



A method for quantifying correlation in the shape of oceanographic profile data

Mark Taylor^{1,2} and Stephanie Henson²

¹School of Ocean and Earth Science, University of Southampton, European Way, Southampton, SO14 3ZH, UK

²Ocean Biogeosciences, National Oceanography Centre, European Way, SO14 3ZH, UK

Correspondence: Mark Taylor (mrt1u21@soton.ac.uk)

Abstract.

Vertical profiles are a common type of oceanographic observation, involving measurements of a variable across a range of depths, and are widely used to identify physical and biogeochemical features of the water column. Recent studies have shown that oceanographic profiles can be represented as functional data objects, where each profile is treated as a single datum and expressed as a function of pressure. This study applies a recently developed technique, which defines a scalar correlation coefficient for functional data, to the analysis of oceanographic profiles. The method represents each profile using basis functions, whose associated weightings are termed basis coefficients, and quantifies dependence through the variability of these coefficients. An important advantage of this method is that the resulting correlation coefficient reflects similarities in overall profile shape, not just correlations between values at specific depths. Two applications of this method are explored: calculating the correlation coefficient between two different oceanographic variables, and estimating the temporal autocorrelation function of a single variable. Each application is demonstrated using two case study datasets: (1) the Coastal Endurance Washington Off-shore Profiler Mooring and (2) biogeochemical-Argo floats. The first case study demonstrates how the method can be used to identify physical drivers of variability in biogeochemical profile structure. The second case study reveals regional differences in relationships between profiled variables and their temporal autocorrelation characteristics. This technique has broad potential for application to data from moorings, autonomous platforms, and ocean models, with possible use in observing system optimisation, data assimilation, and the analysis of vertically structured ocean processes.

1 Introduction

Oceanographic depth profiles describe vertical changes in the physical, chemical, and biological properties of the water column and are widely collected because they help to reveal important features like thermoclines (de Boyer Montégut et al., 2004), subsurface chlorophyll maxima (Cornec et al., 2021a) and oxygen minimum zones (Lovecchio et al., 2022). Measuring profiles of several variables simultaneously can allow for a more integrated interpretation of the vertical environment, such as assessing how biogeochemical phenomena are coupled to the physical structure of the water column (Carranza et al., 2018; Burke et al., 2023; Cao et al., 2024). Historically, ship based observations were the primary source of comprehensive profiling datasets, however they are sparse in space and time, especially in inaccessible regions. However, over the past two decades, the quantity



25 of depth profiles has increased substantially due to the widespread deployment of autonomous observing platforms (Wong et al.,
2020), many of which can measure a suite of physical and biogeochemical variables (Chai et al., 2020; Whitt et al., 2020).
Furthermore, technological developments have enabled autonomous platforms to operate for long periods whilst sampling at
frequencies high enough to observe small-scale processes such as turbulence (Rudnick, 2016). This rapid growth in both the
quantity and diversity of profiling data highlights the need for statistical tools tailored to depth profiles to aid their analysis and
30 improve understanding of ocean vertical structure.

Functional data analysis (FDA) provides one such framework, offering techniques well suited to continuous, shape based
data such as vertical profiles. Functional data take the form of curves and surfaces, where a variable of interest is a function
of an indexing variable (Ramsay and Silverman, 2005). Recent work has demonstrated the benefits of treating oceanographic
profiles as continuous functional data objects, where depth (or equally pressure) is used as an indexing variable (Yarger et al.,
35 2022; Korte-Stapff et al., 2022; Kande et al., 2024). This allows for the essence of the profile shape to be captured within
each datum, alongside the numerical values. This perspective is consistent with mathematical models of the ocean's vertical
structure, which capture depth dependent interactions among key variables (Steele, 1964; Fennel and Boss, 2003; Alhassan
et al., 2024). Although previous studies have utilised this approach in the context of spatio-temporal modelling and interpolation
(Yarger et al., 2022; Korte-Stapff et al., 2022), there are opportunities to explore more fundamental analyses through the lens
40 of FDA. Urbano-Leon et al. (2023) developed a methodology for quantifying the variance of a functional dataset as a single
value (a scalar), when previously only a variance function was used (Ramsay and Silverman, 2005). This approach involves
representing each functional datum as the linear combination of a set of basis functions, whose corresponding weightings are
called basis coefficients. Urbano-Leon et al. (2023) defined the variance and covariance of paired functional datasets as sums
of the variances and covariances of the basis coefficients, and used these quantities to compute a scalar correlation coefficient.
45 The purpose of this work is to present the first applications of this statistical approach to oceanographic profiles.

Correlation coefficients are commonly calculated between scalar oceanographic variables, particularly for surface data
(McGillicuddy Jr et al., 2001; Kahru et al., 2010; Fay and McKinley, 2017) or for metrics derived from profiles such as
mixed layer depth or subsurface chlorophyll maximum depth (Jayaram et al., 2021; Xu et al., 2022). While relationships be-
tween full profiles have been identified (Carranza et al., 2018; Castro-Morales and Kaiser, 2012; Lovecchio et al., 2022), these
50 dependencies have not been quantified with correlation coefficients. Autocorrelation functions (ACFs), which describe the
persistence of a variable across spatial or temporal scales, provide information about the main sources of variability (Abbott
and Letelier, 1998; Delcroix et al., 2005). ACFs are used in a range of applications including interpolation (Meyers et al.,
1991), observing system design (Ford, 2021; Chamberlain et al., 2023; Chu et al., 2024) and data assimilation (Storto et al.,
2018; Mirouze et al., 2016). Sumata et al. (2018) computed ACFs for profiling data in the Arctic, but estimated correlations by
55 binning measurements at each depth rather than treating profiles as single datums. By computing correlation coefficients and
ACFs directly from full profiles, the present approach complements existing methods and naturally incorporates profile shape.

The approach is illustrated with two case studies using simultaneously measured multi-variable profiles to explore the cou-
pling between physical and biogeochemical variables. The first case study uses data from the Coastal Endurance Washington
Offshore Profiler Mooring (CE09OSPM) (Risien et al., 2025), a daily-averaged time series of hydrographic and dissolved oxy-



60 gen measurements at a single site, which makes it suitable for examining local temporal variability. While the annual cycles
of these variables have been estimated (Risien et al., 2025), their seasonal strength was not quantified. The second case study
comprises temperature and chlorophyll profiles from seven Biogeochemical-Argo (BGC-Argo) floats (Claustre et al., 2020),
which typically collect depth profiles every 10 days while drifting with ocean currents. Previous studies have explored rela-
tionships between environmental conditions and vertical chlorophyll structure using BGC-Argo floats (Barbieux et al., 2019;
65 Cornec et al., 2021a, b; Strutton et al., 2023), but correlations between hydrographic and biogeochemical profiles have not
been reported. In this study, correlation coefficients between multiple profiled variables, as well as their temporal ACFs, were
estimated for each case study dataset using the method developed by Urbano-Leon et al. (2023). In the first case study, the
relationship between potential density and dissolved oxygen is quantified, with results indicating that variability in density
may be driven by changes in salinity profiles. The second case study highlights that relationships between temperature and
70 chlorophyll profiles, as well as their temporal ACFs, vary spatially. In summary, this study demonstrates that dependencies
in vertical profile structure can be quantified using a scalar correlation framework. As the availability of profile observations
continues to increase, this technique has broad potential for application across oceanography.

2 Methods

2.1 Scalar correlation for oceanographic profiles

75 Functional datasets are those in which each datum takes the form of a continuous curve or surface which is a function of at
least one other variable. A recent development in FDA includes a method for calculating scalar-valued summary statistics,
specifically the variance and correlation, for functional datasets (Urbano-Leon et al., 2023). The method is described briefly
here, and a simple example is illustrated schematically in Figure 1. Full details and proofs are given in Urbano-Leon et al.
(2023). As background to the formulation below, functional data are represented as linear combinations of orthogonal basis
80 functions, with associated weights known as basis coefficients. Basis functions are smooth reference curves that act as building
blocks for reconstructing profiles. Orthogonality implies that each basis function captures an independent component of the data
that cannot be reproduced by combining the others. Changes in the basis coefficients modify the values of the functional data
and may also change its overall shape. Measuring the variability of the basis coefficients therefore provides a way to quantify
variability in a functional dataset. Extending this to two datasets, together with the covariance of their basis coefficients, yields
85 a natural description of dependence between the datasets.

Suppose that \mathcal{X} and \mathcal{Y} are two datasets each containing n profiles (for example, paired chlorophyll and temperature profiles).
Decompose each profile from \mathcal{X} and \mathcal{Y} into p orthogonal basis functions with $a_{i,j}$ and $b_{i,j}$ denoting the coefficient of the j th
basis function of the i th profile in sets \mathcal{X} and \mathcal{Y} respectively. Calculation of the variability of the basis coefficients requires first
computing their mean values, which for sets \mathcal{X} and \mathcal{Y} are respectively given by

$$90 \quad \overline{A}_j = \frac{1}{n} \sum_{i=1}^n a_{i,j}, \quad \overline{B}_j = \frac{1}{n} \sum_{i=1}^n b_{i,j}. \quad (1)$$

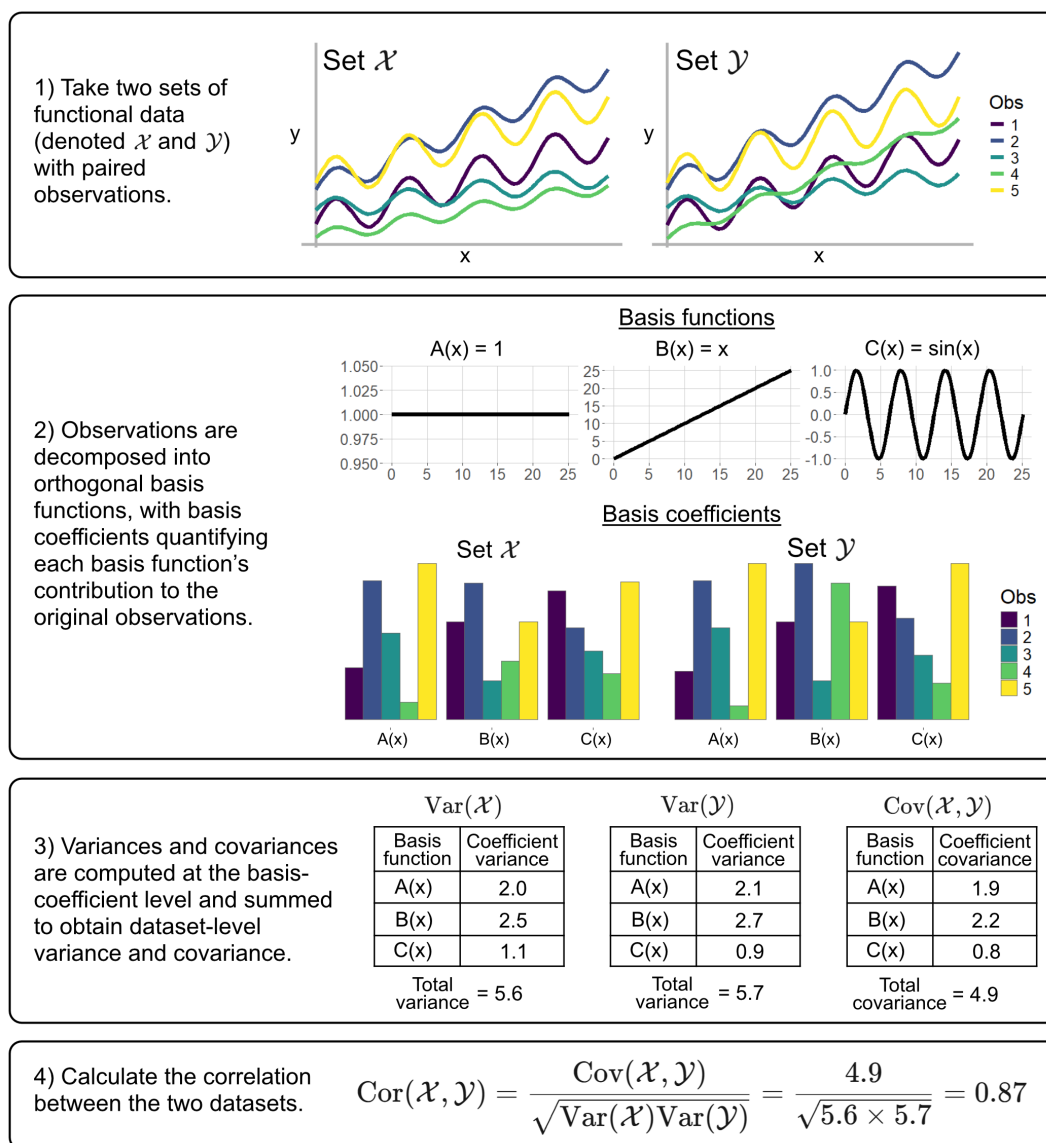


Figure 1. A visual demonstration of the method developed by Urbano-Leon et al. (2023). The figure shows the decomposition of two functional datasets, \mathcal{X} and \mathcal{Y} , into three orthogonal basis functions ($A(x)$, $B(x)$ and $C(x)$) respectively, and the resulting calculation of their correlation. Note that scales of the bar charts were normalised.



The linear combination of the mean basis coefficients represent a mean function for each set, describing a characteristic profile shape. Using these mean basis coefficients, the variance of the coefficients can be calculated for each of the basis functions (denoted V_{a_j} and V_{b_j} respectively). The variances of sets \mathcal{X} and \mathcal{Y} are then simply the sums of the basis coefficient variances V_{a_j} and V_{b_j} respectively.

$$95 \quad \text{Var}(\mathcal{X}) = \sum_{j=1}^p V_{a_j} = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n (a_{i,j} - \overline{A_j})^2 \quad (2)$$

$$\text{Var}(\mathcal{Y}) = \sum_{j=1}^p V_{b_j} = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n (b_{i,j} - \overline{B_j})^2 \quad (3)$$

Similarly, the covariance between sets \mathcal{X} and \mathcal{Y} is defined as the sum of the covariances of each basis coefficient C_j .

$$\text{Cov}(\mathcal{X}, \mathcal{Y}) = \sum_{j=1}^p C_j = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n (a_{i,j} - \overline{A_j})(b_{i,j} - \overline{B_j}) \quad (4)$$

The equation to calculate the correlation between the sets \mathcal{X} and \mathcal{Y} is the same for scalar data.

$$100 \quad \text{Cor}(\mathcal{X}, \mathcal{Y}) = \frac{\text{Cov}(\mathcal{X}, \mathcal{Y})}{\sqrt{\text{Var}(\mathcal{X})\text{Var}(\mathcal{Y})}} \quad (5)$$

The correlation for functional data is restricted to the closed interval $[-1, 1]$, identical to that for scalar data. Mathematically, a correlation of +1 would represent a case where each basis coefficient is perfectly linear and positively correlated. A deviation (from the mean function) in a specific basis component in set \mathcal{X} corresponds to a deviation in the same direction in the equivalent component in the set \mathcal{Y} . In contrast, a correlation of -1 represents a case where each basis is perfectly linear and negatively correlated. A correlation of +1 implies the functional shape is maintained perfectly whereas a correlation of -1 suggests the sets are exactly out of phase, meaning any deviation in a particular component in set \mathcal{X} corresponds to a deviation in the opposite direction from the mean function in set \mathcal{Y} . Note that this does not imply that each pair of functions is a pointwise negative of the other, but instead they vary in opposite directions along the structural features captured by the basis coefficient. Figure 2 shows several examples of datasets containing pairs of oceanographic profiles, and their correlation. The first example displays a near-perfect dependency, meaning almost all deviations from the mean in one set are replicated in the second set. The second shows a slightly weaker positive correlation but a clear relationship is visible by visual inspection. The third shows negligible correlation and the fourth shows a relatively strong negative correlation where deviations from a mean function in the positive direction near the surface in one set corresponds to negative deviations from the mean function near the surface in the other set.

115 2.2 Implementation

In practice, the profiling data from each set were stored in a matrix format, with each column representing a profile and each row representing a depth level. Each profile was converted to basis function coefficients using a fast Fourier transform (FFT). Fourier bases were chosen because many oceanographic profiles vary smoothly with depth, allowing compact representation by

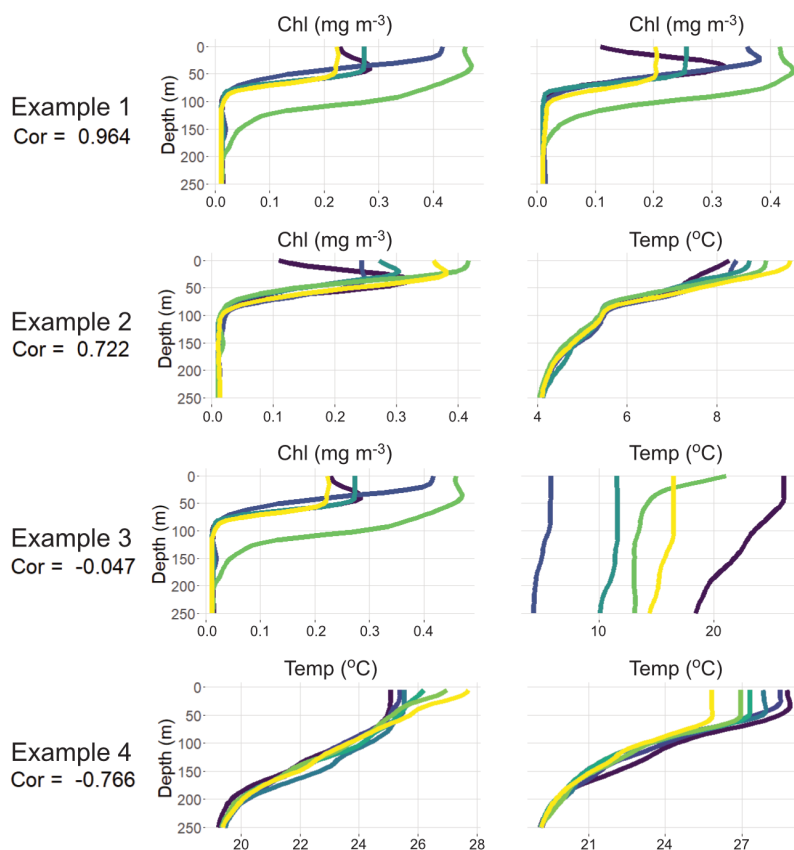


Figure 2. Four examples of paired sets of ocean profiles as functional data and their respective correlations. Colours indicate matching pairs across datasets. Note the direction and relative magnitude of any deviations from a typical profile shape in each profile pair.

sinusoidal functions, and because they performed well in previous work on temperature time series (Urbano-Leon et al., 2023).
120 This transformation requires complete data vectors; therefore, profiles with missing values were removed or gap-filled prior to analysis. Details of the preprocessing steps are provided in the case studies in Section 3. The FFT produces complex-valued coefficients, stored in a matrix where each row now contains coefficients associated with a different basis function. To preserve linearity for subsequent statistical calculations, the variance of each coefficient was defined as the sum of the variances of its real and imaginary components, and the covariance between coefficients as the sum of the covariances of the corresponding
125 real and imaginary parts. This enables direct computation of variance and covariance in the transformed space, whereas some alternative basis expansions yield real-valued coefficients only. Row-wise variances were computed for each coefficient matrix and summed across coefficients to obtain dataset-level variances. Covariances were computed similarly by summing row-wise covariances between paired matrices.

To compute correlations between oceanographic variables, each data matrix contained profiles of a single variable, with
130 simultaneously measured profiles aligned in corresponding columns. Temporal ACFs were estimated by identifying all profile



pairs separated by a specified time lag. For each lag, earlier profiles formed one matrix and later profiles the other, ensuring matched dimensions and indexing. Time lags up to two years (730 days) were evaluated. All analyses were conducted in R version 4.4.1 (R Core Team, 2023). To improve computational efficiency, the search for profile pairs at each temporal lag was implemented in C++, which was substantially faster than the equivalent R implementation, and executed on a workstation with an Intel Core i5 processor. The data for both case studies and code used in the analysis are publicly available at <https://zenodo.org/records/18164579>.

3 Case studies

3.1 Coastal Endurance Washington Offshore Profiler Mooring

3.1.1 Dataset description

140 The Coastal Endurance Washington Offshore Profiler Mooring (CE09OSPM) collected high resolution profiling data from October 2014 to May 2025 at a site 60 km west of Grays Harbor, Washington. Risien et al. (2025) produced a quality controlled, and vertically gridded dataset from these mooring measurements (Figure 3). This dataset comprises 3,244 profiles — each a daily average of up to eight measurements — of temperature, practical salinity, potential density, and dissolved oxygen (DO). It was selected because of its range of physical and biogeochemical variables and its regularly gridded structure in time and
145 depth (0.5 dbar intervals), i.e. it already had the matrix structure described in Section 2.2. Profiles were truncated to 40–400 m due to high rates of missing data outside this depth range. Profiles with any missing values between 40–400 m were removed. As a result, correlation coefficients between physical variables used 3,178 profiles, whereas those involving oxygen used 2,250 profiles. To retain all profile variability, each profile was represented by 721 Fourier basis coefficients, corresponding to each depth measurement in the restricted profiles, prior to the correlation calculations. There are some gaps in the time series due to
150 poor weather and instrument or software failures.

3.1.2 Results

At CE09OSPM, the strongest profile correlations were observed between salinity and potential density (0.90), and between temperature and potential density (-0.79) (Table 1). Potential density and DO were also strongly correlated (-0.75), indicating that oxygen concentrations closely track the vertical structure set by mixing and stratification: as density increases at a
155 given depth, DO tends to decrease. The weakest correlation occurred between temperature and salinity (-0.44), suggesting that density — and thus oxygen variability — is primarily driven by salinity rather than temperature. In turn, this indicates that differences in DO profile shape are governed mainly by salinity-driven changes in water mass structure, which appears appropriate after consulting the similarity in the time series of salinity and DO in Figure 3.

The temporal ACFs of all variables initially decayed exponentially, albeit at different rates, before transitioning to approximately sinusoidal patterns, reflecting seasonal cycles (Figure 4). Temperature exhibited the strongest annual cycle, with $\rho \approx 0.6$
160 at a lag of 365 days and a clear sinusoidal ACF. In contrast, DO had the weakest annual autocorrelation ($\rho \approx 0.22$), indicating

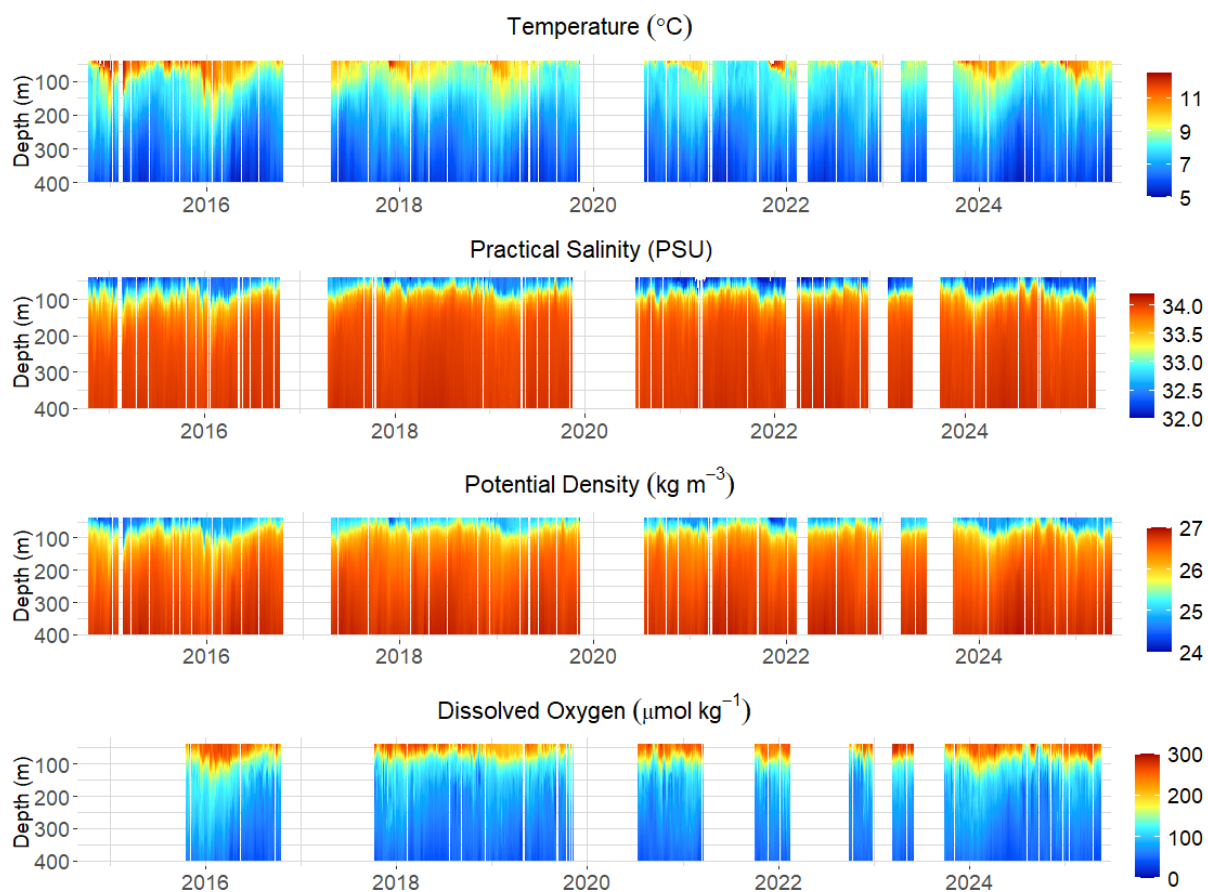


Figure 3. The dataset produced by Risien et al. (2025) comprising time series of temperature, salinity, potential density and dissolved oxygen profiles at the Coastal Endurance Washington Offshore Profiler Mooring (CE09OSPM).

	Temperature	Salinity	Potential density	Dissolved oxygen
Temperature	-	-0.44	-0.79	0.54
Salinity	-0.44	-	0.90	-0.75
Potential density	-0.79	0.90	-	-0.77
Dissolved oxygen	0.54	-0.75	-0.77	-

Table 1. Correlation matrix between profiles of the four oceanographic variables measured at the Coastal Endurance Washington Offshore Profiler Mooring (CE09OSPM).

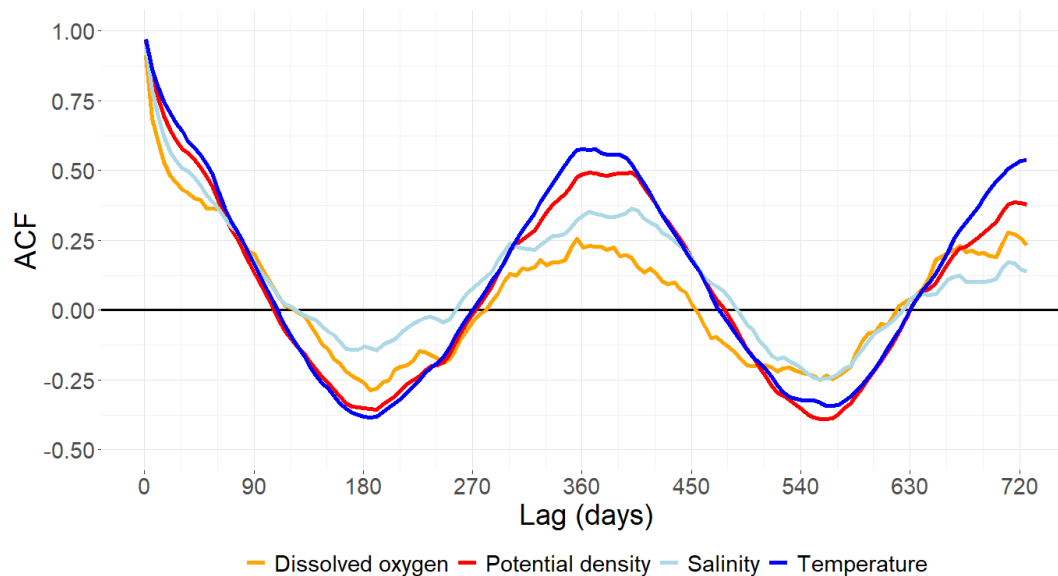


Figure 4. Temporal ACFs for four oceanographic variables measured as profiles at the Coastal Endurance Washington Offshore Profiler Mooring (CE09OSPM).

a less predictable interannual cycle. At short time lags, the DO ACF decayed rapidly, reaching $\rho = 0.6$ after 18 days, whereas temperature took roughly 45 days to reach the same level. These patterns suggest that DO varies over shorter temporal scales than temperature and that its seasonal cycle is weaker and less consistent.

165 3.2 BGC-Argo float profiles

3.2.1 Dataset description

Biogeochemical-Argo (BGC-Argo) floats are autonomous platforms that measure physical and biogeochemical properties throughout the upper 2000 m of the ocean (Claustre et al., 2020). They typically profile every 10 days and drift with the currents, making them semi-Lagrangian. Here, the method was applied to temperature and chlorophyll concentration profiles from seven regularly sampling floats (WMOs 1902385, 4903365, 6901767, 5905107, 5906204, 5904021, and 6901585, shown in Figure 5). The lifespans of floats ranged from 2.83 to 4.98 years. All profiles had undergone standard quality control prior to download (Wong et al., 2025), including correction for non-photochemical quenching of chlorophyll (Schmechtig et al., 2023). Measurements flagged as “probably bad” or “bad” were excluded. Profiles with fewer than 20 measurements between 5 m and 250 m, or those not spanning at least between 20m and 230 m, were removed. Remaining profiles were regrided to 5 m intervals from 5 m to 250 m using linear interpolation, then smoothed using a 15 m moving-median window. Each processed profile was represented by 51 Fourier basis coefficients, one for each regrided depth level.

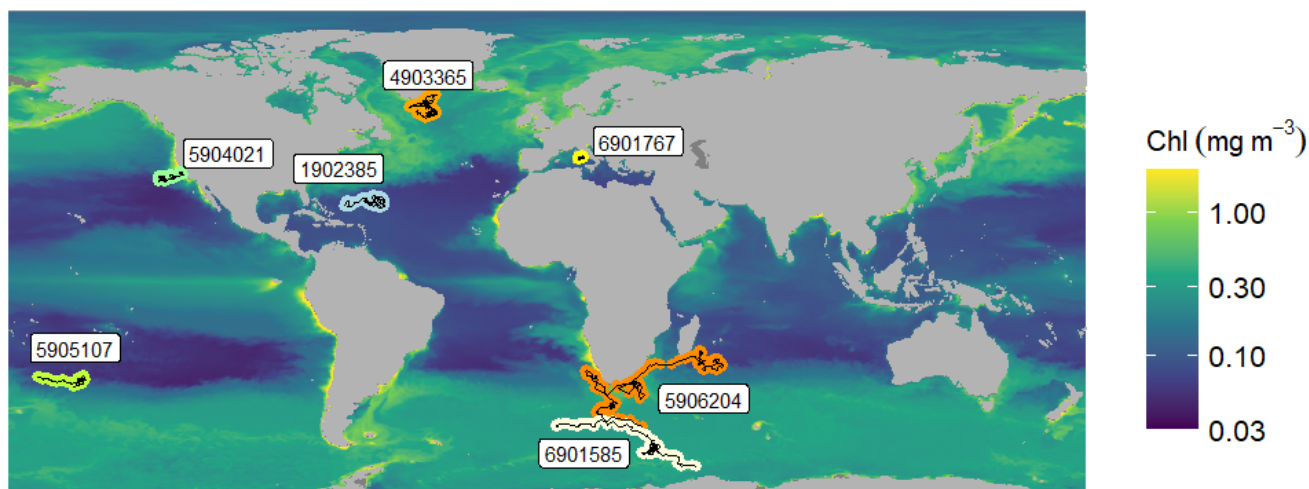


Figure 5. Trajectories of seven BGC-Argo floats used in the second case study. The background map shows the mean surface chlorophyll concentration during 2024, indicating regions of higher productivity.

3.2.2 Results

The correlation between chlorophyll and temperature profiles varied among BGC-Argo floats, both in magnitude and in sign (Table 2). For example, floats 4903365, 6901767, and 1902385 showed moderate positive (0.47), moderate negative (−0.42), and negligible (0.02) correlations, respectively, and these floats were located in different ocean basins. This suggests that the coupling between temperature and chlorophyll profiles may be location dependent. Floats that moved a significant distance during their lifespans (5906204, 5904021 and 6901585) also gave weak to moderate correlations, indicating that floats moving between regions may not affect the strength of correlation between chlorophyll and temperature. The negligible correlation for float 1902385 may be due to year round subsurface chlorophyll maxima despite seasonal changes in surface temperature (Figure 6a).

The temporal ACFs of chlorophyll and temperature profiles varied across floats (Figure 6a-d), with four floats (WMOs 1902385, 4903365, 6901767, and 5905107) showing sinusoidal ACFs for both variables, indicating strong annual cycles. Temperature typically had a stronger seasonal cycle than chlorophyll, with correlation coefficient between 0.55 and 0.9 after a lag of one year, whereas chlorophyll had correlations between 0 and 0.6 after one year. In contrast, floats 5906204, 5904021, and 6901585 did not show sinusoidal ACFs (Figure 6e-g). These floats travelled long distances across water masses, and their temperature ACFs remained above 0.5 for far longer than a seasonal cycle (up to a year for float 6901585). This highlights that float trajectories might influence ACF structure and decorrelation timescales, and that spatial variability might obscure seasonal signals. Over shorter time lags (< 90 days), chlorophyll nearly always had a lower ACF than temperature, indicating a higher proportion of its variability occurs over subseasonal scales than temperature (Figure 6).

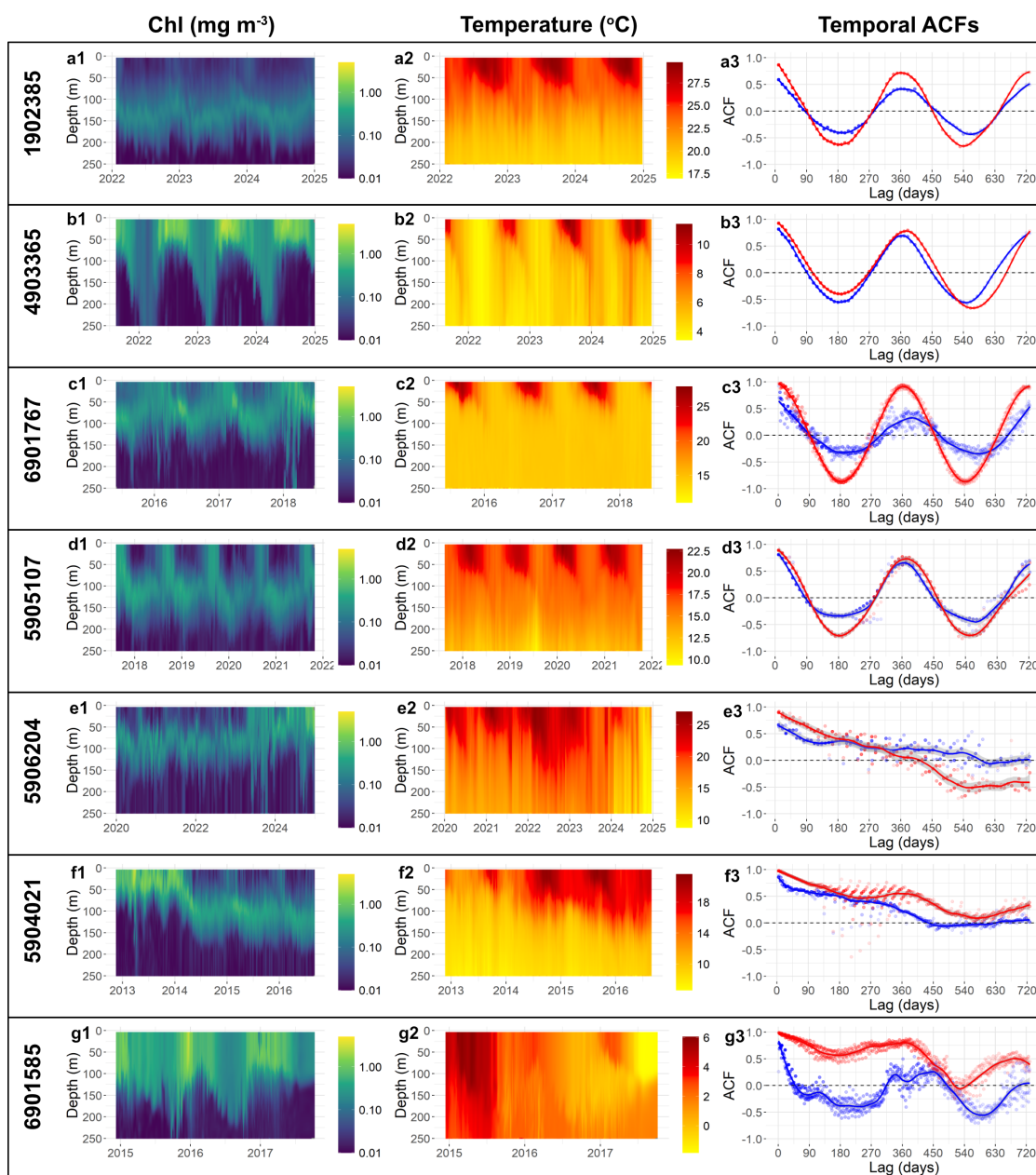


Figure 6. Temporal ACFs of chlorophyll and temperature profiles from a selection of seven BGC-Argo floats. **First and second columns:** semi-Lagrangian sections over the floats’ lifespan of chlorophyll, and temperature, respectively. **Third column:** smooth curves showing the temporal ACFs for temperature (red) and chlorophyll (blue). Point opacity is lower for lags with fewer pairings and the ACFs are weighted towards points with more pairs. ACFs with scattered points indicate greater irregularity in sampling times.



Float ID	Mean latitude (°N)	Lifespan (yrs)	No. of profiles	Cor(Chl,Temp)
1902385	25.6	2.93	108	0.02
4903365	57.1	3.37	124	0.47
6901767	40.6	3.07	206	-0.42
5905107	-34.0	4.20	139	-0.42
5906204	-31.0	4.98	175	-0.34
5904021	33.8	3.79	267	-0.30
6901585	-54.8	2.83	225	0.16

Table 2. Correlation between chlorophyll and temperature profiles collected by seven BGC-Argo floats.

195 4 Discussion

4.1 Profile correlation between multiple variables

The proposed method extends existing statistical tools for oceanographic profile analysis by summarising the similarity between two sets of profiles with a single correlation coefficient, analogous to the scalar-valued case. Importantly, the coefficient reflects both profile shape as well as the absolute magnitude, so dependencies in either characteristic are captured. Given the widespread familiarity with correlation coefficients, this approach provides a natural extension of scalar correlation analyses to full vertical profiles.

In an exploratory data analysis context, the method enables rapid identification of potential dependencies among multiple oceanographic variables and can inform subsequent modelling decisions. In the first case study, simultaneously measured profiles of temperature, salinity, potential density, and DO were analysed. The resulting correlation matrix (Table 1) suggests that salinity-driven density variability is a primary driver of DO variability at the mooring site. The high correlations among hydrographic variables, notably between salinity and potential density, indicate multicollinearity, implying that a model predicting DO profiles may not require all three predictors. These high correlations may be expected given the variables are connected through well established physical relationships (Sharqawy et al., 2010). The second case study shows that the correlation between chlorophyll and temperature varies spatially (Table 2). The difference in correlation coefficients suggests that region specific analyses are required, as a single relationship may not hold across locations. In some regions, factors other than water column structure—such as nutrient availability or light—may exert stronger control on chlorophyll profiles (Cullen, 2015; Uitz et al., 2006; Cornec et al., 2021a).

4.2 Temporal autocorrelation of profiles

Oceanographers frequently investigate the temporal scales of variability in oceanographic time series, for which ACFs are a standard diagnostic tool. Previous approaches have produced ACFs for measurements collected at the surface (Meyers et al., 1991; Delcroix et al., 2005) and within specific depth bins (Sumata et al., 2018). In addition, modes of variability have been



identified using empirical orthogonal functions (EOFs) (Hjelmervik and Hjelmervik, 2014; Bock et al., 2022) or by computing seasonal climatologies (Risien et al., 2025). The advantage of this approach is the quantification of temporal autocorrelation at specific time lags for time series of entire profiles. For example, in the first case study, it is shown that the autocorrelation of DO decays significantly faster than hydrographic variables for the shortest time lags (Figure 4). When combined with weaker annual autocorrelation, this suggests that variability over short time scales is relatively more important for DO profiles, as has been shown for surface chlorophyll measurements (Prend et al., 2022). The second case study builds on this by showing that the similarity between the ACFs of two oceanographic variables can vary spatially. Furthermore, the fact that BGC-Argo floats can move substantial distances over time can lead to temporal variability being dominated by spatial variability. This is demonstrated in Figures 6e-g where the ACFs do not display a sinusoidal curve. Visual inspection of the corresponding time series (Figures 6e-g) confirms that that seasonal cycles are not present and sinusoidal ACFs would not be expected.

4.3 Limitations

Several caveats accompany this approach. First, it captures only linear dependence between profiles, so non linear relationships will not be detected. Second, it does not characterise relationships between functional variables, which requires functional regression models (Ramsay and Silverman, 2005; Morris, 2015). Interpreting correlation scores can also be challenging because the coefficient alone gives no indication of which depths contribute to the relationship, so this method should be used alongside other tools when investigating drivers of profile variability. The degree of profile smoothing may influence autocorrelation estimates, and the method assumes identical vertical resolution and coverage across profiles, often requiring preprocessing or interpolation. An appropriate set of basis functions needs to be chosen; the authors recommend implementing a Fourier basis because they produce smooth function estimates, whilst conveying variation over a range of spatial scales. In addition, long time lag estimates may be affected by sensor drift, which is well documented for autonomous platforms (Wong et al., 2025). Finally, the method becomes less effective for identifying long-term autocorrelation when profiling platforms move large distances, particularly between distinct oceanic regions with different hydrographic or ecological characteristics (Figure 5).

4.4 Potential for future usage

The scalar correlation framework has broad potential across oceanographic research. It can be applied to the growing collection of datasets from moorings, profiling floats and gliders, and to model output or reanalyses to evaluate coupled physical–biogeochemical variability. As profiling data from autonomous platforms continue to expand, this method offers a way to quantify spatio-temporal autocorrelation for multiple variables and across a wide range of scales. The approach is well suited to parallelisation, making it feasible for analysing high-resolution ocean model output or global observational archives.

This framework could support observing system design for programs such as the (BGC-)Argo array (Chamberlain et al., 2023; Mignot et al., 2023; Chu et al., 2024), and extend previous decorrelation length scale studies (Mazloff et al., 2018) by incorporating full vertical profile structure. It could also be applied to high resolution glider datasets (Testor et al., 2019) to evaluate efficient sampling strategies (Patmore et al., 2024) or to compare and cross calibrate nearby platforms. Combined with



250 satellite products such as geostrophic velocities (McKee et al., 2022), it may help assess the role of advection in driving profile variability. Alternatively, it could be beneficial to assess small scale processes through the autocorrelation of climatological anomalies. The method could also be used to validate output from machine learning algorithms (Sauzède et al., 2016; Chen et al., 2022; Pietropoli et al., 2023) or to examine the relationships between the vertical distribution of different phytoplankton communities (Brewin et al., 2022; Miyares et al., 2024).

255 Future developments could include extending this approach to quantify spatial dependence in vertical profiles within spatial or spatio-temporal models (Yarger et al., 2022). The framework could also be integrated with functional regression and clustering methods to identify coherent oceanic regimes (Korte-Stapff et al., 2022; Bock et al., 2022). In addition, the autocorrelation framework could be used to fit functional time series models, such as functional autoregressive models (Chen et al., 2021).

5 Conclusions

260 This study applied the scalar correlation framework for functional data developed by Urbano-Leon et al. (2023) to oceanographic profiling datasets. In this framework, profile variability is represented through the variability of basis function coefficients. This enables the calculation of a scalar-valued correlation coefficient that accounts describes the dependence of profile shape between sets of profiles. Application to two case studies containing profiling data from a stationary mooring and from BGC-Argo floats respectively, demonstrated that the method can reveal physically meaningful relationships and spatial
265 heterogeneity in profile dependencies. Temporal autocorrelation analyses illustrated that different variables exhibit distinct decorrelation timescales and that spatial variability can dominate temporal variability for mobile platforms.

The method has broad potential applications across oceanographic research. It can be applied to the expanding global collection of profiling observations, as well as to ocean model output, to quantify spatio-temporal variability and coupled physical–biogeochemical dynamics. Potential uses include observing system design, evaluation of sampling strategies, assessment
270 of advective versus local variability, validation of machine learning products, and investigation of ecosystem scale decorrelation timescales. A variety of methodological approaches could be integrated with this technique, such as functional regression, clustering, spatial modelling or time series analyses. Together, these extensions would further enhance the ability to quantify and interpret variability in the rapidly growing volume of oceanographic profiling data.

Code and data availability. The datasets and R code used in this study are available on Zenodo (Taylor and Henson, 2026).

275 *Author contributions.* MT conceptualised the work, conducted the analysis and produced the first draft of the manuscript. MT and SH interpreted the results and edited the manuscript. All authors proof read the manuscript before submission.



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

Acknowledgements. Cristhian Leonardo Urbano Leon provided advice on how to interpret the correlation for functional data. This work was supported by the Natural Environmental Research Council (grant number NE/S007210/1). ChatGPT was used in places to improve clarity and grammar.

280



References

- Abbott, M. R. and Letelier, R. M.: Decorrelation scales of chlorophyll as observed from bio-optical drifters in the California Current, *Deep Sea Research Part II: Topical Studies in Oceanography*, 45, 1639–1667, 1998.
- Alhassan, Y., Siekmann, I., and Petrovskii, S.: Mathematical model of oxygen minimum zones in the vertical distribution of oxygen in the ocean, *Scientific Reports*, 14, 22 248, 2024.
- Barbieux, M., Uitz, J., Gentili, B., Pasqueron de Fommervault, O., Mignot, A., Poteau, A., Schmechtig, C., Taillandier, V., Leymarie, E., Penkerch, C., et al.: Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database, *Biogeosciences*, 16, 1321–1342, 2019.
- Bock, N., Cornec, M., Claustre, H., and Duhamel, S.: Biogeographical Classification of the Global Ocean From BGC-Argo Floats, *Global biogeochemical cycles*, 36, e2021GB007 233, 2022.
- Brewin, R. J., Dall’Olmo, G., Gittings, J., Sun, X., Lange, P. K., Raitos, D. E., Bouman, H. A., Hoteit, I., Aiken, J., and Sathyendranath, S.: A conceptual approach to partitioning a vertical profile of phytoplankton biomass into contributions from two communities, *Journal of Geophysical Research: Oceans*, 127, e2021JC018 195, 2022.
- Burke, M., Grant, J., Filgueira, R., and Sheng, J.: Temporal and spatial variability in hydrography and dissolved oxygen along southwest Nova Scotia using glider observations, *Continental Shelf Research*, 254, 104 908, 2023.
- Cao, H., Freilich, M., Song, X., Jing, Z., Fox-Kemper, B., Qiu, B., Hetland, R. D., Chai, F., Ruiz, S., and Chen, D.: Isopycnal sub-mesoscale stirring crucially sustaining subsurface chlorophyll maximum in ocean cyclonic eddies, *Geophysical Research Letters*, 51, e2023GL105 793, 2024.
- Carranza, M. M., Gille, S. T., Franks, P. J., Johnson, K. S., Pinkel, R., and Girton, J. B.: When mixed layers are not mixed. Storm-driven mixing and bio-optical vertical gradients in mixed layers of the Southern Ocean, *Journal of Geophysical Research: Oceans*, 123, 7264–7289, 2018.
- Castro-Morales, K. and Kaiser, J.: Using dissolved oxygen concentrations to determine mixed layer depths in the Bellingshausen Sea, *Ocean Science*, 8, 1–10, 2012.
- Chai, F., Johnson, K. S., Claustre, H., Xing, X., Wang, Y., Boss, E., Riser, S., Fennel, K., Schofield, O., and Sutton, A.: Monitoring ocean biogeochemistry with autonomous platforms, *Nature Reviews Earth & Environment*, 1, 315–326, 2020.
- Chamberlain, P., Talley, L. D., Cornuelle, B., Mazloff, M., and Gille, S. T.: Optimizing the biogeochemical Argo float distribution, *Journal of Atmospheric and Oceanic Technology*, 40, 1355–1379, 2023.
- Chen, J., Gong, X., Guo, X., Xing, X., Lu, K., Gao, H., and Gong, X.: Improved Perceptron of Subsurface Chlorophyll Maxima by a Deep Neural Network: A Case Study with BGC-Argo Float Data in the Northwestern Pacific Ocean, *Remote Sensing*, 14, 632, 2022.
- Chen, Y., Koch, T., Lim, K. G., Xu, X., and Zakiyeva, N.: A review study of functional autoregressive models with application to energy forecasting, *Wiley Interdisciplinary Reviews: Computational Statistics*, 13, e1525, 2021.
- Chu, W. U., Mazloff, M. R., Verdy, A., Purkey, S. G., and Cornuelle, B. D.: Optimizing observational arrays for biogeochemistry in the tropical Pacific by estimating correlation lengths, *Limnology and Oceanography: Methods*, 22, 840–852, 2024.
- Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the global ocean with biogeochemical-Argo, *Annual review of marine science*, pp. 23–48, 2020.
- Cornec, M., Claustre, H., Mignot, A., Guidi, L., Lacour, L., Poteau, A., d’Ortenzio, F., Gentili, B., and Schmechtig, C.: Deep chlorophyll maxima in the global ocean: Occurrences, drivers and characteristics, *Global Biogeochemical Cycles*, 35, e2020GB006 759, 2021a.



- Cornec, M., Laxenaire, R., Speich, S., and Claustre, H.: Impact of mesoscale eddies on deep chlorophyll maxima, *Geophysical research letters*, 48, e2021GL093470, 2021b.
- 320 Cullen, J. J.: Subsurface chlorophyll maximum layers: enduring enigma or mystery solved?, *Annual Review of Marine Science*, 7, 207–239, 2015.
- de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, *Journal of Geophysical Research: Oceans*, 109, 2004.
- Delcroix, T., McPhaden, M. J., Dessier, A., and Gouriou, Y.: Time and space scales for sea surface salinity in the tropical oceans, *Deep Sea Research Part I: Oceanographic Research Papers*, 52, 787–813, 2005.
- 325 Fay, A. R. and McKinley, G. A.: Correlations of surface ocean pCO₂ to satellite chlorophyll on monthly to interannual timescales, *Global Biogeochemical Cycles*, 31, 436–455, 2017.
- Fennel, K. and Boss, E.: Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model, *Limnology and Oceanography*, 48, 1521–1534, 2003.
- 330 Ford, D.: Assimilating synthetic Biogeochemical-Argo and ocean colour observations into a global ocean model to inform observing system design, *Biogeosciences*, 18, 509–534, 2021.
- Hjelmervik, K. and Hjelmervik, K. T.: Time-calibrated estimates of oceanographic profiles using empirical orthogonal functions and clustering, *Ocean Dynamics*, 64, 655–665, 2014.
- Jayaram, C., Bhaskar, T. U., Chacko, N., Prakash, S., and Rao, K.: Spatio-temporal variability of chlorophyll in the northern Indian Ocean: A biogeochemical argo data perspective, *Deep Sea Research Part II: Topical Studies in Oceanography*, 183, 104928, 2021.
- 335 Kahru, M., Gille, S., Murtugudde, R., Strutton, P. G., Manzano-Sarabia, M., Wang, H., and Mitchell, B. G.: Global correlations between winds and ocean chlorophyll, *Journal of Geophysical Research: Oceans*, 115, 2010.
- Kande, Y., Diogoul, N., Brehmer, P., Dabo-Niang, S., Ngom, P., and Perrot, Y.: Demonstrating the relevance of spatial-functional statistical analysis in marine ecological studies: The case of environmental variations in micronektonic layers, *Ecological Informatics*, 81, 102547, 2024.
- 340 Korte-Stapff, M., Yarger, D., Stoev, S., and Hsing, T.: A multivariate functional-data mixture model for spatio-temporal data: inference and cokriging, *arXiv preprint arXiv:2211.04012*, 2022.
- Lovecchio, E., Henson, S., Carvalho, F., and Briggs, N.: Oxygen variability in the offshore northern Benguela Upwelling System from glider data, *Journal of Geophysical Research: Oceans*, 127, e2022JC019063, 2022.
- 345 Mazloff, M., Cornuelle, B., Gille, S., and Verdy, A.: Correlation lengths for estimating the large-scale carbon and heat content of the Southern Ocean, *Journal of Geophysical Research: Oceans*, 123, 883–901, 2018.
- McGillicuddy Jr, D. J., Kosnyrev, V., Ryan, J., and Yoder, J.: Covariation of mesoscale ocean color and sea-surface temperature patterns in the Sargasso Sea, *Deep Sea Research Part II: Topical Studies in Oceanography*, 48, 1823–1836, 2001.
- McKee, D. C., Doney, S. C., Della Penna, A., Boss, E. S., Gaube, P., Behrenfeld, M. J., and Glover, D. M.: Lagrangian-Eulerian statistics of mesoscale ocean chlorophyll from Bio-Argo floats and satellites, *Biogeosciences Discussions*, pp. 1–36, 2022.
- 350 Meyers, G., Phillips, H., Smith, N., and Sprintall, J.: Space and time scales for optimal interpolation of temperature—Tropical Pacific Ocean, *Progress in Oceanography*, 28, 189–218, 1991.
- Mignot, A., Claustre, H., Cossarini, G., D’Ortenzio, F., Gutknecht, E., Lamouroux, J., Lazzari, P., Perruche, C., Salon, S., Sauzède, R., et al.: Using machine learning and Biogeochemical-Argo (BGC-Argo) floats to assess biogeochemical models and optimize observing system design, *Biogeosciences*, 20, 1405–1422, 2023.
- 355



- Mirouze, I., Blockley, E. W., Lea, D. J., Martin, M. J., and Bell, M. J.: A multiple length scale correlation operator for ocean data assimilation, *Tellus A: Dynamic Meteorology and Oceanography*, 68, 29 744, 2016.
- Miyares, M. E., Latasa, M., Cabello, A. M., de la Fuente, P., Guallar, C., Mozetič, P., Riera-Lorente, M., Vidal, M., and Blasco, D.: Relationships between the deep chlorophyll maximum and hydrographic characteristics across the Atlantic, Indian and Pacific oceans, *Scientia Marina*, 88, e092–e092, 2024.
- 360 Morris, J. S.: Functional regression, *Annual Review of Statistics and Its Application*, 2, 321–359, 2015.
- Patmore, R. D., Ferreira, D., Marshall, D. P., du Plessis, M. D., Brearley, J. A., and Swart, S.: Evaluating existing ocean glider sampling strategies for submesoscale dynamics, *Journal of Atmospheric and Oceanic Technology*, 41, 647–663, 2024.
- Pietropolli, G., Manzoni, L., and Cossarini, G.: Multivariate relationship in big data collection of ocean observing system, *Applied Sciences*, 365 13, 5634, 2023.
- Prend, C. J., Keerthi, M. G., Lévy, M., Aumont, O., Gille, S. T., and Talley, L. D.: Sub-Seasonal Forcing Drives Year-To-Year Variations of Southern Ocean Primary Productivity, *Global Biogeochemical Cycles*, 36, e2022GB007 329, 2022.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, 2023.
- 370 Ramsay, J. and Silverman, B.: *Functional Data Analysis*, Springer, 2005.
- Risien, C. M., Desiderio, R. A., Fram, J. P., and Dever, E. P.: Gridded, high-resolution ocean observatories initiative profiler data from the Washington continental slope, 2014–2025, *Data in Brief*, 61, 111 861, 2025.
- Rudnick, D. L.: Ocean research enabled by underwater gliders, *Annual review of marine science*, 8, 519–541, 2016.
- Sauzède, R., Claustre, H., Uitz, J., Jamet, C., Dall’Olmo, G., d’Ortenzio, F., Gentili, B., Poteau, A., and Schmechtig, C.: A neural network-
375 based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient, *Journal of Geophysical Research: Oceans*, 121, 2552–2571, 2016.
- Schmechtig, C., Claustre, H., Poteau, A., D’Ortenzio, F., Schallenberg, C., Trull, T., and Xing, X.: Biogeochemical-Argo quality control manual for chlorophyll-a concentration and chl-fluorescence, Version 3.0, 2023.
- Sharqawy, M. H., Lienhard, J. H., and Zubair, S. M.: Thermophysical properties of seawater: a review of existing correlations and data,
380 *Desalination and water treatment*, 16, 354–380, 2010.
- Steele, J.: A study of production in the Gulf of Mexico, *Journal of Marine Research*, 1964.
- Storto, A., Oddo, P., Cipollone, A., Mirouze, I., and Lemieux-Dudon, B.: Extending an oceanographic variational scheme to allow for affordable hybrid and four-dimensional data assimilation, *Ocean Modelling*, 128, 67–86, 2018.
- Strutton, P. G., Trull, T. W., Phillips, H. E., Duran, E. R., and Pump, S.: Biogeochemical Argo floats reveal the evolution of subsurface
385 chlorophyll and particulate organic carbon in southeast Indian Ocean eddies, *Journal of Geophysical Research: Oceans*, p. e2022JC018984, 2023.
- Sumata, H., Kauker, F., Karcher, M., Rabe, B., Timmermans, M.-L., Behrendt, A., Gerdes, R., Schauer, U., Shimada, K., Cho, K.-H., et al.: Decorrelation scales for arctic ocean hydrography—part i: Amerasian basin, *Ocean Science*, 14, 161–185, 2018.
- Taylor, M. and Henson, S.: Data and code accompanying "Correlation for oceanographic profiling datasets",
390 <https://doi.org/10.5281/zenodo.18164579>, 2026.
- Testor, P., De Young, B., Rudnick, D. L., Glenn, S., Hayes, D., Lee, C. M., Pattiaratchi, C., Hill, K., Heslop, E., Turpin, V., et al.: OceanGliders: a component of the integrated GOOS, *Frontiers in Marine Science*, 6, 422, 2019.



- Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, *Journal of Geophysical Research: Oceans*, 111, 2006.
- 395 Urbano-Leon, C. L., Escabias, M., Ovalle-Muñoz, D. P., and Olaya-Ochoa, J.: Scalar Variance and Scalar Correlation for Functional Data, *Mathematics*, 11, 1317, 2023.
- Whitt, C., Pearlman, J., Polagye, B., Caimi, F., Muller-Karger, F., Copping, A., Spence, H., Madhusudhana, S., Kirkwood, W., Grosjean, L., et al.: Future vision for autonomous ocean observations, *Frontiers in Marine Science*, 7, 697, 2020.
- Wong, A., Keeley, R., and Carval, T.: Argo quality control manual for CTD and trajectory data, 2025.
- 400 Wong, A. P., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., et al.: Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats, *Frontiers in Marine Science*, 7, 700, 2020.
- Xu, W., Wang, G., Cheng, X., Jiang, L., Zhou, W., and Cao, W.: Characteristics of subsurface chlorophyll maxima during the boreal summer in the South China Sea with respect to environmental properties, *Science of The Total Environment*, 820, 153 243, 2022.
- 405 Yarger, D., Stoev, S., and Hsing, T.: A functional-data approach to the Argo data, *The Annals of Applied Statistics*, 16, 216–246, 2022.