# The Thermohaline Modes of the Global Ocean

Etienne Pauthenet, Fabien Roquet, Gurvan Madec, Jean-Baptiste Sallée,
David Nerini

## ▶ To cite this version:

# The Thermohaline Modes of the Global Ocean

Etienne Pauthenet

*Sorbonne Universités, UPMC Université, Paris 06, UMR 7159, LOCEAN-IPSL, Paris, France*

Fabien Roquet

*Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden*

Gurvan Madec and Jean-Baptiste Sallée

*Sorbonne Universités, UPMC Université, Paris 06, UMR 7159, LOCEAN-IPSL, Paris, France*

David Nerini

*Aix-Marseille Université, CNRS/INSU, Université de Toulon, IRD, Mediterranean Institute of Oceanography UM 110, Marseille, France*

## ABSTRACT

The first 2000 m of the global thermohaline structure of the ocean are statistically decomposed into vertical thermohaline modes, using a multivariate functional principal component analysis (FPCA). This method is applied on the Monthly Isopycnal and Mixed-Layer Ocean Climatology (MIMOC). The first three modes account for 92% of the joint temperature and salinity ($T$–$S$) variance, which yields a surprisingly good reduction of dimensionality. The first mode (69% of the variance) is related to the thermocline depth and delineates the subtropical gyres. The second mode (18%) is mostly driven by salinity and mainly displays the asymmetry between the North Pacific and Atlantic basins and the salty circumpolar deep waters in the Southern Ocean. The third mode (5%) identifies the low- and high-salinity intermediate waters, covarying with the freshwater inputs of the upper ocean. The representation of the ocean in the space defined by the first three modes offers a simple visualization of the global thermohaline structure that strikingly emphasizes the role of the Southern Ocean in linking and distributing water masses to the other basins. The vertical thermohaline modes offer a convenient framework for model and observation data comparison. This is illustrated by projecting the repeated Pacific section P16 together with profiles from the Array for Real-Time Geostrophic Oceanography (ARGO) global array of profiling floats on the modes defined with the climatology MIMOC. These thermohaline modes have a potential for water mass identification and robust analysis of heat and salt content.

## 1. Introduction

The global ocean waters are traditionally divided into distinct water masses defined by their origin, their physicochemical properties (in particular their temperature and salinity), and their vertical position. Tracking the position and properties of water masses provides a powerful way for monitoring the ocean circulation and climate variability. The ocean, seen as a network of numerous water masses that are formed, transformed, mixed, and subducted, is however complex to display. A map or section of a physical property can either be represented at a selected depth/longitude/latitude, at a selected minimum or maximum of properties, or at a selected level such as neutral density levels (Jackett and McDougall 1997). The temperature–salinity ($T$–$S$) diagram is a great tool to represent the water masses of the global ocean (Worthington 1981) as the mixing

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JPO-D-19-0120.s1.

*Corresponding author*: Etienne Pauthenet, etienne.pauthenet@locean-ipsl.upmc.fr
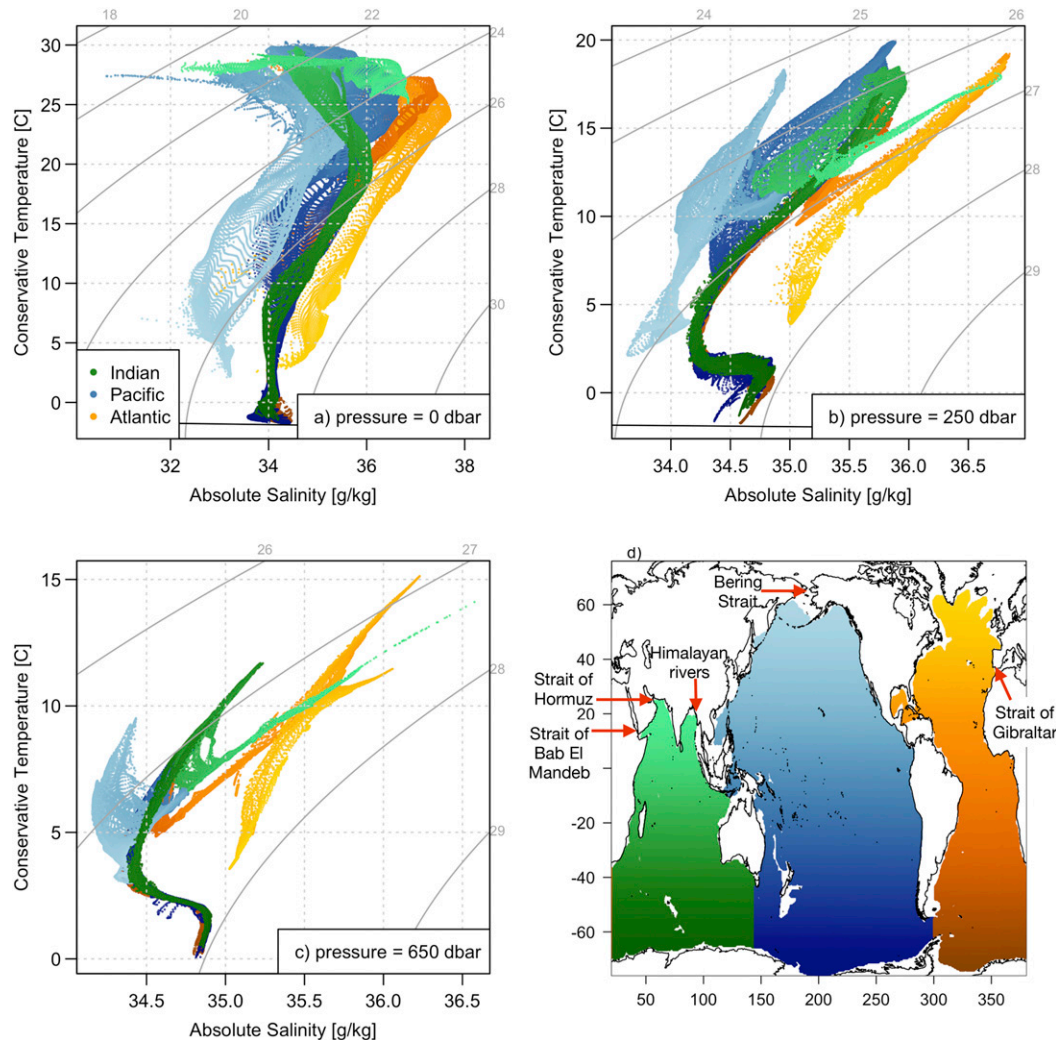
FIG. 1. Temperature–salinity diagram for the (a) surface, (b) 250 dbar, and (c) 650 dbar of the MIMOC climatology. The *T–S* distribution in the intermediate layer in (c) can be closely related to the PC2/PC1 distribution (Fig. 4a). The gray lines are the isopycnals (kg m$^{-3}$), and the horizontal black line is the freezing point of seawater. The data are plotted as follows: Atlantic (orange), Pacific (blue), and Indian (green) with a south–north (dark–light) gradient in (d). The red arrows point to regions also referenced in the Fig. 4.

processes are assumed to occur along straight lines, which facilitates the identification of water mass sources (Fig. 1). The *T–S* space also offers a framework on which the water mass transformations (e.g., Speer 1993; Hieronymus et al. 2014) and the global overturning circulation (e.g., Zika et al. 2012; Döös et al. 2012) can be represented and quantified. However the water masses from different depths and/or regions tend to strongly overlap in *T–S* space (see, e.g., Fig. 1 in Groeskamp et al. 2019). McDougall and Jackett (2007) underline that the ocean occupies little volume in a salinity–temperature–pressure (*S–T–p*) space. Particularly, a neutral density surface becomes a single curve in *S–T–p* space. This simple structure emerges from the

seawater properties and underlines the correlation between properties, as most water masses acquire their characteristics at the surface in specific areas.

In this paper we want to take advantage of the correlations between temperature and salinity at different depths to find the main modes of the thermohaline structure. These modes define a projection in which the ocean thermohaline structure can be displayed. A graphical representation in a reduced dimension space is useful to detect and understand important phenomena as well as to remember and communicate them (Chernoff 1973). A different projection of a dataset can also provide interesting analytical properties, such as identifying outliers or classifying the data. Here we use

a principal component analysis (PCA), also named an empirical orthogonal function (EOF), which is the most standard method in multivariate analysis (Jolliffe and Cadima 2016). A PCA projects the data onto a subspace of reduced dimension that maximizes the variance. PCA has been used to handle several oceanographic variables in order to delineate water masses (de Boer and van Aken 1995; Lindegren and Josefson 1998). Lindegren and Josefson (1998) mentioned the superiority of visualizing the data in the main principal plans of the PCA, instead of the large number of classical two-variable scatterplots. Other studies applied PCA on temperature profiles in the North Atlantic (Maze et al. 2017) and the Southern Ocean (Jones et al. 2019). However they do not analyze the resulting principal components (PCs) distribution but classified them to identify coherent heat patterns and frontal positions. Chang et al. (2011) computed the vertical coupled T–S EOF modes from observations, and estimated salinity profiles from independent temperature and altimetry data by using a weighted least squared procedure to minimize the misfits between the EOF modes and the temperature and altimetry data. Their study aimed at reconstructing salinity at global scale and not at analyzing the T–S EOF patterns. Hjelmervik and Hjelmervik (2013, 2019) also computed the vertical coupled T–S EOF modes and applied an edge detection on the maps of EOF coefficients to locate oceanic fronts in the Arctic and in the North Atlantic. They especially presented a simple way to compare profiles of variable depths in the same EOF framework. Pauthenet et al. (2017) applied a functional PCA (FPCA; Ramsay and Silverman 2005) on T–S profiles from a gridded product (SOSE; Mazloff et al. 2010) and successfully delineated the Southern Ocean circumpolar fronts separating the different regimes of stratification. To compare the vertical shape of thermohaline profiles within the same framework, profiles are first normalized using a suitable basis of B-splines. The PCA is then applied on the coefficients of the B-spline decomposition, hence the term functional.

In this paper we present and describe the modes of the global ocean thermohaline structure down to 2000 m, computed with the method presented in Pauthenet et al. (2017). First, we describe the method and how it is applied on a climatology of temperature and salinity (section 2). In section 3, we give a physical interpretation and description of the main vertical modes' shape. The first three vertical modes contain more than 92% of the T–S variance and define a reduced dimension space on which the thermohaline structure of the ocean can be visualized efficiently. We explore further how the main modes capture the heat and salt content and the dynamic height anomaly. Section 4 describes the spatial distribution of the associated PCs. We finish with concluding remarks, outlining the potential of the method for the study of the ocean T–S variability (section 5 and 6). A list of the acronyms used is provided in the appendix.

## 2. Data and method

### a. MIMOC climatology

The dataset used is Conservative Temperature and Absolute Salinity[1] (T–S) monthly averaged fields from the Monthly Isopycnal and Mixed-Layer Ocean Climatology version 2.2 (MIMOC; Schmidtko et al. 2013). We use the monthly averaged profiles to compute the modes but only present the annual average because the seasonal variations are small on the main modes. These are maps constructed mostly with ARGO float profiles, supplemented by shipboard data, provided at a lateral resolution of 0.5° and with 81 vertical pressure levels from the surface (0 dbar) to 1950 dbar. Our method requires profiles of fixed depth, which constrains us to remove all profiles that are not ranging from 0 to 1950 dbar. We also exclude the fresh Arctic Ocean and the salty Mediterranean Sea, which pull the information away from the main ocean basin, as the PCA is sensitive to outliers. These profiles are excluded by computing the connected component transform (Rosenfeld and Pfaltz 1966) of the binary mask, and by keeping only the profiles that are topologically connected to the global ocean basin. The algorithm also removes marginal seas in the west equatorial Pacific (see Fig. 1d). To differentiate each basin, we define the Indian Ocean to be from 0° to 145°E and south of the Sumatra and Java islands. The Drake Passage between the Atlantic and Pacific is defined at 300°E (Fig. 1d).

### b. In situ data

We use two repetitions of the meridional section P16 sampled in the Pacific at ~150°W by the Climate Variability and Predictability (CLIVAR) program in 2005–06, and the more recent Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP) in 2014–15 (cruise Expocodes are provided in Table S1 in the online supplemental material). All cruise data were acquired from the Carbon Hydrographic Data Office (CCHDO; https://cchdo.ucsd.edu/) data portal.

We also use the 9172 ARGO profiles (Argo 2018) that reach 1950 dbar, sampled over the 2002–15 period, 2° of longitude on each side of the P16 section at ~150°W.

---

[1] Referred to as temperature and salinity in the following.

## c. Functional principal component analysis

We use an FPCA in the multivariate case (Ramsay and Silverman 2005). The specific application of multivariate FPCA to $T$–$S$ profiles is described in a previous study that aimed in describing the Southern Ocean fronts and stratification types (Pauthenet et al. 2017). It was also used to document the seasonal meandering of the Polar Front (PF) in the southern Indian Ocean (Pauthenet et al. 2018). The codes we used to apply FPCA on $T$–$S$ profiles are compiled in an R package and available on Github (https://github.com/EPauthenet/fda.oce).

We first transform the $T$–$S$ profiles into functional curves using a decomposition on a B-spline basis. To avoid spurious extrema of the B-splines, the profiles are first linearly interpolated onto a uniform vertical grid with $100 \times 20$-m levels. The use of functions is a more robust way to handle data profiles and allows us to blend together observations from different oceanographic platforms, such as here with MIMOC data, ARGO profiles, and in situ data. The basis of vertical modes depends on 1) the longitude, latitude, and depth extent of the region to study and 2) the number of B-splines, which controls the vertical smoothness. We choose a number of $K = 40$ cubic B-splines per profile, which is large enough considering that we are analyzing a climatology (i.e., smooth stratification). Each profile can then be expressed with 40 coefficients that are stored end-to-end in a table $\mathbf{X}$ of size $N \times L$, with $N$ being the total number of profiles and $L = 2K$ the number of spline coefficients needed to represent the $T$–$S$ profiles. Let now consider the matrix $\mathbf{C}$ obtained by subtracting from $\mathbf{X}$ the coefficients of the profile averaged over the considered domain $\overline{\mathbf{X}}$. The functional PCA consists in finding the unique decomposition of the cross-covariance matrix $\mathbf{V} = (1/N)\mathbf{C}'\mathbf{C}$:

$$\mathbf{VWMb}_l = \lambda_l \mathbf{b}_l \qquad (1)$$

with the matrix $\mathbf{W}$ that ensures the metric equivalence between the functional problem and its discrete version (for computational details, see Pauthenet et al. 2017). The matrix $\mathbf{M}$ is the weighting matrix used to normalize the coefficients of different units, here temperature and salinity. It is a diagonal matrix with the first $K$ entries being the variance of temperature and the last $K$ entries the variance of salinity. Finally $\mathbf{b}_l$ is the $l$th eigenvector associated to the eigenvalue $\lambda_l$. The first vertical mode corresponds to the eigenvector with the largest eigenvalue.

Profiles can now be projected in a space of small dimension when computing the PCs[2] denoted as

---

[2] For readability, the PC values are noted as PC1, PC2, etc. in the text and figures, and $y_1$, $y_2$, etc. in the equations.

$$\mathbf{y}_l = \mathbf{CM}^{-1/2}\mathbf{W}^{-1/2}\mathbf{b}_l. \qquad (2)$$

After defining the basis of vertical modes, it is computationally inexpensive to project any new profile on the FPCA basis and locate it relatively to the climatology. As in standard PCA, the FPCA allows us to quantify the amount of variance that each vertical mode induces on the water column structure. It can also quantify the amount of variance contained by either temperature or salinity on each mode.

The projection of an eigenvector $\mathbf{b} = (b_1^T, \ldots, b_K^T; b_1^S, \ldots, b_K^S)'$ in the physical $T$–$S$ space will generate two eigenfunctions $(\xi^T, \xi^S)$ that are called vertical modes in this paper. Thus, a $T$–$S$ profile can be reconstructed by summing the $q$ first eigenfunctions weighted by their respective PC values $(y_{n,k})$:

$$\begin{cases} T_n^{(q)} = \overline{T} + \sum_{k=1}^{q} y_{n,k}\xi_k^T \\ S_n^{(q)} = \overline{S} + \sum_{k=1}^{q} y_{n,k}\xi_k^S \end{cases} \qquad (3)$$

with $T^{(q)}$ and $S^{(q)}$ the reconstructed profiles of Conservative Temperature and Absolute Salinity, $n = 1, \ldots, N$ the profile index, $k = 1, \ldots, K$ the mode numbers with $q < K$, $p \in \{0, 1950\}$ dbar the pressure and $\{\overline{T}, \overline{S}\}$ the mean reference temperature and salinity profiles.

The effect of the different modes can be displayed by adding or subtracting the eigenfunctions $(\xi_k^T, \xi_k^S)$ to the mean profiles $\{\overline{T}, \overline{S}\}$ such that

$$\overline{T} \pm \sqrt{\lambda_k}\xi_k^T \quad \text{and} \quad \overline{S} \pm \sqrt{\lambda_k}\xi_k^S. \qquad (4)$$

This is shown in Fig. 2 and is also used to represent the effect of temperature or salinity separately for the first three modes in terms of changes of the potential density mean profile $\overline{\sigma} = \sigma(\overline{T}, \overline{S})$ (Fig. 3). To do so we compute the potential density anomaly $\sigma(t, s) - \overline{\sigma}$ for each depth level, referenced at the surface (0 dbar), for three combinations of $t$ and $s$. First, the total anomaly is computed with both temperature and salinity profiles reconstructed with the corresponding mode $k$ (black, Fig. 3),

$$\begin{cases} t = \overline{T} + \sqrt{\lambda_k}\xi_k^T \\ s = \overline{S} + \sqrt{\lambda_k}\xi_k^S \end{cases}, \qquad (5)$$

the temperature effect with the temperature profile reconstructed with the corresponding mode and the mean salinity profile (orange, Fig. 3),

$$\begin{cases} t = \overline{T} + \sqrt{\lambda_k}\xi_k^T \\ s = \overline{S} \end{cases}, \qquad (6)$$
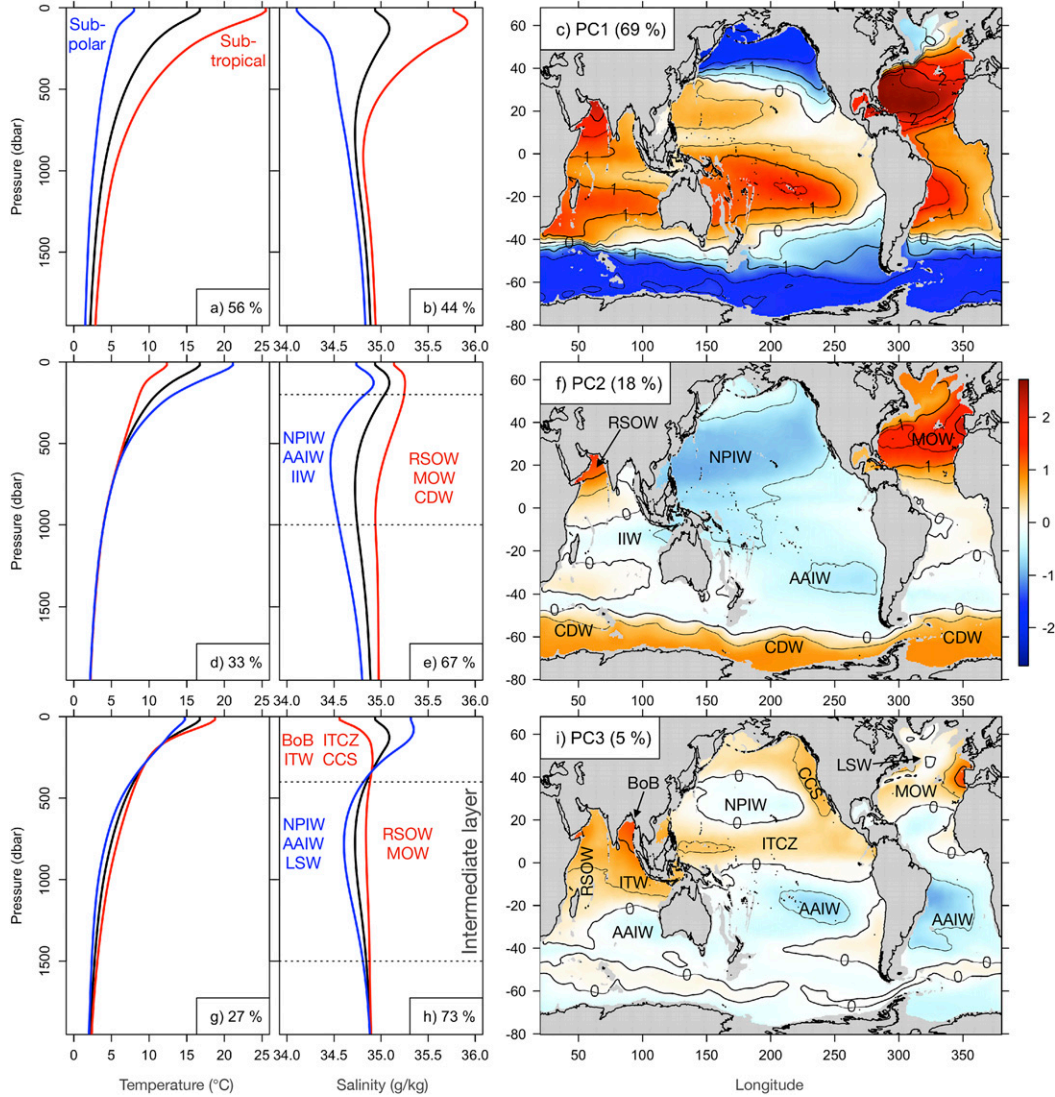
FIG. 2. The first three vertical modes and their PC spatial distributions (annual mean), derived from the MIMOC climatology. Colored curves display the effect of each mode (−1 in blue and +1 in red) on the mean profile [black, see Eq. (4)]. Regions and water masses corresponding to positive or negative PC values are noted at the corresponding depth along the T–S profiles. See the appendix for acronym definitions.

and the salinity effect with the salinity profile reconstructed and the mean temperature profile (blue, Fig. 3),

$$\begin{cases} t = \overline{T} \\ s = \overline{S} + \sqrt{\lambda_k}\xi_k^S \end{cases} \quad (7)$$

### d. Dynamic height anomaly, heat content, and salt content

We can reconstruct profiles and therefore reconstruct integrated variables with a reduced number of modes to understand further the information contained by each

modes. The dynamic height anomaly $\Delta D$ is computed by integrating the specific volume anomaly $\delta = \alpha_{S,T,p} - \alpha_{35,0,p}$ with respect to pressure $p$, following Talley et al. (2011). The integration is done using directly the coefficients of B-splines (Bai 2016). The specific volume is the inverse of the water density ($\alpha = 1/\rho$), which can be computed with reconstructed profiles $\rho(T^{(q)}, S^{(q)}, p)$. So $\Delta D$ is computed for $T^{(q)}$ and $S^{(q)}$ profiles as follows:

$$\Delta D^{(q)} = \frac{1}{10}\int \delta^{(q)}\,dp. \quad (8)$$

The factor 10 is for converting the unit from meters squared per second squared ($m^2\,s^{-2}$) to dynamic meter
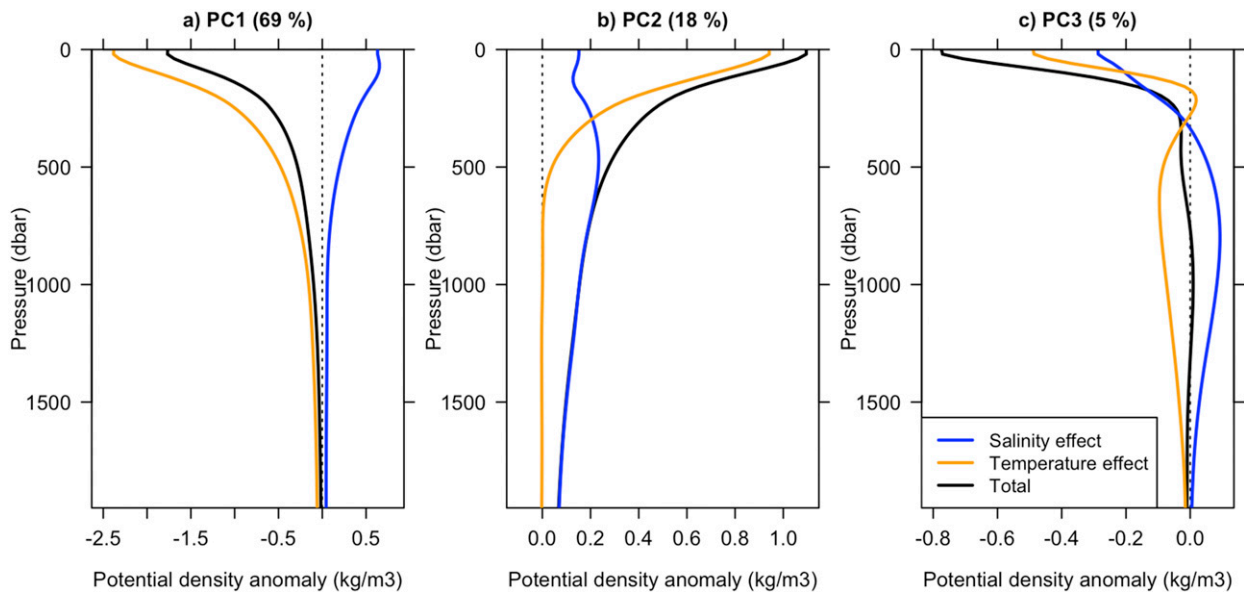
FIG. 3. Effect of adding ($+1$) the eigenfunctions on the potential density profile (kg m$^{-3}$), referenced at the surface pressure. The total anomaly (black) and the anomaly explained by temperature only (orange) and by salinity only (blue) are represented [see Eqs. (5)–(7)].

(dyn m). Note that the $q$th approximation of the dynamic height anomaly cannot be obtained by summing up the contribution of the associated $q$ first modes due to the nonlinearity of the equation of state. Hence, the amount of $\Delta D$ explained by the decomposition will not necessarily increase monotonically with the number of modes used. To obtain the heat content of the reconstructed profiles HC$^{(q)}$ we compute:

$$\mathrm{HC}^{(q)} = \frac{c_p^0}{g} \int T^{(q)} \, dp, \qquad (9)$$

with $c_p^0$ the specific heat for use with Conservative Temperature, taken from the standard TEOS10 (McDougall and Barker 2011), and $g = 9.81 \, \mathrm{m \, s^{-1}}$ the gravitational acceleration. The heat content gives the amount of heat contained by the water column in Joules per square meter (J m$^{-2}$). Note that $T$ is a profile of Conservative Temperature, to represent accurately the conservation of heat (McDougall 2003). Finally, the salinity content of the reconstructed profiles SC$^{(q)}$ can be derived with,

$$\mathrm{SC}^{(q)} = \frac{1}{g} \int S^{(q)} \, dp, \qquad (10)$$

and the result is given in tons per square meter (t m$^{-2}$; $1 \, \mathrm{t} = 1000 \, \mathrm{kg}$), so the salt content is the mass of salt contained by the water column each square meter. The salt content can be related to the more popular freshwater

content (FWC) by FWC $= M - \mathrm{SC}/S_{\mathrm{ref}}$, with $M$ the total mass of the water column and $S_{\mathrm{ref}}$ a reference salinity (Carmack et al. 2008).

## 3. Structure and global distribution of the thermohaline modes

A total of 92% of the variance is captured by the first three vertical modes. The higher modes are representing a smaller amount of variance (PC4: 3%, PC5: 2%, PC6: 1%, and higher modes contain <1%, Fig. S1). The 90% confidence intervals computed on the eigenvalues following Larsen and Warne (2010), show that modes six and higher are degenerated and should not be retained (not shown).

Figure 2 presents the effect of adding (red curve) and subtracting (blue curve) a magnitude one of the first three modes to the mean $T$–$S$ profile. This mean profile (black in the left panels of Fig. 2) is located along the zero contours on the PC maps (Fig. 2, right panels) and at the zero on the PC/PC plots in Fig. 4. The main water masses are identified on the PC maps (Fig. 2) and along three sections of salinity in the Indian, Atlantic, and Pacific (Fig. 5).

In the following we describe the vertical structure of the three first modes and the distribution of the profiles in this three-dimensional PC space. Then we describe the link between the main PCs and the $T$–$S$ field, the potential density profile, heat content, salt content, and dynamical height. We provide a description of the PCs regional distribution and finally an independent hydrographic section
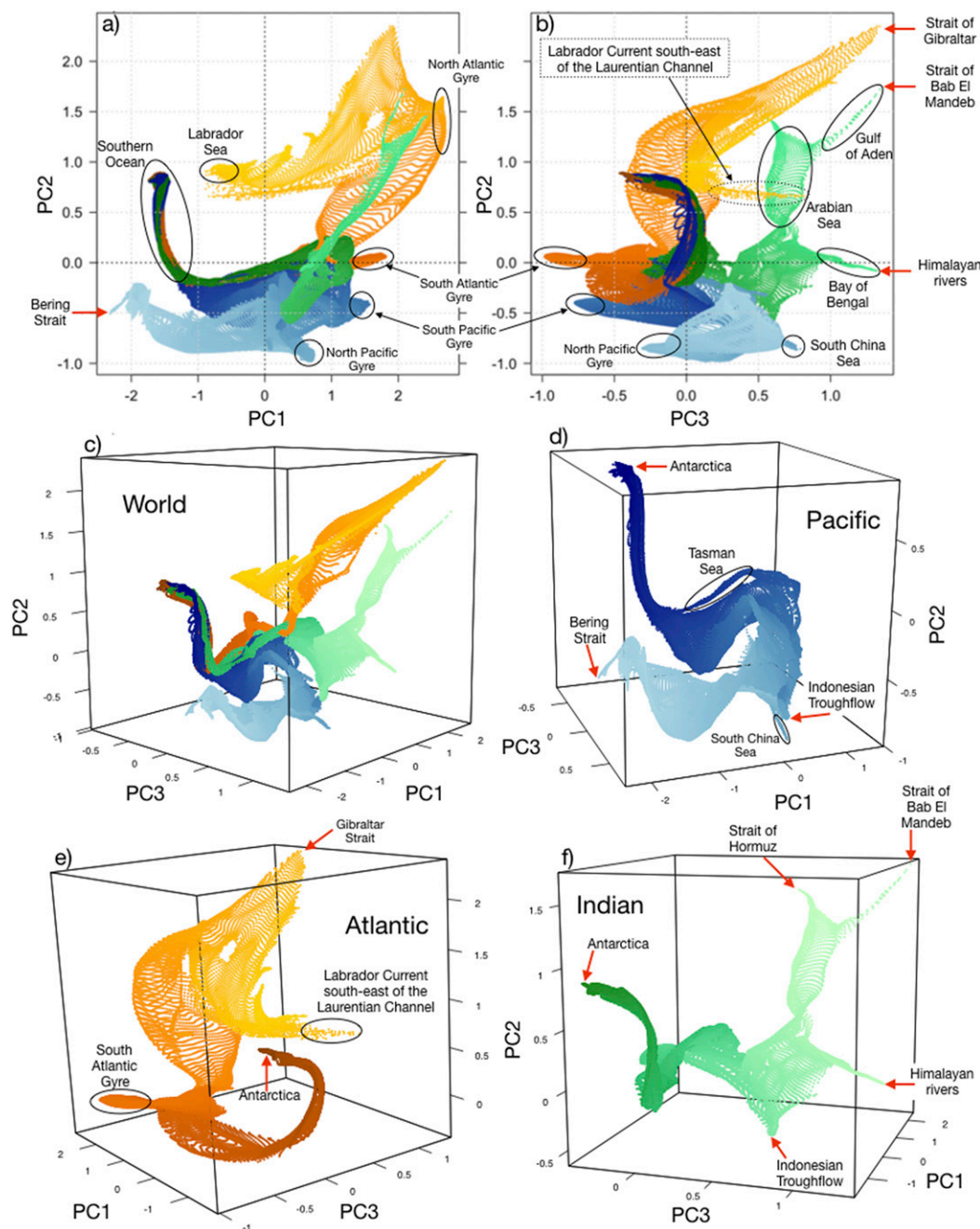
FIG. 4. The three first annual mean PCs colored by basins with a south–north gradient (dark–light), for the Indian (green), Pacific (blue), and Atlantic (orange) (see Fig. 1). It reveals the main thermohaline differences between the Indian (green), Pacific (blue), and Atlantic (orange). A few peculiar regions are circled in black, and red arrows point the link with smaller basins or rivers. Note the rotation of the axis in (c)–(f).

in the Pacific (P16) is projected to illustrate the potential of the method for identifying thermohaline changes.

*a. Vertical mode 1: Thermal mode*

The first vertical mode contains 69% of the variance with a slightly larger contribution of the temperature

(56%, Figs. 2a–c) than the salinity. It captures the temperature gradients between the tropics and the high latitudes for the first 2000 m, gradually less pronounced with depth. This gradient of temperature is covarying with a large gradient of salinity in the upper 750 m (Figs. 2a,b).
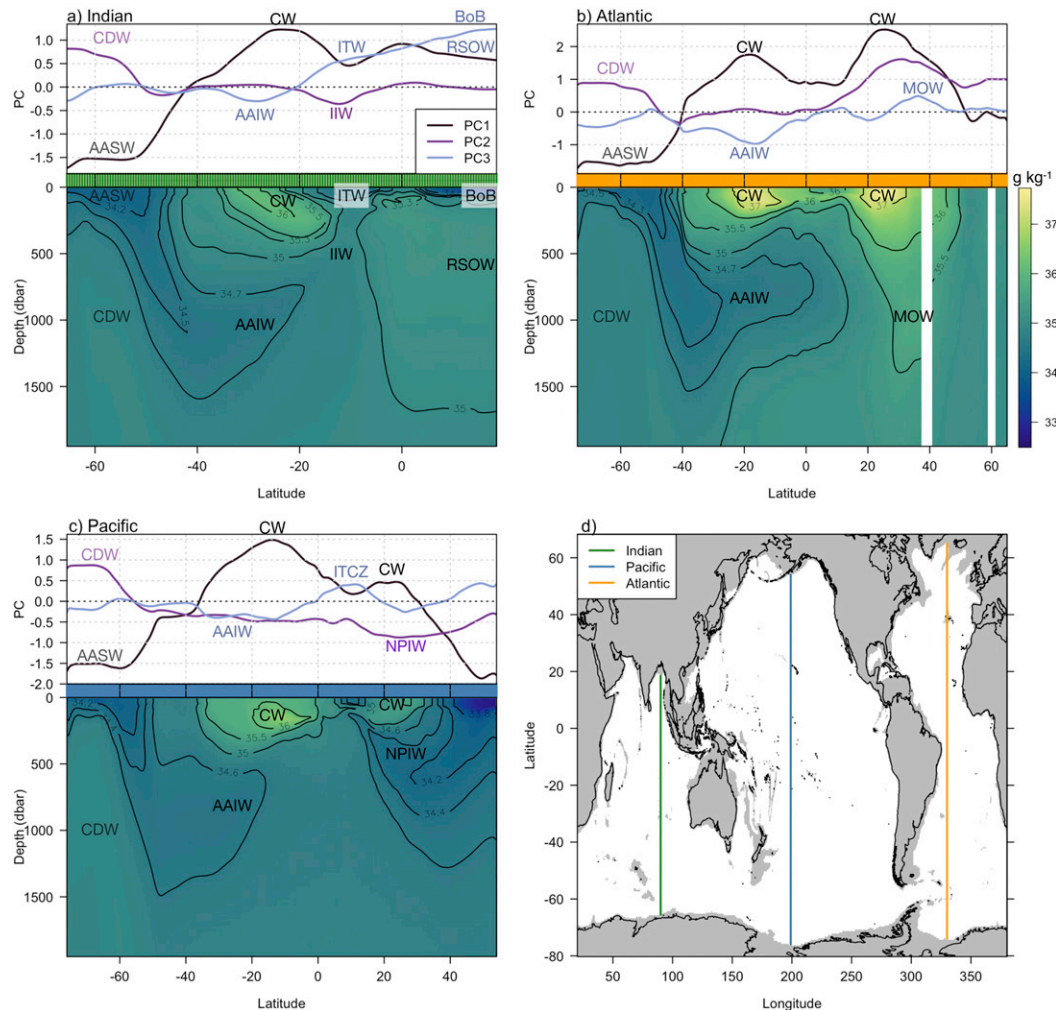
FIG. 5. Salinity sections (annual mean) at (a) 90°E (Indian, green), (b) 199°E (Pacific, blue), and (c) 330°E (Atlantic, orange) with their respective first three PCs. The values of the PCs can be linked to the position of water masses in the upper ocean and intermediate depth. See the appendix for acronym definitions.

We call this mode the "thermal mode" as its spatial distribution is representative of the thermocline depth changes, mainly resulting from the zonal wind patterns (Huang 2015). It captures the subtropical gyres in all oceans and contrast them to the very cold water in the Southern Ocean and Pacific subpolar gyre. This was already defined in the literature as the dominant feature of the global ocean (Fukumori et al. 2018). It is also highly correlated with temperature (see section 3e) and is closely related to integrated heat content (see section 3g). The subtropical thermocline water or Central Waters (CW) can have a large range of temperature and salinity properties, hence they are detected by a wide range of PC1 values, but always by a local maximum of PC1 (Figs. 2c and 5).

## b. Vertical mode 2: Haline mode and temperature change in the first 500 m

The second vertical mode contains 18% of the variance with a contribution of 67% from the salinity (Figs. 2d–f). It captures variations of salinity throughout the whole water column, slightly amplified around 500 m (Fig. 2e). The increase (decrease) of salinity at any depth covaries with a cooling (warming) of the upper 500 m (Fig. 2d).

The large positive values (PC2 ≳ 0.5) are highlighting two types of regions, the Circumpolar Deep Water (CDW) upwelled south of the Southern ACC Front (SACCF) and the salty and warm waters of the Arabian Sea and North Atlantic. The negative values are highlighting the low-salinity intermediate waters

of the North Pacific (NPIW) and Antarctic waters (AAIW). We call this mode the ''haline mode'' because its spatial distribution is representative of the Pacific/Atlantic salinity asymmetry and the salinity-dominated stratification in the Southern Ocean.

## c. Vertical mode 3: High- and low-salinity intermediate waters

The third vertical mode accounts for 5% of the total variance essentially driven by salinity changes (73% of salinity contribution, Figs. 2g–i). It identifies an inversion in the salinity profile at 400 m, separating the upper ocean from the intermediate depth range (400–1500 m, Fig. 2h). It is associated with slightly warmer waters in the upper 100 m and in the intermediate waters (Fig. 2g).

Strongly positive values of PC3 $\gtrsim 1$ are highlighting the Bay of Bengal and the Gulf of Aden in the Indian and the region between the Portugal and the Azores in the Atlantic (Figs. 4b and 2i). The negative values are roughly surrounding the subtropical gyres in the Pacific and the South Atlantic. This is the signal of the higher salinity CW in the major subtropical evaporation centers, overlying the low-salinity intermediate water also detected by PC2 negative values. The PC3 map (Fig. 2i) can be read as a map identifying the low- and high-salinity intermediate waters. It compares well with the previous map found in the literature (Fig. 2 in Talley 2008). Namely, the low-salinity intermediate waters are the NPIW, AAIW, and the Labrador Sea Water (LSW). The high-salinity intermediate waters are the Red Sea and Mediterranean Overflow Waters (RSOW and MOW).

## d. Thermohaline structure in PC space

The three-dimensional (PC1, PC2, PC3) space (Fig. 4) defines a diagram for T–S profiles and the key for reading each axis is the vertical modes shown in Fig. 2. The particularity is that a Euclidean distance in PC space means a difference in terms of both shape of the temperature and salinity profiles (i.e., stratification). Looking at the PC distribution in two or three dimensions is highly relevant because it is their combinations that define the shape of a profile. The T–S profiles appear to almost lie on a single surface for each basin in the three-dimensional space, shaping a ''dragon'' shape with the Southern Ocean as its neck and head, the Indian Ocean as the tail, and the Atlantic and Pacific Oceans as the wings (Fig. 4c).

The Southern Ocean occupies a small area in PC1/PC2 space revealing the highly organized structures south of the Subantarctic Front (SAF). The difference of salinity between the Pacific and the Atlantic is

remarkably well represented on the PC plots (Fig. 4). The North Pacific lies in the bottom-left quarter of the PC1/PC2 plot, opposite to northern Indian and Atlantic located in the top right quarter [except for the Labrador Sea and its low-salinity intermediate water (LSW), Fig. 4]. The Atlantic occupies a large space in the upper-right quarter (Fig. 4a). The north Indian basin is rather thin in PC1/PC2, but PC3 reveals most of its structure with the marginal seas pointing in different directions according to their properties (Fig. 4b).

## e. Link between vertical modes and hydrographic properties

This aesthetic dragon-like distribution (Fig. 4) is similar to the T–S diagram in the intermediate waters (Fig. 1). To assess more precisely at which depth the PC1/PC2 distribution is best represented by T–S we compute the correlations between PC1 or PC2 and temperature or salinity at any depth. PC1 spatial distribution is highly correlated with the distribution of temperature at any depth (>0.75, Fig. 6), which indicates the high coherence of temperature fields with depth. PC1 is also highly correlated with salinity spatial distribution for the upper 400 m, with a maximum of 0.93 at 130 m. PC2 shows no correlation with temperature but a high correlation with salinity under 400 m, with a maximum of 0.82 at 750 m (Fig. 6). These correlations are best illustrated by the similarities between the PC1 against PC2 plot and temperature at 340 m against salinity at 750 dbar (Figs. 6a,b).

## f. Modes' effect on the potential density profile

The effect of each mode of temperature or salinity on the potential density profile is presented in Fig. 3. The potential density anomaly produced by PC1 is essentially driven by temperature, and salinity compensates its effect on density for the whole water column (Fig. 3a). The effect of PC2 on the potential density is driven only by salinity below 500 m, while above the temperature is increasing the anomaly (Fig. 3b). For PC3, temperature and salinity are completely compensating each other below 500 m while both inducing a lightening of the upper layer (warmer and fresher for positive values of PC3, Fig. 3c). This fortifies our statement that a PC3 map can be read as map of high- and low-salinity intermediate waters.

## g. Modes' effect on the dynamic height anomaly, heat content, and salt content

In this section we present the dynamic height anomaly $\Delta D$, heat content, and salt content for reconstructed T–S profiles (Fig. 7). This analysis is done to bring more insight on the information contained by the main modes
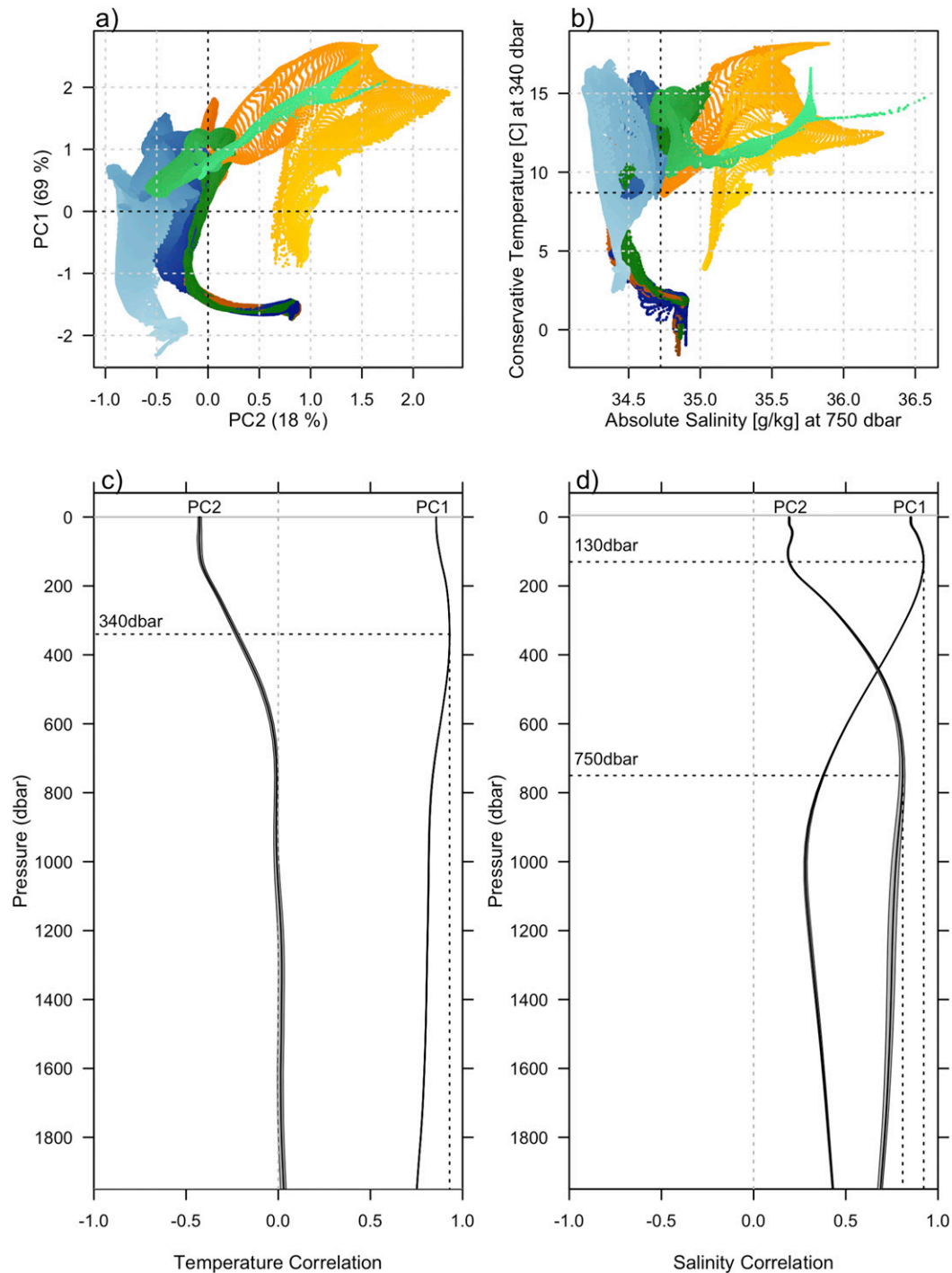
FIG. 6. (a),(b) Correlations of PC1 and PC2 fields (Figs. 2c,f) with (c),(d) temperature and salinity fields at any depth reveal that the highest correlations are between PC1 and temperature at 340 dbar, and between PC2 and salinity at 750 dbar. The seasonal variation is very small and represented with a gray band around the profiles in (c) and (d). This correlation is illustrated by the rotated PC1/PC2 plot (also shown in Fig. 4a) showing similarities with the plot of temperature at 340 dbar and salinity at 750 dbar in (b); the black dotted lines are the mean of temperature and salinity in (b) and 0 lines in (a). The colors are for the Indian (green), Atlantic (orange), and Pacific (blue) (see Fig. 1d).
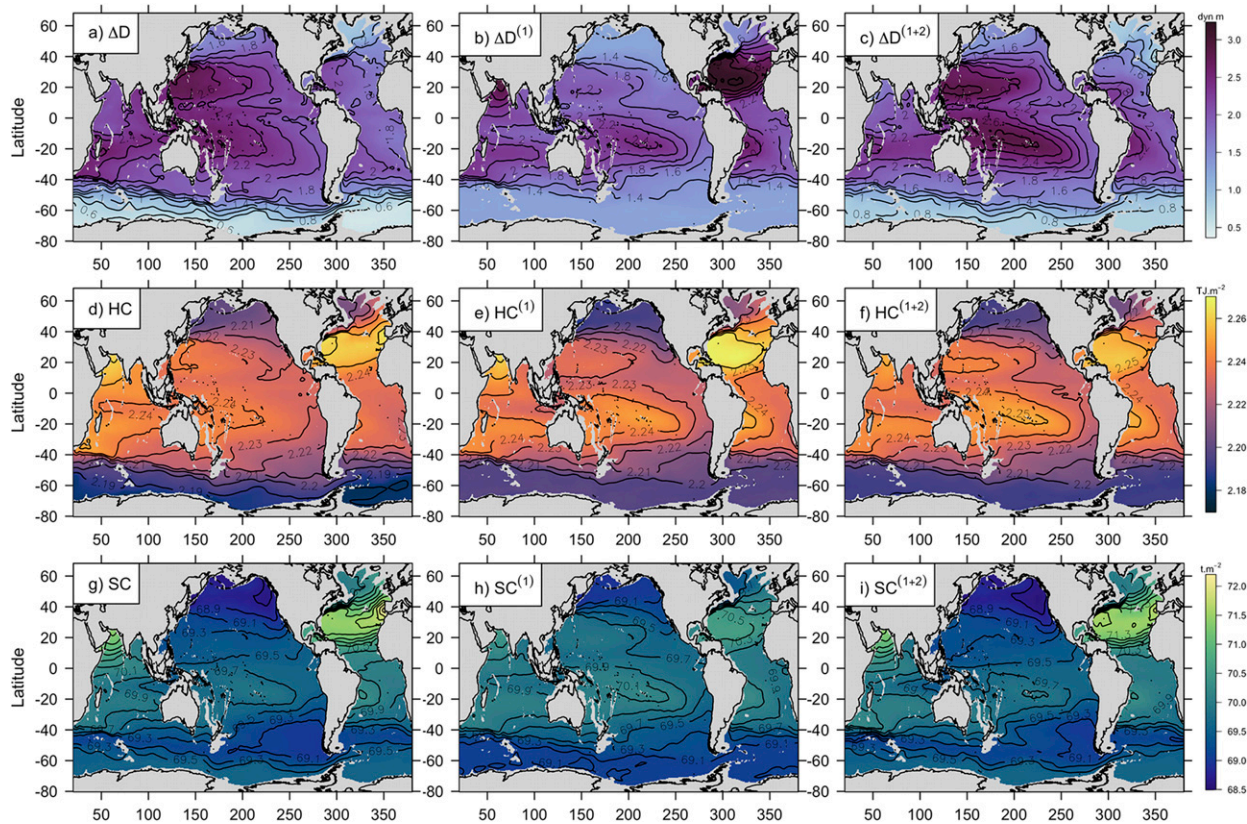
FIG. 7. (a) The dynamic height anomaly $\Delta D$ (dyn m), (d) heat content (HC; TJ m$^{-2}$, i.e., $10^{12}$ J m$^{-2}$), and (g) salt content (SC; t m$^{-2}$, i.e., $10^6$ g m$^{-2}$) of the annual mean MIMOC field are plotted. Also shown are property fields from reconstructed profiles with (b),(e),(h) PC1 alone or (c),(f),(i) PC1 + PC2.

and how they relate to standard metrics. Figure 8 is the zonal root-mean-square error (RMSE) between the total fields of $\Delta D$, heat content, and salt content and the fields from reconstructed profiles up to $q = 6$ modes. The map of $\Delta D^{(1)}$ reveals the subtropical gyre structure but no structures in the Southern Ocean and subpolar Pacific (Fig. 7b). The North Atlantic $\Delta D^{(1)}$ overshoots and is corrected by adding PC2, showing the strong compensation of temperature and salinity in this region (Fig. 7c). The high-latitude structures appear on $\Delta D^{(1+2)}$ (i.e., when adding the first and second mode), where the salinity is dominating the contribution to stratification (Pollard et al. 2002). The heat content is well represented with the first mode only, except for the cold Ross and Weddell gyres (HC$^{(1)}$, Fig. 7e). Its RMSE by latitude is almost the same when estimating it with PC1 only or PC1 + PC2 only (except for the Southern Ocean south of 60°S, Fig. 8b).

In contrast the salt content is poorly represented by PC1 only (Fig. 7h), and PC2 adds the low-salinity trench in the Southern Ocean, opening in the Pacific (Fig. 7i). The error in reconstructed fields of $\Delta D$ and salt content drops drastically when reconstructing

with two modes rather than PC1 only (Fig. 8). The reconstruction with six modes of the dynamic height anomaly, heat content, and salt content has a good accuracy, as the zonal RMSE is below 0.07 dyn m, 1.6 GJ m$^{-2}$, and 18.6 kg m$^{-2}$ (Fig. 8). These analyses comparing the modes' distribution and the hydrographic properties (Figs. 3, 6, 7, and 8) justify naming PC1 the thermal mode, PC2 the haline mode, and PC3 a mode capturing the high- and low-salinity intermediate waters signal.

## 4. Regional description of the modes distribution

### a. Atlantic Ocean

The North Atlantic presents high values in PC1 and PC2 associated with high surface and intermediate depth salinity. The excess of surface salinity in the North Atlantic is essentially due to the net evaporation and the deeper high salinity to the MOW, spreading westward from the Strait of Gibraltar (900 m) to a depth between 1000 and 1500 m (Talley 1996; Candela 2001). PC2 in the tropical and subtropical South Atlantic (between 35°S and the equator) is nearly null
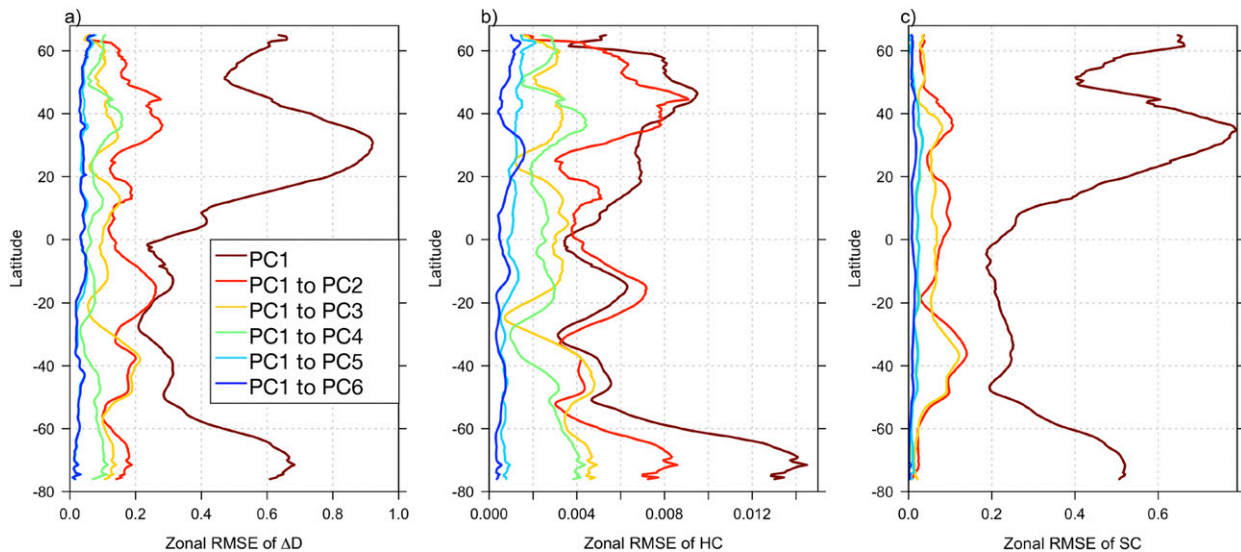
FIG. 8. Zonally averaged RMSE between original and reconstructed fields of (a) dynamic height anomaly $\Delta D$ ( dyn m), (b) heat content (TJ m$^{-2}$), and (c) salt content (t m$^{-2}$) up to six modes.

everywhere ($-0.14 <$ PC2 $< 0.19$), implying a limited effect of the second mode in the region.

The negative PC3 captures the low-salinity intermediate water of the LSW and AAIW (Fig. 2i). The AAIW spreads between the SAF and 30°N, while the LSW is detected in the Labrador Sea and the North Atlantic Current. Positive PC3 marks the high-salinity intermediate water MOW north of 30°N together with the warmer and fresher surface water under the intertropical convergence zone (ITCZ), between the equator and 30°N.

The region southeast of the Laurentian Channel is characterized by a PC3 larger than ~0.2 (Fig. 4) corresponding to fresher water in the upper layer (down to 100 m). Indeed the water over the Scotian Shelf is mainly composed of water coming from the Gulf of St. Lawrence and the Inshore Branch of the Labrador Current (Dever et al. 2016). This region on the poleward flank of the Gulf Stream is also identified as a single class by Maze et al. (2017), who defined eight objective classes containing the ARGO temperature profiles in the North Atlantic (class 4 in Fig. 11 of Maze et al. 2017).

### b. Pacific Ocean

The Pacific wind-driven tropical gyres are accurately delineated by the distribution of PC1 (Fig. 2c). The subpolar waters (negative PC1) are advected southward by the California Current (CC), toward the North Equatorial Current (NEC) that flows westward. The South Pacific Current (SPC) is marked with a broad PC1 transition from PC1 $\gtrsim 1$ at 30°S to PC1 $\lesssim 0$ north of the Subtropical Front (STF; Graham and Boer 2013).

PC2 in the Pacific is rather homogeneously negative, indicating an overall fresher water column than the rest of the global ocean. The minimum of PC2 in the north is contouring the low-salinity NPIW (Fig. 2f) that is marked by a salinity minimum between 300 and 700 m (Talley 1993). The negative PC2 in the southeastern Pacific signals the Pacific AAIW, characterized by a minimum salinity between 700 and 1000 m that originates from the southeast Pacific and stretches northeastward (Hartin et al. 2011). The AAIW and NPIW are represented on the Pacific salinity section of Fig. 5c.

The main pattern of PC3 is a positive band that circles around the north subtropical gyre (Fig. 2i). It highlights the subpolar waters, the California Current System (CCS) along the north American coast and the NEC from its formation in the eastern Pacific to the ITF in the western Pacific. The NEC lies between the equator and 20°N under the rainy ITCZ (Fig. 2i). The CCS holds a southward surface current about 1000 km wide and 500 m deep that advects fresher water from the eastern subpolar gyre (Marchesiello et al. 2003). Positive PC3 values in the Pacific capture the fresher surface waters of the CCS and ITCZ (Fig. 2h).

### c. Indian Ocean

In the Indian Ocean north of 40°S, PC1 values are of $1 \pm 0.37$ standard deviation, which indicates at first order a homogeneous warm and salty water mass distribution, slightly cooler and fresher along the South Equatorial Current (SEC, at ~10°S). The surface waters of the Bay of Bengal and the Arabian Sea are differentiated by PC2. This is expected considering the

evaporative conditions of the Arabian Sea and the excess of freshwater of the Bay of Bengal. Regarding the deeper high salinity in the Arabian Sea, it is connected to two smaller evaporation basins, the Red Sea and the Persian Gulf, both spilling very salty water. The lower layer of the Red Sea deep water spills through the Strait of Bab El Mandeb as the RSOW and spreads southward in the intermediate layer (400–1400 m) of the Arabian Sea (Talley et al. 2011). This is captured by high values of both PC2 and PC3 spreading down to the Mozambique Channel and into the Agulhas Current as already described by Beal et al. (2000). The Persian Gulf outflows very salty water that spreads in the upper 200–350 m in the Arabian Sea (Beal et al. 2000).

The negative PC2 values highlight the SEC that flows westerly with the trade winds north of 22°S. The high values of PC3 are also located in the Bay of Bengal and along the islands of Sumatra and Java, revealing the freshwater outflow coming from several large rivers; the Brahmaputra, Ganga, Irrawaddy, Godavari, and Mahanadi (Sengupta et al. 2006). The Indonesian Throughflow transports ~15 Sv (1 Sv ≡ $10^6$ m$^3$ s$^{-1}$) of relatively warm and freshwater from the western equatorial Pacific to the Indian Ocean (Sprintall et al. 2014). This low-salinity Pacific thermocline water is usually referred to as the Indonesian Throughflow Water (ITW; Gordon et al. 1997). The flow is concentrated in the upper 500 m (Fig. 7; Sprintall et al. 2009), which is consistent with the signature of positive values of PC3 (Figs. 2g,h). So in the Indian, positive values of PC3 are capturing high-salinity intermediate waters and surface freshwater input of the Himalayan rivers combined with the ITW. Further southeast, a minimum of PC3 could signal the presence of AAIW (Figs. 2i and 5a), characterized by a minimum of salinity between 700 and 1000 m (Hartin et al. 2011).

### d. Southern Ocean

The thermohaline structure of the Southern Ocean was described in details in Pauthenet et al. (2017). Despite the fact the decomposition was computed for the Southern Ocean only, the three first modes here are very similar in structure and distribution. The Southern Ocean is highly influenced by the Antarctic Circumpolar Current (ACC) that organizes the thermohaline structure zonally. The stratification changes between the subtropics and the Antarctic waters happen in a steplike manner, with circumpolar regions of similar properties separated by sharp gradients, known as circumpolar fronts (Orsi et al. 1995). The fundamental Southern Ocean fronts can be explained by the interplay of PC1 and PC2. The first mode (thermal) varies only north of the PF, defining the region where

salinity dominates the contribution to stratification (Pollard et al. 2002). Indeed PC1 displays sharp zonal gradients corresponding to the STF and SAF but does not reveal the Polar and Southern ACC Front. The gradients of PC1 are less marked in the Pacific sector due to the multiple jets associated to the STF (Graham and Boer 2013) and the exchange window in the Eastern Pacific that weakens the SAF (Iudicone et al. 2007; Pauthenet et al. 2017). South of the SAF, the structures of the PF and SACCF are best described by the haline mode PC2. These higher values of PC2 correspond to the CDW that upwells south of the PF (Fig. 5). The PC3 distribution in the Southern Ocean is negative everywhere, except for a positive band (0 < PC3 < 0.18) grossly highlighting the ACC between the SAF and the SACCF. This region of positive PC3 is where the Antarctic Surface Water (AASW) in the upper layer reaches ~400 m and overlies the upper CDW (Fig. 5).

## 5. Case study: Repeated P16 section

As a final analysis, we project the Pacific meridional P16 section on the modes (Fig. 9). The goal is to illustrate how the modes can be used to filter the profiles and keep only the fundamental variations. To observe changes with time, we use two repetitions of the section that we named P16$_{old}$ in 2005/06 and P16$_{new}$ in 2014/15. We also select ARGO profiles in a longitude range of ±1° around the section to visualize the seasonal range of the PCs. Figure 9 presents the PC1 and PC2 distribution and the differences between P16$_{old}$ and P16$_{new}$. The average difference between P16$_{old}$ and P16$_{new}$ is 0.046 for PC1 and −0.038 for PC2 (Figs. 9c,e). This implies an average warming between 2005 and 16, which is expected as the ocean is taking up most of the excess heat in the Earth system (e.g., Talley et al. 2016). This average warming is associated with a saltening of the upper ~500 m (PC1) and a freshening of the intermediate layer (PC2) characterized by the NPIW and AAIW along this section. Here it is difficult to disentangle the temperature signal from the salinity signal as the variance is shared almost equally between the two variables (Fig. 2).

Further into the details, there is a positive difference of PC1 (total warming and saltening of the upper ~500 m) north of 35°N (Fig. 9c). Indeed this region experienced the largest marine heatwave ever recorded, during the winter of 2013/14. It is referred to as "the blob" and persisted throughout 2014 and 2015 (Di Lorenzo and Mantua 2016; Bond et al. 2015). A larger PC1 indicates also an increase of the salinity for the upper ~500 m, which has been documented previously with this same section P16 (Kouketsu et al. 2017).
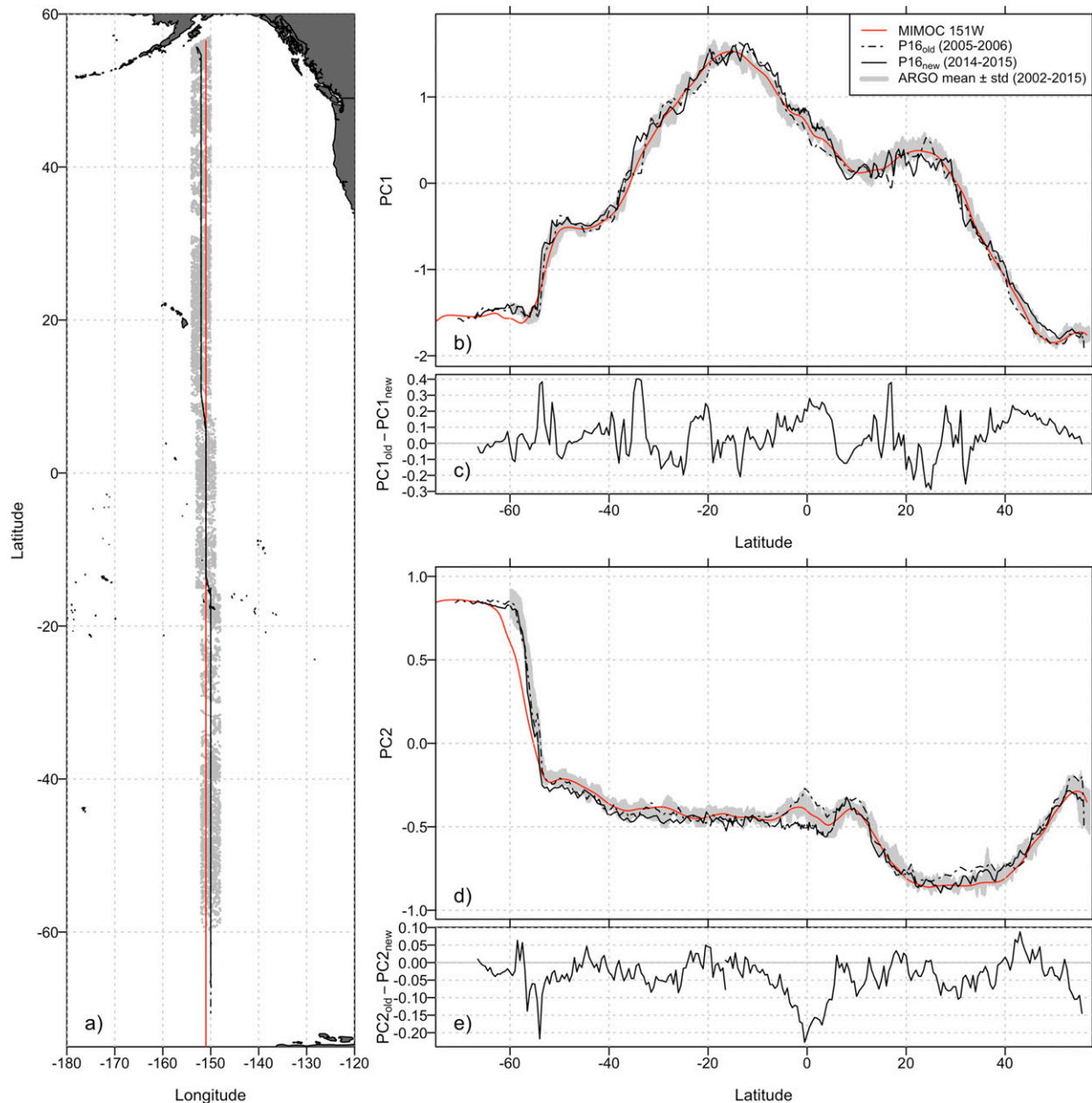
FIG. 9. Projection of the repeated section P16 on the global ocean modes is represented with (b) PC1 and (d) PC2 variations. (a) A map of the section P16 in black, the section of MIMOC at 151°W in red, and the ARGO profiles in gray. P16 was sampled in 2005/06 (dashed line) and 2014/15 (full line) between January and June. Difference between old (2005/06) and new (2014/15) section is plotted for (c) PC1 and (e) PC2. ARGO profiles (gray) around P16 are also projected and represented in (b) and (d) with a gray band of the PC standard deviation added and subtracted from the mean. Annual mean of MIMOC PC values are also plotted for the longitude 150°W (red line).

PC1 is also higher at 0°N and this warm anomaly appears on PC2 too, with a large negative difference (upper-500-m warming and total freshening) around 0°N, which is characteristic of the strong El Niño event in 2015. This extreme event was marked by persistent warm conditions in the equatorial Pacific between March 2014 and May 2016 (Santoso et al. 2017).

Another difference is the PC2 slope around 60°S between MIMOC and the observations (Fig. 9d). It indicates a misrepresentation of the overall salinity and upper-500-m temperature (PC2) in MIMOC relatively to the observations. Indeed MIMOC is not showing the salty Circumpolar Deep Water that is present at ~700 m and 60°S on P16 (see Fig. S2). This mismatch could be

due to the lack of data used for MIMOC in the South Pacific (Fig. 1 in Schmidtko et al. 2013). Overall we see that the time variability of thermohaline properties along the P16 section can be robustly captured by the FPCA method.

## 6. Conclusion and perspectives

The first 2000 m of the thermohaline structure of the global ocean are decomposed into vertical thermohaline modes with the first three modes capturing more than 92% of the observed $T$–$S$ variance. We compute and describe the structure and spatial distribution of these modes bringing new insights on the structure of the ocean by ranking the main features of the thermohaline structure. The first mode captures the structure of the main thermocline and the associated subtropical gyres. The second mode characterizes the salinity contrast between the Atlantic and Indo-Pacific basin, due to larger net evaporation over the Atlantic and salinity transport into the Atlantic (Ferreira et al. 2018). This signal varies with the salty CDW present in the intermediate layer south of the ACC. Finally the third mode reveals the low- and high-salinity intermediate waters together with a surface freshening/saltening due to atmospheric conditions and major rivers outflow.

Bindoff and McDougall (1994) discriminate the changes in hydrographic properties in terms of "pure warming," "pure freshening," and "pure heave" (i.e., vertical displacement of isotherms and isohalines). Each thermohaline mode is not a "pure" modification of a property, yet the changes of PCs through time can be studied according to what they represent in the vertical. Like Maze et al. (2017) proposed, these PCs could be used to define oceanographic domains according to the main profile's shape, which would be more relevant than an arbitrary rectangular box. Gradients in PC maps can be used to study the intensity and variability of the water masses' fronts. A change of stratification is associated with a change of PC. Thus, the abrupt change of PCs, or of a combination of PCs, indicates fronts as shown by Pauthenet et al. (2018), who apply FPCA on $T$–$S$ profiles to locate the Antarctic Polar Front in the Indian sector of the Southern Ocean.

The functional approach allows us to robustly compare profiles from different oceanographic platforms with model outputs. An independent $T$–$S$ profile can be projected on the readily defined modes to find its position relatively to a front, or to assess its general quality. The quality of a profile and its position relative to a climatology or to a front can be assessed objectively, which is useful for data validation. The variation of the PCs may also be used to detect water masses and to

visualize their changes on a reduced dimension space. For instance on P16, we identified the extreme El Niño of 2015, the blob in the North Pacific, and the bias of MIMOC for representing the Southern Pacific salinity around 60°S (Fig. 9d).

A density-based classification such as a Gaussian mixture model could be applied on the PC to identify water masses and fronts objectively, as done by Maze et al. (2017) and Jones et al. (2019) on temperature data only. The PCs could also be estimated from the surface information (sea surface temperature, sea surface height, etc.) using unsupervised learning methods such as self-organizing maps (SOMs). It was shown previously that SOMs are capable of reconstructing temperature profiles (Chen and Tung 2018), salinity profiles (Gueye et al. 2014), or 1000-m velocities (Chapman and Charantonis 2016) from remote sensing. Here we could reconstruct both temperature and salinity profiles by using satellite data to estimate the PC values up to $n$th modes, knowing that six modes capture 98% of the variance, and provide a good estimate of the dynamic height, the heat content, and the salt content (Fig. 8).

Wunsch (2018) discusses how we could better assess the global trends of heat and salt content by "suppressing" the deterministic component of the fields. He employed a singular value decomposition of the property fields and analyzed the residual fields generated by subtracting the main modes, that is, the globally correlated structures. This way of filtering the signal is interesting and we can imagine removing the main modes of our FPCA (e.g., $q \leq 3$) to study the residuals. However considering the result of the projection of the hydrographic section P16 (section 5 and Fig. 9), the temporal variation of the main modes also indicates the fundamental changes in the $T$–$S$ distribution. A more advanced analysis of the PC time series using a seasonal trend decomposition (STL; Cleveland et al. 1990) could also be done to reveal the global variations of each component of the $T$–$S$ field.

The dragon shape (Fig. 4c) is a striking illustration of the Southern Ocean linking the Pacific, Indian, and Atlantic Oceans. Beyond the incredible complexity of the ocean's processes, a beautiful somewhat simple structure is hiding, mainly shaped by the geography of the basins, the ocean forcings, and the equation of state of seawater.

# APPENDIX

## List of Acronyms

| | |
|---|---|
| AAIW | Antarctic Intermediate Water |
| AASW | Antarctic Surface Water |
| ACC | Antarctic Circumpolar Current |
| ARGO | Array for Real-Time Geostrophic Oceanography |
| BoB | Bay of Bengal |
| CCS | California Current System |
| CW | Central Water |
| CDW | Circumpolar Deep Water |
| $\Delta D$ | Dynamical height anomaly |
| EOF | Empirical orthogonal function |
| FPCA | Functional principal component analysis |
| HC | Heat content |
| IIW | Indonesian Intermediate Water |
| ITW | Indonesian Throughflow Water |
| ITCZ | Intertropical convergence zone |
| LSW | Labrador Sea Water |
| MIMOC | Monthly Isopycnal and Mixed-Layer Ocean Climatology |
| MOW | Mediterranean Overflow Water |
| NPIW | North Pacific Intermediate Water |
| PC | Principal component |
| PF | Polar Front |
| RSOW | Red Sea Overflow Water |
| SACCF | Southern ACC Front |
| SAF | Subantarctic Front |
| SC | Salt content |
| STF | Subtropical Front |

## REFERENCES

Argo, 2018: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE, https://doi.org/10.17882/42182.

Bai, Y., 2016: IntegrateBs: Integration for B-Spline. R package, https://cran.r-project.org/web/packages/IntegrateBs/index.html.

Beal, L. M., A. Ffield, and A. L. Gordon, 2000: Spreading of Red Sea overflow waters in the Indian Ocean. *J. Geophys. Res.*, **105**, 8549–8564, https://doi.org/10.1029/1999JC900306.

Bindoff, N. L., and T. J. McDougall, 1994: Diagnosing climate change and ocean ventilation using hydrographic data. *J. Phys. Oceanogr.*, **24**, 1137–1152, https://doi.org/10.1175/1520-0485(1994)024<1137:DCCAOV>2.0.CO;2.

Bond, N. A., M. F. Cronin, H. Freeland, and N. Mantua, 2015: Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophys. Res. Lett.*, **42**, 3414–3420, https://doi.org/10.1002/2015GL063306.

Candela, J., 2001: Mediterranean water and global circulation. *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, International Geophysics Series, Vol. 77, Academic Press, 419–429, https://doi.org/10.1016/S0074-6142(01)80132-7.

Carmack, E., F. McLaughlin, M. Yamamoto-Kawai, M. Itoh, K. Shimada, R. Krishfield, and A. Proshutinsky, 2008: Freshwater storage in the Northern Ocean and the special role of the Beaufort Gyre. *Arctic–Subarctic Ocean Fluxes*, R. R. Dickson, J. Meincke, and P. Rhines, Eds., Springer, 145–169, https://doi.org/10.1007/978-1-4020-6774-7_8.

Chang, Y., A. Rosati, and S. Zhang, 2011: A construction of pseudo salinity profiles for the global ocean: Method and evaluation. *J. Geophys. Res.*, **116**, C02002, https://doi.org/10.1029/2010JC006386.

Chapman, C., and A. A. Charantonis, 2016: Reconstruction of subsurface velocities from satellite observations using iterative self-organizing maps. *IEEE Geosci. Remote Sens. Lett.*, **14**, 617–620, https://doi.org/10.1109/LGRS.2017.2665603.

Chen, X., and K.-K. Tung, 2018: Global surface warming enhanced by weak Atlantic overturning circulation. *Nature*, **559**, 387–391, https://doi.org/10.1038/s41586-018-0320-y.

Chernoff, H., 1973: The use of faces to represent points in k-dimensional space graphically. *J. Amer. Stat. Assoc.*, **68**, 361–368, https://doi.org/10.1080/01621459.1973.10482434.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning, 1990: STL: A seasonal-trend decomposition. *J. Off. Stat.*, **6** (1), 3–73.

de Boer, C. J., and H. M. van Aken, 1995: A study of objective methods for water mass analysis, applied to the Iceland Basin. *Ocean Dyn.*, **47**, 5–22, https://doi.org/10.1007/BF02731988.

Dever, M., D. Hebert, B. Greenan, J. Sheng, and P. Smith, 2016: Hydrography and coastal circulation along the Halifax Line and the connections with the Gulf of St. Lawrence. *Atmos.–Ocean*, **54**, 199–217, https://doi.org/10.1080/07055900.2016.1189397.

Di Lorenzo, E., and N. Mantua, 2016: Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nat. Climate Change*, **6**, 1042–1047, https://doi.org/10.1038/nclimate3082.

Döös, K., J. Nilsson, J. Nycander, L. Brodeau, and M. Ballarotta, 2012: The World Ocean thermohaline circulation. *J. Phys. Oceanogr.*, **42**, 1445–1460, https://doi.org/10.1175/JPO-D-11-0163.1.

Ferreira, D., and Coauthors, 2018: Atlantic-Pacific asymmetry in deep water formation. *Annu. Rev. Earth Planet. Sci.*, https://doi.org/10.1146/annurev-earth-082517-010045

Fukumori, I., P. Heimbach, R. M. Ponte, and C. Wunsch, 2018: A dynamically consistent, multivariable ocean climatology. *Bull. Amer. Meteor. Soc.*, **99**, 2107–2128, https://doi.org/10.1175/BAMS-D-17-0213.1.

Gordon, A. L., S. Ma, D. B. Olson, P. Hacker, A. Ffield, L. D. Talley, D. Wilson, and M. Baringer, 1997: Advection and diffusion of Indonesian throughflow water within the Indian Ocean South Equatorial Current. *Geophys. Res. Lett.*, **24**, 2573–2576, https://doi.org/10.1029/97GL01061.

Graham, R. M., and A. M. Boer, 2013: The dynamical subtropical front. *J. Geophys. Res. Oceans*, **118**, 5676–5685, https://doi.org/10.1002/jgrc.20408.

Groeskamp, S., S. Griffies, D. Iudicone, R. Marsh, G. Nurser, and J. Zika, 2019: The water mass transformation framework for ocean physics and biogeochemistry. *Annu. Rev. Mar. Sci.*, **11**, https://doi.org/10.1146/annurev-marine-010318-095421.

Gueye, M. B., A. Niang, S. Arnault, S. Thiria, and M. Crépon, 2014: Neural approach to inverting complex system: Application to ocean salinity profile estimation from surface parameters. *Comput. Geosci.*, **72**, 201–209, https://doi.org/10.1016/j.cageo.2014.07.012.

Hartin, C. A., R. A. Fine, B. M. Sloyan, L. D. Talley, T. K. Chereskin, and J. Happell, 2011: Formation rates of Subantarctic mode water and Antarctic intermediate water within the South Pacific. *Deep-Sea Res. I*, **58**, 524–534, https://doi.org/10.1016/j.dsr.2011.02.010.

Hieronymus, M., J. Nilsson, and J. Nycander, 2014: Water mass transformation in salinity–temperature space. *J. Phys. Oceanogr.*, **44**, 2547–2568, https://doi.org/10.1175/JPO-D-13-0257.1.

Hjelmervik, K. B., and K. T. Hjelmervik, 2019: Detection of oceanographic fronts on variable water depths using empirical orthogonal functions. *IEEE J. Oceanic Eng.*, https://doi.org/10.1109/JOE.2019.2917456, in press.

Hjelmervik, K. T., and K. B. Hjelmervik, 2013: Estimating temperature and salinity profiles using empirical orthogonal functions and clustering on historical measurements. *Ocean Dyn.*, **63**, 809–821, https://doi.org/10.1007/s10236-013-0623-3, https://doi.org/10.1007/s10236-013-0623-3.

Huang, R. X., 2015: Heaving modes in the world oceans. *Climate Dyn.*, **45**, 3563–3591, https://doi.org/10.1007/s00382-015-2557-6.

Iudicone, D., K. B. Rodgers, R. Schopp, and G. Madec, 2007: An exchange window for the injection of Antarctic Intermediate Water into the South Pacific. *J. Phys. Oceanogr.*, **37**, 31–49, https://doi.org/10.1175/JPO2985.1.

Jackett, D. R., and T. J. McDougall, 1997: A neutral density variable for the world's oceans. *J. Phys. Oceanogr.*, **27**, 237–263, https://doi.org/10.1175/1520-0485(1997)027<0237:ANDVFT>2.0.CO;2.

Jolliffe, I. T., and J. Cadima, 2016: Principal component analysis: A review and recent developments. *Philos. Trans. Roy. Soc. London*, **374**, 20150202, https://doi.org/10.1098/rsta.2015.0202.

Jones, D. C., H. J. Holt, A. J. Meijers, and E. Shuckburgh, 2019: Unsupervised clustering of Southern Ocean Argo float temperature profiles. *J. Geophys. Res. Oceans*, **124**, 390–402, https://doi.org/10.1029/2018JC014629.

Kouketsu, S., S. Osafune, Y. Kumamoto, and H. Uchida, 2017: Eastward salinity anomaly propagation in the intermediate layer of the North Pacific. *J. Geophys. Res. Oceans*, **122**, 1590–1607, https://doi.org/10.1002/2016JC012118.

Larsen, R., and R. T. Warne, 2010: Estimating confidence intervals for eigenvalues in exploratory factor analysis. *Behav. Res. Methods*, **42**, 871–876, https://doi.org/10.3758/BRM.42.3.871.

Lindegren, R., and M. Josefson, 1998: Bottom water formation in the Weddell Sea resolved by principal component analysis and target estimation. *Chemom. Intell. Lab. Syst.*, **44**, 403–409, https://doi.org/10.1016/S0169-7439(98)00072-0.

Marchesiello, P., J. C. McWilliams, and A. Shchepetkin, 2003: Equilibrium structure and dynamics of the California Current System. *J. Phys. Oceanogr.*, **33**, 753–783, https://doi.org/10.1175/1520-0485(2003)33<753:ESADOT>2.0.CO;2.

Maze, G., H. Mercier, R. Fablet, P. Tandeo, M. L. Radcenco, P. Lenca, C. Feucher, and C. Le Goff, 2017: Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean. *Prog. Oceanogr.*, **151**, 275–292, https://doi.org/10.1016/j.pocean.2016.12.008.

Mazloff, M. R., P. Heimbach, and C. Wunsch, 2010: An eddy-permitting Southern Ocean state estimate. *J. Phys. Oceanogr.*, **40**, 880–899, https://doi.org/10.1175/2009JPO4236.1.

McDougall, T. J., 2003: Potential enthalpy: A conservative oceanic variable for evaluating heat content and heat fluxes. *J. Phys. Oceanogr.*, **33**, 945–963, https://doi.org/10.1175/1520-0485(2003)033<0945:PEACOV>2.0.CO;2.

——, and D. R. Jackett, 2007: The thinness of the ocean in S-Θ-*p* space and the implications for mean diapycnal advection. *J. Phys. Oceanogr.*, **37**, 1714–1732, https://doi.org/10.1175/JPO3114.1.

——, and P. Barker, 2011: Getting started with TEOS-10 and the Gibbs Seawater (GSW) oceanographic toolbox. Scientific Committee on Oceanic Research–IAPSO Rep. WG127, 28 pp.

Orsi, A. H., T. Whitworth III, and W. D. Nowlin Jr., 1995: On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Res. I*, **42**, 641–673, https://doi.org/10.1016/0967-0637(95)00021-W.

Pauthenet, E., F. Roquet, G. Madec, and D. Nerini, 2017: A linear decomposition of the Southern Ocean thermohaline structure. *J. Phys. Oceanogr.*, **47**, 29–47, https://doi.org/10.1175/JPO-D-16-0083.1.

——, ——, ——, C. Guinet, M. Hindell, C. R. McMahon, R. Harcourt, and D. Nerini, 2018: Seasonal meandering of the polar front upstream of the Kerguelen Plateau. *Geophys. Res. Lett.*, **45**, 9774–9781, https://doi.org/10.1029/2018GL079614.

Pollard, R. T., M. I. Lucas, and J. F. Read, 2002: Physical controls on biogeochemical zonation in the Southern Ocean. *Deep-Sea Res. II*, **49**, 3289–3305, https://doi.org/10.1016/S0967-0645(02)00084-X.

Ramsay, J., and B. Silverman, 2005: *Functional Data Analysis*. Springer 428 pp.

Rosenfeld, A., and J. L. Pfaltz, 1966: Sequential operations in digital picture processing. *J. Assoc. Comput. Mach.*, **13**, 471–494, https://doi.org/10.1145/321356.321357.

Santoso, A., M. J. Mcphaden, and W. Cai, 2017: The defining characteristics of ENSO extremes and the strong 2015/2016 El Niño. *Rev. Geophys.*, **55**, 1079–1129, https://doi.org/10.1002/2017RG000560.

Schmidtko, S., G. C. Johnson, and J. M. Lyman, 2013: MIMOC: A global monthly isopycnal upper-ocean climatology with mixed layers. *J. Geophys. Res. Oceans*, **118**, 1658–1672, https://doi.org/10.1002/jgrc.20122.

Sengupta, D., G. Bharath Raj, and S. Shenoi, 2006: Surface freshwater from Bay of Bengal runoff and Indonesian throughflow in the tropical Indian Ocean. *Geophys. Res. Lett.*, **33**, L22609, https://doi.org/10.1029/2006GL027573.

Speer, K. G., 1993: Conversion among North Atlantic surface water types. *Tellus* **45A**, 72–79, https://doi.org/10.3402/tellusa.v45i1.14858.

Sprintall, J., S. E. Wijffels, R. Molcard, and I. Jaya, 2009: Direct estimates of the Indonesian Throughflow entering the Indian Ocean: 2004–2006. *J. Geophys. Res.*, **114**, C07001, https://doi.org/10.1029/2008JC005257.

——, A. L. Gordon, A. Koch-Larrouy, T. Lee, J. T. Potemra, K. Pujiana, and S. E. Wijffels, 2014: The Indonesian seas and their role in the coupled ocean–climate system. *Nat. Geosci.*, **7**, 487–492, https://doi.org/10.1038/ngeo2188.

Talley, L. D., 1993: Distribution and formation of North Pacific intermediate water. *J. Phys. Oceanogr.*, **23**, 517–537, https://doi.org/10.1175/1520-0485(1993)023<0517:DAFONP>2.0.CO;2.

——, 1996: North Atlantic circulation and variability, reviewed for the CNLS conference. *Physica D*, **98**, 625–646, https://doi.org/10.1016/0167-2789(96)00123-6.

——, 2008: Freshwater transport estimates and the global overturning circulation: Shallow, deep and throughflow components. *Prog. Oceanogr.*, **78**, 257–303, https://doi.org/10.1016/j.pocean.2008.05.001.

——, G. L. Pickard, W. J. Emery, and J. H. Swift, 2011: *Descriptive Physical Oceanography: An Introduction*. 6th ed., Academic Press, 560 pp.

——, and Coauthors, 2016: Changes in ocean heat, carbon content, and ventilation: A review of the first decade of GO-SHIP global repeat hydrography. *Annu. Rev. Mar. Sci.*, **8**, 185–215, https://doi.org/10.1146/annurev-marine-052915-100829.

Worthington, L., 1981: The water masses of the World Ocean: Some results of a fine-scale census. *Evolution of Physical Oceanography: Scientific Surveys in Honor of Henry Stommel*, B. A. Warren and C. Wunsch, MIT Press, 42–60.

Wunsch, C., 2018: Towards determining uncertainties in global oceanic mean values of heat, salt, and surface elevation. *Tellus*, **70A**, 1–14, https://doi.org/10.1080/16000870.2018.1471911.

Zika, J. D., M. H. England, and W. P. Sijp, 2012: The ocean circulation in thermohaline coordinates. *J. Phys. Oceanogr.*, **42**, 708–724, https://doi.org/10.1175/JPO-D-11-0139.1.