Inferring changes in pico- and nanophytoplankton assemblages from long-term coastal thermohaline observations

Mathilde Couteyen Carpaye[1], David Nerini[1], David Pecqueur[2], Léa Buniak[1], Morgane Didry[1], Jean-Marc Feuerstein[1], Fabrice Garcia[1], Véronique Lagadec[1], Sandra Nunige[1], Christophe Salmeron[2], and Gérald Grégori[1]

[1]Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France
[2]Sorbonne Université, CNRS, Observatoire Océanologique de Banyuls s/mer, Banyuls sur Mer, France

Correspondence: Mathilde Couteyen Carpaye (mathilde.couteyen@mio.osupytheas.fr) and Gérald Grégori (gerald.gregori@mio.osupytheas.fr)

Abstract. Over the past 30 years, the Gulf of Lions in the northwestern Mediterranean Sea has undergone various environmental changes that have impacted marine ecosystems. At the beginning of the 21st century, shifts were observed in small pelagic fish and zooplankton communities. However, little is known about the impact of these changes on phytoplankton communities, especially pico- and nanophytoplankton (hereafter referred as PicoNano), which play a crucial role in the northwestern Mediterranean Sea. One major limitation is the lack of data on these size fractions during the observed shifts. In this study, we show that vertical profiles of temperature and salinity can serve to infer PicoNano assemblages from the thermohaline properties measured in a coastal area, the Bay of Marseille (BoM). We use functional data analysis and clustering methods to identify recurring vertical thermohaline structures over 30 years of temperature and salinity data (1994–2024), measured at low and high frequencies. To address the issue of truncated vertical profiles, we employ a reconstruction method based on functional principal component analysis. Using flow cytometry data collected over the past 11 years (2014–2024), we characterize PicoNano assemblages and the nutrient concentrations in the surface waters associated with each recurring thermohaline structure. We identify three thermohaline structures associated with the most oligotrophic conditions observed in the BoM, all involving a thermally stratified water column. The occurrence of these stratified conditions does not show any significant trend over the past 27 years. However, vertical structures indicative of Rhône River intrusions have been less frequently observed since 2004, while mixed water columns are more frequently highly homogeneous over the same period. We show that highly homogeneous water columns are associated with different PicoNano assemblages compared to mixed but less homogeneous water columns. Overall, these results suggest that thermohaline conditions associated with larger nanoeukaryotes have become less frequent since 2004. This work highlights that easily measurable variables, such as vertical profiles of temperature and salinity, can provide valuable insights into more complex variables—such as nutrient concentrations and PicoNano assemblages—when direct observations are unavailable, a situation commonly encountered in long-term and high-frequency monitoring.

Copyright statement. TEXT

1 Introduction

The Gulf of Lions (GoL) is one of the most productive areas of the Mediterranean Sea due to the Rhône River inputs, wind-driven coastal upwelling activity, bottom morphology and water circulation (Lefevre et al., 1997; Agostini and Bakun, 2002; Petrenko et al., 2005; Hu et al., 2009). It serves as a vital feeding ground for numerous marine species (Bănaru et al., 2013). In this ecosystem, small pelagic fish such as sardine *Sardina pilchardus* Walbaum, 1792 and anchovy *Engraulis encrasicolus* Linnaeus, 1758 play a key role as forage species (Bănaru et al., 2013). In addition to their fundamental role in marine food webs, these species have significant commercial value. Due to their plankton-based diet, small pelagic fish are highly sensitive to environmental variations and subsequent changes in plankton communities (Benson et al., 2002). Their sensitivity to environmental variations and the long-term data they provide, owing to their economic importance, make them effective sentries for environmental changes (Alheit et al., 2009). Since 2008, there has been a significant decrease in the landings of sardine and anchovy in the GoL. A decrease in fish size and weight has led to the cessation or reorientation of fishing activities (Saraux et al., 2019). Studies suggest that predation and overfishing are not the cause, but rather a bottom-up effect resulting from a reduction in the size of their planktonic prey (Brosset et al., 2016; Queiros et al., 2019; Saraux et al., 2019). Indeed, changes in small pelagic fish populations have occurred alongside diverse environmental changes observed in the GoL over the past few decades, which may have impacted the plankton communities (Feuilloley et al., 2020). In the context of climate change, it is crucial to understand how environmental fluctuations impact plankton communities. This knowledge is essential for mitigating the effects of anthropogenic activities and guiding sustainable fisheries' management, since shifts in plankton communities can have notable consequences on the ecosystem and the services it provides.

Monitoring plankton communities requires long-term time series that rely on ship-based measurements, usually managed by a marine laboratory in proximity to the monitored station (Edwards et al., 2010; O'Brien et al., 2017). These time series data are scarce (Edwards et al., 2010; Karl and Church, 2014; von Jackowski, 2024). Fortunately, the Bay of Marseille (BoM) which is a reproduction and fishing area for *S. pilchardus* in the GoL (Pichot et al., 1978; Lee, 1961), is subject to long-term monitoring by multiple French National Observation Services (SNO). In accordance with the dynamics of the small pelagic fish, the monitoring of zooplankton since 2005 has revealed a shift in the size structure of the mesozooplankton communities in 2008 (Garcia et al., 2023). Since 2009, the SNO SOMLIT (Service d'Observation en Milieu Littoral - Coastal Observation Service https://www.somlit.fr/) has used flow cytometry to monitor pico- and nanoplankton populations. Standardized protocols ensuring a consistent time series were established in 2012. Therefore, this dataset provides a reliable time series from 2012 onwards. Hence, the pico- and nano-sized fractions of phytoplankton have been a missing link in our comprehension of the changes observed during the first decade of the century, given that the consistent time series does not extend back to this period. Yet, this is a non-negligible link, as pico- and nanoplankton are believed to be the major contributors to zooplankton biomass in the northwestern Mediterranean Sea (Hunt et al., 2017).

Alongside the changes observed at the scale of the GoL over the past 30 years, the nutrient concentrations monitored by the SNO SOMLIT since 2004, show a tendency towards oligotrophication in the BoM (Lheureux et al., 2023). However, this nutrient depletion may have different causes. Indeed, the nutrient concentration in the surface waters of the BoM is impacted

by various hydrological processes. Firstly, the Mediterranean Sea is characterized by the onset of a seasonal thermocline from spring to autumn (Millot, 1990). The difference in temperature between the surface and bottom divides the water column into two layers, thereby reducing vertical mixing (Millot, 1990). This results in a lower supply of all nutrients to the surface, as it cannot be fuelled by the nutrients regenerated below (Sommer et al., 2002). In contrast, the GoL is subject to wind-induced upwellings all year round. During the stratified period, these upwellings replace warm and nutrient-depleted surface waters with cooler and nutrient-rich deeper waters (Millot, 1979; Bakun and Agostini, 2001). If enhanced stratification caused by an increase in sea surface temperature is observed and expected globally (Sarmiento et al., 1998; Yamaguchi and Suga, 2019; Cheng et al., 2025), the increased upwelling intensity observed in the GoL over recent decades (Feuilloley et al., 2020) may counteract this tendency by occasionally disrupting stratification. Another important source of nutrients in the surface waters of the GoL is the Rhône River (Millot, 1990; Moutin et al., 1998). If its dilution zone is located west of the BoM, the Rhône River plume may intrude the bay under certain conditions of wind and river flow (Fraysse et al., 2014). These intrusion events in the BoM are well documented and have been the subject of several modelling studies (Gatti et al., 2006; Pairaud et al., 2011; Fraysse et al., 2013, 2014). However, the intrusions in the southern part of the bay, where the long-term monitoring station is located, are not well reproduced (Pairaud et al., 2011; Fraysse et al., 2014). Consequently, it remains unclear to which extent the impoverishment of the Rhône River, observed since the early 2000s, has contributed to the negative trend in nutrient concentration evidenced at the monitoring site in the BoM (Feuilloley et al., 2020; Lheureux et al., 2023). Furthermore, during a nutrient-enhanced period in the BoM, intrusions of the Northern Current may considerably decrease the nutrient concentration in the entire water column by introducing warm oligotrophic water (Ross et al., 2016). The intrusions of the Northern Current in the BoM are known to be influenced by wind and stratification (Petrenko, 2003; Petrenko et al., 2005; Barrier et al., 2016), and may consequently have been impacted by the changes in the wind regimes observed over the past 30 years (Feuilloley et al., 2020).

Considering the oligotrophic nature of the Mediterranean Sea, the influence of these hydrodynamic conditions on surface nutrient concentration also shapes phytoplankton assemblages (Para et al., 2010; Siokou-Frangou et al., 2010; Mena et al., 2019; Fuchs et al., 2023). Due to the complex hydrodynamics of the coastal area, it is unclear which changes impacted the availability of nutrients and how this affected the phytoplankton assemblages, leading to the observed changes in the zooplankton communities in the BoM. The longest time series recorded by the SNO SOMLIT in the BoM concerns temperature and salinity, both at the surface (1 m depth) and as vertical profiles (down to 50 m deep). These time series began in 1994, are reliable, and cover the period during which the environmental and biological changes observed in the GoL were observed. Most hydrodynamic conditions described above also impact the thermohaline vertical structure of the water column. In this context, our study aims to identify these hydrodynamic conditions in terms of vertical temperature and salinity profiles (TS profiles).

Specifically, we aim at using the SNO SOMLIT and COAST-HF time-series to evaluate the effectiveness of TS profiles in identifying hydrodynamic conditions that affect the availability of nutrients and pico- and nanophytoplankton (PicoNano) assemblages at the surface, in the BoM. We employ functional data analysis and clustering methods to address the issue of missing data in time series and identify recurring vertical thermohaline structures within the water column. Once these recurring

vertical structures have been identified, we match them to the hydrodynamic conditions responsible for these profiles. We then compare the concentrations of surface nutrients and the PicoNano assemblages between the identified vertical structures. We also address the reliability of the identification of such vertical structures using automated measurements. This study improves our understanding of local environmental variations at a coastal northwestern Mediterranean site and enables us to empirically infer the potential changes in surface nutrient concentrations and phytoplankton assemblages when these data are unavailable. This work points out the insightfulness of exploiting easily measurable and fairly reliable variables, such as temperature and salinity, throughout the entire water column for long-term monitoring in coastal areas.

2 Material & Methods

2.1 Study site and period

The long-term monitoring station in the BoM is the Frioul site (43°14.50'N, 5°17.23'E), which is located 5 km from the continental coast and is 60 m deep (Appendix A1). Since 1994, it has been surveyed twice a month by the SNO SOMLIT, and since 2011, at high frequency by the SNO COAST-HF (Coastal OceAn observing SysTem – High Frequency, https://coast-hf.fr/). The study period ran from 1994 to 2024. The time series used in this study are listed in Table 1. Whenever possible, the time series were used in their entirety. When this was not possible, the length of the time series was determined based on known changes in protocols and standardization and obvious anomalies identified visually. The SNO SOMLIT created a new standardized protocol for nutrients in November 2011. Therefore, nutrient data were used from 2012 onwards. PicoNano flow cytometry data have been standardized since 2012. Due to a change of flow cytometer, a change point was evident in all the sideward light scatter (SSC) time series in 2014, but not in the time series of abundance. Therefore, only the abundance and SSC data from 2014 onwards were used. Due to the low frequency of the SNO SOMLIT monitoring, 691 couples of TS profiles were collected from 1994 to 2024. However, owing to the high frequency of the COAST-HF monitoring, 928 couples of TS profiles were collected from 2014 to 2024. Therefore, to constitute a substantial dataset, when working on vertical profiles, we combined the low- and high-frequency data.

2.2 Environmental Data

2.2.1 High-frequency hydrological data

The SOLEMIO buoy of the SNO COAST-HF was deployed in September 2011, close to the Frioul archipelago (43°14.7428' N, 5°17.1950' E), less than 500 m away from the long-term monitoring station surveyed by the SNO SOMLIT in the southern Bay of Marseille (Appendix A1). This system enables high-frequency surface data acquisition (hourly) and vertical profiling (once a day) throughout the water column in a marine environment subject to harsh meteorological and hydrodynamic conditions, including the dominant Mistral wind, waves up to 4 m, and strong currents. These conditions being more frequent and intense in winter, the buoy is generally deployed from April to November.

Table 1. List of time series used in the study. HF: high frequency; LF: low frequency; PicoNano: pico- and nanophytoplankton flow cytometry data.

Time series	Variables	Depth	Observation System	Site	Time Span	Sampling Frequency	Time of Year
HF surface	temperature (°C), salinity	1 m	COAST-HF	SOLEMIO	2014–2023	1 hour	April–November
HF vertical profiles	temperature (°C), salinity	5–50m	COAST-HF	SOLEMIO	2014–2023	1 day	April–November
LF vertical profiles	temperature (°C), salinity	1–50m	SOMLIT	Frioul	1994–2024	15 days	all year round
Surface nutrient concentration and chlorophyll *a*	NH_4^+ (µM), NO_3^- (µM), NO_2^- (µM), PO_4^{3-} (µM), $Si(OH)_4$ (µM), chl *a* ($\mu g.L^{-1}$)	1 m	SOMLIT	Frioul	2012–2024	15 days	all year round
Surface PicoNano	abundance ($cell.mL^{-1}$), SSC	1 m	SOMLIT	Frioul	2014–2024	15 days	all year round
Wind	speed ($m.s^{-1}$), direction (degrees)		Météo-France	Marignane	1997–2024	3 hours	all year round

A multiparameter probe from NKE Instrumentation, a SMATCH (from June 9th 2011 to September 23rd 2021) and then a WIMO continuously measured temperature (°C), salinity (Practical Salinity), turbidity (NTU), and chlorophyll *a* (chl *a*, $\mu g.L^{-1}$) at hourly intervals. Data are transmitted daily to the Mediterranean Institute of Oceanography (MIO) server and integrated in real time into the national Coriolis Côtier database (https://data.coriolis-cotier.org/platform/EXIN0004). A PROVOR-type profiler (NKE Instrumentation), adapted for fixed-point deployment on a mooring line, performs daily vertical profiles along an inductive cable connecting a surface buoy to a 50 m deep anchor weight. The system has been operational since 2014. Measured variables include those recorded by the surface probe, supplemented by phycoerythrin-induced fluorescence. Only temperature and salinity measurements were used in this study (Garcia et al., 2025). The vertical profiles data were available from 2014 to 2023. More details on the buoy can be found in Garcia et al. (2016).

2.2.2 Low-frequency hydrological data

Low-frequency hydrological data were retrieved for the Frioul coastal site surveyed by the SNO SOMLIT from the SOMLIT database (SNO SOMLIT, 2025). Sampling was conducted every two weeks. Additional information on the measured variables, sampling protocols and analytical protocols is available on the SNO SOMLIT website (SNO SOMLIT, 2025) and in Cocquempot et al. (2019). The following environmental variables measured at the surface (1 m) were retrieved : water temperature (°C) and salinity, ammonium (NH_4^+; µM), nitrate (NO_3^- ; µM), nitrite (NO_2^-; µM), phosphate (PO_4^{3-}; µM), orthosilicic

acid ($Si(OH)_4$; µM), and chl *a* ($\mu g.L^{-1}$). Chlorophyll *a* was used as a relative measure for comparisons among observations rather than as an absolute estimate of chl *a* concentration. The temperature and salinity vertical profiles were obtained with the Conductivity-Temperature-Depth (CTD) Sea-Bird SBE19 probe from 1 m to 50 m deep. Since 2006, standardized protocols have been used and data-quality control has been routinely performed (Breton et al., 2023). The detailed analysis protocols are available on the SNO SOMLIT website (SNO SOMLIT, 2025). Chlorophyll *a* was used as a relative measure, allowing comparisons among observations, rather than as an absolute estimate of chl *a* concentration.

2.2.3 Wind data

Wind data were retrieved from the Météo-France SYNOP time series (Météo-France, 2026). The closest station is Marignane, located 22 km further northwest from the Frioul site (Appendix A1). Although some meteorological surveys have been conducted closer to the Frioul site, none lasted continuously for the last 25 years. The SYNOP time series have been continuously recording meteorological variables including wind speed and direction from 1997, every 3 hours. Wind speed was given as the mean speed recorded over 10 minutes in $m.s^{-1}$. Wind direction was given as the mean direction over 10 minutes in degrees, rounded to 10.

2.3 Biological Data

Along environmental variables, picoplankton (0.2 < size < 2-3 µm) and nanoplankton (2-3 < size < 20 µm) communities were characterized by flow cytometry for SOMLIT surface water samples. Phytoplankton samples collected within the SNO SOMLIT were analysed at the BioPIC platform (Observatoire Océanologique de Banyuls, France), using a FACS Canto II (BD Biosciences), with the same operator throughout the study period. Flow cytometry analyses followed the protocols described in Marie et al. (1999, 2014), and the gating strategy followed the scheme described in Marie et al. (2000), which is also presented in Appendix B1. Five groups of pico- and nanophytoplankton were distinguished based on their optical properties: RedPicoProk, OraPicoProk, OraNano, RedPico and RedNano. RedPicoProk is associated with the cyanobacteria *Prochlorococcus* Chisholm et al., 1992, which are generally smaller than 0.7 µm (Morel et al., 1993). OraPicoProk is associated with the cyanobacteria *Synechococcus* Nägeli, 1849, which generally range from 0.8 to 1.1 µm in size (Morel et al., 1993). RedPico includes eukaryotic picophytoplankton smaller than 3 µm in size. OraNano comprises phytoplankton ranging from 3 to 20 µm in size and containing phycoerythrin, including members of the class *Cryptophyceae* Fritsch, 1927. RedNano regroup eukaryotic nanophytoplankton ranging from 2-3 µm to 20 µm that lack phycoerythrin. A more detailed description of the groups has been provided by Thyssen et al. (2022) or can be found on the Natural Environment Research Council (NERC) Vocabulary Server (https://vocab.nerc.ac.uk/collection/F02/current/). The eukaryotic groups may encompass multiple taxa and are therefore considered functional groups based on cell size (Le Quéré et al., 2005; Fuchs et al., 2022). Phytoplankton cell size influences many ecological processes, such as sinking rate, nutrient uptake, light harvesting, and grazing pressure (Kiørboe, 1993). As a result, it is considered one of the most relevant functional traits (Litchman and Klausmeier, 2008).

For each group, the abundance ($cell.mL^{-1}$), and the mean SSC of the cells were determined. The SSC is affected by the internal structure of cells and can be used as a proxy for the size of cells above 1 µm (Dubelaar et al., 2007). It is also a

good indicator of the size of picophytoplanktonic cells, as it is usually positively correlated with cell diameter (Calvo-Díaz and Morán, 2006).

The SNO SOMLIT data used in this study can be visualized using the SOMLITMed Explorer Shiny application available at https://mcouteyen.github.io/SOMLITMed_Explorer/ (Couteyen Carpaye, 2026b).

2.4 Statistical Analysis

All data analyses were performed using the R free software environment (version 4.4.0) for statistical computing and graphics (R Core Team, 2024). The significance level for all statistical analyses was set at $\alpha = 0.05$.

2.4.1 Outliers identification

The consistency of the time series and outliers were identified using exploratory visualizations, including distribution-based plots (box plots and density plots) and temporal representations (time series plots). All time series and vertical profiles were checked visually using an R shiny app developed for this purpose.

2.4.2 Smoothing of the vertical profiles

Functional data analysis (FDA) was applied to the TS profiles. This branch of statistics is specifically designed for the analysis of data that can be represented as continuous curves, such as vertical profiles. By representing observations as functions, FDA overcomes challenges associated with irregular sampling grids and enables dimensionality reduction. The first step in functional data analysis is the conversion of discrete data into functions. To do so, penalized B-splined smoothing was used on every vertical profiles to obtain a continuous function from 1 m to 50 m deep using the fda package. A vertical profile is expressed by a linear combination of the basis functions, in this case, B-splines. This method is extensively described by Ramsay and Silverman (2005). The number of basis functions was set to 100 after exploring values ranging from 30 to 250. This number was sufficiently large to ensure that smoothing was primarily controlled by the smoothing parameter. The latter was arbitrarily fixed at 10^{-3} to preserve the resolution of the measuring instruments (0.001 °C for temperature and 0.001 for salinity). This parameter can be determined using various methods, such as cross-validation, but the efficiency of these methods is disputable (Ramsay and Silverman, 2005).

2.4.3 Vertical profile reconstruction

Combining the SOMLIT and COAST-HF datasets would increase substantially the number of observations available. However, HF vertical profiles collected automatically were truncated compared to the real water column height owing to the installation constraints and dimensions of the buoy. Figure 1 shows that most data from the surface down to 10 m deep and from 45 m to 50 m deep were missing. This may result in the loss of crucial information contained within the upper part of the vertical profile, such as thermocline formation and the freshwater discharge. Truncated vertical profiles were completed by the surface observations measured by the probes of the automated buoy (SMATCH or WIMO) in the same 1-hour interval when available.

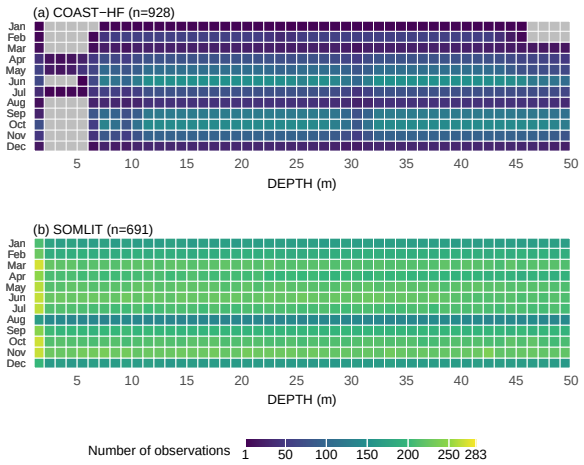


Figure 1. Number of observations available by month and depth at the Frioul site, aggregated from vertical profiles for the 1994–2024 period. These data are presented for (a) the COAST-HF dataset (2014–2024) and (b) the SOMLIT dataset (1994–2024). The number n of vertical profiles in each dataset is indicated in the title of the plots. The legend is common for both heatmaps, with colours indicating the number of observations, ranging from low (purple) to high (yellow), and grey indicating 0 observations.

To combine the SOMLIT and COAST-HF datasets, the first step was to assess the potential bias between the two TS profiles datasets. First, all profiles collected on the same day by both the SNOs were selected. The profiles were smoothed using B-spline smoothing over the 10–40 m depth range, and the smoothed values were evaluated every 0.5 m. The residuals between the two datasets were then calculated at each depth. Finally, linear regression was applied to the residuals to assess their linearity.

The second step involved reconstructing the truncated profiles to create a dataset of complete vertical profiles ranging from 1 to 50 m deep. Penalized B-spline smoothing is highly effective for predicting curves with an irregular sampling grid, but tends to be inaccurate when the data are truncated at either end. To ensure better accuracy, a method based on principal component analysis through conditional expectation (PACE), as described by Yao et al. (2005), was used to predict the vertical profiles with missing ends. PACE uses the entire dataset to identify the mean trend, covariance matrix, and principal modes of variations (eigenfunctions) with Functional Principal Component Analysis (FPCA). These are then combined with the individual features to predict the individual curve. Our method differed from that used by Yao et al. (2005) because we did not use the entire dataset to characterize our sampled profiles. Instead, our dataset was partitioned into complete and incomplete profiles, and only complete profiles were used. A profile was considered complete if it started above 1.5 m and ended below 49.5 m in depth. Most of the vertical profiles from the SOMLIT dataset were complete (Fig. 1) and with some profiles from the COAST-HF dataset, they constituted a sample of more than 750 trajectories from which the mean curve, covariance matrix, eigenvalues, and eigenfunctions were estimated. The FPCA scores of each incomplete profile for the first K eigenfunctions were then estimated with PACE (Yao et al., 2005), and each profile was predicted on $t \in [1, 50]$ m. The choice of K was based on the number of

principal components retained in the FPCA on the complete vertical profiles, for which the variability explained plateaued. Here, K was set to 30, which explained 99% of the variance for both temperature and salinity. The parameter σ^2, which controls the fit between the prediction and observed data, was set to $\sigma^2 = 0.01$ for temperature and salinity. The different steps of the vertical profile reconstruction method and the procedure used to set σ^2 are described in Appendix C. The fda package was used, and the script developed is publicly available (Couteyen Carpaye, 2026a).

To evaluate the efficiency of the prediction method, a simulation was performed using the SOMLIT complete vertical profiles. Twenty percents of the dataset (150 profiles) were randomly truncated between 3 and 10 m at the surface and 40 and 48 m at the bottom. The remaining 80% were kept intact and used for the estimation of the mean curve, covariance matrix, and eigenfunctions. The mean squared errors (MSE) were computed for the predictions of the truncated profiles against the intact discrete measurements. The average MSE was computed for 100 Monte Carlo simulations.

The same simulation was performed by adding the measurement at the surface (1 m) for every truncated profile to replicate the surface sensor measurement.

Two variants of the reconstruction method were tested to determine whether they would improve the accuracy of the prediction. The first variant was conditional FPCA (cond. FPCA). This method consist in weighting the observations based on their proximity to the prediction. In our case, the proximity was based on the time of the year and the observations were weighted with a Gaussian kernel with a 50 days window. For each prediction, the mean curve, covariance matrix and eigenfunctions were determined conditionally to the date of the prediction. This method is more time-consuming, as a new FPCA is computed for each prediction. The fda package was used along scripts developed based on the description of the method by Cardot (2007).

The second variant was multivariate FPCA (MFPCA). For this method, the information carried by both temperature and salinity profiles are used through the cross-covariance matrix. In this method, the MFPCA scores are common for a couple of temperature-salinity profiles, but each variable is associated to its mean curve, covariance matrix and eigenfunctions. The fda package was used along scripts developed based on the description of the method by Nerini et al. (2021).

2.4.4 Classification of similar TS profiles

The goal was to identify recurring hydrodynamic conditions that would affect the thermohaline vertical structure of the water column differently. To achieve this, pairs of TS profiles with similar shapes were grouped together using a clustering method.

After reconstruction, the functions representing the data for all the temperature and salinity vertical profiles from 1 to 50 m were pooled together. In this dataset, the predominant signals were the seasonal variation of temperature and the salinification in the GoL (Sammartino et al., 2022; Liu et al., 2025). To reveal the obscured information and to focus on the variations of temperature and salinity along the water column, each vertical profile was self-centred as described in Appendix E. The resultant vertical profiles represented a variation of temperature (ΔT) or salinity (ΔS) on the $[1, 50]$ m depth interval.

To reduce the dimensions of the dataset, MFPCA was conducted on the n pairs of self-centred TS profiles. The first four principal component scores were retained, accounting for 93.66% of the total variance. These components represent the four principal modes of variations observed among self-centred temperature profiles and self-centred salinity profiles. Additionally, it reduced the dimensions of the dataset from $2 \times n \times 100$ to $n \times 4$.

Before applying model-based clustering on the MFPCA scores, the Euclidean distances d of each observation to the centroid of the dataset were computed. To prevent the extreme observations to drive the clustering, the most distant values from the centroid ($d > 4$) were set aside. Model-based clustering was then applied to the remaining MFPCA scores with the mclust package. The Mclust method was used and the clustering was evaluated for two to ten clusters and for all models available. Initialization was made with Hierarchical Agglomerative Clustering (HAC) and the model was selected based on the higher Bayesian Information Criterion (BIC). The model selected was the unconstrained model VVV representing nine clusters as ellipsoid with varying volume, shape, and orientation. Classification was accepted when the level of uncertainty was lower than 0.25, i.e. when the posterior conditional probability of an observation coming from the cluster was higher than 0.75 (Scrucca et al., 2023). The observations that were previously set aside were classified with the resulting model. The nine clusters were interpreted as nine vertical structures, characterized by variations in temperature and salinity throughout the water column. For clusters indicative of freshwater intrusions, profiles for which bottom salinity (50 m) was not at least 0.2 higher than surface salinity (1 m) were considered inconsistent with a freshwater intrusion and excluded from these clusters. An a priori threshold of 0.2 was used as an operational criterion to ensure that profiles assigned to these clusters exhibited a minimum vertical salinity gradient consistent with a freshwater intrusion.

To evaluate the reliability of the clustering on the prediction of the truncated vertical profiles, a simulation was made with the SOMLIT complete vertical profiles. Twenty percents of the dataset (150 profiles) were randomly truncated between 3 and 10 m at the surface and 40 and 48 m at the bottom. The remaining 80% were left intact and used for the estimation of the mean curve, covariance matrix, and eigenfunctions. The predictions of the truncated profiles were classified with the clustering model previously obtained. And the predicted clusters were compared to the clusters that had been assigned to the complete profiles before degradation. Confusions matrices were computed for 100 Monte Carlo runs.

The same simulation was performed by adding the measurement at the surface (1 m) for every truncated profile to replicate the surface sensor measurement.

2.4.5 Vertical structure comparisons

Comparisons between vertical structures were made using only the TS profiles from the SOMLIT dataset, which were sampled consistently throughout the year (Table 1). The observed density of each vertical structure over the year was computed with the circular package, which can deal with periodic data.

Nutrient concentration, chl *a* and PicoNano assemblages from surface samples were compared among vertical structures for the 2012–2024 period. During this period, nutrient data, chl *a* and flow cytometry data were available each time temperature and salinity profiles were measured. Thus, each vertical structure identified from the TS profiles was matched to the nutrient concentration, chl *a* and flow cytometry data obtained on the same day. During this period, the number of observations among clusters was greater than 12, except for vertical structure 8, for which three observations were available. Consequently, vertical structure 8 was excluded from the comparisons.

From 2012 to 2024, nutrient concentrations (NH_4^+, NO_3^-, NO_2^-, PO_4^{3-}, $Si(OH)_4$) were systematically available alongside the TS profiles of the SOMLIT dataset, with no missing values. A mean-normalized nutrient index was computed for each date.

It was computed as the sum of nutrient concentrations normalized by their respective means. As the mean-normalized nutrient index was not normally distributed, it was compared among vertical structures using the non-parametric Kruskal–Wallis test and Dunn's post hoc test with Bonferroni adjustment using the rstatix package. Descriptive discriminant analysis (DDA), as described in Saporta (1990), was used to distinguish vertical structures based on nutrient concentrations. Variables were centred and reduced before the DDA.

Prior to analysis, chlorophyll *a* concentrations were log-transformed. Following transformation, the assumptions of normality and homoscedasticity of the residuals were met. Chlorophyll *a* concentrations were compared among vertical structures using a one-way ANOVA and Tukey's post hoc test with the stats package.

From 2012 to 2013, no sideward light scatter data were usable due to the reasons discussed in Sect 2.1. To avoid biasing the analysis by imputing values, these years were not included. A second type of missing data occurred when the abundance of a group was zero, in which case SSC values are not defined. Because RedPicoProk and OraNano abundances were frequently zero, the corresponding SSC variables (RedPicoProkSSC and OraNanoSSC) were excluded from the analysis. Because temperature and light are known to strongly influence the distribution of pico- and nanophytoplankton species (Partensky et al., 1999a; Flombaum et al., 2013), a segregation by season could reduce the bias induced by the seasonality of these variables and the vertical structures. Since there were too few observations per season, winter-spring and summer-autumn were compared among them. Winter-spring period was set from December 1st to May 31st, when the temperature was typically below 16 °C. The summer-autumn period was set from June 1st to November 30th, when the temperature ranged from 14 °C to 27 °C. For both winter-spring and summer-autumn periods, DDA was used to explore and summarize differences among the vertical structures based on the abundance and SSC of each PicoNano group. Variables were centred and reduced before the discriminant analysis.

2.4.6 Wind analysis

Vertical structures exhibited varying degrees of vertical mixing. As wind forcing is considered the primary driver of vertical mixing in the Gulf of Lion (Pairaud et al., 2011), the intensity of north-westerly winds (320°, 330°, 340°, and 350°) was compared between mixing regimes. These comparisons were made solely using the SOMLIT dataset because reconstructed COAST-HF profiles may have been misclassified. Because the SOMLIT sampling was conducted no later than 2:00 pm, wind intensity data recorded on the same day between 12:00 am and 2:00 pm were used for the analysis. As data were not normally distributed, one-sided Wilcoxon rank-sum tests were performed for each wind direction to assess whether wind intensity was greater in well-mixed than in mixed waters. A one-sided test was used based on the a priori hypothesis that stronger winds enhance vertical mixing. Because the objective was to identify potentially relevant wind directions within the north-west sector rather than to test a single global hypothesis, the analysis was considered exploratory and no correction for multiple comparisons was applied. P-values should therefore be interpreted as indicative of potential direction-specific patterns rather than as confirmatory statistical evidence.

An exploratory analysis of wind data over the past thirty years revealed a shift in the dominant wind direction within the north-western sector. For each year between 1998 and 2024, the 95th percentile of wind intensity was calculated for each wind

direction of the north-western sector (between 270° and 360°), rather than using maximum values, to reduce the influence of extreme observations. The dominant wind direction for each year was defined as the mean direction corresponding to the highest 95th percentile. A regression tree was then applied to the annual dominant wind direction as a function of year to identify potential shifts in wind direction over time, using the tree package. A regression tree was fitted 100 times, and for each run, k-fold cross-validation with $k = 5$ was used to select the tree complexity that minimized deviance. A tree with four terminal nodes was retained for the final analysis, as this level of complexity was selected by the cross-validation procedure in about 85% of the runs.

2.4.7 Generalized Additive Model

As for the comparisons, the model was fitted only to TS profiles from the SOMLIT dataset, which were sampled consistently throughout the year. Vertical structures exhibiting similar hydrological characteristics and PicoNano assemblages (abundance and SSC) were grouped together to reduce the number of categories before model fitting. For each aggregated category, annual counts and proportions of observations were computed for each complete year from 1998 to 2024. The 1994–1997 data were not used because the entire years were not covered. To assess potential non-linear effects of time on the observation frequency of each category, a multinomial generalized additive model (GAM) was fitted using the VGAM package (Yee, 2010). Vertical couples that could not be categorized into a vertical structure were regrouped in a not categorized (NC) category. The NC category was set as the reference group, as it was never zero and remained reasonably constant over time. A vector cubic spline smoother with four degrees of freedom (df) was applied to the time variable, as higher df did not improve the mean log-likelihood obtained through leave-one-out cross-validation. A detailed description of the statistical model is provided in Appendix D. Ninety-five percent pointwise confidence intervals were computed by resampling individual observations with replacement within each year and deriving empirical quantiles from the predicted probabilities.

As the assignment of observations to clusters depended on the uncertainty threshold used for model-based clustering, the sensitivity of the multinomial GAM results to this threshold was assessed using values of 0.2, 0.25, 0.3 and 0.4 (Appendix D).

3 Results

3.1 Vertical profile reconstruction

To assess the efficiency of the vertical profiles reconstruction methods, the average mean squared error was computed over 100 Monte Carlo runs, based on the prediction of 150 SOMLIT TS profiles that were randomly truncated. For temperature profiles, the most efficient method was the cond. FPCA, which had an MSE 5% lower than the FPCA method (Table 2). For salinity profiles, the FPCA was the most efficient method, with an MSE 4% lower than the MFPCA and 12% lower than the cond. FPCA (Table 2).

The FPCA method produced fairly reliable predictions for both temperature and salinity profiles in 75% of cases (see Appendix F1). However, in 5% of cases, the difference between the predicted and observed surface measurements was notable

Table 2. Average mean squared error (MSE) for 100 Monte Carlo runs using n = 150 randomly truncated vertical profiles of temperature and salinity. Comparisons are made between profiles without surface measurements and those supplemented by a measurement at a depth of 1 m. Predictions are generated using three variants of PACE described in the Methods section: univariate FPCA (FPCA), multivariate FPCA (MFPCA), and conditional FPCA (cond. FPCA). The units for MSE are $°C^2$ for temperature and dimensionless for salinity.

Variable	Method	Without surface measure	With surface measure
Temperature	FPCA	0.0260	0.00889
	MFPCA	0.0267	0.0103
	cond. FPCA	0.0247	0.00875
Salinity	FPCA	0.00251	0.000965
	MFPCA	0.00260	0.000993
	cond. FPCA	0.00286	0.00108

(>0.8 °C for temperature, >0.2 for salinity). These poor predictions primarily impacted profiles with a shallow stratification, most discernible in the truncated section. The poor prediction of temperature profiles mainly affected those observed in summer and early autumn, particularly in June and September, when the thermocline was often shallow (less than 5 m deep). Poor salinity predictions mainly affected profiles observed in June, but could potentially occur from April to December. Adding a surface measurement to the truncated profiles considerably improved the prediction, reducing the MSE by 66% (Table 2).

The COAST-HF dataset did not show significant deviations from the SOMLIT dataset. The relationships between the vertical profiles of the two datasets on the same day were highly linear for both temperature (slope = 1, intercept = 0.06, R^2 = 0.98) and salinity (slope = 0.98, intercept = 0.76, R^2 = 0.92). When supplemented with the surface sensor measurement at the depth of 1 m, 15.8% of the 928 COAST-HF profiles were complete, while 87.2% of the 691 SOMLIT profiles were complete. Combining both datasets produced consistent predictions for incomplete profiles using PACE, for both temperature and salinity (Fig. 2). Even when surface measurements were not available, the projection of the predicted scores of the incomplete profiles with PACE overlapped well with the scores of the complete profiles that were used to determine the FPCA factorial planes. This can be seen in Fig. 2, where the first factorial planes for the temperature and salinity profiles depict 98% and 93% of the total variance, respectively, of the datasets constituted by complete profiles.

3.2 Identification of vertical structures from TS profiles

A multivariate functional principal component analysis (MFPCA) was conducted on pairs of temperature and salinity profiles. The first four principal components accounted for 93.66% of the total variance. A total of 1,619 TS profiles were clustered into nine different vertical structures with model-based clustering based on the first four PC scores of the MFPCA (Fig. 3). Approximately 25% of the TS profiles could not be classified as the uncertainty was higher than 0.25.

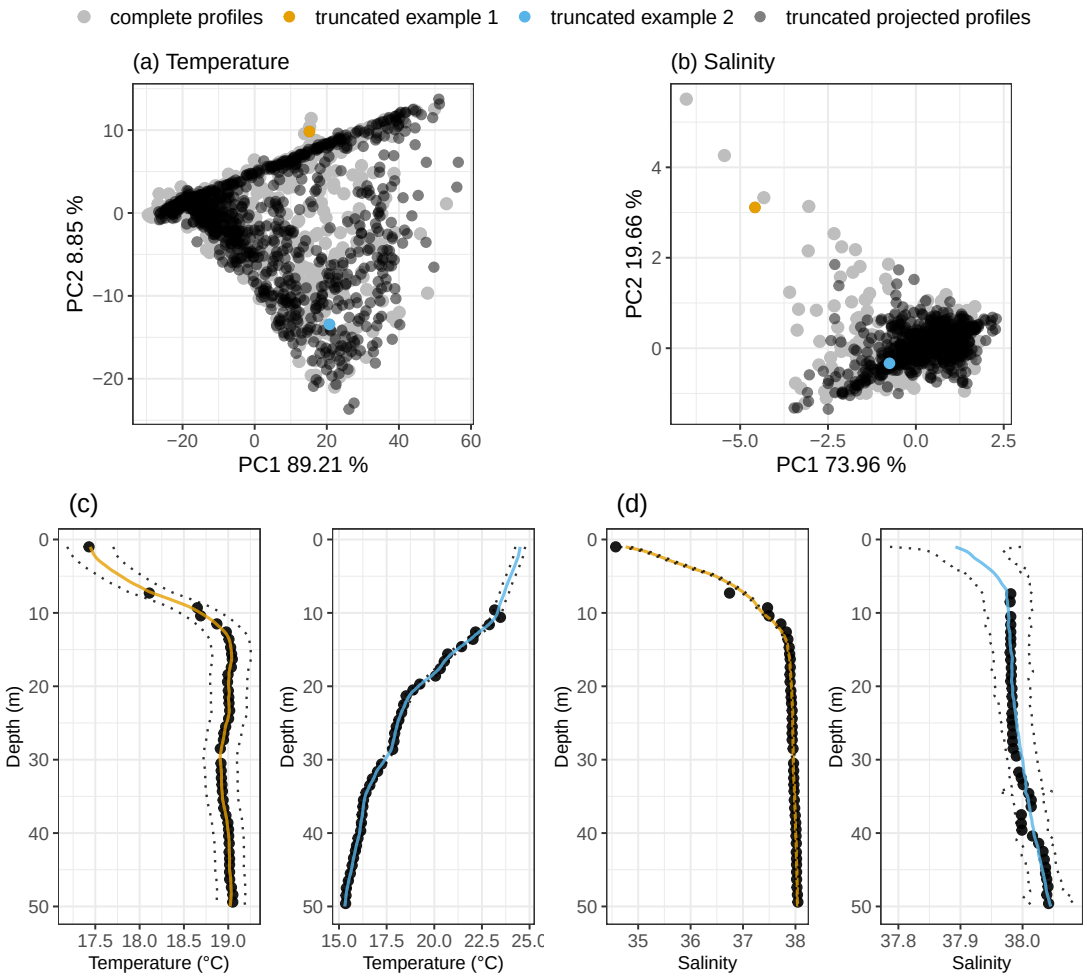


Figure 2. Projection of predicted scores for truncated vertical profiles (n = 867) using PACE on the first factorial plane of the FPCA derived from complete vertical profiles (n = 752). Predicted scores are represented as black dots, while scores from complete profiles are shown as grey dots. (a) Temperature first factorial plane; (b) salinity first factorial plane. Orange and blue dots indicate two truncated profiles, with their predictions displayed below as orange and blue curves. (c) Example predictions of the temperature profile; (d) example predictions of the salinity profile. Black dots represent observed measurements; solid lines indicate the predictions, while dotted lines represent the associated 95% asymptotic pointwise confidence interval.

Vertical structures 6, 9 and 7 exhibited mixed waters, with varying degrees of mixing. Based on the temperature and salinity differences between the depths of 1 m and 50 m, the most homogeneous water columns were labelled 'well-mixed waters', while those that were less homogeneous were labelled 'mixed waters', as detailed thereafter.

Vertical structures 6 and 9 indicated well-mixed waters, with a temperature difference of no more than 0.8 °C and a salinity difference of no more than 0.3 between the surface and a depth of 50 m (Appendix G1). These structures were mainly encoun-

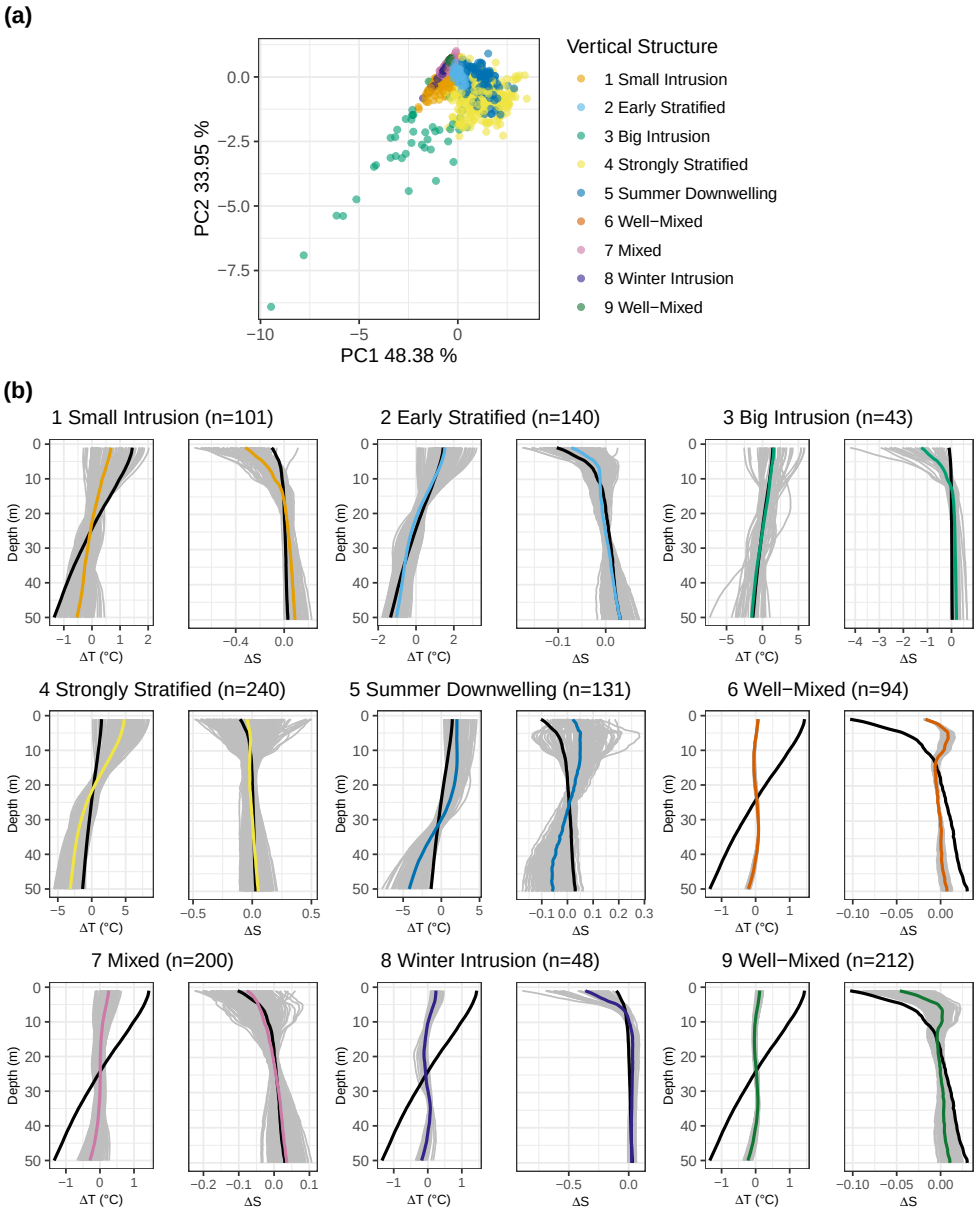


Figure 3. (a) First factorial plane of the MFPCA on the 1,619 pairs of temperature (ΔT) and salinity (ΔS) self-centred vertical profiles (TS profiles) from SOMLIT and COAST-HF datasets. Colours highlight the nine vertical structures determined by model-based clustering on the first four principal component scores. Vertical structures were assigned when uncertainty was lower than 0.25 and 410 TS profiles were not assigned to a vertical structure. (b) Profiles associated to each of the nine vertical structures, grey lines represent the vertical profiles of temperature (left) and salinity (right) and coloured lines represent the mean TS profiles of the vertical structure. Black lines represent the mean TS profiles of the whole dataset.

tered in winter, when north-westerly winds were strong (up to 13 $\mathrm{m.s^{-1}}$) and were not observed from June to September (Fig. 4, Appendix G2).

Vertical structure 7 was indicative of mixed waters, with a temperature and salinity difference between the surface and a depth of 50 m slightly higher than that of well-mixed waters (up to 1.6 °C and 0.34, respectively). Mixed waters were encountered year-round when north-westerly winds were moderate to strong (5–13 $\mathrm{m.s^{-1}}$). They were scarce in summer.

Vertical structure 4 was indicative of strongly stratified waters with a temperature difference reaching 13 °C between the surface and the bottom. This vertical structure was only encountered from April to November, when winds were weak (<8 $\mathrm{m.s^{-1}}$).

Vertical structure 5 was indicative of summer downwellings, characterized by a homogeneous surface layer approximately 30 m thick with warmer temperatures. This structure was mainly encountered in September, when winds were weak (<8 $\mathrm{m.s^{-1}}$), and was absent in winter.

Vertical structures 1, 3, and 8 were indicative of freshwater intrusions. Vertical structure 8 indicated winter intrusions, which occurred when the temperature was homogeneous throughout the water column and the wind blew from an east-south-east direction. Vertical structure 3 was indicative of big intrusions, with a significant decrease in salinity at the surface, reaching down to 4.4 lower than at the bottom. These intrusions occurred throughout the year, peaking in May, with no clear pattern in the temperature profiles. The prevailing winds were moderate north-westerly winds (8–10 $\mathrm{m.s^{-1}}$). Vertical structure 1 was labelled small intrusion because the salinity decrease at the surface was less important than in vertical structure 3 (less than 1). These occurred year-round, but primarily in spring and fall, when the water column showed little to no stratification and winds were weak (<8 $\mathrm{m.s^{-1}}$).

3.3 Insights into vertical structures from low-frequency monitoring

The low-frequency data from the SNO SOMLIT were used to compare nutrients, chl *a* and PicoNano assemblages between vertical structures for the 2012–2024 period. There was little data available on winter intrusions, as they were rare and were only captured three times by the low-frequency survey between 2012 and 2024. Consequently, they were not included in the comparison.

3.3.1 Surface nutrients & chlorophyll *a*

The mean-normalized nutrient index of surface waters differed significantly between vertical structures (Kruskal–Wallis, H = 63.7, df = 7, $p < 0.001$). The richest surface waters were found in well-mixed and mixed water columns. The total normalized nutrient concentrations showed significant differences under these conditions compared to those observed when the water column was stratified. Namely, under summer downwelling conditions, early stratified and strongly stratified conditions, where nutrient concentrations were lowest (Fig. 4). The nutrient composition of big intrusions was highly variable. However, they were generally richer in $\mathrm{PO_4^{3-}}$ and $\mathrm{NO_3^-}$. Small intrusions, meanwhile, were poorer in $\mathrm{NO_3^-}$ and $\mathrm{PO_4^{3-}}$. Stratified waters (strongly stratified, early stratified and summer downwelling) were the poorest in $\mathrm{PO_4^{3-}}$ and $\mathrm{NO_3^-}$. The nutrient composition of well-mixed waters from vertical structure 9 was closer to that of mixed waters than of well-mixed waters from vertical

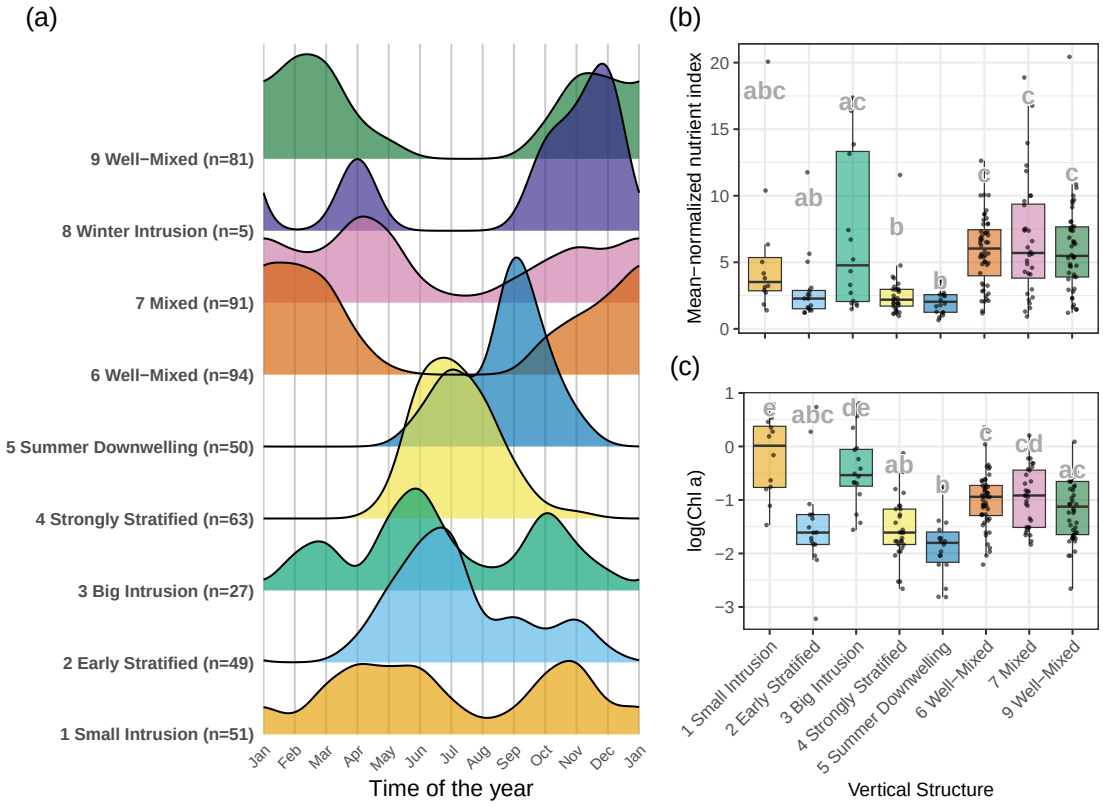


Figure 4. (a) Observed densities of each vertical structure throughout the year. The y-axis indicates the number of observations n in each vertical structure. (b) Box plot comparing mean-normalized nutrient index in surface waters among vertical structures; letters indicate significant differences between structures according to Dunn's post hoc test ($\alpha = 0.05$, Bonferroni adjustment). The mean-normalized nutrient index is defined as the sum of nutrient concentrations normalized by their respective means. (c) Box plot comparing log-transformed chlorophyll-*a* concentrations in surface waters among vertical structures; letters indicate significant differences between vertical structures based on Tukey's post hoc test ($\alpha = 0.05$). Because little data were available on winter intrusions, they were not included in the comparison.

structure 6 which were richer in NO_2^- (Appendix G3, G4). There was no high variability between vertical structures for either $Si(OH)_4$ or NH_4^+.

Log-transformed chlorophyll *a* concentration in surface waters differed significantly among clusters (one-way ANOVA, F = 13.93, df = 7, 213, $p < 0.001$). Chlorophyll *a* concentration was the highest during small intrusions, even though nutrients were not necessarily elevated compared to the other clusters. Chlorophyll *a* concentrations were the lowest in stratified conditions (early stratified, strongly stratified and summer downwelling) where nutrient concentrations were also the lowest.

3.3.2 Surface pico- and nanophytoplankton assemblages

The surface PicoNano assemblage data associated with each vertical structure were distinguished using DDA in winter–spring and summer–autumn. The first two DDA axes captured most of the differences among vertical structures, accounting for 92.2% of the overall separation in winter–spring and 77.2% in summer–autumn (Fig. 5).

During winter and spring we were able to compare the PicoNano assemblages of the two vertical structures of well-mixed waters, mixed waters and small intrusions. The two vertical structures identifying well-mixed waters (6 and 9) had very similar assemblages. These were characterized by a lower abundance of OraNano, as well as a lower SSC in RedNano. In contrast, the PicoNano assemblages of the mixed waters differed slightly from those of the two well-mixed waters vertical structures. The differences between well-mixed waters and mixed waters were mainly driven by the abundance of OraPicoProk and the SSC mean intensity of RedNano. Big intrusion events were characterized by higher SSC intensity of RedNano, greater abundance of OraNano, and lower abundance of OraPicoProk.

During summer and autumn, the assemblages in mixed and well-mixed waters were the opposite of those observed in strongly stratified waters and big intrusions. Strongly stratified waters and big intrusions were characterized by greater abundance of OraPicoProk and higher SSC intensity of RedNano, whereas mixed and well-mixed waters exhibited greater abundance of RedPicoProk and higher SSC intensity of RedPico. Early stratified waters and summer downwellings exhibited intermediate assemblages. Additionally, big intrusions, well-mixed waters, and early stratified waters showed greater abundance of RedPico.

3.3.3 Occurrence of vertical structures over the past three decades

For a more robust multinomial GAM analysis across the 27 years (1998–2024), the nine vertical structures were grouped into five categories. The resulting categories were Rhône intrusion, stratified water, well-mixed water, mixed water, and not categorized for non-identified vertical structures. Rhône intrusion included big intrusion, small intrusion, and winter intrusion. Stratified water included early stratified, strongly stratified, and summer downwelling. Mixed water only included mixed water. Well-mixed water included well-mixed water from vertical structure 6 and 9. The temporal evolution of these regrouped vertical structures was analysed using a multinomial GAM, with non-identified vertical structures used as the reference category. Based on a visual inspection of the temporal distribution, the proportion of non-identified vertical structures appeared to remain relatively constant over the study period. Relative to this reference, the proportion of Rhône intrusion varied significantly and non-linearly over time (df = 3.0, χ^2 = 19.5, p < 0.001), with a marked decrease between 2003 and 2004. Stratified conditions did not vary significantly over the study period (df = 2.9, χ^2 = 7.1, p = 0.061). In contrast, well-mixed and mixed waters exhibited opposite non-linear temporal dynamics, both significantly affected by time (well-mixed: df = 2.8, χ^2 = 16.9, p < 0.001; mixed: df = 3.0, χ^2 = 14.8, p = 0.002). While the relative proportion of well-mixed waters increased between 2003 and 2005, the proportion of mixed waters decreased. The proportion of well-mixed waters remained highest between 2004 and 2018 before slightly declining thereafter. These results were robust to the uncertainty threshold selected for model-based

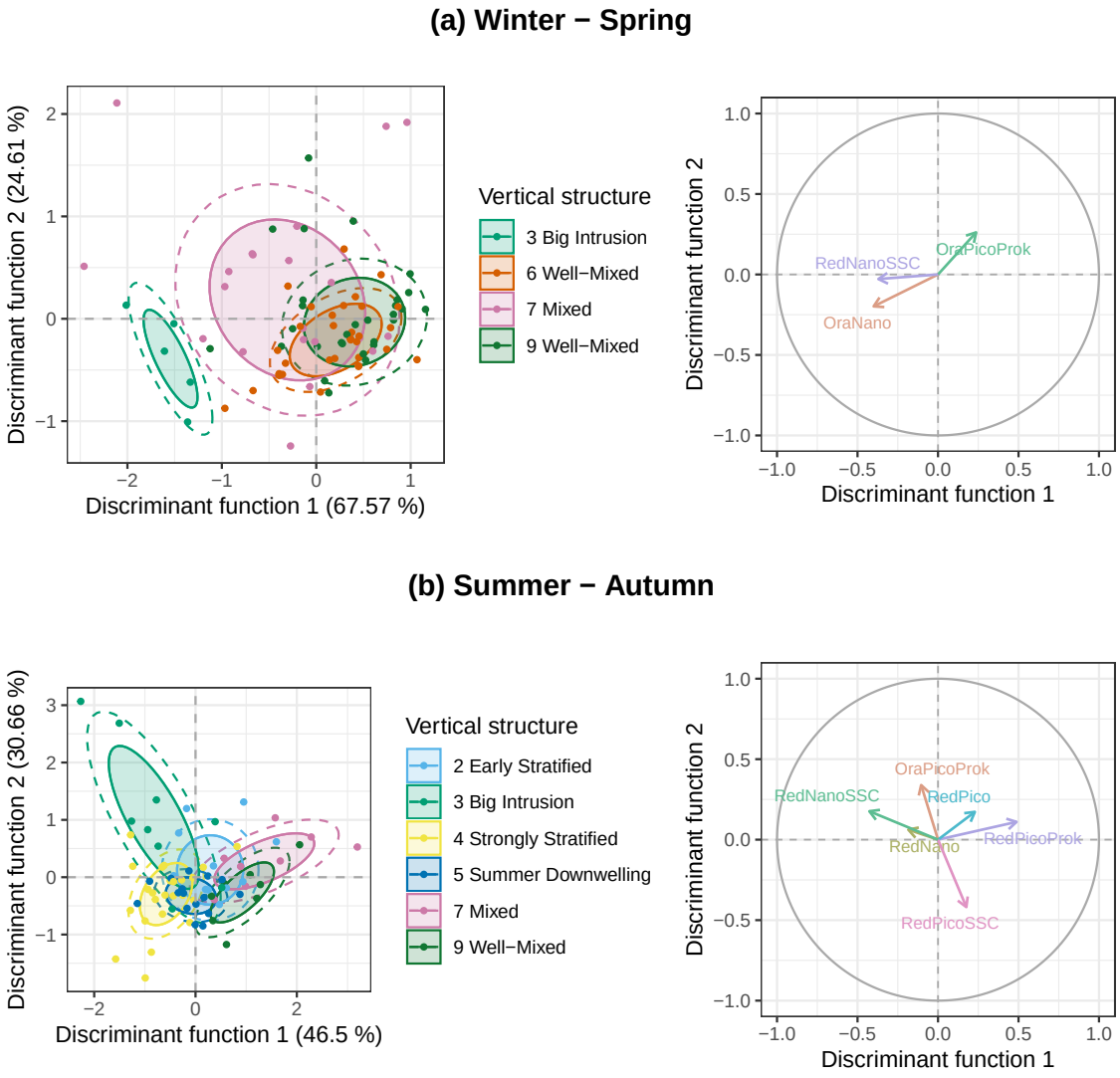


Figure 5. Discrimination of vertical thermohaline structures based on abundance and mean sideward light scatter (SSC) of pico- and nanophytoplankton flow cytometry (FCM) groups associated to each pair of temperature and salinity profiles measured on the same day. Winter–spring and summer–autumn were treated separately. (a) Descriptive discriminant analysis (DDA) for winter–spring; (b) DDA for summer–autumn. The plots on the left show the first discriminant plane of the DDA discriminating vertical structures, and the plots on the right show the corresponding correlation circles. Axis labels show the percentage of discriminant inertia explained. On the discriminant planes, solid lines show 50% normal data ellipses and dashed lines show 75% normal data ellipses. The common nomenclature is used for FCM groups (https://vocab.nerc.ac.uk/collection/F02/current/).

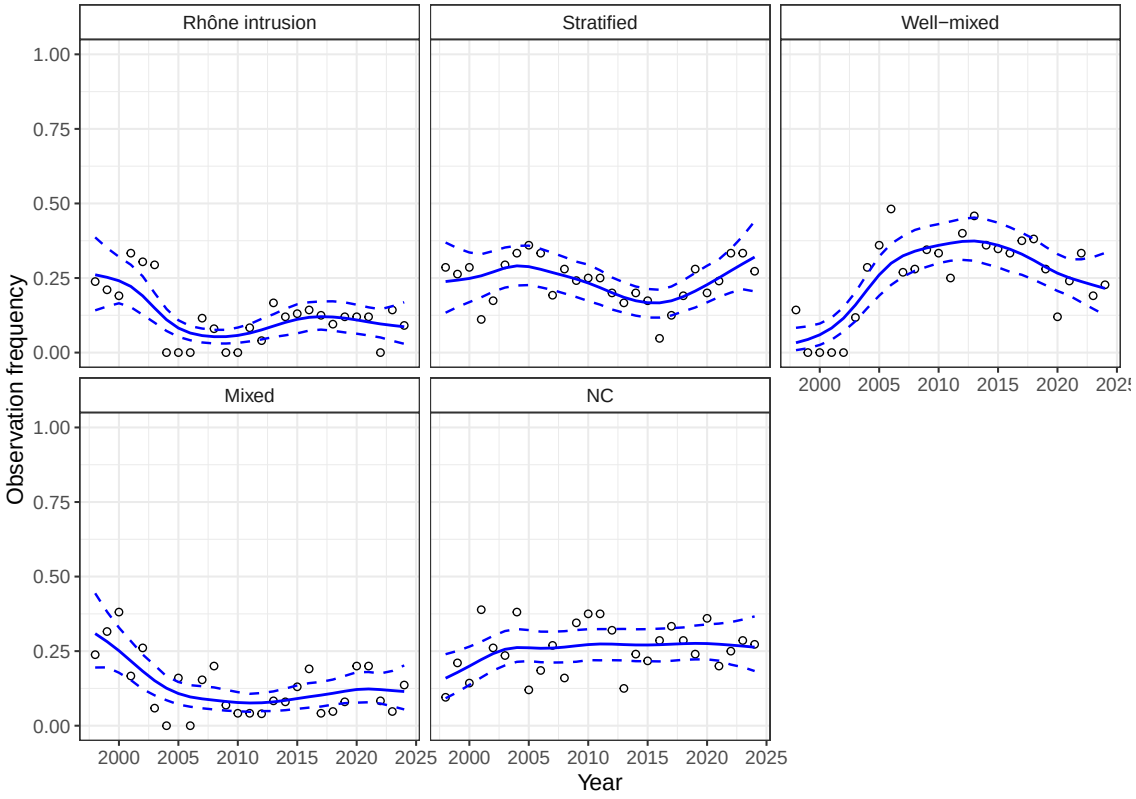


Figure 6. Evolution of the annual observation frequency of aggregated vertical structures. Solid blue curves represent the multinomial Generalized Additive Model (GAM) fitted to the observed annual frequencies of each vertical structure, while dashed curves indicate the corresponding 95% pointwise confidence intervals estimated with bootstrapping. Annual counts of observations ranged from $n = 17$ to $n = 29$ across the years studied. NC: not categorized vertical structure.

clustering (Appendix D1, D2). Based on the distribution of the vertical structures over the years, a predominance of well-mixed waters over mixed waters was mainly observed during winter.

As water mixing is closely related to wind, we examined the wind data to see if it could provide any useful information. Shifts in the dominant wind direction within the north-western sector were observed over the past three decades. A first shift occurred in 2003, from 350° to 340°, followed by a second shift in 2008, from 340° to 320° (Appendix H1). Well-mixed waters were generally associated with stronger north-westerly winds than mixed waters. To explore which wind direction was most strongly associated with well-mixed conditions, wind intensity measured before each observation was compared between the two vertical mixing regimes for wind directions of 320°, 330°, 340°, and 350°. Among these directions, winds from 330° showed the lowest p-value (one-sided Wilcoxon rank-sum test, W = 380, p = 0.031), suggesting a potential association between stronger winds from this direction and the occurrence of well-mixed waters. However, after correction for multiple comparisons using Hommel's method, this difference was no longer statistically significant.

3.4 Classification efficiency and reliability of vertical structures sampled at high frequency

To evaluate the potential of studying different vertical structures based on a high-frequency sampling, we first needed to assess the reliability of classifying truncated TS profiles that had been reconstructed. The vertical structures were predicted over 100 Monte Carlo runs based on the prediction of 150 SOMLIT TS profiles randomly truncated. The predicted vertical structures were then compared to the expected vertical structures of the TS profiles before alteration (Fig. 7).

Non-stratified vertical structures (6–9) were poorly identified. Generally, they were classified as either early stratified or mixed waters. Strongly stratified waters were the most accurately identified vertical structure, as they typically exhibited a deeper thermocline than the upper truncated portion. Almost 88% of strongly stratified waters were correctly classified. However, small intrusions, big intrusions, early stratified, summer downwellings and winter intrusions were sometimes misclassified as strongly stratified waters. Adding surface measurements considerably improved the classification of big intrusions. Nevertheless, well-mixed waters and mixed waters were still often misidentified as early stratified waters. Summer downwellings were correctly identified 41% of the time, but 34% were misidentified as big intrusions and 20% as strongly stratified. Approximately 88% of strongly stratified waters, 20% of summer downwellings, 4% of big intrusions, 16% of early stratified waters, 29% of small intrusions, and 45% of winter intrusions were found in strongly stratified waters. Winter intrusions were, however, rarely observed. The summer downwelling and small intrusions surface waters had nutrient concentrations comparable to those of strongly stratified waters: surface waters during summer downwellings were as nutrient-poor, while nutrient concentrations in small intrusions were not significantly different (see Sect. 3.3.1). Consequently, vertical profiles identified as strongly stratified mostly comprised nutrient-poor surface waters, including those initially classified as other vertical structures.

4 Discussion

This study demonstrates that vertical profiles of temperature and salinity provide an informative means of identifying environmental variations associated with nutrient availability and pico- and nanophytoplankton assemblage distribution at the surface. Once these relationships have been established from sufficiently long periods of concurrent physical, biogeochemical, and biological observations, historical physical observations can potentially be used to extend this interpretation to periods where direct biological and chemical observations are limited. By combining low-frequency TS profiles from the SNO SOMLIT with the high-frequency measurements from the SNO COAST-HF, we established a coherent dataset suitable for model-based clustering. Using multivariate functional principal component analysis, we treated vertical TS profiles as continuous functions of depth to identify the main variations in profile shape, which could then be interpreted visually (Fig. 3). Applying MBC to MFPCA scores enabled to distinguish recurring thermohaline structures associated with specific hydrodynamic conditions. These vertical structures consistently differentiated oligotrophic surface waters linked to a stratified water column paired with the absence of freshwater intrusions from richer surface waters associated with a mixed water column (Fig. 4). They also revealed characteristic PicoNano assemblages in surface waters across these regimes. Notably, even under comparable nutrient concentrations, the identified vertical structures captured changes in the assemblages driven by the hydrodynamic context (Fig. 5). Over the past 30 years, the frequency of observations of each hydrodynamic condition suggested a decrease in Rhône in-

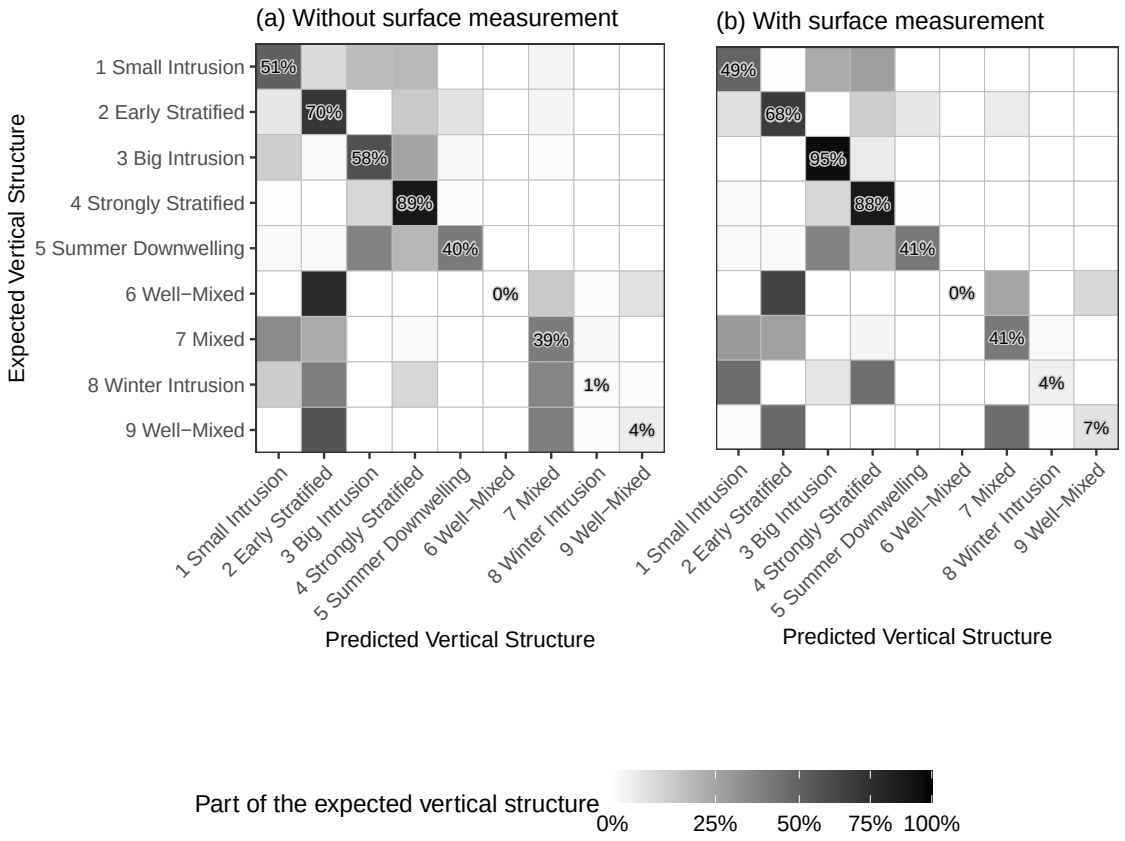


Figure 7. Row-normalized confusion matrices of the TS profiles predicted vertical structures after being randomly truncated at both ends, from 100 Monte Carlo simulations on 150 TS profiles. (a) TS profiles reconstructed without adding a surface measurement at 1 m. (b) TS profiles reconstructed after adding a surface measurement at 1 m.

trusions in the southern part of the BoM, as well as an increase in highly homogeneous water, favouring different PicoNano assemblages in winter (Fig. 6, Fig. 5). Together, these findings support the use of TS profiles as an effective proxy for hydrodynamic conditions and their biogeochemical consequences. Here, we assessed the relationships between vertical profiles and the biogeochemistry of surface waters. Nevertheless, the value of using entire vertical profiles in oceanography to link the physical properties of the water column and biogeochemistry has also been highlighted in Taylor and Henson (2026), which presents an example linking temperature and chlorophyll vertical profiles from Biogeochemical-Argo floats. Here, we show that the relationship between biology and vertical profiles characterizing water column structure can also be extended to cellular-level observations derived from flow cytometry data, even within the complex hydrodynamic environment of a coastal site.

4.1 Vertical structures identification and characterization

During summer and autumn, weak winds and surface warming gradually strengthen the stratification of the water column, while strong north-westerly winds and associated upwelling enhance mixing. Once winds calm down, stratification quickly reestablishes itself (Appendix I1). Early stratification, evidenced by a shallow thermocline, was associated with an intermediate surface picophytoplankton assemblage, between the higher relative abundance of *Synechococcus* (OraPicoProk) observed in strongly stratified waters and the higher relative abundances of *Prochlorococcus* (RedPicoProk) and phytoplanktonic picoeukaryotes (RedPico) observed in mixed and well-mixed waters (Fig. 5b). This gradual shift in the composition of the pico fraction was accompanied by an increase in the mean size of phytoplanktonic nanoeukaryotes (RedNano) as stratification became more pronounced and nutrient availability decreased. These contrasts among mixed, early stratified, and strongly stratified waters were mainly captured by the first DDA axis, which summarized 46.5% of the overall separation among vertical structures. These patterns align with typical Mediterranean stratified-season observations: *Synechococcus* dominate the surface, while *Prochlorococcus* and picoeukaryotes peak deeper in the water column (Partensky et al., 1999b; Latasa et al., 2010; Mella-Flores et al., 2011; Marrec et al., 2018; Mena et al., 2019; Vallina et al., 2023). The higher abundance of *Prochlorococcus* and picoeukaryotes in mixed waters may reflect the upward transport of populations from deeper waters into the surface layer. Alternatively, the higher nutrient availability associated with mixing may favour a rapid response of picoeukaryotes to nutrient inputs, as suggested by high-frequency observations (Thyssen et al., 2026); however, the short response timescales resolved in such observations cannot be directly compared with the temporal resolution of our dataset. Since nanoeukaryotes size does not vary with depth (Mena et al., 2019), their enlargement under strong stratification could be driven by reduced division under nutrient depletion (Latasa and Berdalet, 1994; Mouriño-Carballido et al., 2016). However, this is not accompanied by a decline in their abundance. On the contrary, nanoeukaryotes are larger and more abundant under stratified conditions than under mixed conditions. In mixed waters, where turbulence may enhance predator–prey encounter rates (Kiørboe, 1993; Prairie et al., 2012), grazing could counter faster growth and contribute to these patterns. Another possible interpretation is that mixing may favour different and smaller species of nanoeukaryotes. However, our data are not sufficient to determine whether changes in species composition underlie this pattern. Another notable point is that the shift between *Prochlorococcus* and *Synechococcus* may appear inconsistent with the idea that *Prochlorococcus* thrive in oligotrophic conditions (Partensky et al., 1999b; Zubkov et al., 2000; Johnson et al., 2006). However, high light and UV stress can negatively affect certain *Prochlorococcus* strains and ecotypes (Partensky et al., 1999b; Flombaum et al., 2013) and reduce their detectability by flow cytometry due to photoinhibition (Mella-Flores et al., 2011; Thyssen et al., 2022). Therefore, our data likely reflect only the low-light–adapted fraction of the *Prochlorococcus* population. Because high-light–adapted ecotypes dominate Mediterranean surface waters in summer and autumn (Mella-Flores et al., 2011), the apparent decline of detected *Prochlorococcus* in more oligotrophic conditions may correspond to the increase of undetected high-light–adapted cells with a very low amount of photosynthetic pigments, under the limit of sensitivity of flow cytometry. Accordingly, the patterns described here for *Prochlorococcus* should be interpreted as applying to the detectable low-light-adapted fraction rather than to the total *Prochlorococcus* population.

The use of vertical hydrological structures allowed us to distinguish different degrees of mixing intensity induced by strong north-westerly winds, which were associated with contrasting assemblages of pico- and nanophytoplankton. Model-based clustering identified two highly homogeneous structures and one slightly less homogeneous structure (Fig. 3). Assemblage similarity within each vertical structure was more closely related to vertical homogeneity than to nutrient concentrations (Fig. 5a, Appendix G3), indicating a potential role of turbulence or convective intensity in shaping these communities. Well-mixed waters appeared to be less favourable for large nanoeukaryotes and OraNano than slightly less homogeneous, but still mixed, waters. It has been suggested by previous studies that turbulence may negatively impact phytoplankton growth (Hondzo and Lyn, 1999; Berdalet et al., 2007; Llaveria et al., 2009). However, the species considered and the intensity of turbulence examined in these studies may differ from those in our system. Nonetheless, the smaller nanoeukaryotes observed in well-mixed waters in winter and spring is in accordance with the observation of the dominance of small nanoeukaryotes of the genus *Minidiscus* Hasle, 1973, during an intense convection event in the northwestern Mediterranean Sea (Leblanc et al., 2018). This genus is known to exist at the Frioul coastal site (Leblanc et al., 2018). A finer taxonomic resolution of pico- and nanoeukaryotes would help to better constrain the mechanisms underlying these patterns. Molecular techniques such as metabarcoding provide valuable approaches for better resolving the taxonomy of these small size fractions (Abad et al., 2016; Hubert et al., 2025; Lescot et al., 2026).

Under stratified conditions, downwelling events deepen the thermocline and generate a thick, vertically homogeneous surface layer of warmer water. Such events can strongly impact benthic communities by transporting warm surface waters downward (Estournel et al., 2025). Although strongly stratified waters and summer downwelling boundaries were partly blurred, strongly stratified waters were generally characterized by warmer surface temperatures and a shallower thermocline than summer downwellings (Fig. 3). Downwellings are typically induced by southeasterly winds, which are rare in summer but become more frequent in autumn (Millot, 1990). We observed that downwellings in summer did not significantly alter surface nutrient concentrations. Pico- and nanophytoplankton assemblages associated with downwelling conditions closely resembled those observed during early stratification, but were characterized by fewer and larger picoeukaryotes (Fig. 5b). As discussed above, this may reflect a negative effect of increased turbulence on picoeukaryote growth, resulting in limited division and increased cell size (Berdalet et al., 2007), as southeasterly winds are known to generate strong swells (Millot, 1990). However, in our dataset, strong southeasterly winds were not clearly identified during the summer downwelling events (Appendix G2). This may have been due to wind attenuation at the Marignane meteorological station caused by shielding from the Côte Bleue massif. Offshore pelagic observations from the SNO MOOSE (Mediterranean Ocean Observing System for the Environment, https://www.moose-network.fr/) could provide complementary information on the potential contribution of offshore-to-coastal advection to the observed coastal assemblages.

Three vertical structures associated with freshwater intrusions were identified. The marked salinity decrease between surface and bottom waters (generally >0.4) indicated that these structures were representative of Rhône River intrusions, given that bottom salinity at the station remained close to 38.15 and that the Rhône dilution area is typically characterized by salinities below 37.8 (Gatti et al., 2006). Following the classification of Fraysse et al. (2014), who described small, big, and short-lived intrusions, we distinguished small and big intrusions based on the magnitude of the salinity decrease, in agreement with their

observations. Big intrusions were generally associated with nutrient-enriched surface waters and elevated chl *a* concentration (Fig. 4). Our small intrusions occurred when the water column was barely stratified, in contrast to Fraysse et al. (2014) who found that small intrusions were linked to the presence of a shallow thermocline. Which may suggest that the vertical structures identified here may represent different stages of an intrusion rather than distinct intrusion types. The small intrusions we observed could correspond to big intrusions undergoing destruction by strong south-easterly winds that enhance mixing. However, these winds may be underestimated at the Marignane station, as discussed above. Even so, this interpretation is consistent with the lower nutrient concentrations but higher chl *a* concentration we observed during small intrusions, potentially indicating a bloom initiated during a big intrusion (Para et al., 2010; Fraysse et al., 2014) (Fig. 4). During summer–autumn, big intrusions favoured higher abundances of *Synechococcus* and small picoeukaryotes, as well as larger nanoeukaryotes (Fig. 5b), consistent with previous observations linking *Synechococcus* and picoeukaryotes to mesotrophic conditions (Campbell and Vaulotaf, 1993; Partensky et al., 1999a; Zubkov et al., 2000). Compared to well-mixed conditions, the reduced salinity, increased turbidity within the Rhône plume, and weaker vertical mixing, likely contributed to the distinct assemblages observed. The general characteristics of pico– and nanophytoplankton assemblages during intrusions were consistent across seasons, except concerning *Synechococcus*. Compared to big intrusions, well-mixed waters were more favourable to *Synechococcus* in winter and spring than in summer and autumn. Too few observations were available to characterize small intrusions.

We were unable to identify Northern Current intrusions in the BoM. Indeed, temperature and salinity profiles alone are often insufficient, as the current lacks a constant thermohaline contrast with coastal waters and exhibits high intrinsic variability (Gatti, 2008). To better characterize the Northern Current intrusions and their influence on the PicoNano assemblages, it would be necessary to include an Acoustic Doppler Current Profiler (ADCP) in the Bay or to make use of modelling.

Overall, our results suggested that a high abundance of *Prochlorococcus* low-light adapted and picoeukaryotes at the surface in summer was mainly attributable to vertical mixing, consistent with their low tolerance to light (Zubkov et al., 2000). Nanoeukaryotes and picoeukaryotes appeared sensitive to nutrient variations and the degree of mixing of the water column. However, with our data, we were unable to establish whether it affected their growth rate or taxonomic composition. OraNano abundance was correlated with nanoeukaryotes size in winter and spring, suggesting that the conditions favouring larger nanoeukaryotes also benefited phycoerythrin-rich nanoeukaryotes (e.g. cryptophytes) (Fig. 5). The study of the PicoNano assemblages in the context of the hydrological vertical structures enabled us to consider both the physical and chemical background. The consistency of the observed assemblages with the literature supports the reliability of the vertical structures we identified as indicators of the hydrodynamic conditions at play and their impact on pico- and nanophytoplankton assemblages. However, some vertical structures partly overlapped, and although cluster assignments were associated with relatively low uncertainty (<0.25), this uncertainty should be kept in mind. Although the biweekly sampling can capture some signatures of short-lived events such as Rhône River intrusions and upwelling that can induce mixed water-column conditions during summer, the time elapsed between the onset of these events and our observations is unknown. Therefore, their effects on phytoplankton assemblage composition cannot be fully characterized from our dataset. Resolving the short-term biological and biogeochemical responses to such events would require higher-frequency sampling, ideally combined with a Lagrangian approach to track

water-mass evolution and the integration of satellited data (e.g. MODIS, Sentinel-3, and PACE) to better capture these transient but ecologically important phenomena (Fuchs et al., 2023; Thyssen et al., 2026).

4.2 Reliability of vertical structures identification from high-frequency profiles

Vertical profiles measured automatically by the profiler installed on a mooring under the SOLEMIO buoy in the BoM are truncated due to installation constraints. These profiles are supplemented by a surface measurement from another sensor installed on the mooring at a depth of 1m. However, one or both instruments may occasionally malfunction, undergo maintenance, or be dismantled in case of bad weather conditions during winter. These situations can result in missing or truncated data, which are common issues in high-frequency monitoring (Schmitt and Lefebvre, 2016). To deal with the truncated profiles, we used a reconstruction method based on functional data analysis to predict the vertical profiles along the entire water column (1–50 m deep). Following the approach of Fonvieille et al., we tried to strengthen the prediction by conditioning it with time and by using the information held by both temperature and salinity profiles simultaneously. However, the complexification of the method did not increase the prediction significantly (Table 2). If the time conditioning may have produced better results for the temperature profiles because of the seasonality of the temperature, it was not the case for salinity. The latter being mainly impacted by Rhône intrusions that are not correlated with the day of the year. Furthermore, the multivariate approach was not better suited, given that the cross-covariance between temperature and salinity was negligible. This is not surprising in a coastal environment, where continental fresh water inputs and wind mixing are the main drivers of the vertical structure of salinity and temperature. While adding environmental variables should theoretically refine our results, the high variability of the coastal zone combined with our limited number of observations meant that these complex models were more sensitive to local noise, ultimately making them less reliable than a simpler, more general approach.

The predicted truncated profiles were consistent with the observed profiles in the BoM (Fig. 2) and allowed us to build a relatively large dataset. However, if we used them for the clustering of the TS profiles, we did not include them in the comparisons between vertical structures or in temporal analysis. The reconstruction method mostly produced reliable predictions but were strongly biased in rare cases (5% of cases), particularly affecting profiles with shallow stratification (Appendix F1). The biases induced by the reconstruction introduced uncertainty into the classification of the high-frequency profiles. Consequently, we needed to evaluate how the error in prediction impacted the classification of the TS profiles for high-frequency applications. Our results revealed that vertical profiles collected at high frequency could be used to identify strongly stratified waters, keeping in mind that they may include summer downwellings and small intrusions. As such, they could be used as a reliable indicator of oligotrophic waters associated with enhanced stratification (Sect 3.4). The measurements by the surface probe allowed the correct classification of most of the big intrusions (95%) while only half were correctly classified without them. For other vertical structures, the misclassification rate being higher than 50% may be too elevated to draw any reliable conclusion (Figure 7). High-frequency vertical profiles could be a useful way to record the onset of stratification each year. Combined with wind data, these profiles could be used to assess the necessary wind conditions to offset strong stratification. These data could provide insight into the duration of the most oligotrophic conditions in the BoM, which were encountered when the water column was strongly stratified or during summer downwellings (Fig. 4). This is particularly relevant since

these oligotrophic conditions are expected to become more frequent due to enhanced stratification linked to climate change (Moore et al., 2018). To fully exploit these data, a consistent high-frequency time series covering the period from the beginning of March to the end of November is needed, as this is the period when the water column can become highly stratified (Fig. 4). However, gaps in the time series, caused for example by instrument malfunction, or removal, currently prevent a continuous record from being obtained throughout this period. Maintaining continuous high-frequency observations from March to November over several consecutive years would therefore substantially improve the potential of these data to characterize the dynamics of thermal stratification.

4.3 Environmental changes in the Bay of Marseille over the past three decades and implications on nutrient concentrations and pico- and nanophytoplankton

The occurrence of the different vertical structures over the past three decades (1998–2024) was assessed using only data from the bi-weekly observations of the SNO SOMLIT. In accordance with Feuilloley et al. (2020), vertical structures corresponding to a stratified water column did not show any significant trend over the past 27 years (Fig. 6). However, the sampling frequency (every two weeks) used by the SNO SOMLIT may be insufficient to detect any long-term trends regarding the stratification.

The number of Rhône intrusions observed at the Frioul site in the BoM significantly decreased at the beginning of the century. The number of intrusions observed each year from 2004 onwards was half the number observed in the preceding years (Fig. 6). We could not determine whether this was due to a decrease in the frequency, spread or duration of the intrusions, or a combination of these factors. To our knowledge, this decrease has not been documented before. However, this is in accordance with the oligotrophication observed in the BoM (Lheureux et al., 2023; Garcia et al., 2023). The impoverishment of the Rhône River along with the fewer or shorter intrusions since 2004 surely contributed to the decreasing trend in nutrients concentration (Feuilloley et al., 2020; Lheureux et al., 2023). This decrease in Rhône intrusions is concomitant with the decrease in the occurrence of mixed water column and the increase in occurrence of well-mixed water column. These observations coincided with a shift in the dominant wind direction within the north-western sector, from 350° to 340° in 2004 (Appendix H1). A potential tendency was observed for well-mixed conditions to occur under relatively stronger winds from 330° compared with mixed conditions. The increased occurrence of well-mixed conditions, particularly during winter and spring, may therefore be related to a westward shift in the dominant wind direction. However, confirming this relationship would require dedicated modelling studies to further investigate how wind direction modulates vertical mixing intensity. Between 2008 and 2018, the recurrent predominance of winds from 330° was consistent with the increased occurrence of well-mixed waters. Coupling modelling approaches with long-term observational data could further improve our understanding of the mechanisms driving changes in vertical water-column structure in the Bay of Marseille. The fewer Rhône intrusions and the prevalence of well-mixed waters suggest that the conditions associated with an increase in the mean size of nanoeukaryotes may have become less frequent. Interestingly, Garcia et al. (2024) independently reported a decrease in the lipid content in zooplankton 300-500 µm size fraction, mainly composed by copepods, since 2009. On the one hand, a phytoplankton-based diet is known to be associated to a higher lipid content in copepods (Lee et al., 2006). On the other hand, nanoplankton are believed to be the main contributors to the zooplanktonic biomass in the northwestern Mediterranean Sea in winter (Hunt et al., 2017). Copepods may

feed on nanophytoplankton through direct herbivory or on heterotrophic nanoflagellates (Sommer et al., 2002). However, most copepods do not efficiently feed on the smallest particles (Sommer et al., 2002). As a result, the smallest nanophytoplankton may not be directly accessible to them. One possible hypothesis is that the suspected increased prevalence of small eukaryotic nanophytoplankton in winter may have reduced the contribution of larger nanophytoplankton to copepod diets, potentially resulting in a relatively greater contribution of heterotrophic nanoflagellates to their diet. This could have contributed to the observed decrease in their lipid content. The hypothesis of a shift from autotrophic prey to heterotrophic and mixotrophic prey has already been suggested by Garcia et al. (2024). If this hypothesis is correct, such changes could also have affected the quality of the favourite preys ingested by *S. pilchardus* which are copepods (Chen et al., 2021; Nikolioudakis et al., 2012), leading to the decrease in the body condition of these small pelagic fish for which the BoM is a reproduction and fishing area (Pichot et al., 1978; Lee, 1961; Queiros et al., 2021; Garcia et al., 2024).

If the physical conditions induced by the intensity of the vertical mixing are responsible for the PicoNano assemblages we observed, these are more likely to stay consistent over time. However, the PicoNano assemblages that we observed during the Rhône intrusions may be more influenced by nutrient input. Thus, as the nutrient composition of the Rhône River has changed over the past thirty years (Feuilloley et al., 2020; Lheureux et al., 2023), the PicoNano assemblages observed during the last twelve years may not be representative of the assemblages that we would have observed thirty years back.

5 Conclusions

In the context of climate change, understanding ecosystem responses to environmental variability is increasingly critical, motivating intensified efforts to document ecosystems across spatial and temporal scales (Sunagawa et al., 2020). High-frequency ocean monitoring, however, requires automation, data management, maintenance, resilient equipment, and sufficient funding (Glenn et al., 2000), meaning many Essential Ocean Variables (EOVs) cannot yet be surveyed at high frequency. Consequently, ship-based time series remain crucial to monitor ecosystem changes (von Jackowski, 2024). Long-term studies demand consistent, high-quality monitoring, but maintaining certain variables at the same level of quality is challenging. Observations may vary due to methods, instruments, calibration, or human operators, necessitating rigorous quality control (Strayer et al., 1986; Taverniers et al., 2004; von Jackowski, 2024). Therefore, some time series may not be fully comparable over their entire duration (Strayer et al., 1986; Glenn et al., 2000; von Jackowski, 2024).

To address these challenges, a common strategy is to monitor variables that are easy to maintain and rapidly analysable, forming a core time series supplemented by short- to medium-term studies (Strayer et al., 1986; Wiebe et al., 1987). This approach allows identification of recurring environmental conditions and evaluation of their impacts on biological communities. In this study, we demonstrated that vertical profiles of temperature and salinity effectively indicate hydrodynamic conditions influencing pico- and nanophytoplankton assemblages in a northwestern Mediterranean coastal site, where nutrient availability is primarily driven by wind mixing and freshwater inputs. Our results suggest potential changes in the nanoeukaryote functional group, consistent with shifts observed in zooplankton and small pelagic fish over the past 30 years.

This study highlights the value of monitoring variables that are simple to measure and maintain over time, such as temperature and salinity, to build reliable long-term time series. It also underscores the challenges of sustaining high-quality, high-frequency observations over extended periods. Overall, these findings demonstrate that long-term scientific insight depends more on the consistency and reliability of well-curated time series than on data volume alone.

Code and data availability. Physico-chemical and flow cytometry data of the SNO SOMLIT for the Frioul coastal site are accessible from the SEANOE portal (Savoye et al., 2026). The SNO SOMLIT data used in this study can be explored through the SOMLITMed Explorer Shiny application (Couteyen Carpaye, 2026b). Météo-France SYNOP data are accessible from the French public data catalogue (Météo-France, 2026). High-frequency surface temperature and salinity data, as well as vertical profiles, are available from the SEANOE portal (Garcia et al., 2025). The R script implementing the vertical profile reconstruction method used in this study is publicly available (Couteyen Carpaye, 2026a). The following R packages were used : dplyr (Wickham et al., 2023), ggplot2 (Wickham, 2016), fda (Ramsay, 2024), VGAM (Yee, 2010), tree (Ripley, 2025), mclust (Scrucca et al., 2023), rstatix (Kassambara, 2023), lubridate (Grolemund and Wickham, 2011), shiny (Chang et al., 2024), circular (Agostinelli and Lund, 2025)

Appendix A: Study area

The study area is the Bay of Marseille, located in the northwestern Mediterranean Sea. Figure A1 indicates the locations of the SNO SOMLIT and the SNO COAST-HF monitoring sites and the Météo-France meteorological station.

Appendix B: Flow cytometry gating strategy

The flow cytometry gating strategy used by the SNO SOMLIT is described in Figure B1.

Appendix C: Methodological workflow for vertical profile reconstruction

The spline smoothing step requires the setting of the number of basis functions and a smoothing parameter, which is described in Sect 2.4.2. The PACE estimation step requires an estimation of the variance parameter σ^2, which controls the balance between fidelity to the observed data and smoothness of the reconstructed profiles. The choice of σ^2 is described below (Sect C1).

The vertical profile reconstruction method follows 6 steps:

1. Vertical profiles are split in two groups : complete (first observation above 1.5 m deep and last observation below 49.5 m deep) and incomplete.
2. Complete profiles are smoothed over the 1–50 m depth range using penalized B-splines.
3. Functional principal component analysis (FPCA) is applied to the complete smoothed profiles.



Figure A1. Map of the study area, the Bay of Marseille, located in the Gulf of Lions in the northwestern Mediterranean Sea, showing the monitoring sites. Land cover data were obtained from CORINE Land Cover (CLC). Bathymetry data were obtained from GEBCO 2022, elevation contours from IGN, and river data from BD Carthage.

4. The mean values μ and the covariance matrix are estimated from the coefficients of the complete profiles. Eigenvalues and associated eigenfunctions (λ_k, ϕ_k) are estimated from the spectral decomposition of that covariance matrix.

5. For each incomplete profile i, the functional principal component (FPC) scores $\xi_{ik}, k = 1, \ldots, K$ associated with the first K eigenfunctions are estimated using PACE (Yao et al., 2005) assuming the variance σ^2 of the measurement errors is known (see below).

6. Each incomplete profile i is reconstructed over the depth interval $t \in [1, 50]$ m as

$$\hat{X}_i^K(t) = \mu(t) + \Sigma_{k=1}^K \xi_{ik} \phi_k(t)$$

C1 Variance parameter σ^2 selection

In PACE, the variance σ^2 of the measurement errors controls the balance between fidelity to the observations and smoothness of the reconstructed profiles. Small values of σ^2 lead to reconstructions that closely follow the observations, whereas larger

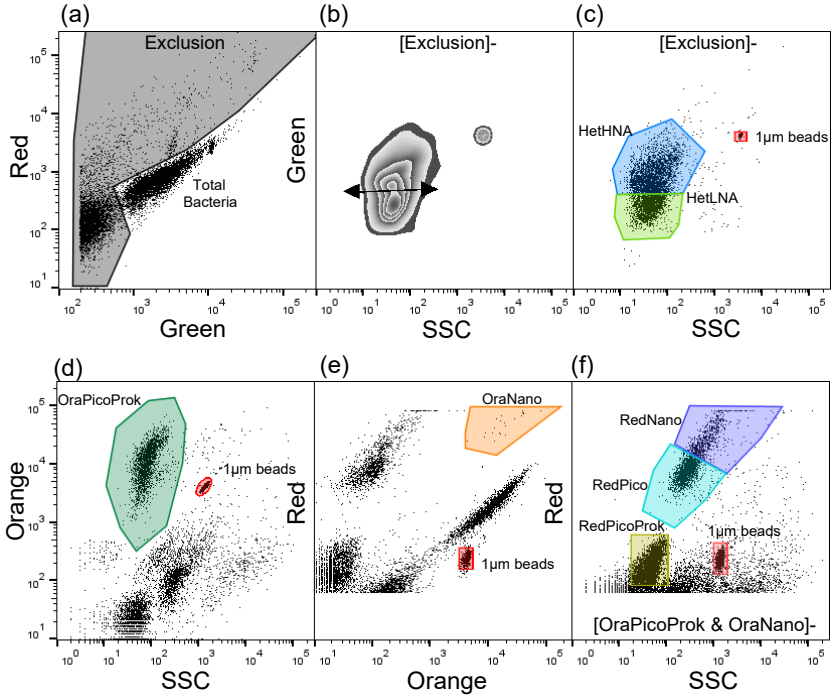


Figure B1. Flow cytometry gating strategy for SNO SOMLIT analysis. The top panels show the optical resolution of heterotrophic prokaryotes (hereafter referred as bacteria) after staining of their nucleic acid using the Sybrgreen dye: (a) Exclusion gate removes red-fluorescent cells (autotrophs) to isolate heterotrophic bacterial populations. (b) Zebra plot displaying two bacterial groups distinguished by their green fluorescence intensity (HetLNA: low nucleic acid content and HetHNA: high nucleic acid content). The zebra plot facilitates the visualisation of the boundary between the two groups. A zebra plot is a bivariate plot type specific to FlowJo (version 10 onward), combining features of contour and density plots. (c) Same data as in (b), displayed as a dot plot, with each dot representing a cell. The bottom panels show the optical resolution of pico- and nanophytoplankton based on the autofluorescence only (no staining): (d) *Synechococcus sp.* (OraPicoProk) (with orange fluorescence induced by phycoerythrin). (e) *Cryptophyceae* (OraNano) (dual red and orange fluorescence from chlorophyll a and phycoerythrin). OraPicoProk and OraNano are quantified separately and excluded from the red/SSC biparametric plot. (f) Pico- and nanoeukaryotic phytoplankton and *Prochlorococcus* are distinguished based on their scatter and red fluorescence intensities. *Prochlorococcus* are very small and very dim, sometimes difficult to separate from the background noise. Normalization: 1 µm reference beads were added to all samples to standardize fluorescence and scatter intensities and monitor instrument performance over time.

values produce smoother reconstructions but may underfit the data. Yao et al. (2005) proposed an estimator for σ^2. However, in our dataset, this estimator did not minimize the reconstruction error.

To select σ^2, simulations were performed by degrading 20% of the complete profiles, following the procedure described in Sect. 2.4.3. Predictions of the degraded profiles over the depth interval $t \in [1, 50]$ m were computed using the $\hat{\sigma}^2$ estimator from Yao et al. (2005), as well as a range of σ^2 values. The mean squared error between observed and reconstructed profiles was then computed for temperature and salinity, for each σ^2 value, over 20 Monte Carlo runs.

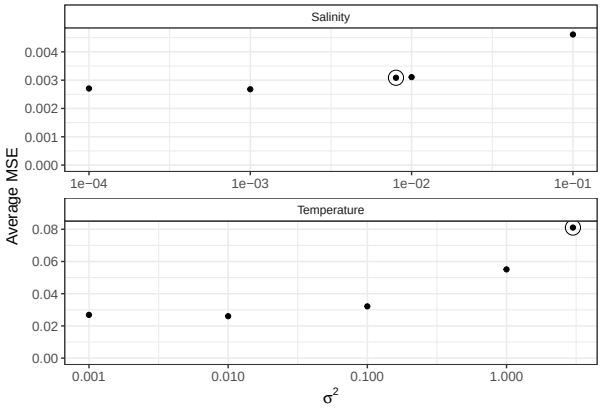


Figure C1. Average mean squared error (MSE) for 20 Monte Carlo Runs with $n = 150$ randomly truncated vertical profiles of temperature and salinity, for different values of the variance parameter σ^2 used in the predictions with principal component analysis through conditional expectation (PACE). Circles indicate the average MSE obtained with the σ^2 computed with the estimation of Yao et al. (2005). The X-axis uses alog-scale.

The lowest average MSE values were obtained for $\sigma_T^2 = 0.01$ for temperature and $\sigma_S^2 = 0.001$ for salinity (Fig. C1). However, to account for the instrumental accuracy of salinity measurements (0.1), we adopted a uniform $\sigma^2 = 0.01$ for both variables. This value aligns closely with the $\hat{\sigma}_S^2$ estimate proposed by Yao et al. (2005) and it provides a way to balance data fidelity and profile smoothness while considering instrumental constraints.

Appendix D: Multinomial Generalized Additive Model

For each year t, let

$$\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{tK})$$

denote the vector of observed counts for the $K = 5$ aggregated vertical structure categories (Rhône intrusion, stratified water, well-mixed water, mixed water, and not categorized, NC), with total count

$$N_t = \sum_{j=1}^{K} Y_{tj}$$

We assumed that $\mathbf{Y}_t$ followed a multinomial distribution:

$$\mathbf{Y}_t \sim \text{Multinomial}(N_t, \boldsymbol{\pi}(t))$$

where $\boldsymbol{\pi}(t) = (\pi_1(t), \ldots, \pi_K(t))$ represents the vector of category probabilities at year t.

To model temporal variation in category probabilities, we used a baseline-category logit formulation. The NC category ($j = K$) was chosen as the reference category. For each non-reference category $j = 1, \dots K-1$, the log-odds relative to the reference were modelled as smooth functions of time:

$$log\left(\frac{\pi_j(t)}{\pi_K(t)}\right) = \beta_{0j} + f_j(t)$$

The functions $f_j(t)$ were represented using cubic smoothing splines:

$$f_j(t) = \sum_{m=1}^{M} \beta_{jm} B_m(t)$$

where $B_m(t)$ are spline basis functions and $M = 4$ denotes the basis dimension. The vector smoother treats the $K-1$ log-odds as a multivariate response and fits the spline jointly.

Parameters are estimated by maximizing the penalized multinomial log-likelihood:

$$\ell(\boldsymbol{\beta}) - \sum_{j=1}^{K-1} \lambda_j \int [f_j''(t)]^2 dt$$

where $\ell(\boldsymbol{\beta})$ is the multinomial log-likelihood and λ_j are smoothing parameters controlling the trade-off between fit and smoothness (Yee and Wild, 1996).

Since the composition of the vertical structures depended on the uncertainty threshold used for model-based clustering, the multinomial GAM was fitted using different uncertainty thresholds (0.2, 0.25, 0.3 and 0.4). The sensitivity analysis showed that the estimated temporal patterns and the statistical significance of the GAM smooth terms were robust across the range of uncertainty thresholds tested. The only exception was the temporal trend in the occurrence of stratified waters, whose statistical significance varied with the selected threshold.

Appendix E: Vertical centred profiles

Each vertical profile i is represented as a linear combination of basis functions,

$$f_i(t) = \boldsymbol{\alpha}_i^\top \boldsymbol{\phi}(t), \qquad t \in [1{,}50]\ \mathrm{m}.$$

Each profile i was centred by subtracting its mean, $\bar{f}_i$, as follows

$$f_i(t) - \bar{f}_i = (\boldsymbol{\alpha}_i - \boldsymbol{\mu})^\top \boldsymbol{\phi}(t),$$

where the profile mean is defined as

$$\bar{f}_i = \frac{1}{50-1} \int_1^{50} f_i(t)\,\mathrm{d}t,$$

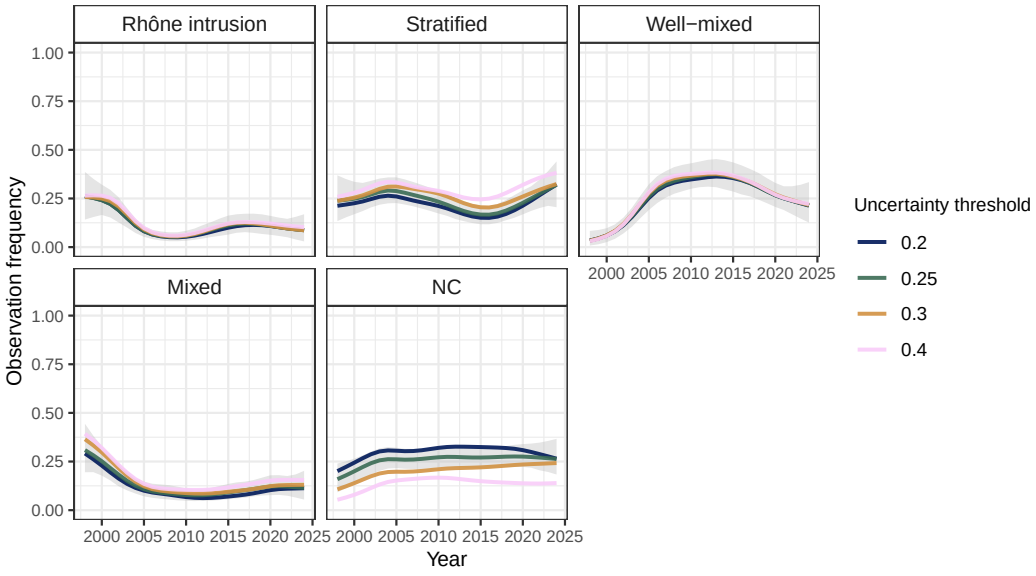


Figure D1. Evolution of the annual observation frequency of aggregated vertical structures obtained from model-based clustering using different uncertainty thresholds. Solid curves represent the multinomial Generalized Additive Models (GAMs) fitted to the observed annual frequencies of each vertical structure for each uncertainty threshold. For each vertical structure, the grey ribbon represents the 95% pointwise confidence interval estimated by bootstrapping for the uncertainty threshold of 0.25 used in the study.

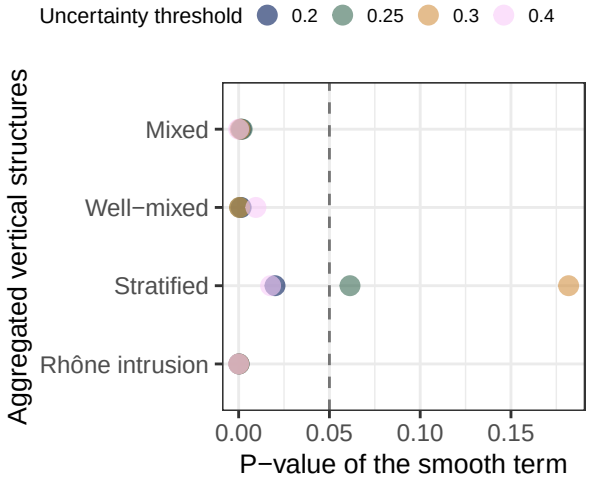


Figure D2. P-values of the (χ^2) tests for the smooth terms of the multinomial Generalized Additive Models (GAMs) fitted to the observed annual frequencies of each aggregated vertical structure obtained from model-based clustering, using different uncertainty thresholds.

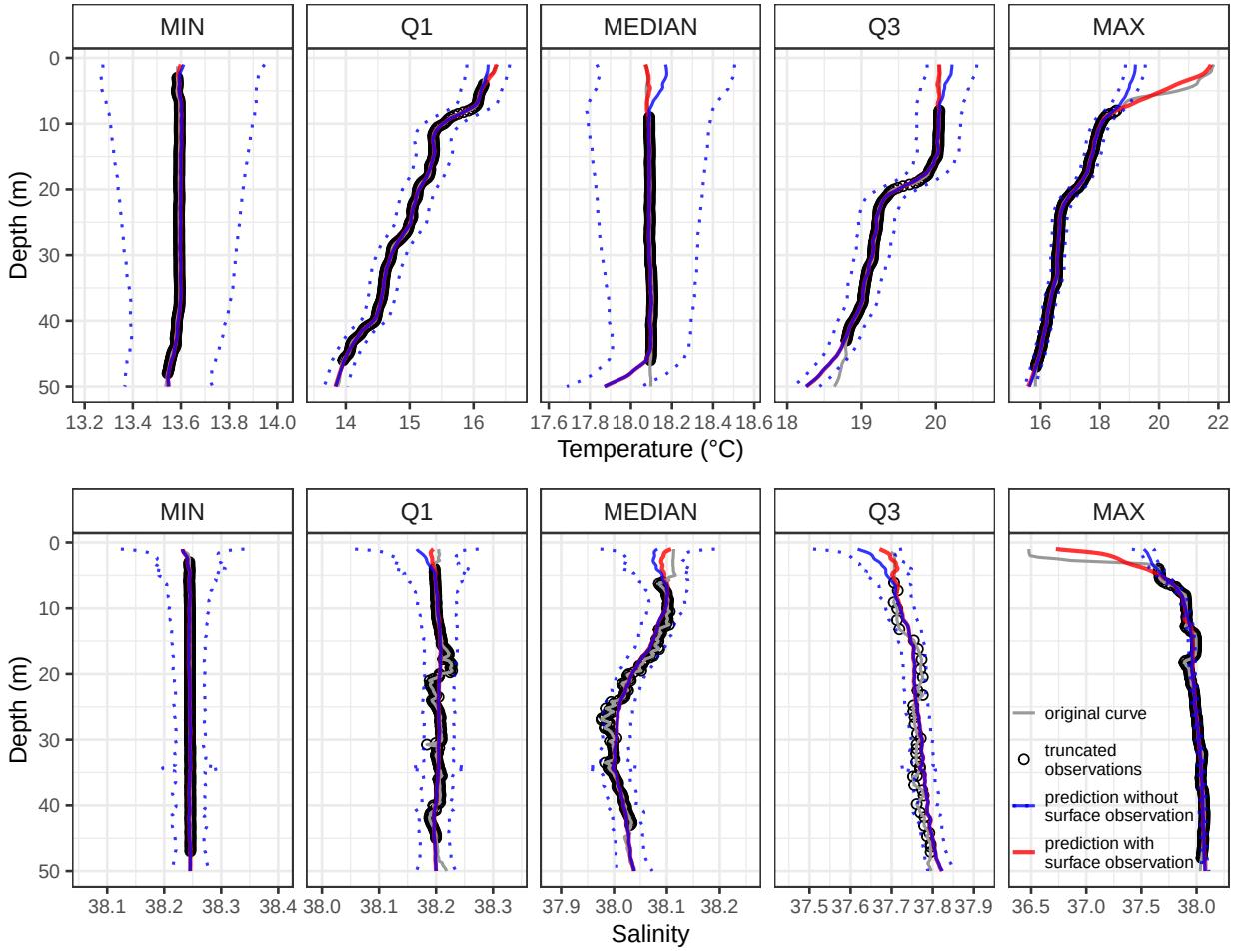


Figure F1. Predictions of truncated SOMLIT vertical profiles with principal component analysis through conditional expectation (PACE, blue solid line) and 95% asymptotic pointwise confidence intervals (dotted line) associated with minimal (MIN), maximal (MAX) and quantiles 0.25 (Q1), 0.5 (MEDIAN), 0.75 (Q3) of mean squared error (MSE) computed for 152 predicted profiles. Also shown are original complete curves (grey solid line), truncated observations (black dots) and predictions with additional surface observation (red solid line).

and $\boldsymbol{\mu}$ denotes a constant vector with entry $\bar{f}_i$.

Appendix F: Simulation of the reconstruction of degraded SOMLIT vertical profiles

The SOMLIT dataset was randomly degraded to evaluate the efficiency of the reconstruction method described in Appendix C. Twenty per cent of the profiles were randomly truncated between 3 and 10 m at the surface and 40 and 48 m at the bottom. The remaining 80% were left intact. Figure F1 shows the temperature and salinity profiles associated to the best and worst predictions and the profiles associated to the quantiles 0.25, 0.5, 0.75 of the MSE.

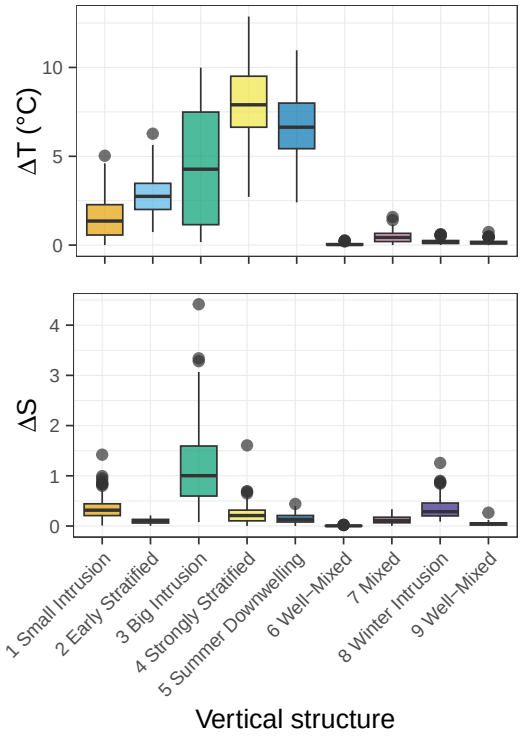


Figure G1. Box plots comparing temperature (top, ΔT) and salinity (bottom, ΔS) differences between 1 m and 50 m deep among the vertical structures. The central line of the box plots indicates the median, and the lower and upper hinges correspond to the first and third quartiles. Whiskers extend to the most extreme values within 1.5×IQR (interquartile range) from the hinges, while points beyond are plotted individually as outliers.

Appendix G: Vertical structure comparisons

Nine vertical structures were identified using MBC based on temperature and salinity vertical profiles. To identify the hydrodynamic conditions associated with these structures and to compare them with one another, the differences in temperature and salinity between the surface and the bottom were computed (Fig. G1). Winds blowing between 12:00 am and 2:00 pm on the sampling date were compared using wind roses (Fig. G2). Nutrient concentrations were compared using descriptive discriminant analysis (Fig. G3) and are also presented as box plots (Fig. G4). The following figures are referred to in Sect 3.2.

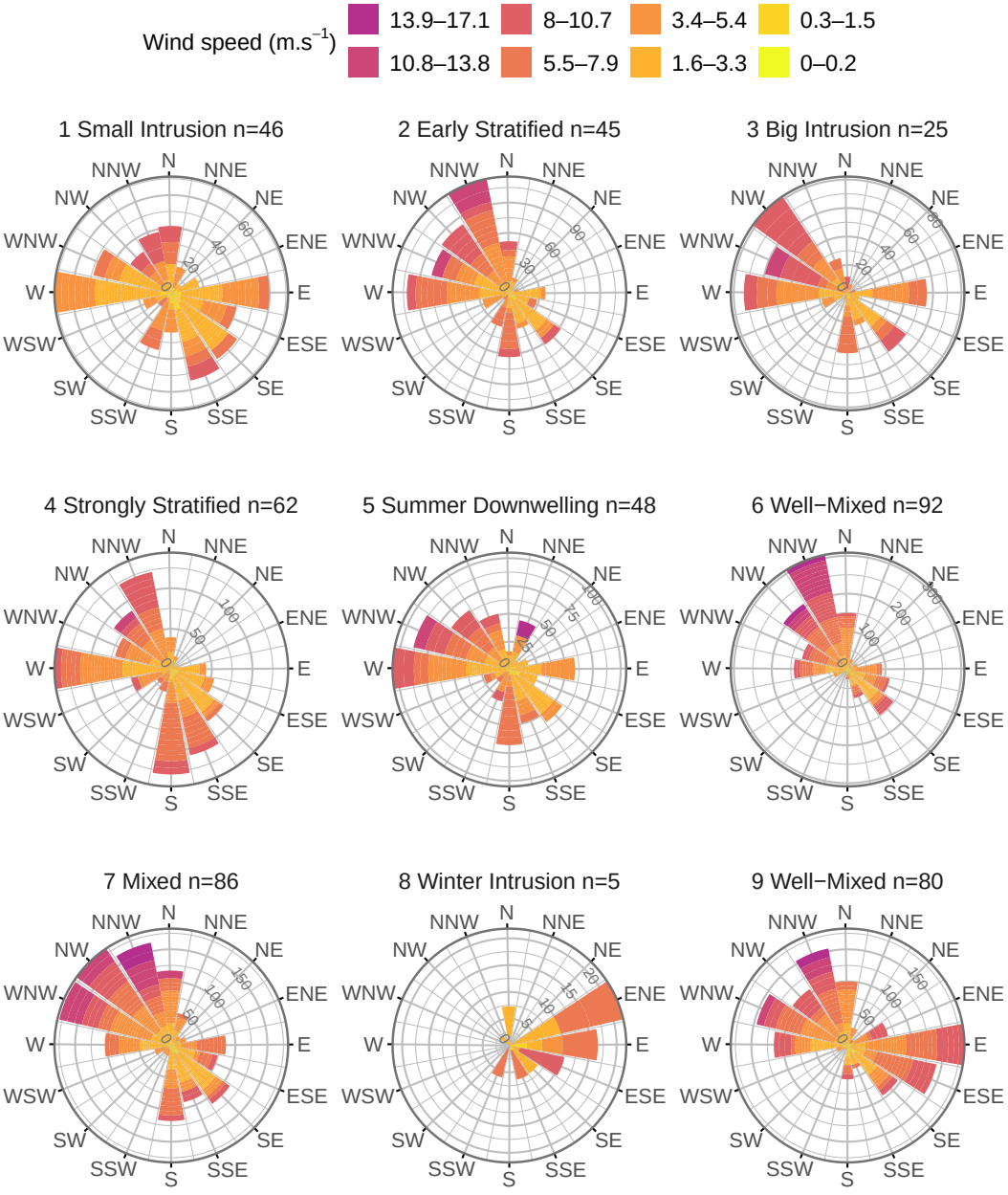


Figure G2. Stacked wind-rose plot showing the wind observations made between 12:00 am and 2:00 pm on each of the days that the vertical structures were observed. The number of days n each vertical structure was observed is indicated on each wind-rose title. The height of each bar represents the raw number of measurements within a directional sector, while colours correspond to wind-speed categories.

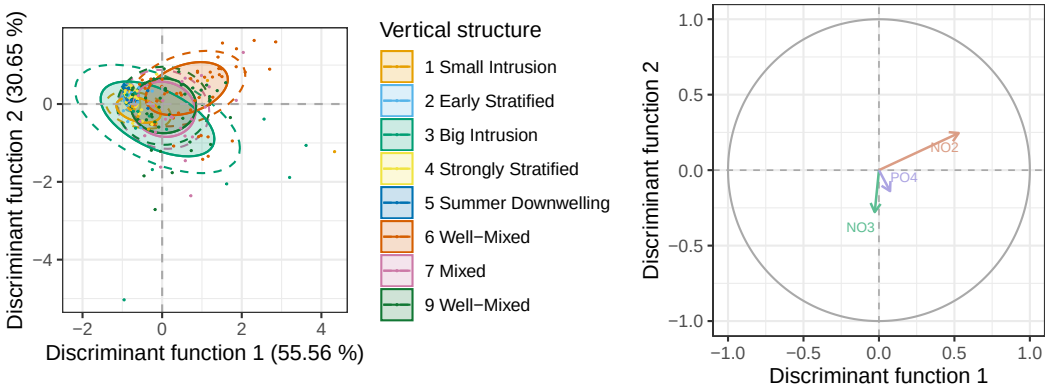


Figure G3. First discriminant plane of the descriptive discriminant analysis (DDA) of nutrient concentration discriminating vertical structures, together with the corresponding correlation circles. Axis labels show the percentage of discriminant inertia explained. On the discriminant planes, solid lines show 50% normal data ellipses and dashed lines show 75% normal data ellipses. (NO3: nitrate, NO2: nitrite, PO4: phosphate)

Appendix H: Dominant wind shift

The wind direction associated with the maximum 95th percentile of wind speed for the north-western sector was computed for each year between 1998 and 2024. A regression tree revealed three shifts in dominant wind direction (Fig. H1). The first shift, from 350° to 340°, occurred between 2003 and 2004. The second shift, from 340° to 325°, occurred between 2008 and 2009. The third shift, from 325° to 320°, occurred between 2017 and 2018.

Appendix I: Wind conditions and water-column structure during October–November 2020

Wind speed and direction were examined from 14 October to 3 November 2020, a period for which vertical profiles from the SNO COAST-HF high-frequency monitoring were available. The vertical structures identified from these profiles are shown alongside the wind data (Figure I1). Although these structures should be interpreted with caution (see Sect. 4.2), this example illustrates that strong north-westerly winds coincided with mixed conditions, whereas stratified conditions were observed within two days of the winds weakening.

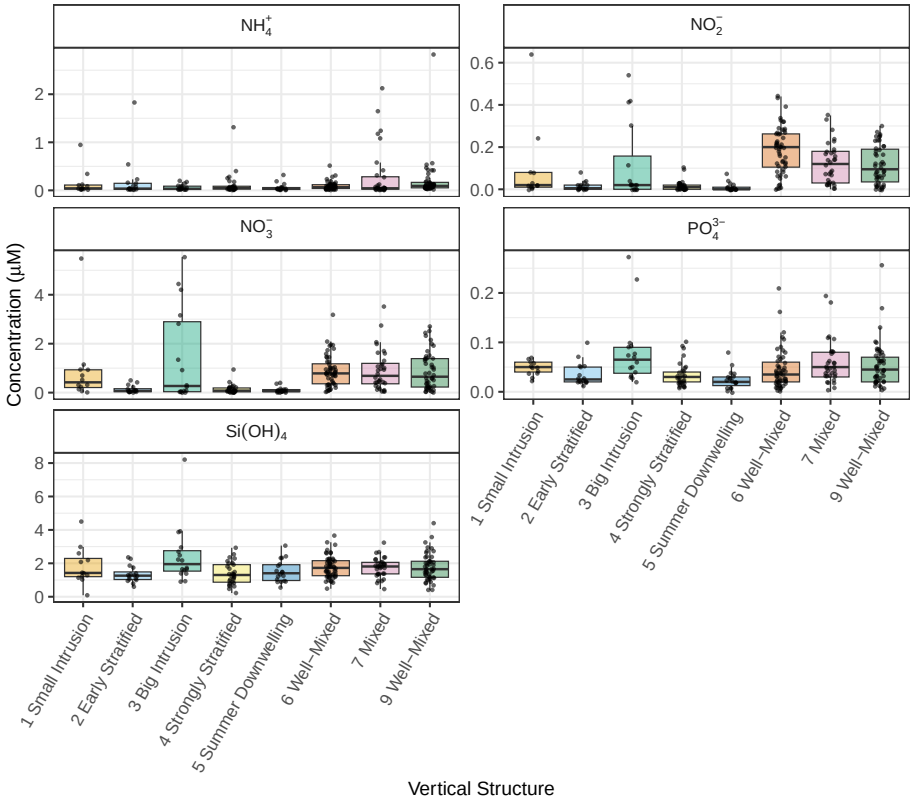


Figure G4. Box plots showing nutrient concentrations for each vertical structure.

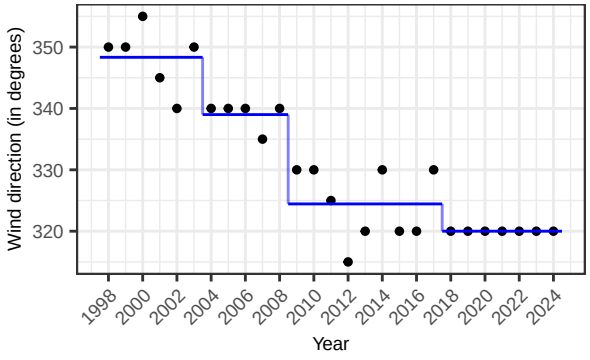


Figure H1. Piecewise constant regression (Classification and Regression Trees, CART (Ripley, 2025), blue line) of the mean annual dominant wind direction (black dots) in the northwestern quadrant (270°–360°) at the Marignane Météo-France station. The dominant direction is defined as the wind direction associated with the maximum 95th percentile of wind speed. A tree with four terminal nodes was selected by cross-validation, minimizing the deviance in approximately 85% of the 100 runs.

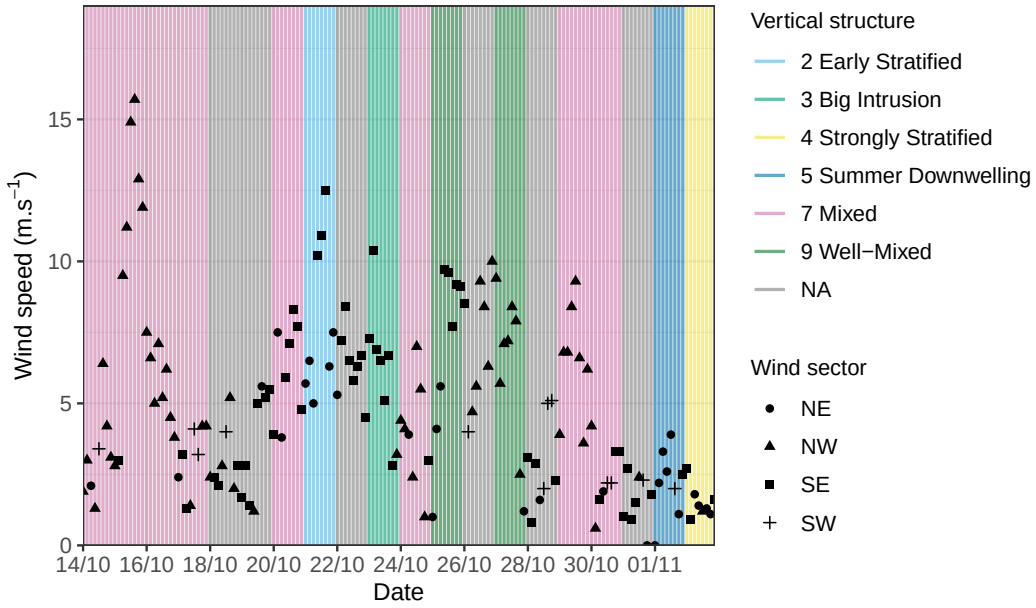


Figure I1. Wind conditions and vertical water-column structures from 14 October to 3 November 2020. Wind speed and direction measured during this period are shown together with the vertical structures identified from high-frequency observations. Wind direction sectors (NW, NE, SE, SW) are indicated for reference.

Author contributions. MCC, GG, and DN conceptualized the study. GG is responsible of project management. MCC, DP, FG, VL, SN, and CS contributed to the data curation. LB, MD, JMF, FG, VL, SN, DP, and CS contributed to data collection. MCC performed the formal analysis, visualization and wrote the manuscript draft under the supervision of GG and DN. All authors reviewed the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors wish to thank Patrick Raimbault, Nicole Garcia, Michel Lafont for their contribution to the long-term monitoring of the Bay of Marseille. We thank the crew of the R/V *Antedon II* (CNRS INSU, French National Oceanographic Fleet), Damien Schmutz, Arnaud Mahaud and Bruce Thomas ensuring the observations at the Frioul SOMLIT site. We also thank the crew of the R/V *Astroides* (OSU Institut Pythéas) and the divers operating on the SOLEMIO COAST-HF buoy, Dorian Guillemain, Frédéric Zuberer, and Deny Malengros. Finally, we acknowledge the French national observation services SOMLIT and COAST-HF, within the ILICO research infrastructure.

References

Abad, D., Albaina, A., Aguirre, M., Laza-Martínez, A., Uriarte, I., Iriarte, A., Villate, F., and Estonba, A.: Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy, Mar. Biol., 163, 149, https://doi.org/10.1007/s00227-016-2920-0, 2016.

Agostinelli, C. and Lund, U.: R package circular: Circular Statistics (version 0.5-2), CRAN [code], https://CRAN.R-project.org/package=circular, 2025.

Agostini, V. N. and Bakun, A.: 'Ocean triads' in the Mediterranean Sea: physical mechanisms potentially structuring reproductive habitat suitability (with example application to European anchovy, Engraulis encrasicolus), Fish. Oceanogr., 11, 129–142, https://doi.org/10.1046/j.1365-2419.2002.00201.x, 2002.

Alheit, J., Roy, C., and Souad, K.: Decadal-scale variability in populations, in: Climate Change and Small Pelagic Fish, edited by Checkley, D. M. J., Alheit, J., Oozeki, Y., and Roy, C., pp. 202–262, Cambridge University Press, https://doi.org/10.1017/CBO9780511596681, 2009.

Bakun, A. and Agostini, V. N.: Seasonal patterns of wind-induced upwelling/downwelling in the Mediterranean Sea, Sci. Mar., 65, 243–257, https://doi.org/10.3989/scimar.2001.65n3243, 2001.

Barrier, N., Petrenko, A., and Ourmières, Y.: Strong intrusions of the Northern Mediterranean Current on the eastern Gulf of Lion: insights from in-situ observations and high resolution numerical modelling, Ocean Dyn., 66, 313–327, https://doi.org/10.1007/s10236-016-0921-7, 2016.

Benson, A. J., McFarlane, G. A., Allen, S. E., and Dower, J. F.: Changes in Pacific Hake (Merluccius productus) migration patterns and juvenile growth related to the 1989 regime shift, Can. J. Fish. Aquat. Sci., 59, 1969–1979, https://doi.org/10.1139/f02-156, 2002.

Berdalet, E., Peters, F., Koumandou, V. L., Roldán, C., Guadayol, O., and Estrada, M.: Species-specific physiological response of dinoflagellates to quantified small-scale turbulence, J. Phycol., 43, 965–977, https://doi.org/10.1111/j.1529-8817.2007.00392.x, 2007.

Breton, E., Savoye, N., Rimmelin-Maury, P., Sautour, B., Goberville, E., Lheureux, A., Cariou, T., Ferreira, S., Agogué, H., and Alliouane, S.: Data quality control considerations in multivariate environmental monitoring: experience of the French coastal network SOMLIT, Front. Mar. Sci., 10, 1135 446, https://doi.org/10.3389/fmars.2023.1135446, 2023.

Brosset, P., Bourg, B. L., Costalago, D., Bănaru, D., Beveren, E. V., Bourdeix, J. H., Fromentin, J. M., Ménard, F., and Saraux, C.: Linking small pelagic dietary shifts with ecosystem changes in the Gulf of Lions, Mar. Ecol. Prog. Ser., 554, 157–171, https://doi.org/10.3354/meps11796, 2016.

Bănaru, D., Mellon-Duval, C., Roos, D., Bigot, J. L., Souplet, A., Jadaud, A., Beaubrun, P., and Fromentin, J. M.: Trophic structure in the Gulf of Lions marine ecosystem (north-western Mediterranean Sea) and fishing impacts, J. Mar. Syst., 111-112, 45–68, https://doi.org/10.1016/j.jmarsys.2012.09.010, 2013.

Calvo-Díaz, A. and Morán, X. A. G.: Seasonal dynamics of picoplankton in shelf waters of the southern Bay of Biscay, Aquat Microb Ecol, 42, 159–174, https://doi.org/10.3354/ame042159, 2006.

Campbell, L. and Vaulotaf, D.: Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA), Deep Sea Res. I: Oceanogr. Res. Pap., 40, 2043–2060, https://doi.org/10.1016/0967-0637(93)90044-4, 1993.

Cardot, H.: Conditional functional principal components analysis, Scand. J. Stat., 34, 317–335, https://doi.org/10.1111/j.1467-9469.2006.00521.x, 2007.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B.: shiny: Web Application Framework for R, CRAN [code], https://CRAN.R-project.org/package=shiny, r package version 1.10.0, 2024.

Chen, C. T., Carlotti, F., Harmelin-Vivien, M., Guilloux, L., and Bănaru, D.: Temporal variation in prey selection by adult European sardine (Sardina pilchardus) in the NW Mediterranean Sea, Prog. Oceanogr., 196, 102 617, https://doi.org/10.1016/j.pocean.2021.102617, 2021.

Cheng, L., Li, G., Long, S. M., Li, Y., von Schuckmann, K., Trenberth, K. E., Mann, M. E., Abraham, J., Du, Y., Cheng, X., Liu, H., Xu, Z., Liu, M., Peng, Q., Gong, X., Ma, Z., and Yuan, H.: Ocean stratification in a warming climate, https://doi.org/10.1038/s43017-025-00715-5, 2025.

Cocquempot, L., Delacourt, C., Paillet, J., Riou, P., Aucan, J., Castelle, B., Charria, G., Claudet, J., Conan, P., Coppola, L., Hocdé, R., Planes, S., Raimbault, P., Savoye, N., Testut, L., and Vuillemin, R.: Coastal ocean and nearshore observation: A French case study, Front. Mar. Sci., 6, 324, https://doi.org/10.3389/fmars.2019.00324, 2019.

Couteyen Carpaye, M.: Reconstruction of truncated vertical profiles using functional data analysis, Zenodo [code], https://doi.org/10.5281/zenodo.23598020, 2026a.

Couteyen Carpaye, M.: SOMLITMed Explorer, Zenodo [code], https://doi.org/10.5281/zenodo.23039092, 2026b.

Dubelaar, G. B., Casotti, R., Tarran, G. A., and Biegala, I. C.: Phytoplankton and their analysis by flow cytometry, in: Flow Cytometry with Plant Cells: Analysis of Genes, Chromosomes and Genomes, edited by Doležel, J., Greilhuber, J., and Suda, J., pp. 287–322, WILEY-VCH, https://doi.org/10.1002/9783527610921.ch13, 2007.

Edwards, M., Beaugrand, G., Hays, G. C., Koslow, J. A., and Richardson, A. J.: Multi-decadal oceanic ecological datasets and their application in marine policy and management, Trends Ecol. Evol., 25, 602–610, https://doi.org/10.1016/j.tree.2010.07.007, 2010.

Estournel, C., Estaque, T., Ulses, C., Barral, Q.-B., and Marsaleix, P.: Extreme sensitivity of the northeastern Gulf of Lion (western Mediterranean) to subsurface heatwaves: Physical processes and devastating impacts on ecosystems in the summer of 2022, Ocean Sci., 21, 1487–1503, https://doi.org/doi.org/10.5194/os-21-1487-2025, 2025.

Feuilloley, G., Fromentin, J. M., Stemmann, L., Demarcq, H., Estournel, C., and Saraux, C.: Concomitant changes in the environment and small pelagic fish community of the Gulf of Lions, Prog. Oceanogr., 186, 102 375, https://doi.org/10.1016/j.pocean.2020.102375, 2020.

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., and Martiny, A. C.: Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus, Proc. Natl. Acad. Sci. U.S.A., 110, 9824–9829, https://doi.org/10.1073/pnas.1307701110, 2013.

Fonvieille, N., Guinet, C., Picard, B., and Nerini, D.: Conditional multivariate functional PCA for the reconstruction of temperature and salinity profiles partially sampled by deep-diving marine mammals, arXiv [preprint], arXiv:2608.05376, 5 August 2026.

Fraysse, M., Pinazo, C., Faure, V., Fuchs, R., and Lazzari, P.: Development of a 3D Coupled Physical-Biogeochemical Model for the Marseille Coastal Area, PLoS One, 8, e80 012, https://doi.org/10.1371/journal.pone.0080012, 2013.

Fraysse, M., Pairaud, I., Ross, O. N., Faure, V., Pinazo, C., and Faure, V. M.: Intrusion of Rhone River diluted water into the Bay of Marseille: Generation processes and impacts on ecosystem functioning, J. Geophys. Res. Oceans, 119, 6535–6556, https://doi.org/10.1002/2014JC010022, 2014.

Fuchs, R., Thyssen, M., Creach, V., Dugenne, M., Izard, L., Latimier, M., Louchart, A., Marrec, P., Rijkeboer, M., and Grégori, G.: Automatic recognition of flow cytometric phytoplankton functional groups using convolutional neural networks, Limnol. Oceanogr.: Methods, 20, 387–399, https://doi.org/10.1002/lom3.10493, 2022.

Fuchs, R., Rossi, V., Caille, C., Bensoussan, N., Pinazo, C., Grosso, O., and Thyssen, M.: Intermittent Upwelling Events Trigger Delayed, Major, and Reproducible Pico-Nanophytoplankton Responses in Coastal Oligotrophic Waters, Geophys. Res. Lett., 50, e2022GL102 651, https://doi.org/10.1029/2022GL102651, 2023.

Garcia, F., Fornier, M., Garcia, N., Lafont, M., Lagadec, V., Pairaud, I., and Raimbault, P.: Observation à long terme des caractéristiques hydrologiques du littoral marseillais et de l'influence rhodanienne, in: Mesures à haute résolution dans l'environnement côtier, edited by Schmitt, F. and Lefebvre, A., pp. 61–71, CNRS Editions, ISBN 978-2-271-08592-4, 2016.

Garcia, F., Raimbault, P., Lafont, M., Guillemain, D., Deny, M., Lagadec, V., Garcia, N., and Grégori, G.: SOLEMIO data and metadata from Coriolis Data Centre., SEANOE [data set], https://doi.org/10.17882/88495, 2025.

Garcia, T., Bănaru, D., Guilloux, L., Cornet, V., Gregori, G., and Carlotti, F.: Temporal changes in zooplankton indicators highlight a bottom-up process in the Bay of Marseille (NW Mediterranean Sea), PLoS One, 18, e0292 536, https://doi.org/10.1371/journal.pone.0292536, 2023.

Garcia, T., Carlotti, F., Lepoint, G., Guilloux, L., Tesán-Onrubia, J. A., Grassi, B., Russias, V., and Bănaru, D.: The decadal variability of elemental, isotopic, and biochemical compositions of coastal Mediterranean zooplankton responds to environmental changes, Mar. Biol., 171, 119, https://doi.org/10.1007/s00227-024-04430-5, 2024.

Gatti, J.: Intrusions du Courant Nord Méditerranéen sur la partie est du plateau continental du Golfe du Lion, Ph.D. thesis, Université Aix-Marseille II, 2008.

Gatti, J., Petrenko, A., Devenon, J.-L., Leredde, Y., and Ulses, C.: The Rhone river dilution zone present in the northeastern shelf of the Gulf of Lion in december 2003, Cont. Shelf Res., 26, 1794–1805, https://doi.org/10.1016/j.csr.2006.05.012, 2006.

Glenn, S. M., Dickey, T. D., Parker, B., and Boicourt, W.: Long-term Real-time Coastal Ocean Observation Networks, Oceanography, 13, 24–34, https://doi.org/10.5670/oceanog.2000.50, 2000.

Grolemund, G. and Wickham, H.: Dates and Times Made Easy with lubridate, Journal of Statistical Software, 40, 1–25, https://doi.org/10.18637/jss.v040.i03, 2011.

Hondzo, M. and Lyn, D.: Quantified small-scale turbulence inhibits the growth of a green alga, Freshw. Biol., 41, 51–61, https://doi.org/10.1046/j.1365-2427.1999.00389.x, 1999.

Hu, Z. Y., Doglioli, A. M., Petrenko, A. A., Marsaleix, P., and Dekeyser, I.: Numerical simulations of eddies in the Gulf of Lion, Ocean Model., 28, 203–208, https://doi.org/10.1016/j.ocemod.2009.02.004, 2009.

Hubert, Z., Artigas, L. F., Li, L.-L., Dédécker, C., and Monchy, S.: Exploring the regional diversity of eukaryotic phytoplankton in the English Channel by combining high-throughput approaches, MicrobiologyOpen, 14, e70 097, https://doi.org/10.1002/mbo3.70097, 2025.

Hunt, B. P., Carlotti, F., Donoso, K., Pagano, M., D'Ortenzio, F., Taillandier, V., and Conan, P.: Trophic pathways of phytoplankton size classes through the zooplankton food web over the spring transition period in the north-west Mediterranean Sea, J. Geophys. Res. Oceans, 122, 6309–6324, https://doi.org/10.1002/2016JC012658, 2017.

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., and Chisholm, S. W.: Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients, Science, 311, 1737–1740, https://doi.org/10.1126/science.1118052, 2006.

Karl, D. M. and Church, M. J.: Microbial oceanography and the Hawaii Ocean Time-series programme, Nat. Rev. Microbiol., 12, 699–713, https://doi.org/10.1038/nrmicro3333, 2014.

Kassambara, A.: rstatix: Pipe-Friendly Framework for Basic Statistical Tests, CRAN [code], https://CRAN.R-project.org/package=rstatix, r package version 0.7.2, 2023.

Kiørboe, T.: Turbulence, Phytoplankton Cell Size, and the Structure of Pelagic Food Webs, in: Advances in Marine Biology, edited by Blaxter, J. and Southward, A., vol. 29, pp. 1–72, Academic Press, https://doi.org/10.1016/S0065-2881(08)60129-7, 1993.

Latasa, M. and Berdalet, E.: Effect of nitrogen or phosphorus starvation on pigment composition of cultured Heterocapsa sp, J. Plankton Res., 16, 83–94, https://doi.org/10.1093/plankt/16.1.83, 1994.

Latasa, M., Scharek, R., Vidal, M., Vila-Reixach, G., Gutiérrez-Rodríguez, A., Emelianov, M., and Gasol, J. M.: Preferences of phytoplankton groups for waters of different trophic status in the northwestern Mediterranean sea, Mar. Ecol. Prog. Ser., 407, 27–42, https://doi.org/10.3354/meps08559, 2010.

Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., and Wolf-Gladrow, D.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Glob. Change Biol., 11, 2016–2040, https://doi.org/https://doi.org/10.1111/j.1365-2486.2005.1004.x, 2005.

Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu De Madron, X., Bowler, C., Malviya, S., Thyssen, M., Grégori, G., Rembauville, M., Grosso, O., Poulain, J., Vargas, C. D., Pujo-Pay, M., and Conan, P.: Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export, Nat. Commun., 9, 953, https://doi.org/10.1038/s41467-018-03376-9, 2018.

Lee, J. Y.: La sardine du golfe du lion (Sardina pilchardus sardina REGAN), Revue des Travaux de l'Institut des Pêches Maritimes, 25, 418–513, 1961.

Lee, R. F., Hagen, W., and Kattner, G.: Lipid storage in marine zooplankton, Mar. Ecol. Prog. Ser., 307, 273–306, https://doi.org/10.3354/meps307273, 2006.

Lefevre, D., Minas, H. J., Minas, M., Robinson, . C., Le, P. J., Williams, B., and Woodwards, E. M. S.: Review of gross community production, primary production, net community production and dark community respiration in the Gulf of Lions, Deep Sea Res. II: Top. Stud. Oceanogr., 44, 801–832, https://doi.org/10.1016/S0967-0645(96)00091-4, 1997.

Lescot, M., Lezzoche, N., Laux, L., Romac, S., Guilloux, L., Chevillon, E., Bodson, C., Desnos, C., Elineau, A., Jalabert, L., Monferrer, N. L., Sandin, M. M., Vannier, T., Vernette, C., Villar, E., Lombard, F., Carlotti, F., Irisson, J. O., Guidi, L., Bosse, A., Testor, P., Coppola, L., and Not, F.: A holistic perspective on planktonic communities across the Northwestern Mediterranean Sea, Front. Mar. Sci., 13, 1755 855, https://doi.org/10.3389/fmars.2026.1755855, 2026.

Lheureux, A., David, V., Del Amo, Y., Soudant, D., Auby, I., Bozec, Y., Conan, P., Ganthy, F., Grégori, G., Lefebvre, A., Leynart, A., Rimmelin-Maury, P., Souchu, P., Vantrepote, V., Blondel, C., Cariou, T., Crispi, O., Cordier, M. A., Crouvoisier, M., Duquesne, V., Ferreira, S., Garcia, N., Gouriou, L., Grosteffan, E., Le Merrer, Y., Meteigner, C., Retho, M., Tournaire, M. P., and Savoye, N.: Trajectories of nutrients concentrations and ratios in the French coastal ecosystems: 20 years of changes in relation with large-scale and local drivers, Sci. Total Environ., 857, 159 619, https://doi.org/10.1016/j.scitotenv.2022.159619, 2023.

Litchman, E. and Klausmeier, C. A.: Trait-based community ecology of phytoplankton, Annu. Rev. Ecol. Evol. Syst., 39, 615–639, https://doi.org/10.1146/annurev.ecolsys.39.110707.173549, 2008.

Liu, C., Liang, X., and Yu, L.: Salinity trends and mass balances in the Mediterranean Sea: revisit the role of air-sea freshwater fluxes and oceanic exchange, Ocean Sci., 21, 2069–2083, https://doi.org/10.5194/os-21-2069-2025, 2025.

Llaveria, G., Figueroa, R. I., Garcés, E., and Berdalet, E.: Cell cycle and cell mortality of Alexandrium minutum (Dinophyceae) under small-scale turbulence conditions, J. Phycol., 45, 1106–1115, https://doi.org/10.1111/j.1529-8817.2009.00740.x, 2009.

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

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

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

Marrec, P., Grégori, G., Doglioli, A. M., Dugenne, M., Penna, A. D., Bhairy, N., Cariou, T., Nunige, S. H., Lahbib, S., Rougier, G., Wagener, T., and Thyssen, M.: Coupling physics and biogeochemistry thanks to high-resolution observations of the phytoplankton community structure in the northwestern Mediterranean Sea, Biogeosciences, 15, 1579–1606, https://doi.org/10.5194/bg-15-1579-2018, 2018.

Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahé, F., Bariat, L., Courties, C., Marie, D., Ras, J., Mauriac, R., Jeanthon, C., Bendif, E. M., Ostrowski, M., Scanlan, D. J., and Garczarek, L.: Is the distribution of Prochlorococcus and Synechococcus ecotypes in the Mediterranean Sea affected by global warming?, Biogeosciences, 8, 2785–2804, https://doi.org/10.5194/bg-8-2785-2011, 2011.

Mena, C., Reglero, P., Hidalgo, M., Sintes, E., Santiago, R., Martín, M., Moyà, G., and Balbín, R.: Phytoplankton community structure is driven by stratification in the oligotrophic mediterranean sea, Front. microbiol., 10, 1698, https://doi.org/10.3389/fmicb.2019.01698, 2019.

Millot, C.: Wind induced upwellings in the Gulf of Lions, Oceanol. Acta, 2, 261–274, 1979.

Millot, C.: The Gulf of Lions' hydrodynamics, Cont. Shelf Res., 10, 885–894, https://doi.org/10.1016/0278-4343(90)90065-T, 1990.

Moore, J. K., Fu, W., Primeau, F., Britten, G. L., Lindsay, K., Long, M., Doney, S. C., Mahowald, N., Hoffman, F., and Randerson, J. T.: Sustained climate warming drives declining marine biological productivity, Science, 359, 1139–1143, https://doi.org/10.1126/science.aao6379, 2018.

Morel, A., Ahn, Y.-H., Partensky, F., Vaulot, D., and Claustre, H.: Prochlorococcus and Synechococcus: a comparative study of their optical properties in relation to their size and pigmentation, J. Mar. Res., 51, 617–649, https://doi.org/10.1357/0022240933223963, 1993.

Mouriño-Carballido, B., Hojas, E., Cermeño, P., Chouciño, P., Fernández-Castro, B., Latasa, M., Marañón, E., Morán, X. A. G., and Vidal, M.: Nutrient supply controls picoplankton community structure during three contrasting seasons in the northwestern Mediterranean Sea, Mar. Ecol. Prog. Ser., 543, 1–19, https://doi.org/10.3354/meps11558, 2016.

Moutin, T., Raimbault, P., Golterman, H. L., and Coste, B.: The input of nutrients by the Rhône river into the Mediterranean Sea: Recent observations and comparison with earlier data, in: Proceedings of the Third International Joint Conference on Limnology and Oceanography, pp. 237–246, Oceans, Rivers and Lakes: Energy and Substance Transfers at Interfaces, Nantes, France, https://doi.org/10.1007/978-94-011-5266-2_19, 1998.

Météo-France: Données SYNOP essentielles OMM, https://www.data.gouv.fr/datasets/archive-synop-omm, last accessed: 16 December 2024, 2026.

Nerini, D., Manté, C., and Monestiez, P.: Extending functional kriging when data are multivariate curves : some technical considerations and operational solutions, in: Geostatistical Functional Data Analysis, edited by Mateu, J. and Giraldo, R., pp. 73–103, Wiley, https://doi.org/10.1002/9781119387916.ch4, 2021.

Nikolioudakis, N., Isari, S., Pitta, P., and Somarakis, S.: Diet of sardine Sardina pilchardus: An 'end-to-end' field study, Mar. Ecol. Prog. Ser., 453, 173–188, https://doi.org/10.3354/meps09656, 2012.

O'Brien, T. D., Lorenzoni, L., Isensee, K., and Valdés, L.: What are Marine Ecological Time Series telling us about the ocean? A status report., Tech. rep., IOC-UNESCO, 2017.

Pairaud, I. L., Gatti, J., Bensoussan, N., Verney, R., and Garreau, P.: Hydrology and circulation in a coastal area off Marseille: Validation of a nested 3D model with observations, J. Mar. Syst., 88, 20–33, https://doi.org/10.1016/j.jmarsys.2011.02.010, 2011.

Para, J., Coble, P. G., Charrière, B., Tedetti, M., Fontana, C., and Sempéré, R.: Fluorescence and absorption properties of chromophoric dissolved organic matter (CDOM) in coastal surface waters of the northwestern Mediterranean Sea, influence of the Rhône River, Biogeosciences, 7, 4083–4103, https://doi.org/10.5194/bg-7-4083-2010, 2010.

Partensky, F., Blanchot, J., and Vaulot, D.: Differential distribution and ecology of Prochlorococcus and Synechococcus in oceanic waters : a review, in: Marine cyanobacteria, edited by Charpy, L. and Larkum, A., pp. 457 – 475, Musée océanographique, Monaco, ISBN 2-7260-0210-2, 1999a.

Partensky, F., Hess, W. R., , and Vaulot, D.: Prochlorococcus, a marine photosynthetic prokaryote of global significance, Microbiol. Mol. Biol. Rev., 63, 106–127, https://doi.org/10.1128/MMBR.63.1.106-127.1999, 1999b.

Petrenko, A.: Variability of circulation features in the Gulf of Lion NW Mediterranean Sea. Importance of inertial currents, Oceanol. Acta, 26, 323–338, https://doi.org/10.1016/S0399-1784(03)00038-0, 2003.

Petrenko, A., Leredde, Y., and Marsaleix, P.: Circulation in a stratified and wind-forced Gulf of Lions, NW Mediterranean Sea: In situ and modeling data, Cont. Shelf Res., 25, 7–27, https://doi.org/10.1016/j.csr.2004.09.004, 2005.

Pichot, P., Aldebert, Y., Carries, C., and Pichot, Y.: La pêche de la sardine en Méditerranée française, Science et Pêche, 277, 1–16, 1978.

Prairie, J. C., Sutherland, K. R., Nickols, K. J., and Kaltenberg, A. M.: Biophysical interactions in the plankton: A cross-scale review, Limnol. Oceanogr.: Fluids and Environments, 2, 121–145, https://doi.org/10.1215/21573689-1964713, 2012.

Queiros, Q., Fromentin, J. M., Gasset, E., Dutto, G., Huiban, C., Metral, L., Leclerc, L., Schull, Q., McKenzie, D. J., and Saraux, C.: Food in the sea: Size also matters for pelagic fish, Front. Mar. Sci., 6, 385, https://doi.org/10.3389/fmars.2019.00385, 2019.

Queiros, Q., Saraux, C., Dutto, G., Gasset, E., Marguerite, A., Brosset, P., Fromentin, J. M., and McKenzie, D. J.: Is starvation a cause of overmortality of the Mediterranean sardine?, Mar. Environ. Res., 170, 105 441, https://doi.org/10.1016/j.marenvres.2021.105441, 2021.

R Core Team: R: A Language and Environment for Statistical Computing, https://www.R-project.org/, 2024.

Ramsay, J.: fda: Functional Data Analysis, CRAN [code], https://CRAN.R-project.org/package=fda, r package version 6.2.0, 2024.

Ramsay, J. O. and Silverman, B. W.: Functional data analysis, Springer, springer series in statistics edn., ISBN 038740080X, 2005.

Ripley, B.: tree: Classification and Regression Trees, CRAN [code], https://CRAN.R-project.org/package=tree, r package version 1.0-45, 2025.

Ross, O. N., Fraysse, M., Pinazo, C., and Pairaud, I.: Impact of an intrusion by the Northern Current on the biogeochemistry in the eastern Gulf of Lion, NW Mediterranean, Estuar. Coast. Shelf Sci., 170, 1–9, https://doi.org/10.1016/j.ecss.2015.12.022, 2016.

Sammartino, M., Aronica, S., Santoleri, R., and Buongiorno Nardelli, B.: Retrieving Mediterranean Sea Surface Salinity Distribution and Interannual Trends from Multi-Sensor Satellite and In Situ Data, Remote Sens., 14, 2502, https://doi.org/10.3390/rs14102502, 2022.

Saporta, G.: L'analyse discriminante, in: Probabilités, analyse des données et statistique, pp. 403–428, Editions Technip, ISBN 2710805650, 1990.

Saraux, C., Beveren, E. V., Brosset, P., Queiros, Q., Bourdeix, J. H., Dutto, G., Gasset, E., Jac, C., Bonhommeau, S., and Fromentin, J. M.: Small pelagic fish dynamics: A review of mechanisms in the Gulf of Lions, Deep Sea Res. II: Top. Stud. Oceanogr., 159, 52–61, https://doi.org/10.1016/j.dsr2.2018.02.010, 2019.

Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J., and Manabe, S.: Simulated response of the ocean carbon cycle to anthropogenic climate warming, Nature, 393, 245–249, https://doi.org/10.1038/30455, 1998.

Savoye, N., Lizon, F., Breton, E., Claquin, P., Joly, O., Sultan, E., Jung, J.-L., Bozec, Y., Boulart, C., Rimmelin-Maury, P., Leynaert, A., Agogué, H., Pineau, P., Del Amo, Y., Conan, P., Mostajir, B., Grégori, G., Mousseau, L., and Mendès, F.: SOMLIT (Service d'Observation

en Milieu Littoral) time series (French Research Infrastructure ILICO): long-term core parameter monitoring of French coasts, SEANOE [data set], https://doi.org/10.17882/100323, 2026.

Schmitt, F. and Lefebvre, A.: Mesures à haute résolution dans l'environnement marin côtier, CNRS Editions, ISBN 978-2-271-08592-4, 2016.

Scrucca, L., Fraley, C., Murphy, T. B., and Raftery, A. E.: Model-Based Clustering, Classification, and Density Estimation Using mclust in R, Chapman and Hall/CRC, 1st edn., https://doi.org/10.1201/9781003277965, 2023.

Siokou-Frangou, I., Christaki, U., Mazzocchi, M. G., Montresor, M., Ribera D'Alcala, M., Vaque, D., and Zingone, A.: Plankton in the open mediterranean Sea: A review, Biogeosciences, 7, 1543–1586, https://doi.org/10.5194/bg-7-1543-2010, 2010.

SNO SOMLIT: SOMLIT: Service d'Observation en Milieu Littoral, https://somlit.fr, data retrieved incrementally between 2024 and 2025; last accessed 25 August 2025, 2025.

Sommer, U., Stibor, H., Katechakis, A., Sommer, F., and Hansen, T.: Pelagic food web configurations at different levels of nutrient richness and their implications for the ratio fish production:primary production, Hydrobiologia, 484, 11–20, https://doi.org/10.1023/A:1021340601986, 2002.

Strayer, D., Glitzenstein, J. S., Jones, C. G., Kolasa, J., Likens, G. E., McDonnell, M. J., Parker, G. G., and Pickett, S. T.: Long-term ecological studies: an illustrated account of their design, operation, and importance to ecology, Tech. rep., The Institute of Ecosystem Studies, 1986.

Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Babin, M., Boss, E., Cochrane, G., de Vargas, C., Follows, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels, S., Karp-Boss, L., Karsenti, E., Lescot, M., Not, F., Ogata, H., Pesant, S., Poulton, N., Raes, J., Sardet, C., Sieracki, M., Speich, S., Stemmann, L., Sullivan, M. B., Wincker, P., Eveillard, D., and Lombard, F.: Tara Oceans: towards global ocean ecosystems biology, Nature Reviews Microbiology, 18, 428–445, https://doi.org/10.1038/s41579-020-0364-5, 2020.

Taverniers, I., Loose, M. D., and Bockstaele, E. V.: Trends in quality in the analytical laboratory. I. Traceability and measurement uncertainty of analytical results, Trends Anal. Chem., 23, 480–490, https://doi.org/10.1016/S0165-9936(04)00733-2, 2004.

Taylor, M. and Henson, S.: A method for quantifying correlation in the shape of oceanographic profile data, Ocean Sci., 22, 1377–1390, https://doi.org/10.5194/os-22-1377-2026, 2026.

Thyssen, M., Grégori, G., Créach, V., Lahbib, S., Dugenne, M., Aardema, H. M., Artigas, L. F., Huang, B., Barani, A., Beaugeard, L., Bellaaj-Zouari, A., Beran, A., Casotti, R., Del Amo, Y., Denis, M., Dubelaar, G. B., Endres, S., Haraguchi, L., Karlson, B., Lambert, C., Louchart, A., Marie, D., Moncoiffé, G., Pecqueur, D., Ribalet, F., Rijkeboer, M., Silovic, T., Silva, R., Marro, S., Sosik, H. M., Sourisseau, M., Tarran, G., Oostende, N. V., Zhao, L., and Zheng, S.: Interoperable vocabulary for marine microbial flow cytometry, Front. Mar. Sci., 9, 975 877, https://doi.org/10.3389/fmars.2022.975877, 2022.

Thyssen, M., Marrec, P., Taupier-Letage, I., Ismail, S. B., Ismail, M. A. B., Denis, M., Lahbib, S., Bhairy, N., Hassen, M. B., Boudriga, I., and Sammari, C.: Early winter pico-nanophytoplankton bloom in the Bonifacio cyclonic gyre after an extreme flooding, Prog. Oceanogr., 247, 103 790, https://doi.org/10.1016/j.pocean.2026.103790, 2026.

Vallina, S. M., Gaborit, C., Marrase, C., Gasol, J. M., Bahamon, N., Follows, M. J., Gland, G. L., and Cermeño, P.: Seasonal dynamics of phytoplankton community assembly at the Blanes Bay Microbial Observatory (BBMO), NW Mediterranean Sea, Prog. Oceanogr., 219, 103 125, https://doi.org/10.1016/j.pocean.2023.103125, 2023.

von Jackowski, A.: Navigating the current landscape of ocean observations: an overview from platform infrastructures to networks related to ocean time series, Front. Mar. Sci., 11, 1363 863, https://doi.org/10.3389/fmars.2024.1363863, 2024.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, 1st edn., https://doi.org/10.1007/978-0-387-98141-3, 2016.

Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D.: dplyr: A Grammar of Data Manipulation, CRAN [code], https://CRAN.R-project.org/package=dplyr, r package version 1.1.4, 2023.

Wiebe, P. H., Miller, C. B., McGowan, J. A., and Knox, R. A.: Long Time Series Study of Oceanic Ecosystems, Eos, 68, 1178–1190, https://doi.org/10.1029/EO068i044p01178, 1987.

Yamaguchi, R. and Suga, T.: Trend and Variability in Global Upper-Ocean Stratification Since the 1960s, J. Geophys. Res. Oceans, 124, 8933–8948, https://doi.org/10.1029/2019JC015439, 2019.

Yao, F., Müller, H. G., and Wang, J. L.: Functional data analysis for sparse longitudinal data, JASA, 100, 577–590, https://doi.org/10.1198/016214504000001745, 2005.

Yee, T. W.: The VGAM Package for Categorical Data Analysis, J. Stat. Softw., 32, 1–34, https://doi.org/10.18637/jss.v032.i10, 2010.

Yee, T. W. and Wild, C. J.: Vector Generalized Additive Models, J. R. Statist. Soc. B, 58, 481–493, https://doi.org/10.1111/j.2517-6161.1996.tb02095.x, 1996.

Zubkov, M. V., Sleigh, M. A., Burkill, P. H., and Leakey, R. J. G.: Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons, Prog. Oceanogr., 45, 3690–0386, https://doi.org/10.1016/S0079-6611(00)00008-2, 2000.