management. Toxins 14, 786. https://doi.org/10.3390/toxins14110786.

Díaz, P.A., Reguera, B., Moita, T., Bravo, I., Ruiz-Villarreal, M., Fraga, S., 2019a. Mesoscale dynamics and niche segregation of two *dinophysis* species in Galician-Portuguese coastal waters. Toxins 11, 37. https://doi.org/10.3390/toxins11010037.

Díaz, P.A., Ruiz-Villarreal, M., Mouriño-Carballido, B., Fernández-Pena, C., Riobó, P., Reguera, B., 2019b. Fine scale physical-biological interactions during a shift from relaxation to upwelling with a focus on *Dinophysis acuminata* and its potential ciliate prey. Prog. Oceanogr. 175, 309–327. https://doi.org/10.1016/j.pocean.2019.04.009.

Díaz, P.A., Ruiz-Villarreal, M., Pazos, Y., Moita, T., Reguera, B., 2016. Climate variability and *Dinophysis acuta* blooms in an upwelling system. Harmful Algae 53, 145–159. https://doi.org/10.1016/j.hal.2015.11.007.

de Boyer Montégut, C., Madec, G., Fischer, A.S., Lazar, A., Iudicone, D., 2004. Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology. J. Geophys. Res. Oceans 109. https://doi.org/10.1029/2004JC002378.

Demuth, H.B., Beale, M.H., De Jess, O., Hagan, M.T., 2014. Neural Network Design, second ed. Martin Hagan, Stillwater, OK, USA.
Derot, J., Yajima, H., Jacquet, S., 2020. Advances in forecasting harmful algal blooms using machine learning models: a case study with *Planktothrix rubescens* in Lake Geneva. Harmful Algae 99, 101906. https://doi.org/10.1016/j.hal.2020.101906.
Detoni, A.M.S., Navarro, G., Padín, X.A., Ramirez-Romero, E., Zoffoli, M.L., Pazos, Y., Caballero, I., 2024. Potentially toxigenic phytoplankton patterns in the northwestern Iberian Peninsula. Front. Mar. Sci. 11. https://doi.org/10.3389/fmars.2024.1330090.
Dunn, O.J., 1961. Multiple comparisons among means. J. Am. Stat. Assoc. 56, 52–64. https://doi.org/10.1080/01621459.1961.10482090.
Dunn, P.K., Smyth, G.K., 1996. Randomized quantile residuals. J. Comput. Graph Stat. 5, 236–244. https://doi.org/10.1080/10618600.1996.10474708.
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. J. Geophys. Res. Oceans 124, 5827–5843. https://doi.org/10.1029/2019JC015131.
El Hourany, R., Mejia, C., Faour, G., Crépon, M., Thiria, S., 2021. Evidencing the impact of climate change on the phytoplankton community of the Mediterranean Sea through a bioregionalization approach. JGR Oceans 126. https://doi.org/10.1029/2020JC016808 e2020JC016808.
Escalera, L., Reguera, B., Moita, T., Pazos, Y., Cerejo, M., Cabanas, J.M., Ruiz-Villarreal, M., 2010. Bloom dynamics of *Dinophysis acuta* in an upwelling system: *in situ* growth *versus* transport. Harmful Algae 9, 312–322. https://doi.org/10.1016/j.hal.2009.12.002.
Escudero, M., Castillo, S., Querol, X., Avila, A., Alarcón, M., Viana, M.M., Alastuey, A., Cuevas, E., Rodríguez, S., 2005. Wet and dry African dust episodes over eastern Spain. J. Geophys. Res. Atmos. 110. https://doi.org/10.1029/2004JD004731.
Atlantic-Iberian Biscay Irish- Ocean Physics Reanalysis, 2023a. E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). https://doi.org/10.48670/moi-00029 (accessed on 17 Jun 2023).
Global Total (COPERNICUS-GLOBCURRENT), 2023b. Ekman and Geostrophic currents at the Surface and 15m. E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). https://doi.org/10.48670/mds-00327 (accessed on 4 Apr 2023).
Atlantic-Iberian Biscay Irish- Ocean BioGeoChemistry NON ASSIMILATIVE Hindcast, 2023c. E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). https://doi.org/10.48670/moi-00028 (accessed on 10 Apr 2023).
Fay, A.R., McKinley, G.A., 2014. Global open-ocean biomes: mean and temporal variability. Earth Syst. Sci. Data 6, 273–284. https://doi.org/10.5194/essd-6-273-2014.
Fendereski, F., Vogt, M., Payne, M.R., Lachkar, Z., Gruber, N., Salmanmahiny, A., Hosseini, S.A., 2014. Biogeographic classification of the Caspian Sea. Biogeosciences 11, 6451–6470. https://doi.org/10.5194/bg-11-6451-2014.
Ferreira, A., Brotas, V., Palma, C., Borges, C., Brito, A.C., 2021. Assessing phytoplankton bloom phenology in upwelling-influenced regions using ocean color remote sensing. Remote Sens. 13, 675. https://doi.org/10.3390/rs13040675.
Ferreira, A., Garrido-Amador, P., Brito, A.C., 2019. Disentangling environmental drivers of phytoplankton biomass off Western iberia. Front. Mar. Sci. 6. https://doi.org/10.3389/fmars.2019.00044.
Fiúza, A.F. de G., de Macedo, M.E., Guerreiro, M.R., 1982. Climatological space and time variation of the Portuguese coastal upwelling. Oceanol. Acta 5, 31–40.
García Lafuente, J., Ruiz, J., 2007. The Gulf of Cádiz pelagic ecosystem: a review. Prog. Oceanogr. 74, 228–251. https://doi.org/10.1016/j.pocean.2007.04.001.
García-Lafuente, J., Delgado, J., Criado-Aldeanueva, F., Bruno, M., del Río, J., Miguel Vargas, J., 2006. Water mass circulation on the continental shelf of the Gulf of Cádiz. Deep Sea Res. Part II Top. Stud. Oceanogr. 53, 1182–1197. https://doi.org/10.1016/j.dsr2.2006.04.011.
Garel, E., Laiz, I., Drago, T., Relvas, P., 2016. Characterisation of coastal counter-currents on the inner shelf of the Gulf of Cadiz. J. Mar. Syst. 155, 19–34. https://doi.org/10.1016/j.jmarsys.2015.11.001.
Glibert, P.M., Burford, M.A., 2017. Globally changing nutrient loads and harmful algal blooms: recent advances, new paradigms, and continuing challenges. Oceanography (Wash. D. C.) 30, 58–69. https://doi.org/10.5670/oceanog.2017.110.
Gobler, C.J., Sunda, W.G., 2012. Ecosystem disruptive algal blooms of the brown tide species, *Aureococcus anophagefferens* and *Aureoumbra lagunensis*. Harmful Algae 14, 36–45. https://doi.org/10.1016/j.hal.2011.10.013.
Gregr, E.J., Bodtker, K.M., 2007. Adaptive classification of marine ecosystems: identifying biologically meaningful regions in the marine environment. Deep Sea Res. Oceanogr. Res. Pap. 54, 385–402. https://doi.org/10.1016/j.dsr.2006.11.004.
Guallar, C., Delgado, M., Diogène, J., Fernández-Tejedor, M., 2016. Artificial neural network approach to population dynamics of harmful algal blooms in alfacs Bay (NW Mediterranean): case studies of *Karlodinium* and *Pseudo-nitzschia*. Ecol. Model. 338, 37–50. https://doi.org/10.1016/j.ecolmodel.2016.07.009.
Harred, L.B., Campbell, L., 2014. Predicting harmful algal blooms: a case study with *Dinophysis ovum* in the Gulf of Mexico. J. Plankton Res. 36, 1434–1445. https://doi.org/10.1093/plankt/fbu070.
Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J., 2020. Statistical Methods in Water Resources. U.S. Geological Survey Techniques and Methods. https://doi.org/10.3133/tm4A3.
Hilbe, J.M., 2011. Negative Binomial Regression, second ed. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511973420.
Hoepffner, N., Dowell, M., 2005. Assessing the Dynamics of Ecological Provinces in the European Seas. Joint Research Centre (JRC), European Commission. https://publications.jrc.ec.europa.eu/repository/handle/JRC30857. (Accessed 21 January 2026).
Hofmann Elizondo, U., Righetti, D., Benedetti, F., Vogt, M., 2021. Biome partitioning of the global ocean based on phytoplankton biogeography. Prog. Oceanogr. 194, 102530. https://doi.org/10.1016/j.pocean.2021.102530.
Hugman, R., Stigter, T.Y., Monteiro, J.P., Costa, L., Nunes, L.M., 2015. Modeling the spatial and temporal distribution of coastal groundwater discharge for different water use scenarios under epistemic uncertainty: case study in South Portugal. Environ. Earth Sci. 73, 2657–2669. https://doi.org/10.1007/s12665-014-3709-4.
Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice, second ed. OTexts https://otexts.org/fpp2/. (Accessed 15 January 2026).
IPMA, 2013. Plano de ação Sistema Nacional de Monitorização de Moluscos Bivalves. https://www.ipma.pt/bin/docs/institucionais/p.accao_snmb_2013.pdf.
Karna, A., Gibert, K., 2022. Automatic identification of the number of clusters in hierarchical clustering. Neural Comput. Appl. 34, 119–134. https://doi.org/10.1007/s00521-021-05873-3.
Kenitz, K.M., Anderson, C.R., Carter, M.L., Eggleston, E., Seech, K., Shipe, R., Smith, J., Orenstein, E.C., Franks, P.J.S., Jaffe, J.S., Barton, A.D., 2023. Environmental and ecological drivers of harmful algal blooms revealed by automated underwater microscopy. Limnol. Oceanogr. 68, 598–615. https://doi.org/10.1002/lno.12297.
Kheireddine, M., Mayot, N., Ouhssain, M., Jones, B.H., 2021. Regionalization of the Red Sea based on phytoplankton phenology: a satellite analysis. J. Geophys. Res. Oceans 126. https://doi.org/10.1029/2021JC017486 e2021JC017486.
Kohonen, T., 1990. The self-organizing map. Proc. IEEE 78, 1464–1480. https://doi.org/10.1109/5.58325.
Kohonen, T., 2001. Self-Organizing Maps, third ed. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-56927-2.
Kohonen, T., 2013. Essentials of the self-organizing map. Neural Netw. 37, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018.
Konik, M., Angelica Peña, M., Hirawake, T., Hunt, B.P.V., Suseelan Vishnu, P., Eisner, L. B., Bracher, A., Xi, H., Marchese, C., Costa, M., 2024. Bioregionalization of the subarctic Pacific based on phytoplankton phenology and composition. Prog. Oceanogr. 228, 103315. https://doi.org/10.1016/j.pocean.2024.103315.
Krug, L.A., Platt, T., Barbosa, A.B., 2018a. Delineation of ocean surface provinces over a complex marine domain (off SW Iberia): an objective abiotic-based approach. Reg. Stud. Mar. Sci. 18, 80–96. https://doi.org/10.1016/j.rsma.2018.01.003.
Krug, L.A., Platt, T., Sathyendranath, S., Barbosa, A.B., 2017a. Ocean surface partitioning strategies using ocean colour remote sensing: a review. Prog. Oceanogr. 155, 41–53. https://doi.org/10.1016/j.pocean.2017.05.013.
Krug, L.A., Platt, T., Sathyendranath, S., Barbosa, A.B., 2017b. Unravelling region-specific environmental drivers of phytoplankton across a complex marine domain (off SW Iberia). Rem. Sens. Environ. 203, 162–184. https://doi.org/10.1016/j.rse.2017.05.029.
Krug, L.A., Platt, T., Sathyendranath, S., Barbosa, A.B., 2018b. Patterns and drivers of phytoplankton phenology off SW Iberia: a phenoregion based perspective. Prog. Oceanogr. 165, 233–256. https://doi.org/10.1016/j.pocean.2018.06.010.
Kwok, C.-F., Qian, G., Kuleshov, Y., 2023. Analyzing error bounds for seasonal-trend decomposition of Antarctica temperature time series involving missing data. Atmosphere 14, 193. https://doi.org/10.3390/atmos14020193.
Lambert, D., 1992. Zero-Inflated poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14. https://doi.org/10.1080/00401706.1992.10485228.
Lane, J.Q., Raimondi, P.T., Kudela, R.M., 2009. Development of a logistic regression model for the prediction of toxigenic *Pseudo-nitzschia* blooms in monterey Bay, California. Mar. Ecol. Prog. Ser. 383, 37–51. https://doi.org/10.3354/meps07999.
Lebedev, K.V., Yoshinari, H., Maximenko, N.A., Hacker, P.W., 2007. YoMaHa'07: velocity data assessed from trajectories of Argo floats at parking level and at the sea surface. International Pacific Research Center.
Leloup, J.A., Lachkar, Z., Boulanger, J.-P., Thiria, S., 2007. Detecting decadal changes in ENSO using neural networks. Clim. Dyn. 28, 147–162. https://doi.org/10.1007/s00382-006-0173-1.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R. News 2, 18–22.
Liefer, J.D., MacIntyre, H.L., Novoveská, L., Smith, W.L., Dorsey, C.P., 2009. Temporal and spatial variability in *Pseudo-nitzschia* spp. in Alabama coastal waters: a "hot spot" linked to submarine groundwater discharge? Harmful Algae 8, 706–714. https://doi.org/10.1016/j.hal.2009.02.003.
Lima, M.J., Relvas, P., Barbosa, A.B., 2022. Variability patterns and phenology of harmful phytoplankton blooms off southern Portugal: looking for region-specific environmental drivers and predictors. Harmful Algae 116, 102254. https://doi.org/10.1016/j.hal.2022.102254.
Liu, Y., Weisberg, R.H., Mooers, C.N.K., 2006. Performance evaluation of the self-organizing map for feature extraction. J. Geophys. Res. Oceans 111. https://doi.org/10.1029/2005JC003117.
Lloyd-Smith, J.O., 2007. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PLoS One 2, e180. https://doi.org/10.1371/journal.pone.0000180.
Longhurst, A.R., 2007. Ecological Geography of the Sea. Elsevier. https://doi.org/10.1016/B978-0-12-455521-1.X5000-1.
Loureiro, S., Newton, A., Icely, J.D., 2005. Microplankton composition, production and upwelling dynamics in Sagres (SW Portugal) during summer of 2001. Sci. Mar. 69, 323–341. https://doi.org/10.3989/scimar.2005.69n3323.
Loureiro, S., Reñé, A., Garcés, E., Camp, J., Vaqué, D., 2011. Harmful algal blooms (HABs), dissolved organic matter (DOM), and planktonic microbial community dynamics at a near-shore and a harbour station influenced by upwelling (SW Iberian Peninsula). J. Sea Res. 65, 401–413. https://doi.org/10.1016/j.seares.2011.03.004.

Marañón, E., 2015. Cell size as a key determinant of Phytoplankton metabolism and community structure. Ann. Rev. Mar. Sci 7, 241–264. https://doi.org/10.1146/annurev-marine-010814-015955.
Marchese, C., Hunt, B.P.V., Giannini, F., Ehrler, M., Costa, M., 2022. Bioregionalization of the coastal and open oceans of British Columbia and Southeast Alaska based on Sentinel-3A satellite-derived phytoplankton seasonality. Front. Mar. Sci. 9. https://doi.org/10.3389/fmars.2022.968470.
Marzidovšek, M., Francé, J., Podpečan, V., Vadnjal, S., Dolenc, J., Mozetič, P., 2024. Explainable machine learning for predicting diarrhetic shellfish poisoning events in the Adriatic Sea using long-term monitoring data. Harmful Algae 139, 102728. https://doi.org/10.1016/j.hal.2024.102728.
Mayot, N., D'Ortenzio, F., Ribera d'Alcalà, M., Lavigne, H., Claustre, H., 2016. Interannual variability of the Mediterranean trophic regimes from ocean color satellites. Biogeosciences 13, 1901–1917. https://doi.org/10.5194/bg-13-1901-2016.
McGillicuddy, D.J., Anderson, D.M., Lynch, D.R., Townsend, D.W., 2005. Mechanisms regulating large-scale seasonal fluctuations in *Alexandrium fundyense* populations in the Gulf of Maine: results from a physical–biological model. Deep Sea Res. Part II Top. Stud. Oceanogr. 52, 2698–2714. https://doi.org/10.1016/j.dsr2.2005.06.021.
McGillicuddy, D.J. Jr., Townsend, D.W., He, R., Keafer, B.A., Kleindinst, J.L., Li, Y., Manning, J.P., Mountain, D.G., Thomas, M.A., Anderson, D.M., 2011. Suppression of the 2010 *Alexandrium fundyense* bloom by changes in physical, biological, and chemical properties of the Gulf of Maine. Limnol. Oceanogr. 56, 2411–2426. https://doi.org/10.4319/lo.2011.56.6.2411.
Mermer, O., Zhang, E., Demir, I., 2025. A comparative study of ensemble machine learning and explainable AI for predicting harmful algal blooms. Big Data and Cognitive Computing 9, 138. https://doi.org/10.3390/bdcc9050138.
Moita, M.T., 2001. Estrutura, Variabilidade E Dinâmica do Fitoplâncton Na Costa De Portugal Continental [Structure, Variability and Dynamics of Phytoplankton off the Coast of Continental Portugal]. University of Lisbon. PhD dissertation. https://www.ipma.pt/resources.www/docs/publicacoes.site/teses/Teresa_Moitathesis.pdf. (Accessed 30 June 2025).
Moita, M., Sobrinho-Gonçalves, L., Oliveira, P., Palma, S., Falcão, M., 2006. A bloom of *Dinophysis acuta* in a thin layer off North-West Portugal. Afr. J. Mar. Sci. 28, 265–269. https://doi.org/10.2989/18142320609504160.
Moita, M.T., Pazos, Y., Rocha, C., Nolasco, R., Oliveira, P.B., 2016. Toward predicting *Dinophysis blooms* off NW Iberia: a decade of events. Harmful Algae 53, 17–32. https://doi.org/10.1016/j.hal.2015.12.002.
Muñoz, M., Reul, A., Plaza, F., Gómez-Moreno, M.-L., Vargas-Yañez, M., Rodríguez, V., Rodríguez, J., 2015. Implication of regionalization and connectivity analysis for marine spatial planning and coastal management in the Gulf of Cadiz and Alboran Sea. Ocean Coast Manag. 118, 60–74. https://doi.org/10.1016/j.ocecoaman.2015.04.011.
Navarro, G., Ruiz, J., 2006. Spatial and temporal variability of phytoplankton in the Gulf of Cádiz through remote sensing images. Deep Sea Res. Part II Top. Stud. Oceanogr. 53, 1241–1260. https://doi.org/10.1016/j.dsr2.2006.04.014.
NOAA, 2020. NCEI Blended Seawinds Product (NBS). NOAA National Centers for Environmental Information (NCEI). https://doi.org/10.25921/mxt4-b075.
O'Reilly, J.E., Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. Rem. Sens. Environ. 229, 32–47. https://doi.org/10.1016/j.rse.2019.04.021.
Oliver, M.J., Glenn, S., Kohut, J.T., Irwin, A.J., Schofield, O.M., Moline, M.A., Bissett, W.P., 2004. Bioinformatic approaches for objective detection of water masses on continental shelves. J. Geophys. Res. Oceans 109. https://doi.org/10.1029/2003JC002072.
Oliver, M.J., Irwin, A.J., 2008. Objective global ocean biogeographic provinces. Geophys. Res. Lett. 35. https://doi.org/10.1029/2008GL034238.
Ottaviani, E., Magri, S., Gaetano, P.D., Cuneo, C., Chiantore, M., Asnaghi, V., Pedroncini, A., 2020. A machine learning based approach to predict *ostreopsis* cf. *ovata* bloom events from Meteo- marine forecasts. Chem. Eng. Trans. 82, 409–414. https://doi.org/10.3303/CET2082069.
Palma, S., Mouriño, H., Silva, A., Barão, M.I., Moita, M.T., 2010. Can *Pseudo-nitzschia* blooms be modeled by coastal upwelling in Lisbon Bay? Harmful Algae 9, 294–303. https://doi.org/10.1016/j.hal.2009.11.006.
Park, J., Patel, K., Lee, W.H., 2024. Recent advances in algal bloom detection and prediction technology using machine learning. Sci. Total Environ. 938, 173546. https://doi.org/10.1016/j.scitotenv.2024.173546.
Peliz, A., Dubert, J., Marchesiello, P., Teles-Machado, A., 2007. Surface circulation in the Gulf of Cadiz: model and mean flow structure. J. Geophys. Res. Oceans 112. https://doi.org/10.1029/2007JC004159.
Peliz, Á., Santos, A.M.P., Oliveira, P.B., Dubert, J., 2004. Extreme cross-shelf transport induced by eddy interactions southwest of Iberia in winter 2001. Geophys. Res. Lett. 31. https://doi.org/10.1029/2004GL019618.
Peliz, Á., Teles-Machado, A., Marchesiello, P., Dubert, J., Lafuente, J.G., 2009. Filament generation off the strait of Gibraltar in response to gap winds. Dynam. Atmos. Oceans 46, 36–45. https://doi.org/10.1016/j.dynatmoce.2008.08.002.
Piló, D., Barbosa, A.B., Teodósio, M.A., Encarnação, J., Leitão, F., Range, P., Krug, L.A., Cruz, J., Chícharo, L., 2018. Are submarine groundwater discharges affecting the structure and physiological status of rocky intertidal communities? Mar. Environ. Res. 136, 158–173. https://doi.org/10.1016/j.marenvres.2018.02.013.
Pitcher, G.C., Jiménez, A.B., Kudela, R.M., Reguera, B., 2017. Harmful algal blooms in eastern boundary upwelling systems: a GEOHAB core research project. Oceanography (Wash. D. C.) 30, 22–35. https://doi.org/10.5670/oceanog.2017.107.
Pramlall, S., Jackson, J.M., Marchese, C., Suchy, K.D., Hunt, B.P.V., Costa, M., 2024. Mapping phenoregions and phytoplankton seasonality in Northeast Pacific marine coastal ecosystems via a satellite-based approach. Prog. Oceanogr. 228, 103336. https://doi.org/10.1016/j.pocean.2024.103336.
Prieto, L., Navarro, G., Rodríguez-Gálvez, S., Huertas, I.E., Naranjo, J.M., Ruiz, J., 2009. Oceanographic and meteorological forcing of the pelagic ecosystem on the Gulf of Cadiz shelf (SW Iberian Peninsula). Cont. Shelf Res. 29, 2122–2137. https://doi.org/10.1016/j.csr.2009.08.007.
Reguera, B., Velo-Suárez, L., Raine, R., Park, M.G., 2012. Harmful *Dinophysis* species: a review. Harmful Algae 14, 87–106. https://doi.org/10.1016/j.hal.2011.10.016.
Relvas, P., Barton, E.D., Dubert, J., Oliveira, P.B., Peliz, Á., da Silva, J.C.B., Santos, A.M.P., 2007. Physical oceanography of the western Iberia ecosystem: latest views and challenges. Prog. Oceanogr. 74, 149–173. https://doi.org/10.1016/j.pocean.2007.04.021.
Relvas, P., Barton, E.D., 2002. Mesoscale patterns in the Cape São Vicente (Iberian Peninsula) upwelling region. J. Geophys. Res. Oceans 107. https://doi.org/10.1029/2000JC000456, 28-1-28–23.
Reygondeau, G., Dunn, D.C., 2019. Pelagic biogeography. In: Encyclopedia of Ocean Sciences, Third. Academic Press, pp. 588–598. https://doi.org/10.1016/B978-0-12-409548-9.11633-1.
Robinson, I.S., 2010. Discovering the Ocean from Space: the Unique Applications of Satellite Oceanography. Springer.
Ruiz-Villarreal, M., García-García, L.M., Cobas, M., Díaz, P.A., Reguera, B., 2016. Modelling the hydrodynamic conditions associated with *Dinophysis* blooms in Galicia (NW Spain). Harmful Algae 53, 40–52. https://doi.org/10.1016/j.hal.2015.12.003.
Sarma, N.S., Baliarsingh, S.K., Lotliker, A.A., Pandi, S.R., Samanta, A., Srichandan, S., 2023. Sea surface temperature and phytoplankton abundance as crucial proxies for green noctiluca bloom monitoring in the Northeastern Arabian Sea: a case study. Ocean Sci. J. 58, 2. https://doi.org/10.1007/s12601-022-00096-6.
Sathyendranath, S., Jackson, T., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Clements, O., Cipollini, P., Danne, O., Dingle, J., Donlon, C., Grant, M., Groom, S., Krasemann, H., Lavender, S., Mazeran, C., Mélin, F., Müller, D., Steinmetz, F., Valente, A., Zühlke, M., Feldman, G., Franz, B., Frouin, R., Werdell, J., Platt, T., 2023. ESA ocean colour climate change initiative (Ocean_Colour_cci): monthly climatology of global ocean colour data products at 4 km resolution. https://catalogue.ceda.ac.uk/uuid/690fdf8f229c4d04a2aa68de67beb733/. (Accessed 25 April 2023).
Sherman, K., 1991. The large marine ecosystem concept: research and management strategy for living marine resources. Ecol. Appl. 1, 349–360. https://doi.org/10.2307/1941896.
Sherman, K., 1994. Sustainability, biomass yields, and health of coastal ecosystems: an ecological perspective. Mar. Ecol. Prog. Ser. 112, 277–301. https://doi.org/10.3354/meps112277.
Shumway, S.E., Burkholder, J.M., Morton, S.L., 2018. Harmful Algal Blooms: a Compendium Desk Reference. John Wiley & Sons.
Silva, A., Palma, S., Oliveira, P.B., Moita, M.T., 2009. Composition and interannual variability of phytoplankton in a coastal upwelling region (Lisbon Bay, Portugal). J. Sea Res. 62, 238–249. https://doi.org/10.1016/j.seares.2009.05.001.
Smayda, T.J., Trainer, V.L., 2010. Dinoflagellate blooms in upwelling systems: seeding, variability, and contrasts with diatom bloom behaviour. Prog. Oceanogr. 85, 92–107. https://doi.org/10.1016/j.pocean.2010.02.006.
Sánchez, R.F., Relvas, P., 2003. Spring–summer climatological circulation in the upper layer in the region of cape st. Vincent, Southwest Portugal. ICES J. Mar. Sci. 60, 1232–1250. https://doi.org/10.1016/S1054-3139(03)00137-1.
Spalding, M.D., Agostini, V.N., Rice, J., Grant, S.M., 2012. Pelagic provinces of the world: a biogeographic classification of the world's surface pelagic waters. Ocean Coast Manag. 60, 19–30.
Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A., Robertson, J., 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. Bioscience 57, 573–583. https://doi.org/10.1641/B570707.
Suominen, T., Westerholm, J., Kalliola, R., Attila, J., 2021. Partition of marine environment dynamics according to remote sensing reflectance and relations of dynamics to physical factors. Remote Sens. 13, 2104. https://doi.org/10.3390/rs13112104.
Taylor, B.B., Torrecilla, E., Bernhardt, A., Taylor, M.H., Peeken, I., Röttgers, R., Piera, J., Bracher, A., 2011. Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance. Biogeosciences 8, 3609–3629. https://doi.org/10.5194/bg-8-3609-2011.
Thomas, M.K., Fontana, S., Reyes, M., Kehoe, M., Pomati, F., 2018. The predictability of a lake phytoplankton community, over time-scales of hours to years. Ecol. Lett. 21, 619–628. https://doi.org/10.1111/ele.12927.
Thompson, P.A., O'Brien, T.D., Paerl, H.W., Peierls, B.L., Harrison, P.J., Robb, M., 2015. Precipitation as a driver of phytoplankton ecology in coastal waters: a climatic perspective, Estuarine. Coastal and Shelf Science 162, 119–129. https://doi.org/10.1016/j.ecss.2015.04.004.
Torres Palenzuela, J.M., González Vilas, L., Bellas, F.M., Garet, E., González-Fernández, Á., Spyrakos, E., 2019. *Pseudo-nitzschia* blooms in a coastal upwelling system: remote sensing detection, toxicity and environmental variables. Water 11. https://doi.org/10.3390/w11091954, 1954.
Trainer, V.L., Moore, S.K., Hallegraeff, G., Kudela, R.M., Clement, A., Mardones, J.I., Cochlan, W.P., 2020. Pelagic harmful algal blooms and climate change: lessons from nature's experiments with extremes. Harmful Algae 91, 101591. https://doi.org/10.1016/j.hal.2019.03.009.
Trainer, V.L., Pitcher, G.C., Reguera, B., Smayda, T.J., 2010. The distribution and impacts of harmful algal bloom species in eastern boundary upwelling systems. Prog. Oceanogr. 85, 33–52. https://doi.org/10.1016/j.pocean.2010.02.003.

Utermöhl, H., 1958. Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. Mitteilungen Internationale Vereinigung Für Theoretische Und Angewandte Limnologie 9, 1–38. https://doi.org/10.1080/05384680.1958.11904091.

Valbi, E., Ricci, F., Capellacci, S., Casabianca, S., Scardi, M., Penna, A., 2019. A model predicting the PSP toxic dinoflagellate *Alexandrium minutum* occurrence in the coastal waters of the NW Adriatic Sea. Sci. Rep. 9, 4166. https://doi.org/10.1038/s41598-019-40664-w.

Vale, P., Sampayo, M.A. de M., 2003. Seasonality of diarrhetic shellfish poisoning at a coastal lagoon in Portugal: rainfall patterns and folk wisdom. Toxicon 41, 187–197. https://doi.org/10.1016/S0041-0101(02)00276-3.

Vargas-Yáñez, M., Reeves-Bueno, A., Fernández-Topham, S., Moya, F., Ballesteros, E., Alonso, C., Pérez-Sánchez, T., Romero-Fernández, P., Sánchez-Aguado, S., Sánchez-Leal, R.F., García-Martínez, M.C., 2024. A method to establish marine bio-regions in the pelagic ecosystem based on phytoplanktonic communities. Aplication to the southern Spanish coast. Ocean Coast Manag. 247, 106930. https://doi.org/10.1016/j.ocecoaman.2023.106930.

Velasco-Senovilla, E., Díaz, P.A., Nogueira, E., Rodríguez, F., Garrido, J.L., Ruiz-Villarreal, M., Reguera, B., 2023. The niche of a stress-tolerant specialist, *Dinophysis acuminata,* in a coastal upwelling system. Harmful Algae 125, 102427. https://doi.org/10.1016/j.hal.2023.102427.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. SOM Toolbox for Matlab 5. Helsinki University of Technology.

Vidal, T., Calado, A.J., Moita, M.T., Cunha, M.R., 2017. Phytoplankton dynamics in relation to seasonal variability and upwelling and relaxation patterns at the mouth of Ria de Aveiro (West Iberian Margin) over a four-year period. PLoS One 12, e0177237. https://doi.org/10.1371/journal.pone.0177237.

Weithoff, G., Beisner, B.E., 2019. Measures and approaches in trait-based phytoplankton community ecology – from freshwater to marine ecosystems. Front. Mar. Sci. 6. https://doi.org/10.3389/fmars.2019.00040.

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, third ed. Academic Press.

Zhang, H.-M., Bates, J.J., Reynolds, R.W., 2006. Assessment of composite global sampling: sea surface wind speed. Geophys. Res. Lett. 33. https://doi.org/10.1029/2006GL027086.

Zhao, Q., Huang, H., Costello, M., 2025. No clear relationships between environmentally classified ecosystems and species diversity. Mar. Ecol. Prog. Ser. 764, 1–14. https://doi.org/10.3354/meps14877.

Zhou, Y., Yan, W., Wei, W., 2021. Effect of sea surface temperature and precipitation on annual frequency of harmful algal blooms in the East China Sea over the past decades. Environ. Pollut. 270, 116224. https://doi.org/10.1016/j.envpol.2020.116224.

Zou, L., Chen, H.T., Zhang, J., 2000. Experimental examination of the effects of atmospheric wet deposition on primary production in the Yellow Sea. J. Exp. Mar. Biol. Ecol. 249, 111–121. https://doi.org/10.1016/S0022-0981(00)00186-6.

Zuur, A.F., Ieno, E.N., 2021. The World of Zero-Inflated Models Volume 1: Using GLM. Highland Statistics Ltd.