A Linear Decomposition of the Southern Ocean Thermohaline Structure🖉

E TIENNE P AUTHENET AND F ABIEN R OQUET

Department of Meteorology (MISU), Stockholm University, Stockholm, Sweden

G URVAN M ADEC

LOCEAN Sorbonne Universités (UPMC, University Paris 6)/CNRS/IRD/MNHN, Paris, France

D AVID N ERINI

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

ABSTRACT

The thermohaline structure of the Southern Ocean is deeply influenced by the presence of the Antarctic Circumpolar Current (ACC), where water masses of the World Ocean are advected, transformed, and redistributed to the other basins. It remains a challenge to describe and visualize the complex 3D pattern of this circulation and its associated tracer distribution. Here, a simple framework is presented to analyze the Southern Ocean thermohaline structure. A functional principal component analysis (PCA) is applied to temperature θ and salinity S profiles to determine the main spatial patterns of their variations. Using the Southern Ocean State Estimate (SOSE), this study determines the vertical modes describing the Southern Ocean thermohaline structure between 5 and 2000 m. The first two modes explain 92% of the combined θ–S variance, thus providing a surprisingly good approximation of the thermohaline properties in the Southern Ocean. The first mode (72% of total variance) accurately describes the north–south property gradients. The second mode (20%) mostly describes salinity at 500 m in the region of Antarctic Intermediate Water formation. These two modes present circumpolar patterns that can be closely related with standard frontal definitions. By projecting any given hydrographic profile onto the SOSE-based modes, it is possible to determine its position relative to the fronts. The projection is successfully applied on the hydrographic profiles of the WOCE SR3 section. The Southern Ocean thermohaline decomposition provides an objective way to define water mass boundaries and their spatial variability and has useful application for comparing model output with observations.

1. Introduction

The Southern Ocean has a latitude band with no meridional boundary, allowing the Antarctic Circumpolar Current (ACC) to flow from west to east around Antarctica. This nearly zonal ACC is organized in three major fronts that deeply influence the distribution of properties (Talley et al. 2011): from north to south, the Subantarctic Front (SAF), the Polar Front (PF), and the Southern ACC Front (SACCF). The Southern Ocean is traditionally bordered by the Subtropical Front (STF) to the north, forming a major water mass boundary between the warm, salty subtropical surface waters and the cold, fresh subantarctic surface waters (Orsi et al. 1995; Belkin and Gordon 1996). South of the southern boundary of the ACC (SBdy) lies the three subpolar gyre systems in the Weddell–Enderby Basin, the Australian–Antarctic Basin, and the Ross Basin. The location and intensity of ACC fronts have a significant influence on the heat, salt, and nutrient exchange in the World Ocean (Rintoul and Sokolov 2001) as part of the global meridional overturning circulation (MOC).

Several criteria have been proposed in the literature to determine the position and characteristics of ACC

Corresponding author e-mail: Etienne Pauthenet, etienne. pauthenet@misu.su.se

fronts. Historically, they have relied on phenological indicators (Wyrtki 1960; Deacon 1982; Pollard et al. 2002) such as the position of minimum temperature or salinity maximum layers. The tracking of specific hydrographic values at given depths were later preferred as they are generally less ambiguous (Park et al. 1993; Orsi et al. 1995; Belkin and Gordon 1996). Refined atlases of hydrographic properties in the Southern Ocean were used in the early 1990s to draw a circumpolar picture of ACC fronts (Orsi et al. 1995), which remains to date a widely accepted reference. The density gradient due to temperature and salinity gradients across ACC fronts generally drives a geostrophic current; thereby, a front is often associated with a swift deep-reaching current with a clear signature in sea surface height (SSH). Sokolov and Rintoul (2007) proposed frontal definitions using specific SSH values for the Australian–Antarctic Basin sector, later generalized to the entire Southern Ocean by Sokolov and Rintoul (2009). In their SSH-based view, they defined eight circumpolar fronts instead of the three proposed by Orsi et al. (1995), significantly complicating the picture of ACC fronts. These fronts are prone to merging and separations as part of the intense eddy activity observed along the ACC. A number of frontal definitions were later proposed based on this same idea (Boning et al. 2008; Sallée et al. 2008; Langlais et al. 2011; Chapman 2014; Kim and Orsi 2014), and it was found that SSH satellite observations were very useful to track fronts and eddies.

However, the notion of well-defined circumpolar fronts structuring the ACC has been increasingly debated lately, as it is unable to account for some major features of the ACC. Distinct differences are observed in the ACC structure between the different basins, with fronts being more intense in the Atlantic and Indian sector than in the Pacific sector. In several sectors, such as around the Kerguelen and Campbell Plateaus, surface and deep signatures of ACC fronts can diverge by several hundreds of kilometers (Park et al. 2008; Smith et al. 2013). Also, water mass fronts are not always unambiguously associated with frontal jets that can be tracked from the surface. Strong gradients in temperature and salinity can be density compensating (e.g., Graham et al. 2012), while the magnitude of gradients largely depends on the local strength of the topographic control or the mesoscale variability and can thus significantly vary with time. The idea of an ACC structured in circumpolar fronts conveys an essentially barotropic view, acting as an effective barrier to any meridional exchange, yet the ACC has a strong baroclinic signature that supports the southern MOC (Marshall and Speer 2012; Peña-Molino et al. 2014).

The distribution of temperature and salinity gives some clue on how the 3D structure is organized. The temperature distribution at the surface presents circumpolar contours similar to the STF, SAF, and PF, but farther south the gradients are too weak to control the stratification (Dong et al. 2006). The salinity distribution better describes the fronts of the Antarctic zone (AZ; SACCF; SBdy). Pollard et al. (2002) observed that the main features of the ACC zonation can be related to the stratification, specifically whether the stratification just below the mixed layer is controlled by the temperature or salinity vertical gradients or both acting together. Similarly, Carmack (2007) pointed out that the upper layers of high-latitude seas are permanently stratified by salinity (beta oceans), while the upper layers of subtropical seas are permanently stratified by temperature (alpha oceans). More recently, Stewart and Haine (2016) proposed maps of the distribution of alpha, beta, and transitions zone oceans. Comparing the salinity distribution at the surface and at ~600 m best illustrates the complex nonbarotropic nature of the ACC (Fig. 1). The salinity distribution at the surface closely covaries with the surface temperature distribution, but at ~600 m, it presents a very different structure. A low-salinity trench present in the Indian and Atlantic basins is opening in the Pacific, which provides a clear signature of the subduction of the Antarctic Intermediate Water (AIW). In the Pacific the low-salinity signal extends north, creating an exchange window with subtropical waters (Iudicone et al. 2007). In this region, actual observations diverge most dramatically from standard definitions of SAF and STF, challenging the reality of their circumpolar nature.

In this study, we analyze the shape of temperature and salinity profiles using an objective statistical method based on functional principal component analysis (PCA) in the multivariate case (Ramsay and Silverman 2005). Classical PCAs are widely used in climate science in the form of empirical orthogonal functions (EOFs). They have also been used to handle several oceanographic variables simultaneously in order to delineate water masses (de Boer and van Aken 1995; Lindegren and Josefson 1998). Yet, in this paper, the PCA is applied on a spline functional model of temperature and salinity profiles instead of applying it directly to the profiles themselves (hence the term functional). The extension of univariate functional PCA to multivariate functional data is of high practical relevance to reveal joint variation of the different variables (Happ and Greven 2015). A multivariate functional PCA provides an effective way to understand the structure of a multivariate system such as the 3D temperature–salinity structure of the Southern Ocean.

Functional PCA have become increasingly important in recent years and have been applied on various datasets and in many fields. For example, in medical sciences, it can be used to explore human cornea curvature images (Locantore et al. 1999) or brain scan images (Viviani et al.
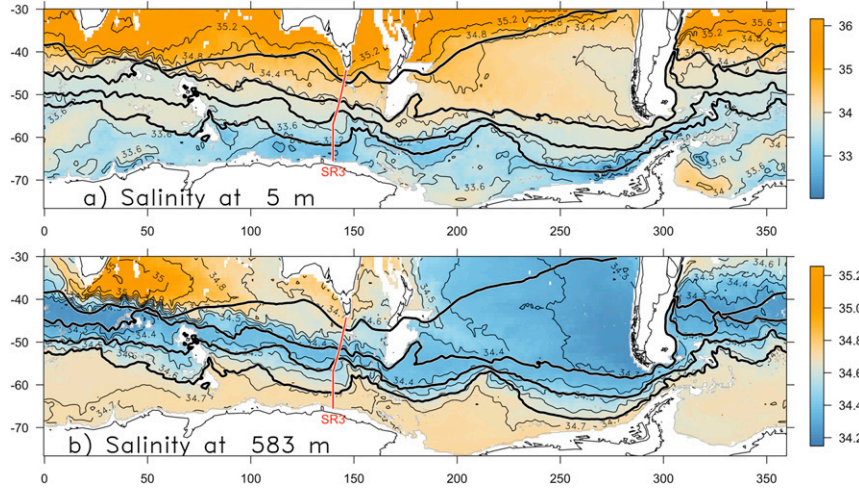
FIG. 1. Salinity structures between the surface and 600 m are very different. These maps are averages over the year 2007 of SOSE at 5 and 583 m for the location where bathymetry is deeper than 2000 m. The 2000-m bathymetric contour is in gray. The WOCE SR3 section (140°E) is in red, and the climatological locations of the STF (Orsi et al. 1995), SAF, PF, and SACCF (Kim and Orsi 2014) are in black.

2005). It also helped to describe handwriting dynamics (Ramsay 2000; Hosseinkashi and Shafie 2009), characterize Internet pageview dynamics (Moyer et al. 2015), classify the size distribution of a zooplankton community (Nerini and Ghattas 2007), or explore the variations of cadmium concentrations in Pacific oysters (Feng et al. 2011). Here, for the first time to our knowledge, a functional PCA is used in physical oceanography.

This paper is organized as follows: The first section describes the hydrographic datasets. The second describes the functional PCA method and how it has been applied on Southern Ocean hydrographic data. The third section presents the results of the decomposition, focusing in particular on the two first modes of variability that explain together more than 90% of the hydrographic variance in the first 2000 m of the Southern Ocean. The two last sections conclude on the potential of the functional PCA method and discuss the implications of our results for our understanding of the Southern Ocean stratification and frontal structure.

2. Data

a. The Southern Ocean state estimate

The main dataset used is potential temperature and practical salinity (θ–S) profiles from the Southern Ocean State Estimate (SOSE).[1] The SOSE product is a 4D estimate of the Southern Ocean physical state obtained by constraining an eddy-permitting simulation based on the Massachusetts Institute of Technology General Circulation Model (MITgcm; Marshall et al. 1997) to remain close to available Southern Ocean observations (Mazloff et al. 2010). The SOSE product has been preferred to hydrographic data or more standard climatological products because it is provided on a regular spatial and temporal grid, and yet it is by construction physically and kinematically consistent (Wunsch and Heimbach 2007).

The model grid has a $\frac{1}{6}°$ resolution, encompassing the Southern Ocean south of 25°S. We considered only data south of 30°S to avoid possible boundary effects. Moreover, 30°S is the most inclusive definition to fully encompass all Southern Ocean phenomena northward of the STF in each basin (Talley et al. 2011). It is possible to run a functional PCA on the total SOSE product, but such a resolution is not needed, as the focus is on large-scale variations. Therefore, we use only one data point every five grid points in both latitude and longitude, thus reducing the number of grid points by a factor 25. SOSE provides a 5-day mean that we average two by two, giving 36 outputs every 10 days for the period between the 1 January 2007 and 26 December 2007.

The vertical grid is composed of 42 unevenly spaced levels extending from the surface (5 m) to the bottom (5575 m). The statistical method requires us to work with profiles at a fixed-depth range. Considering that ACC fronts rarely cross regions shallower than 2000 m (e.g., Sokolov and Rintoul 2009) and that the bulk of

---

[1] For simplicity, we will refer to potential temperature as "temperature" in the following.

TABLE 1. Nomenclature table of the symbols used in the method (section 2) along with the size and a brief description of the vector/matrix they refer to.

Symbol	Size	Definition
$x_n(z)$	IR, $z \in [z_{min}, z_{max}]$	Continuous curve of temperature θ or salinity S of the nth location [Eq. (1)].
$\{\phi_1, \ldots, \phi_k\}$	K	B-spline basis of K functions [Eq. (1)].
$\mathbf{c}_n$	$2K$	Vector containing the K coefficients of decomposition in B-splines of the concatenated θ and S profiles of the nth location [Eq. (1)].
$\mathbf{X}$	$N \times L$	Matrix containing the N vectors $\mathbf{c}_n$ of length $L = 2K$.
$\mathbf{C}$	$N \times L$	$\mathbf{X}$ centered.
$\mathbf{V}$	$L \times L$	Crossed covariance matrix of $\mathbf{C}$ [Fig. 3; Eqs. (6) and (7)].
$\mathbf{M}$	$L \times L$	Weighting matrix used to balance the variables of different units (°C and psu) [Eq. (5)].
$\mathbf{W}$	$L \times L$	Metric equivalence between the functions $x_n(z)$ and the discrete version $\mathbf{c}_n$ [Eq. (5)].
λ_l	1	lth eigenvalue of the eigenvalue problem [Eq. (5)].
$\mathbf{b}_l$	L	lth eigenvector of the eigenvalue problem [Eq. (5)].
$\boldsymbol{\beta}_l$	L	Normalized version of the eigenvector $\mathbf{b}_l$ [Eq. (10)].
$\boldsymbol{\xi}_l$	IR, $z \in [z_{min}, z_{max}]$	Eigenfunction (vertical mode) of the lth mode [Fig. 4; Eq. (9)].
$\mathbf{y}_l$	N	lth principal components [Figs. 5 and 7; Eq. (11)].

existing hydrographic data (Argo data) is about 2000 m deep, it was decided to apply the method on θ–S profiles between 5 and 2000 m (28 levels). Regions where the bathymetry is less than 2000 m (e.g., Kerguelen, Falkland, and Campbell Plateaus or the Antarctic continental shelf) were discarded from the analysis.

b. WOCE SR3 section

The World Ocean Circulation Experiment (WOCE) coordinated numerous hydrographic sections in the World Ocean in the 1990s. We used the WOCE SR3 section (Rintoul et al. 1997; Sokolov and Rintoul 2002) as a case study. It has been broadly used and is located at a longitude representative of the zonally averaged Southern Ocean. This section was sampled from Tasmania to Antarctica along the 140°E meridian on board the *Aurora Australis* between December 1994 and February 1995. Only the profiles deeper than 2000 m were retained, consisting of 44 stations with temperature and salinity sampled every 2 m (Fig. 1).

3. Method

The goal is to analyze the simultaneous variations of θ and S profiles as a function of depth z. For each spatial position, an observation is defined as two vertical profiles of two different tracers (θ and S), described as continuous functional curves using a decomposition in a B-spline basis. Then we analyze the curves obtained with the specific method coined functional PCA by Ramsay and Silverman (2005). The terminology EOF is used among oceanographers and atmospheric scientists. EOF and PCA are the same methods, except that the term EOF is mostly used for analysis that seeks the main spatial structures in a two-dimensional dataset (space) with time as the sampling dimension. In our case we seek spatial structures in a one-dimensional dataset (depth) with space and time as the sampling dimension. The PCA allows us both to decouple the phenomena that govern the structure of the curves and to quantify the amount of variance that they induce on the water column structure. The analysis is applied on the coefficients of the B-spline expansion model, hence in a functional space. It transforms the original variables into new variables that are uncorrelated linear combinations of the original ones and concentrate the variance of the system. These new variables are principal components (PCs) and represent the most significant modes of data variation. In this section, the reader can refer to the nomenclature table (Table 1) that state every matrix, vector, and scalar with its size and a short definition.

a. Transformation of discrete profiles into functions

Depth can differ from one dataset to another, making obsolete most of the conventional statistical methods dealing with large data tables. To compare datasets they must be on a common vertical grid. This is achieved by linearly interpolating the SR3 and SOSE data onto a uniform vertical grid with 100×20 m levels. Then the profiles are projected on a chosen basis $\{\phi_1, \ldots, \phi_K\}$ of finite dimension K, which allows us to express a profile x_n^i as a linear combination:

$$x_n^i(z) = \sum_{k=1}^{K} c_{n,k}^i \phi_k(z), \quad i \in \{\theta, S\}, \quad n = 1, \ldots, N, \quad (1)$$

where the $c_{n,k}^i$ are coefficients estimated by regression using data on hand. In other words, the discrete profiles are smoothed, turning them into continuous curves $x_n(z)$ that could be compared at any depth. For this study the functions ϕ_k form a B-spline basis. The B-splines are
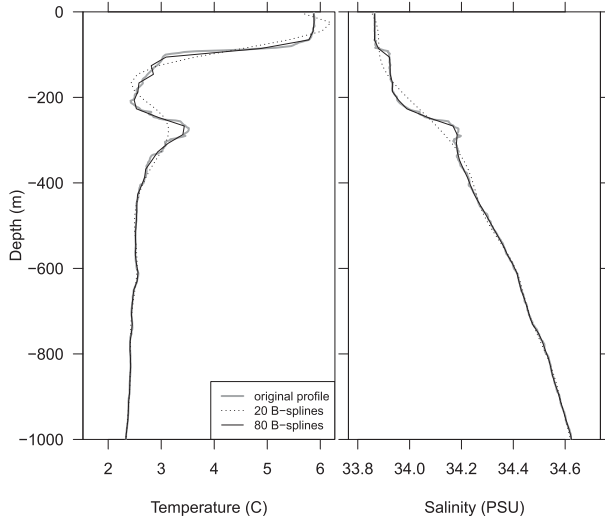
FIG. 2. Example of a profile projection into a B-spline basis of 20 and 80 functions revealing the smoothness of the estimation due to the number of functions selected. The profile is from the WOCE SR3 section at 52.67°S and 142.41°E. Only depths between 0 and 1000 m are shown because deeper vertical gradients are small and therefore the estimate is good. The gray line is the original profile, and the projections in a 20 (dotted line) or 80 (thin black line) B-spline basis are displayed on it.

polynomial segments of degree three joined end to end at argument values (knots), adjusted on the data under continuity and derivability constraints. Other choices of basis can be relevant as well, but this basis is best suited for nonperiodic functions and has the advantages of very fast computation and great flexibility (Ramsay et al. 2013). The number of segments controls the flexibility of the spline (Fig. 2). We used by default $K = 80$ spline functions to represent a 2000-m-deep profile.

The proposed method can keep the vertical complexity of the data (high number of basis functions) or disregard it by smoothing the profile (low number of basis function), which is useful when the data are highly corrupted by noise (Nerini et al. 2010). A projection on a basis of $K = 20$ functions would be sufficient to get a good estimate of the SOSE profiles, given their small, vertical resolution and smooth curvature. But the number of basis functions to use depends also on the dataset we want to project on the modes afterward. A smaller number of basis functions is relevant when using observations with sparse vertical data points and noise. Here, the case study (WOCE SR3 section) is composed of clean and high-resolution profiles (one sample every 2 m). These profiles can be estimated with a high number of basis functions to capture finescale without adding another source of error. We find that using 80 B-splines by profile allows us to get a negligible loss in the SR3 data, the absolute mean error being 10^{-3} psu for the salinity and 7×10^{-3} °C for the temperature. The

estimation error of the SOSE profiles with $K = 80$ B-splines is close to zero. This projection procedure from numerous raw data to a basis containing a reasonable number K of coefficients can also be useful as a first step toward dimension reduction.

b. Functional principal component analysis

For each spatial position, θ and S curves are projected into a B-spline basis with K coefficients. We therefore have a data table **X** of size N rows and $L = 2 \times K$ columns composed with the coefficients of that decomposition. Each row constitutes a vector $\mathbf{c}_n$ such that

$$\mathbf{c}_n = (c_{n,1}^\theta, \ldots, c_{n,K}^\theta; c_{n,1}^S, \ldots, c_{n,K}^S)' \qquad (2)$$

is composed with $L = 2 \times K$ elements that merge both coefficients of temperature and salinity profiles. The mean observation $\bar{\mathbf{c}}$ can be computed with

$$\bar{\mathbf{c}} = (\bar{c}_1^\theta, \ldots, \bar{c}_K^\theta; \bar{c}_1^S, \ldots, \bar{c}_K^S)', \qquad (3)$$

where $\bar{c}_k^i$, $i \in \{S, \theta\}$, $k = 1, \ldots, K$ corresponds to the mean of the N coefficients of column k in **X**. As in Eq. (1), a mean functional profile can be obtained for both salinity and temperature variables with

$$\bar{x}^i(z) = \sum_{k=1}^K \bar{c}_k^i \phi_k(z), \quad i \in \{\theta, S\}. \qquad (4)$$

Let us now consider the anomaly of coefficients matrix **X** denoted as matrix **C** where the mean observation $\bar{\mathbf{c}}$ has been subtracted from each row of **X**. The functional PCA consists of finding the unique decomposition of the matrix **VWM** by solving the following eigenvalue problem:

$$\mathbf{VWM}\mathbf{b}_l = \lambda_l \mathbf{b}_l, \qquad (5)$$

where $\mathbf{b}_l$ is the lth eigenvector associated with the eigenvalue λ_l. These L eigenvectors can then be ranked by order of their associated eigenvalues, so that the first vertical mode corresponds to the eigenvector with the largest eigenvalue and so on. The matrix **V** is the crossed covariance matrix of size $L \times L$ of the coefficients in **C** computed with

$$\mathbf{V} = \frac{1}{N}\mathbf{C}'\mathbf{C}. \qquad (6)$$

The structure of the data matrix **C** (rows composed with centered coefficients merged from θ and S decomposition) provides a block-structured covariance matrix **V** that can also be expressed as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{\theta\theta} & \mathbf{V}_{\theta S} \\ \mathbf{V}_{S\theta} & \mathbf{V}_{SS} \end{pmatrix}, \qquad (7)$$

where matrix $\mathbf{V}_{S\theta}$ means the covariance matrix of size $K \times K$ between coefficients of variable S and variable θ.

As B-splines do not form an orthonormal basis, the matrix $\mathbf{W}$ of size $L \times L$ ensures the metric equivalence between the functional problem (working on functions) and its discrete version (working on coefficients). It is constructed by block as follows:

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_\theta & 0 \\ 0 & \mathbf{W}_S \end{pmatrix}, \qquad (8)$$

where matrices $\mathbf{W}_i$ are constituted with scalar products of the basis functions $\{\phi_1, \ldots, \phi_k\}$, which are integrals in that case. They can easily be computed or approximated numerically. In that special case, $\mathbf{W}_\theta = \mathbf{W}_S$ because the same B-spline basis has been used for θ and S profile basis expansion. In the same way, the block diagonal weighting matrix $\mathbf{M}$ of size $L \times L$ is used to balance the coefficient values when decomposing both the variance of temperature and salinity data. It acts as a normalization step usual in standard PCA when variables do not have the same units.

As in Eq. (1), an eigenvector $\mathbf{b} = (b_1^\theta, \ldots, b_k^\theta; b_1^S, \ldots, b_k^S)'$ with associated eigenvalue λ will generate two eigenfunctions (ξ^θ, ξ^S) (called vertical modes in this paper) that can be expressed in the physical θ–S space with

$$\xi^i(z) = \sum_{k=1}^{K} \beta_k^i \phi_k(z), \quad i \in \{\theta, S\}, \qquad (9)$$

where coefficients

$$\boldsymbol{\beta} = (\beta_1^\theta, \ldots, \beta_k^\theta; \beta_1^S, \ldots, \beta_k^S)' = \mathbf{W}^{-1/2}\mathbf{M}^{-1/2}\mathbf{b} \qquad (10)$$

are a normalized version of the eigenvector $\mathbf{b}$ with respect to the metric. Here, we used the fact that $\mathbf{W}$ is a symmetric, positive, definite matrix that has a unique Cholesky decomposition $\mathbf{W} = (\mathbf{W}^{1/2})'\mathbf{W}^{1/2}$, where $\mathbf{W}^{1/2}$ is an invertible upper triangular matrix; $\mathbf{M}$ is a diagonal matrix, so its inverse square root is readily defined.

In the same way, observations can be projected in a space of small dimension when computing the PCs denoted as

$$\mathbf{y}_l = \mathbf{C}\mathbf{M}^{-1/2}\mathbf{W}^{-1/2}\mathbf{b}_l. \qquad (11)$$

For example, the N row coordinates of matrix $(\mathbf{y}_1, \mathbf{y}_2)$ provide the best 2D mapping of the profiles in the PCs space, associated with the highest projected variance $\lambda_1 + \lambda_2$; an observation is no more a B-spline curve but a curve in the space of the eigenfunctions.

c. Computation of the Southern Ocean modes

To compute the vertical modes of the Southern Ocean, we use θ–S profiles from the year 2007 of SOSE, as explained in section 3a. The crossed covariance matrix $\mathbf{V}$ of the SOSE profiles is represented in Fig. 3. Temperature is highly correlated (>0.83) with itself at any depth (block $\mathbf{V}_{\theta\theta}$). Surface salinity is anticorrelated with deep salinity; the transition is located at $\sim$500 m (contour 0 of the block $\mathbf{V}_{SS}$). The block $\mathbf{V}_{\theta S}$ (or $\mathbf{V}_{S\theta}$) reveals that temperature from 0 to 500 m is highly correlated (0.8) with surface salinity, whereas salinity at 500 m has a null correlation with temperature at any depth. The functional PCA finds the unique decomposition of this matrix $\mathbf{V}$ that decorrelates objectively the structures we just described.

Modes have been determined using the SOSE product; however, they define a basis on which any observations can be projected. Once the observations are projected in a space of small dimension $q < L$, it is possible to reconstruct each salinity and temperature curve with only these q dimensions using the attached eigenfunctions computed with SOSE. As a case study the WOCE SR3 hydrographic profiles are projected on the modes and reconstructed with only one or two dimensions in section 4b.

4. Results

a. Vertical modes

A total of 98.83% of the variance is captured by the six modes attached to the first six highest eigenvalues (Table 2). Only the first two PCs summarize as much as 92.41% of the variance. The effect of the different modes can be displayed by adding or subtracting the eigenfunctions (ξ_l^θ, ξ_l^S) to the notional mean profiles $(\bar{x}^\theta, \bar{x}^S)$, such that

$$\bar{x}^i \pm \sqrt{\lambda_l}\xi_l^i, \quad i \in \{\theta, S\}, \qquad (12)$$

as shown in Fig. 4. This representation is equivalent to the display of variables in standard PCA. It presents the effect of adding ($+$ curve) and subtracting ($-$ curve) a suitable multiple of each mode.

The first mode explains 72.52% of the variance and involves a modification of the whole water column in both temperature and salinity. The variation in temperature explained by PC1 goes from an almost unstratified, cold water column when PC1 is positive to a warm, strongly stratified column when PC1 is negative. The variation of salinity shows an inversion at 610 m. The $-$ curve presents fresher waters in the first 200 m, while the $+$ curve is maximum at 200 m and minimum at 1000 m. Temperature and salinity are contributing almost equally to this PC with a slight advantage of the temperature (61.37%).
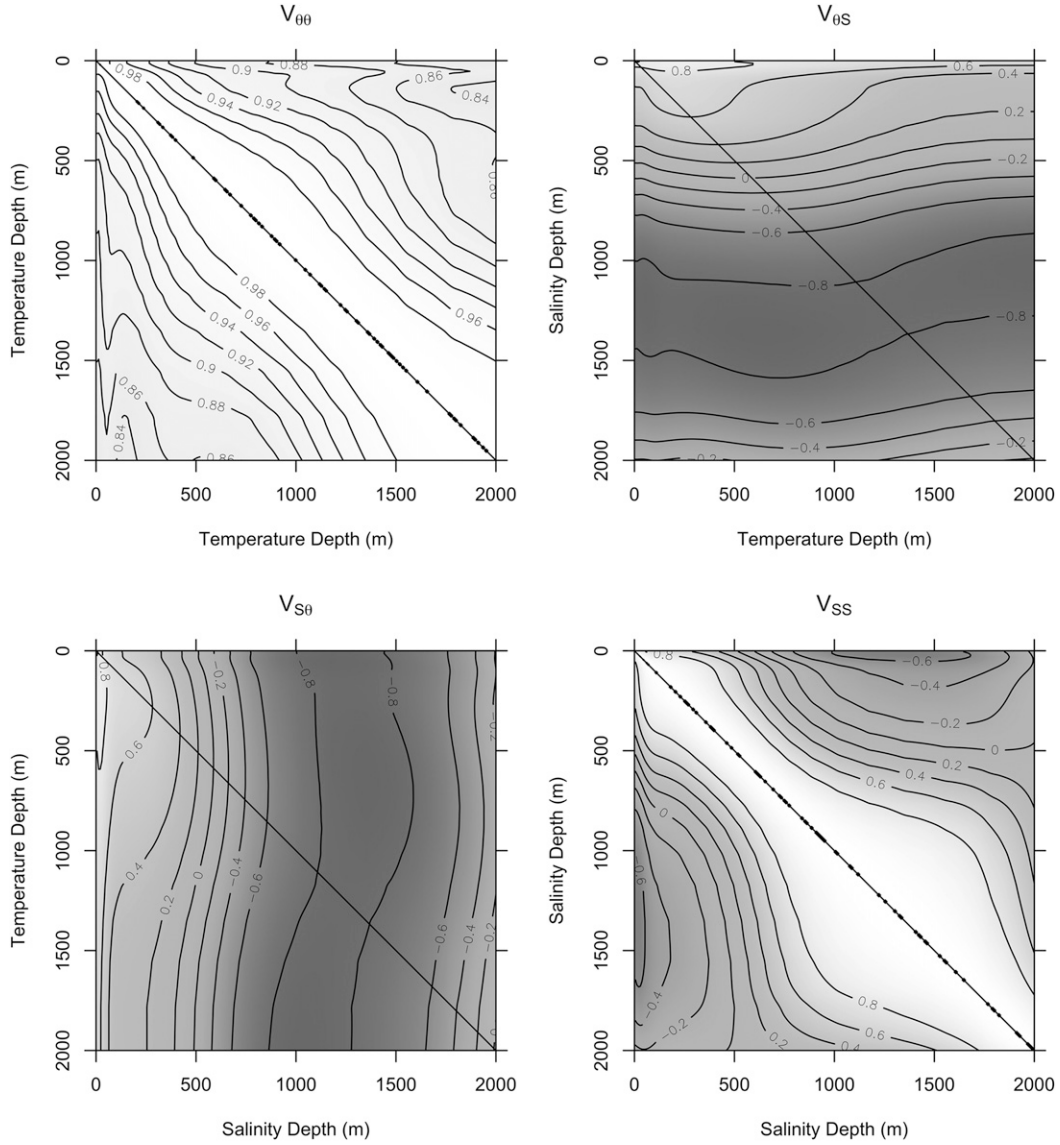
FIG. 3. Functional (normalized) representation of the block matrix **V** when SOSE profiles are lying in their B-spline functional space. On each panel, correlation between vertical profiles of pairwise variables (θ and S) can be read at any depth.

The second mode represents 19.89% of the total variance. It is largely driven by salinity with a relative contribution of 85.7%. When the PC2 value increases, the temperature decreases at any depth, with slightly larger variations in the top 150 m. The mode PC2 influences salinity from the near surface to ~1500 m, increasing with PC2 values. The maximum effect of PC2 on the salinity is found around 500-m depth. Note that

TABLE 2. The first six eigenvalues and their corresponding variance proportions computed on 20 or 80 basis functions for θ and S profiles from the year 2007 of SOSE.

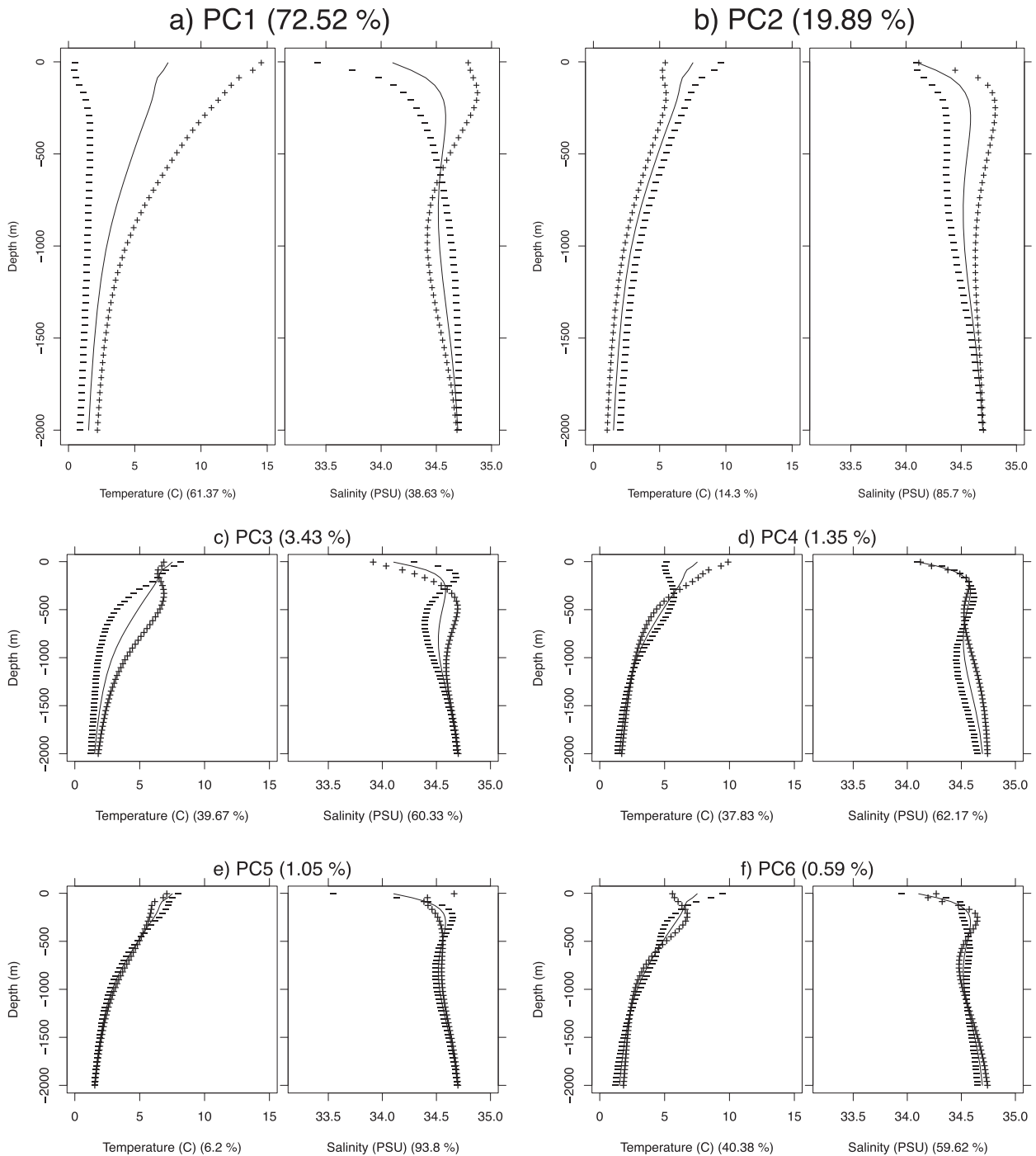|		PC1	PC2	PC3	PC4	PC5	PC6
20 basis functions	Eigenvalues	1.4529	0.3999	0.0689	0.0272	0.0199	0.0115
	Variance %	72.65	19.99	3.44	1.36	0.99	0.58
80 basis functions	Eigenvalues	1.4504	0.3978	0.0687	0.0271	0.0210	0.0117
	Variance %	72.52	19.89	3.43	1.35	1.05	0.59

FIG. 4. Representation of the effects of the first six vertical modes on temperature and salinity mean profiles. The curves show the mean profile (solid) and the effects of adding (+) and subtracting (−) eigenfunctions. The percentages next to the header titles are the amount of variance explained by the mode displayed. The percentages in the horizontal axis label are the variance contained by each variable (θ and S) on the mode displayed. For modes three to six each eigenfunction is multiplied by a factor of 2 to enhance the deformation of + and − curves.

the − curve has less salinity variation near the surface than the + curve, which means that larger PC2 values are associated with stronger salinity gradients just below the mixed layer.

Higher modes represent a smaller fraction of the total variance (7.59%), although they might be of large relative importance in some sectors of vertical profiles such as near the surface. PC3 has a significant influence on
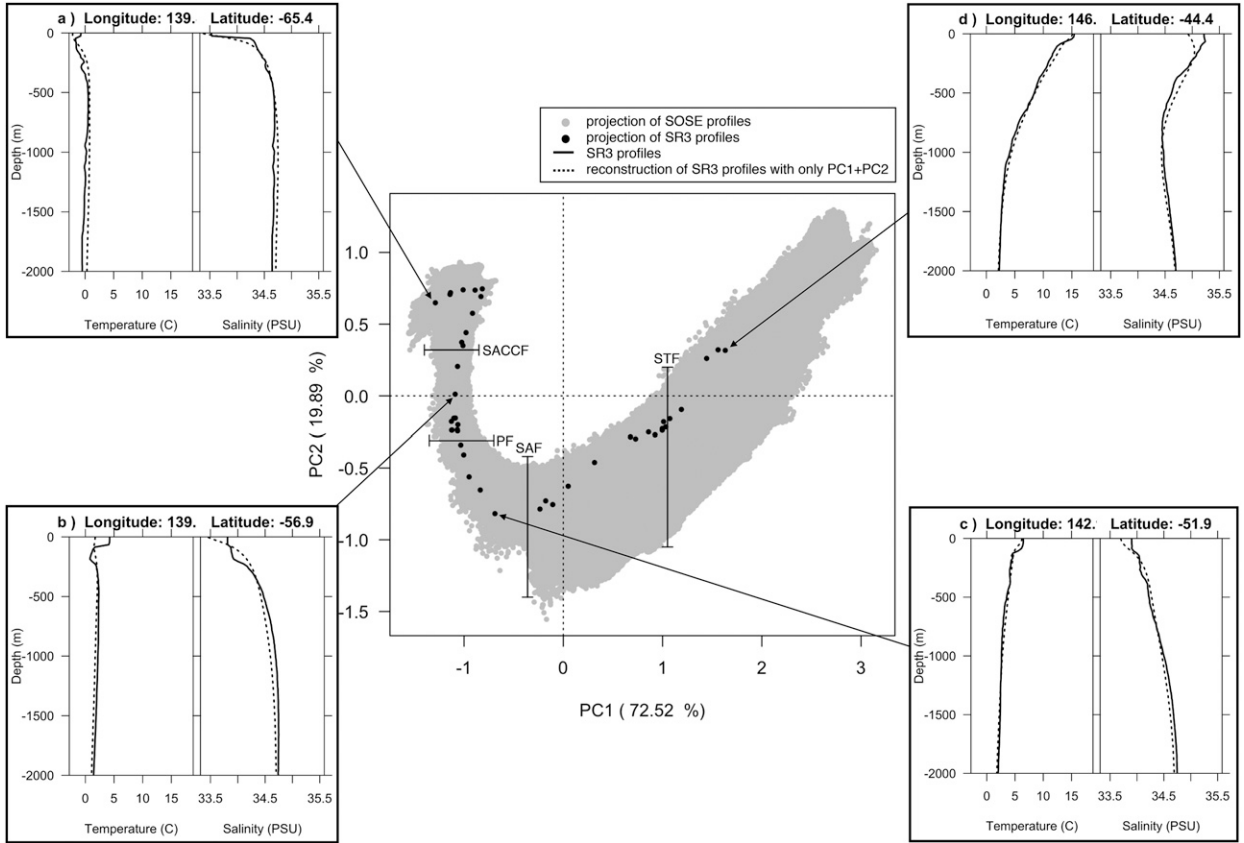
FIG. 5. 2D mapping of the functional PCA (main panel) and four examples of SR3 profiles projected on the SOSE modes. Every individual profile (two curves) is represented by a single point: gray for the SOSE profiles and black for the WOCE SR3 profiles. The black segments mark values of PC1 for the northern fronts (STF and SAF) and PC2 for the southern fronts (PF and SACCF) defined in section 5. The four panels around the main panel display specific SR3 profiles (black lines) and their reconstruction with PC1 + PC2 only (dotted black lines).

temperature below 250 m, and it has an inversion at 300 m for salinity, meaning that increasing the PC3 value freshens the surface while increasing salinity at intermediate depth. PC4 and PC6 have large signatures in surface temperature, while PC5 instead has a very strong signature in surface salinity.

A test of the sensitivity to the number of B-spline basis functions has been performed by redoing the PCA using only $K = 20$ basis functions (see Table 2). There is almost no effect on the variance explained of each of the first six modes, showing that the PCA is robust to the B-spline decomposition. The variance associated with PC5 and PC6 with a close fit (80 B-splines) is slightly higher than with 20 splines ($1.05 > 0.99$ and $0.59 > 0.58$). These changes are very small, but it implies that an increased vertical resolution of the model might put a little more variance to the higher modes. An analysis that extends deeper than 2000 m will not produce different modes as there are smoother changes of properties there. If a functional PCA is done on profiles that exclude the mixed layer (e.g., profiles that extend from 400 to

2000 m), the variance explained of the modes higher than PC4 is redistributed on PC1 to PC4.

Because PC1 and PC2 modes alone are able to capture more than 90% of the total variance, it is natural to attempt a description of the thermohaline structure of the Southern Ocean based only on these two modes. We can map the PC1/PC2 coordinate, where the origin corresponds to the notional mean θ–S profile, and each point in this plane represents a single θ–S profile obtained as a linear combination of the first two modes. The distribution of SOSE profiles in the PC1/PC2 map is displayed in the central panel of Fig. 5, showing a characteristic comma shape. To illustrate how temperature and salinity profiles vary on the PC1/PC2 map along the comma shape, we will now turn to the case of WOCE SR3 hydrographic profiles.

b. A case study: The WOCE SR3 section

Figure 5 allows us to visualize the shape of any hydrographic observations relatively to the modes. Here, the black points are the profiles of the WOCE SR3

section that we first transformed into functions (Fig. 2) and then projected on the SOSE modes to identify the features of θ and S responsible for the strongest components of variability: PC1 and PC2. In Fig. 5, we add black segments representing the fronts defined in section 5 as a visual aid. On the PC1/PC2 plane, each point contains the whole vertical structure of temperature and salinity profiles. Here, it displays a continuous comma shape, sorting the stations from the south [−1, 1] to the north [3, 1]. With the help of SR3 profiles (Fig. 5, black dots), we identify the subpolar region south of the SBdy [−1, 0.6] forming an angle with the Antarctic zone. Then the latter is delineated by PC2 only, PC1 being almost constant at −1 from the SBdy until the PF. The regime changes drastically in the polar frontal zone (PFZ, between PF and SAF) from a constant PC1 on the left side of the plot to PC1 and PC2 increasing together on the right side. The PFZ is the area of transition where temperature and salinity contribute equally to stratification (Pollard et al. 2002). North of the SAF temperature dominates in contribution to stratification, whereas south of the PF salinity dominates. On the PCs space the PFZ spreads on a large area (−1 to −0.2 for PC1) but only on ~4° of latitude in the section (Fig. 6). Finally the SAF and STF positions are consistent with the projection of SR3.

Hydrographic observations projected on the SOSE-based modes can be reconstructed with only the main modes of variation as illustrated here with the SR3 section. Four SR3 profiles are plotted with their PC1 + PC2 reconstruction in Figs. 5a–d. This mainly shows that the information kept by the modes higher than two is mostly surface stratification. More precisely, the reconstruction of the SR3 section with only PC1 and PC1 + PC2 is shown on Figs. 6d–g. PC1 does not take into account the surface stratification in temperature (Fig. 6d) characteristic of the Antarctic Surface Water (AASW). It only contains the difference in stratification that changes at ~52°S, both in temperature and salinity. The reconstructed temperature profiles are almost homogenous in the south at ~2°C and stratified in the north with warmer waters at the surface. The reconstructed salinity shows an inversion of stratification similar to the one described by the vertical mode (Fig. 4b): fresher surface waters in the south and past 50° saltier surface waters with a minimum at ~1000 m. Adding PC2 reveals more details in the temperature structure of the AASW, but it mainly solves the characteristic low-salinity tongue that was totally missed by PC1 alone (Figs. 6f,g). PC2 strengthens also the salinity stratification at the surface of the AASW and adds more details in the high-salinity surface waters north of

52°S. The temperature markers for the southernmost fronts (PF, SACCF, and SBdy) are poorly estimated by PC1 + PC2. Whereas the reconstruction of salinity marks these southernmost fronts by larger slopes, the northernmost fronts (SAF and STF) are equally marked in the reconstruction of both variables, with larger slopes of temperature and changes in the salinity from the surface to 750 m. Figure 6a is complementary of the main panel of Fig. 5. It displays the PC1 and PC2 of SR3 profiles and mark four values (red segments) that match best with climatology contours of fronts (see section 5). The reconstruction of the section with PC1 + PC2 + PC3 mainly shows an improvement of the temperature field under the winter water between 53° and 62°S (Fig. S1 in the supplemental material).

c. Spatial distribution of modes

By displaying the spatial distribution of the PCs averaged over the year 2007 (Fig. 7), we can identify the different regimes of water masses, defined by the modes. The first two PCs have similar shape from 30°S to the PF but with different amplitude. In the Antarctic zone, PC1 is almost constant with a value of −1 and is therefore not discriminant anymore. It is PC2 that reveals additional structures that are consistent with the most recent maps of mean dynamic topography (Rio et al. 2011). We identify, for example, the Fawn Trough Current (Roquet et al. 2009), the Weddell and Ross Gyres and even the multiple jets at 140°E described by Sokolov and Rintoul (2002). The South Pacific Gyre is revealed by null values of PC1 and minimum values of PC2 with contours directed in opposite directions; PC1 contours are going north with longitude, while PC2 contours are going south with longitude. Finally the boundary made by the STF south of the Agulhas system is intense in both PC1 and PC2.

Higher PCs show less zonal and more complex structures. PC3 is a variation related to the depth of the pycnocline (Fig. 4c). The positive values are the deepest pycnocline (Agulhas system and Indian Basin), whereas the negative values are the shallowest pycnocline (Weddell Sea, Pacific and Atlantic subtropical waters). The ACC shows a slightly positive signal, while the Pacific and Ross Gyre are null (Fig. 7c). This third mode especially differentiates the southern part of the three basins with positive values in the Indian, null values in the Pacific, and negative values in the Atlantic.

The negative values of PC4 associates cooler surface waters with fresher waters deeper than 600 m. In contrast, PC5 shows a strong signal of low salinity (− curve) at the surface driving the mode by 93.8% (Fig. 4e). This can be related to the seasonal ice
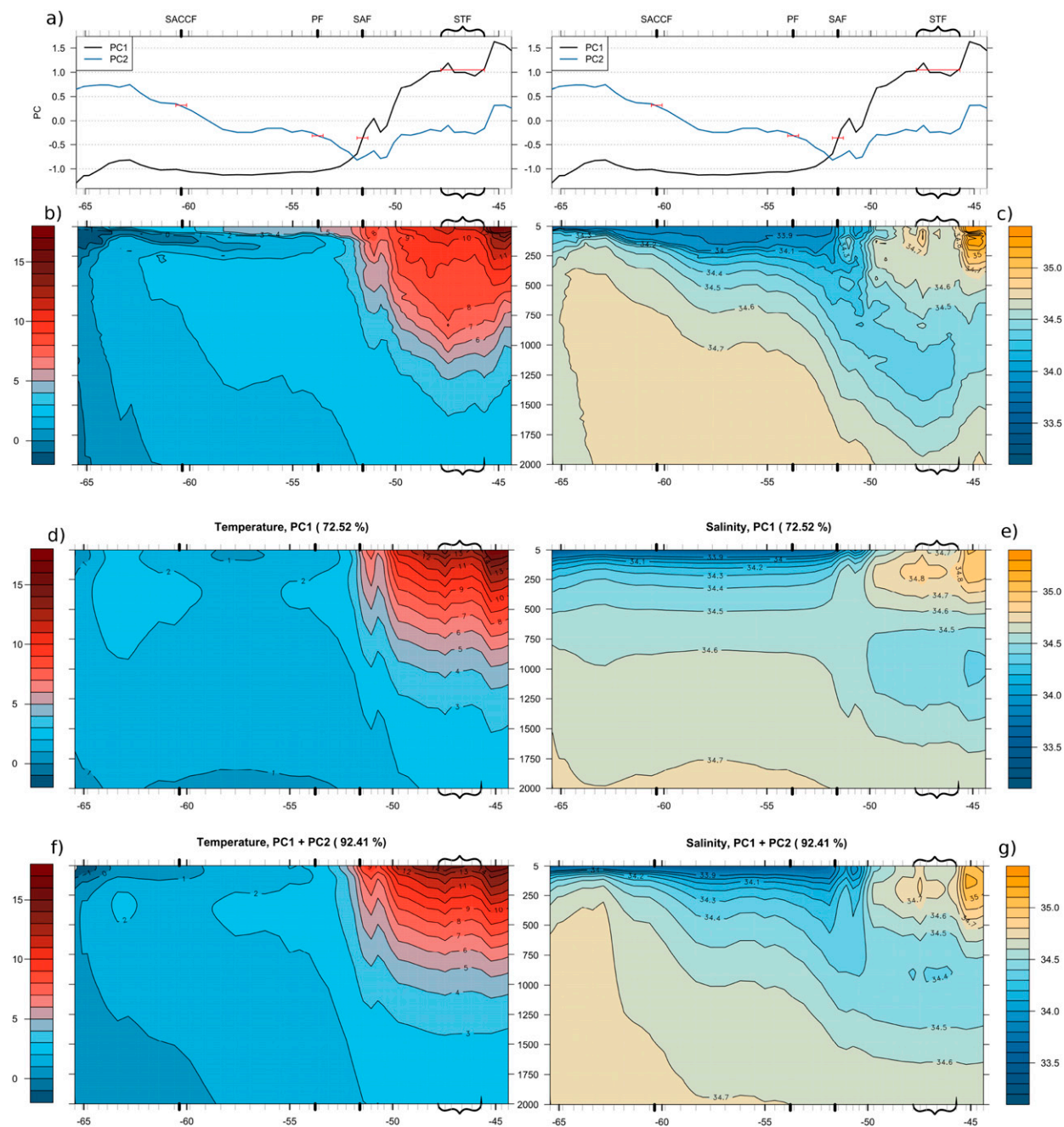
FIG. 6. Linear interpolation of the WOCE SR3 profiles along the latitudes for (b) temperature and (c) salinity and their reconstruction with (d),(e) PC1 only and (f),(g) PC1 + PC2 gives a great insight into the meaning of PC1 and PC2. The contour intervals are 1°C and 0.1 psu. The small gray ticks are the position of the stations along the section. (a) Identical plots of PC1 (black) and PC2 (blue) for the SR3 profiles against latitude. The red segments mark values of PC1 for the northern fronts (STF and SAF) and PC2 for the southern fronts (PF and SACCF) that best match the climatology contours of fronts (see section 5). The values selected are reported in Table 3, and the thick ticks show the corresponding location of fronts along the SR3 section. The STF is contained in a bracket because several profiles are close to the criterion PC1 = 1.05 (see Fig. 5).

melting and thus to surface release of freshwater in the ACC, as the map of PC5 shows negative values around the ice limit (Fig. 7e). The maximum values of PC5 highlight the Weddell and Ross Gyres as well as

the sluggish zone west of the Kerguelen Plateau. PC6 is not obvious to interpret but could be a seasonal thermocline mode as seen on its mode representation (Fig. 4f).
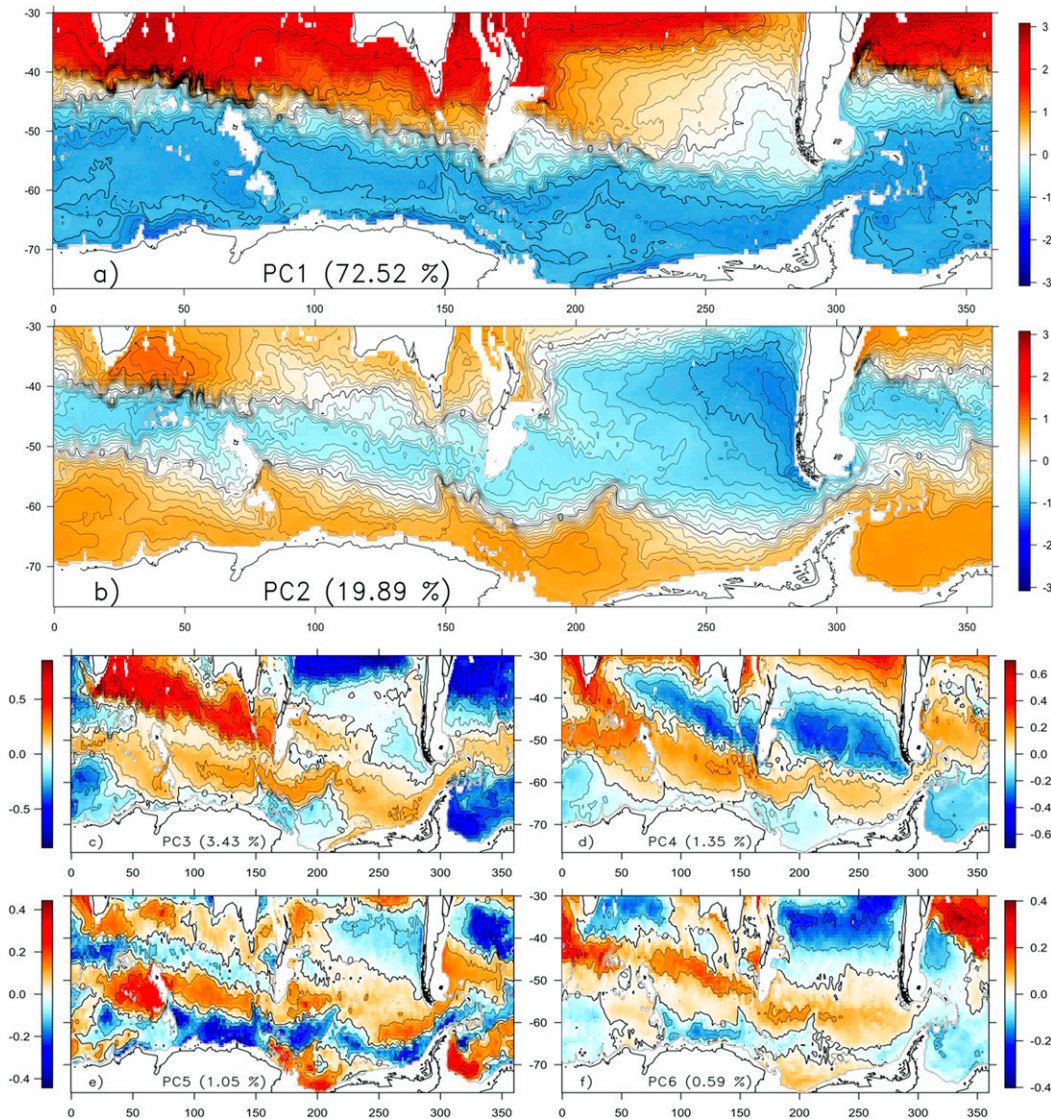
FIG. 7. Spatial distributions of PC1 to PC6 averaged over the year 2007 represent the distribution of water masses according to our PC decomposition (see Fig. 4). The contour interval is 0.1 and the contours −1, 0, 1, and 2 are thick and labeled. The percentages next to the titles in Antarctica are the amounts of variance explained by the PC displayed. The 2000-m bathymetric contour is in gray.

d. Relationship between modes and hydrographic properties

One can wonder which unique criterion, such as a tracer at a given depth, could best describe the whole vertical structure of the ocean. Thus, the correlation between the spatial distribution of the two first modes and temperature or salinity at any depth is computed (Fig. 8). The black profile is the mean, and the gray envelope is bordered by the maximum and minimum correlation for any time step. The seasonal variability is illustrated with the gray envelope around the temporal

average. These variations of correlation during the year are null except for the salinity at the surface, showing that the seasonal cycle has no influence on the correlations. PC1 and temperature have a correlation >0.76 at any depth and a maximum of 0.98 at 250 m, while PC2 and temperature have a low correlation at any depth (between −0.57 and −0.15). The correlation between PC1 and salinity is high (0.95) at the surface, vanishes at 610 m, and decreases until a minimum (−0.8) at 1355 m. This vertical variation is coherent with the expression of PC1 on salinity profiles (Fig. 4). The correlation between PC2 and salinity shows a maximum at 610 m and
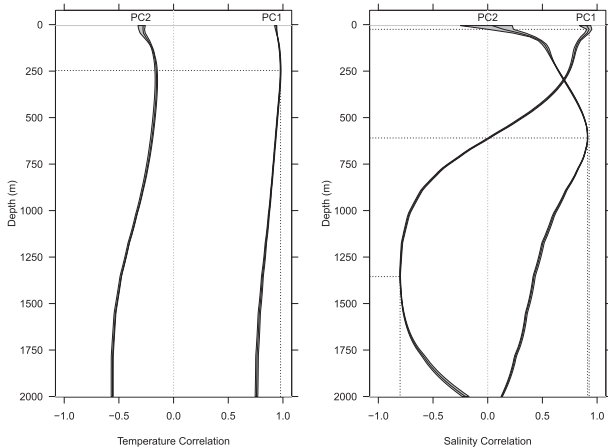
FIG. 8. Correlation between PC1 or PC2 and θ or S fields at any depth. The black line is the mean temporal value, and the limits of the gray envelope are the seasonal minimum and maximum at any depth. The dashed segments mark the minima and maxima of correlation.
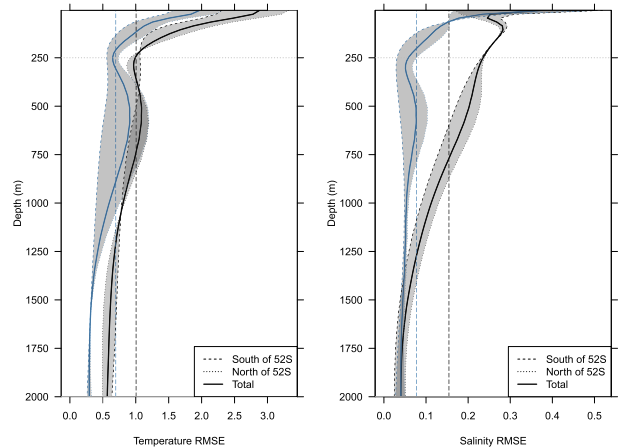
FIG. 9. RMSE of the profiles reconstructed with PC1 (black lines) and PC1 + PC2 (blue lines), computed at every depth for the whole region (solid lines), only south of 52°S (shaded lines) and only north of 52°S (dotted lines). The horizontal dotted gray lines are at 250 m to illustrate the minimum of RMSE for temperature and the choice of plotting only from 250 to 2000 m in Fig. 10. The vertical shaded lines are the global RMSE for temperature (1° and 0.7°C) and salinity (0.15 and 0.08 psu).

null values at surface and depth, in opposition with the correlation of PC1 and salinity. This illustrates the orthogonality between modes. To summarize, temperature at 250 m and salinity at 25 and 1355 m are the best correlated fields with PC1 (72.52% of the total variance), whereas the salinity at 610 m correlates very well with PC2 (19.89% of the total variance). This is coherent with Fig. 3 revealing a null correlation between salinity at ~600 m and temperature at any depth.

At the surface the salinity is the minimum in the south following the ice limit and the field is positively correlated with PC1 (Fig. 1a). Then at 610 m the salinity minimum shifts northward, revealing this low-salinity channel characteristic of PC2 (Fig. 1b). Note that the salinity minimum at 610 m extends in the whole South Pacific basin, which is not the case for the PC2 field (Fig. 7b). This is why the correlation between these two is only 0.91 and not 1. It also delineates well the Agulhas system and leakage. Thus, PC1 dissociates well the subtropical and subantarctic zone (i.e., STF), as seen in the temperature at 250 m and in the salinity at 25 m. It also reveals the SAF, best seen in the temperature at 250 m and in the salinity at 1355 m. PC2 adds the PF and SACCF, which appear in the salinity at 610 m.

e. Analysis of residuals

In this section, we describe how the residuals are distributed between the profiles reconstructed with PC1 and PC1 + PC2 only and the original profiles of θ and S. We use the root-mean-square error (RMSE) to represent these residuals on the vertical (Fig. 9) and on the horizontal (Fig. 10). The total RMSEs (Fig. 9, solid lines) are at their maximum at the surface and reach a

first minimum at 250 m for temperature. PC2 clearly improves the salinity estimation from 25 to 1500 m and does not at all modify the surface, while the temperature profiles are corrected on the whole water column, as expected from Fig. 4b. A second maximum is found at ~600 m, more marked on PC1 + PC2 RMSEs (blue). These two maxima (surface and 600 m) can be isolated by computing the RMSE on profiles north or south of 52°S. Indeed the 600-m bump is not found on the RMSE south of 52°S for temperature and salinity (dashed lines). Moreover the salinity RMSE south of 52°S obviously contains the surface error, which is the seasonal variability due to the ice, contained by PC5 (Figs. 4e and 7e). In contrast RMSE north of 52°S contains the 600-m bulb and most of the seasonal variability for temperature. So the maximum of RMSE at 600 m is concentrated north of 52°S and the Fig. 10 indicates that it is mainly due to the Subantarctic Mode Water (SAMW) that lies north of the SAF in the Australian–Antarctic sector. This area is also well highlighted by a positive PC3 in Fig. 7c. The 600-m maximum of temperature RMSE is not corrected by PC2 north of 52°S. The RMSE south of 52°S and deeper than 250 m is concentrated on the Weddell Sea (Fig. 10c). Figure 10 indicates that PC2 corrects well the profiles south of the STF under 250 m, as expected. The ACC band is almost not affected by this error (Figs. 10c,d), and PC1 + PC2 gives a good estimation of its vertical structure while disregarding the surface seasonal variability. Therefore, PC1 + PC2 is sufficient
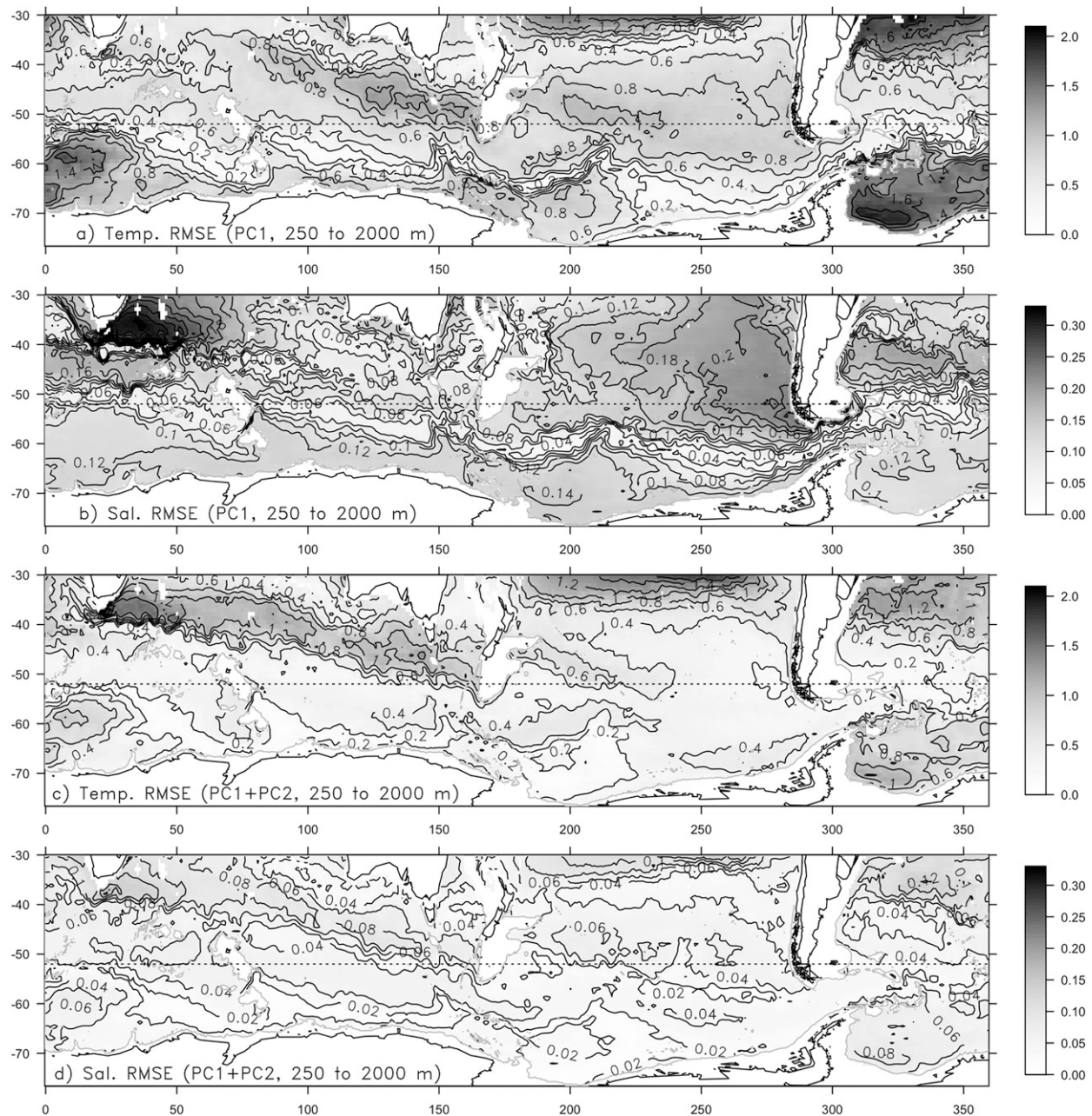
FIG. 10. Vertical RMSE of the profiles reconstructed with (a),(b) PC1 and (c),(d) PC1 + PC2 for temperature (°C) and salinity (psu) from 250 to 2000 m. The dotted line is at 52°S, illustrating the regional separation used in Fig. 9.

to detect ACC fronts, but it might be less accurate for the STF in the Indian and Atlantic basins.

5. Front detection

One potential application of the PC decomposition is to provide an objective way to detect fronts. It is apparent that some PC contours can be closely related to the climatological frontal path. Here, we want to

provide more quantitative information on how well PC contours can be used as proxies of the frontal position. The climatological position of four Southern Ocean fronts will be considered here. The reference STF position is taken from Orsi et al. (1995), and the SAF, PF, and SACCF are from Kim and Orsi (2014), who define fronts using SSH and validate them with Argo profiles. For the four circumpolar fronts, we first extract the PC1 and PC2 values of grid points that are
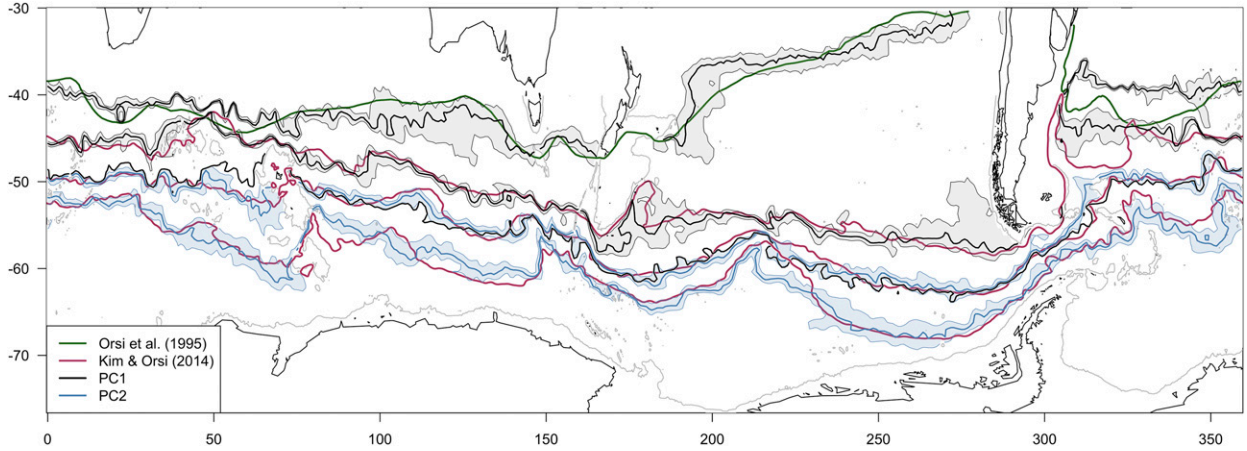
FIG. 11. Selection of PC contours (Table 3) that reveals the uncertainties around each front definition. Contours of PC1 (black) and PC2 (blue) are fitted on the SAF, PF, and SACCF (maroon; Kim and Orsi 2014), and STF (green; Orsi et al. 1995). PC1 reveals the STF, SAF, and PF (−1 contour) and then PC2 is used for PF and SACCF. The median isoline is shown in the middle of the colored band closed by the first and third quartile isolines. The 2000-m bathymetric contour is in gray.

closest to the climatological front position. Then we compute the median of extracted PC values, as well as the interquartile range (IQR), which is the difference between the third and first quartile (Fig. 11 and Table 3). We also extract the gradients of PC1 and PC2 along each circumpolar front. The average value of the PC gradient along the fronts is displayed in Table 3 and reveals a larger gradient for PC1 along the STF and SAF and a larger one for PC2 along the PF and SACCF. This means that, on average, the PC1 contours provide better estimators of the STF and SAF positions, yet PC2 contours work better for the PF and SACCF.

The values of PCs selected (Table 3) are shown in Fig. 5 (black segments) and Fig. 6 (red segments). The STF (PC1 = 1.05) is crossing several SR3 profiles on Figs. 5 and 6, which could indicate that the section is aligned with the fronts or the presence of a meander and eddies. Rintoul and Sokolov (2001) describe a recirculation between 44° and 48°S with water from the Tasman Sea flowing westward at 44°S. The SAF defined by PC1 = −0.36 is located at the right place according to the property indicators of Orsi et al. (1995), that is, $S <$ 34.2 under 300 m and $\theta > 4°–5°C$ at 400 m (see Figs. 6a–c). The PC values for the PF and SACCF are also coherent with Orsi et al. (1995). The PF (PC2 = −0.31) is found near the subsurface $\theta = 2°C$ and the SACCF for $\theta = 0°C$ along θ min at 150 m and $S > 34.73$ along S max at 800 m. Thus, the fronts' position defined with the vertical modes identified on the year 2007 of SOSE corresponds to the property indicators of fronts in SR3 that is done in 1994/95. This is a good indicator of the robustness of PC1 and PC2.

Figure 11 gives a visual idea of the match between selected contours of PC and standard frontal definitions. This method implies that the zonation is defined by the same criteria at every longitude. The contours on the PC maps are not tight all along the ACC (Figs. 4 and 11), which implies a weakening of the water masses' boundaries in some places.

The STF band is only narrow in the Atlantic and along the Agulhas return current, while the east Indian Basin and Pacific basin present smaller gradients (wider gray band). On the west side of the Pacific, the front forms a right angle that has also been described with SSH contouring (Sokolov and Rintoul 2009), and in the west Atlantic it is found ~5° south of Orsi et al. (1995). The wide STF area south of Australia may be due to the poor estimation done by PC1 in this area (Fig. 10a). It could also relate to the multiple STF jets in

TABLE 3. Values of PCs selected to reveal uncertainties around each front definition. The table contains the median and IQR of the closest values of PC1 and PC2 for every point of the STF (Orsi et al. 1995), SAF, PF, and SACCF (Kim and Orsi 2014). The IQR is the difference between the third and first quartiles. The column "gradient" contains the mean of the gradient along each standard fronts. The values in bold represent the colored bands displayed in Fig. 11. The median values are also displayed in Fig. 5 (black segments) and Fig. 6a (red segments).

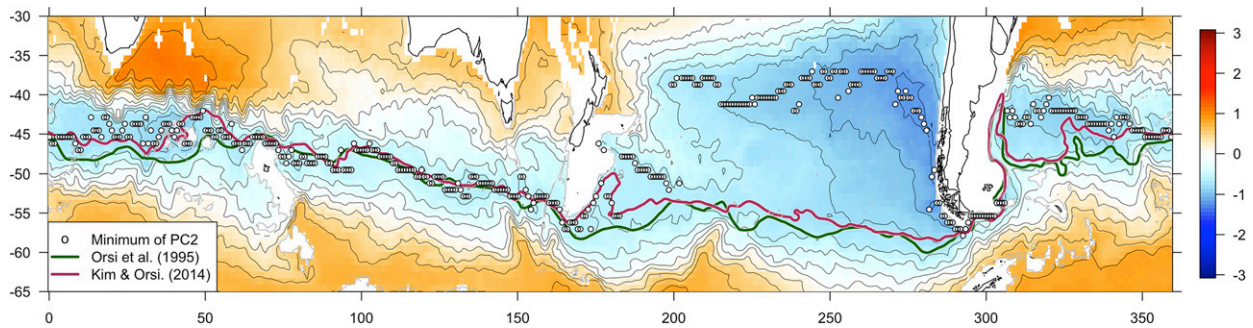|	PC1			PC2		
	Median	IQR	Gradient	Median	IQR	Gradient
STF	**1.05**	**0.44**	**0.17**	−0.27	0.60	0.11
SAF	**−0.36**	**0.33**	**0.12**	−0.65	0.16	0.09
PF	−1.00	0.09	0.08	**−0.31**	**0.19**	**0.09**
SACCF	−1.08	0.08	0.06	**0.32**	**0.26**	**0.09**

FIG. 12. Minimum of PC2 by longitude (white dots) coincides with the SAF from Orsi et al. (1995) (green) and Kim and Orsi (2014) (maroon) on the Indian Ocean. The 2000-m bathymetric contour is in gray.

this area as suggested by James et al. (2002). The SAF is well followed by PC1 with some discrepancies: PC1 contours surround the Crozet Archipelago and get wider southeast of the Campbell Plateau and in the east Pacific and Argentine Basin. The largest difference of path is on the Argentine Basin where PC1 = −0.36 is ∼5° north of the SAF of Kim and Orsi (2014). However, a southernmost contour of PC1 or PC2 (Figs. 7a,b) gives this particular spike specific of the Falkland Current, which means that a different dynamic must be occurring between 300° and 320°E for the SAF. The correlation between PC2 and salinity at 610 m is interesting; this is the depth at which the subsurface salinity minimum appears north of the SAF on SR3 (Fig. 6b) and where temperature begins to dominate the contribution to stratification. Therefore, the minimum of PC2 (or salinity at 610 m) is a good indicator of the SAF if this latter is only defined as the northern limit of temperature domination in contribution to stratification. Indeed on PC2 the SAF is found in the middle of the minimum value channel (Fig. 12) between the Kerguelen and Challenger Plateaus, otherwise it follows contours that mark the southern limit of the PC2 low-value channel (in Atlantic and Pacific basins). In the Pacific and Atlantic, this subsurface minimum of salinity is farther north, far away from the velocity-based fronts' definition (Fig. 12). This raises up the concept of "not so circumpolar" fronts as Pollard et al. (2002, p. 1) said, "It is not the fronts that are circumpolar, but the total ACC transport and scalar properties of the salinity and temperature fields."

The PF is well delimited by the −1 contour of PC1 (black contour on Fig. 11), yet the whole Antarctic zone has an approximate PC1 value of −1. Thus, the PF is a northern limit to where the thermal mode is not discriminant anymore to differentiate water masses. The PF is a transition between the thermal mode in the north and the haline mode in the south. This is also consistent with Pollard et al. (2002) describing the PF as the northern limit of salinity domination to stratification. As mentioned above, PC2 reveals the frontal structures in the Antarctic zone, starting at the PF. Indeed PC2 fits remarkably well the PF in tight contours except west of the Kerguelen Plateau where it gets wider, crosses through the plateau, and reappears more concentrated. The −1 contour of PC1 is passing completely north of Kerguelen and reappears farther south, like PC2 contours. In other words, around Kerguelen the haline mode depicts the PF south of the island like Kim and Orsi (2014) and Park et al. (2014), whereas the thermal mode follows a path north of the plateau similar to the definition of Orsi et al. (1995) and many others. The other notable difference is between south Georgia and the Falklands Islands where the contour of PC1 coincides with Kim and Orsi (2014), while PC2 contours are found north of that (Fig. 11).

The SACCF is driven by an intense subsurface salinity gradient and is globally well adjusted on PC2; the zone of discrepancies is located where the distribution of reoccurrence rates of Kim and Orsi (2014) frontal indicators are the largest. We must add that PC3 and PC4 are accurate indicators of the SB with some differences between the two. PC3 presents an intrusion on the eastern side of the Ross Sea (Fig. 7, 0 contour of PC3), while the 0 contour of PC4 follows very well the SB described by Orsi et al. (1995) without the intrusion of ACC water in the Ross Gyre.

6. Conclusions

In this study, we used functional PCA to decompose the vertical structure of Southern Ocean hydrographic profiles in order to establish a simple ranking of them. The originality of the analysis is to integrate the functional nature of θ–S profiles. The functional approach allows us to match any profiles from model or observation, and the result of the decomposition brings an

interesting view on the Southern Ocean circulation. It compares the shape of the profiles and identifies which information is contained in each variable, temperature, or salinity. The main modes of variation are directly related to the joint θ–S distribution. They do not capture the seasonal variance because the profiles reach 2000 m, revealing larger spatial variations at depth than in the surface seasonal layer.

We found that the main mode of variation is the north–south θ–S gradient opposing Antarctic to subtropical waters (PC1, 72.52%). This primarily thermal mode has a spatial distribution directly related to temperature at 250 m and salinity at 25 m. It explains well the STF and SAF but does not vary much in the Antarctic zone, delineating the PF. The second mode (19.89%) is rather associated with a variation of salinity between 25 and 1500 m. This haline mode reveals the AZ structures (PF and SACCF) and is best correlated to salinity at 610 m. These PC maps (Fig. 7) combined with the vertical modes (Fig. 4) are a new type of useful indicators to describe water structure and define fronts. Here, we did not give a fully objective method to locate frontal structure as we are based on climatological fronts (section 5). Yet, the PC values used to depict the frontal path are consistent with Orsi et al. (1995) classical property indicators of ACC fronts.

The PF is the only circumpolar front where PC1 and PC2 contours coincide (Fig. 7). PC1 only vary north of the PF and does not vary anymore south of it, allowing PC2 to dominate the Antarctic zone. Pollard et al. (2002) proposed a zonation primarily controlled by the ACC and second by the changing balance of stratification, with temperature dominating the stratification north of the SAF, contributing equally with salinity in the PFZ, and south of the PF, salinity dominates over temperature. It is interesting to note that our objective statistical method gives PCs with spatial patterns that can be connected to the zonation of Pollard et al. (2002).

In some areas PC1 and PC2 are acting against each other, revealing the strong baroclinicity of the flow. This is particularly the case west of the Kerguelen Plateau, southeast of the Challenger Plateau, in the southeast Pacific and Argentine Basin (Fig. 7). These areas also correspond to wider fronts on Fig. 11 and high values of baroclinic cross-stream transport (Peña-Molino et al. 2014, their Fig. 7). Indeed, the water mass fronts that appear on the maps of PC (Figs. 7, 11) vary in width with longitude, revealing weakening and strengthening of the water mass boundaries especially for the SAF in the Pacific basin.

The residuals associated with modes higher than two (section 4e) can be dissociated in two types: the surface seasonal variability and the subtropical waters high

gradients at $\sim$600 m (Figs. 9, 10). For example, the fifth mode captures the freshwater input to the surface from melting ice. Further investigations of higher modes are needed to interpret which physical processes govern them. This would be useful to study the temporal variability of the ocean structures.

The functional PCA decomposition is particularly powerful in the Southern Ocean probably because of the well-organized nature of its thermohaline structure. The ACC organizes the interbasin flow by receiving, mixing, and distributing waters of the three other basins, which is strikingly well captured by the functional PCA. With 3.4% of the total variance, the third mode differentiates well the structure of the three subtropical basins, and a 3D representation of the PCA similar to the Fig. 5, but with PC3 as the third dimension, would show a clear separation between basin properties north of the SAF, but we leave that for a future study. In theory nothing would prevent us from adding more than two variables in the analysis (e.g., oxygen and nutrients), and the maximum depth of profiles could be fixed at a different value than 2000 m. A different maximum depth (e.g., 300 m) or a more restricted geographic area (e.g., considering a smaller latitude range) would give more detailed and local information about the water mass distribution.

Note that the SOSE-based modes are not specific to the SOSE hydrographic distribution. Of course, a similar calculation of thermohaline modes based directly on available hydrographic profiles could be performed; however, it would probably not produce significantly different modes both because SOSE has been constrained with available observations and because we focused here on robust large-scale features of the Southern Ocean thermohaline structure, which is the existence and depth of minima below the mixed layer. The SOSE-based mode can thus be used as Southern Ocean modes to classify any θ–S observations and locate them relatively to the main ACC fronts as we did for the WOCE SR3 section.

Being able to determine objectively the type of a cast in terms of water mass properties is a valuable asset for many oceanographic problems. For instance, it can be extremely useful for data calibration. In frontal regions, the usual calibration methods are uncertain (Wong et al. 2003), and our study could improve that by locating any profile relatively to a climatology in a PC space, our advantage being to take into account the whole vertical shape. Another perspective of this work is to correlate PC modes to SSH variations in order to reconstruct temperature and salinity profiles from altimetry observations similarly to the gravest empirical mode projections (Meijers et al. 2011), with the advantage that

functional PCA do not need to assume an equivalent barotropic structure of the Southern Ocean. This method also has the potential to robustly assess changes in heat and salt content and could be used to study the spatiotemporal variability of the fronts.

It is the first time to our knowledge that functional PCA is applied to a problem in physical oceanography. Compared to classical PCA methods, that is, EOF in the standard meteorological terminology, functional PCA adds more flexibility, enabling us to simultaneously analyze datasets with unstructured sampling or heterogeneous data types. More generally, it is clear that functional PCA presents a large potential for data analysis in climate sciences.

REFERENCES

Belkin, I. M., and A. L. Gordon, 1996: Southern Ocean fronts from the Greenwich meridian to Tasmania. *J. Geophys. Res.*, **101**, 3675–3696, doi:10.1029/95JC02750.

Boning, C. W., A. Dispert, M. Visbeck, S. R. Rintoul, and F. U. Schwarzkopf, 2008: The response of the Antarctic Circumpolar Current to recent climate change. *Nat. Geosci.*, **1**, 864–869, doi:10.1038/ngeo362.

Carmack, E. C., 2007: The alpha/beta ocean distinction: A perspective on freshwater fluxes, convection, nutrients and productivity in high-latitude seas. *Deep-Sea Res. II*, **54**, 2578–2598, doi:10.1016/j.dsr2.2007.08.018.

Chapman, C. C., 2014: Southern Ocean jets and how to find them: Improving and comparing common jet detection methods. *J. Geophys. Res. Oceans*, **119**, 4318–4339, doi:10.1002/2014JC009810.

Deacon, G., 1982: Physical and biological zonation in the Southern Ocean. *Deep-Sea Res.*, **29A**, 1–15, doi:10.1016/0198-0149(82)90058-9.

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

Dong, S., J. Sprintall, and S. T. Gille, 2006: Location of the Antarctic polar front from AMSR-E satellite sea surface temperature measurements. *J. Phys. Oceanogr.*, **36**, 2075–2089, doi:10.1175/JPO2973.1.

Feng, C. X., J. Cao, and L. Bendell, 2011: Exploring spatial and temporal variations of cadmium concentrations in Pacific oysters from British Columbia. *Biometrics*, **67**, 1142–1152, doi:10.1111/j.1541-0420.2010.01534.x.

Graham, R. M., A. M. Boer, K. J. Heywood, M. R. Chapman, and D. P. Stevens, 2012: Southern Ocean fronts: Controlled

by wind or topography? *J. Geophys. Res.*, **117**, C08018, doi:10.1029/2012JC007887.

Happ, C., and S. Greven, 2015: Multivariate functional principal component analysis for data observed on different (dimensional) domains. ArXiv. [Available online at https://arxiv.org/abs/1509.02029.]

Hosseinkashi, Y., and K. Shafie, 2009: Persian handwriting analysis using functional principal components. *J. Stat. Res. Iran*, **6**, 141–160.

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, doi:10.1175/JPO2985.1.

James, C., M. Tomczak, I. Helmond, and L. Pender, 2002: Summer and winter surveys of the subtropical front of the southeastern Indian Ocean 1997–1998. *J. Mar. Syst.*, **37**, 129–149, doi:10.1016/S0924-7963(02)00199-9.

Kim, Y., and A. Orsi, 2014: On the variability of Antarctic Circumpolar Current fronts inferred from 1992–2011 altimetry. *J. Phys. Oceanogr.*, **44**, 3054–3071, doi:10.1175/JPO-D-13-0217.1.

Langlais, C., S. Rintoul, and A. Schiller, 2011: Variability and mesoscale activity of the Southern Ocean fronts: Identification of a circumpolar coordinate system. *Ocean Modell.*, **39**, 79–96, doi:10.1016/j.ocemod.2011.04.010.

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, doi:10.1016/S0169-7439(98)00072-0.

Locantore, N., and Coauthors, 1999: Robust principal component analysis for functional data. *Test*, **8**, 1–73, doi:10.1007/BF02595862.

Marshall, J., and K. Speer, 2012: Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nat. Geosci.*, **5**, 171–180, doi:10.1038/ngeo1391.

——, A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, **102**, 5753–5766, doi:10.1029/96JC02775.

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

Meijers, A. J. S., N. L. Bindoff, and S. R. Rintoul, 2011: Estimating the four-dimensional structure of the Southern Ocean using satellite altimetry. *J. Atmos. Oceanic Technol.*, **28**, 548–568, doi:10.1175/2010JTECHO790.1.

Moyer, D. C., S. L. Carson, T. K. Dye, R. T. Carson, and D. Goldbaum, 2015: Determining the influence of Reddit posts on Wikipedia pageviews. *Ninth Int. AAAI Conf. on Web and Social Media*, Oxford, United Kingdom, 75–82. [Available online at https://opus.lib.uts.edu.au/bitstream/10453/43941/1/Moyer%2615InfluenceRedditOnWiki.pdf.]

Nerini, D., and B. Ghattas, 2007: Classifying densities using functional regression trees: Applications in oceanology. *Comput. Stat. Data Anal.*, **51**, 4984–4993, doi:10.1016/j.csda.2006.09.028.

——, P. Monestiez, and C. Manté, 2010: Cokriging for spatial functional data. *J. Multivar. Anal.*, **101**, 409–418, doi:10.1016/j.jmva.2009.03.005.

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, doi:10.1016/0967-0637(95)00021-W.

Park, Y.-H., L. Gamberoni, and E. Charriaud, 1993: Frontal structure, water masses, and circulation in the Crozet Basin. *J. Geophys. Res.*, **98**, 12 361–12 385, doi:10.1029/93JC00938.

——, F. Roquet, I. Durand, and J. L. Fuda, 2008: Large-scale circulation over and around the northern Kerguelen Plateau. *Deep-Sea Res. II*, **55**, 566–581, doi:10.1016/j.dsr2.2007.12.030.

——, I. Durand, E. Kestenare, G. Rougier, M. Zhou, F. d'Ovidio, C. Cotté, and J. Lee, 2014: Polar front around the Kerguelen Islands: An up-to-date determination and associated circulation of surface/subsurface waters. *J. Geophys. Res. Oceans*, **119**, 6575–6592, doi:10.1002/2014JC010061.

Peña-Molino, B., S. R. Rintoul, and M. R. Mazloff, 2014: Barotropic and baroclinic contributions to along-stream and across-stream transport in the Antarctic Circumpolar Current. *J. Geophys. Res. Oceans*, **119**, 8011–8028, doi:10.1002/2014JC010020.

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, doi:10.1016/S0967-0645(02)00084-X.

Ramsay, J. O., 2000: Functional components of variation in handwriting. *J. Amer. Stat. Assoc.*, **95**, 9–15, doi:10.1080/01621459.2000.10473894.

——, and B. W. Silverman, 2005: *Functional Data Analysis*. 2nd ed. Springer, 426 pp.

——, H. Wickham, S. Gravers, and G. Hooker, 2013: fda: Functional data analysis, version 2.4.4. R package. [Available online at https://cran.r-project.org/web/packages/fda/index.html.]

Rintoul, S. R., and S. Sokolov, 2001: Baroclinic transport variability of the Antarctic Circumpolar Current south of Australia (WOCE repeat section SR3). *J. Geophys. Res.*, **106**, 2815–2832, doi:10.1029/2000JC900107.

——, J. Donguy, and D. Roemmich, 1997: Seasonal evolution of upper ocean thermal structure between Tasmania and Antarctica. *Deep-Sea Res. I*, **44**, 1185–1202, doi:10.1016/S0967-0637(96)00125-2.

Rio, M. H., S. Guinehut, and G. Larnicol, 2011: New CNES-CLS09 global mean dynamic topography computed from the combination of grace data, altimetry, and in situ measurements. *J. Geophys. Res.*, **116**, C07018, doi:10.1029/2010JC006505.

Roquet, F., Y. H. Park, C. Guinet, F. Bailleul, and J. B. Charrassin, 2009: Observations of the fawn trough current over the Kerguelen Plateau from instrumented elephant seals. *J. Mar. Syst.*, **78**, 377–393, doi:10.1016/j.jmarsys.2008.11.017.

Sallée, J., K. Speer, and R. Morrow, 2008: Response of the Antarctic Circumpolar Current to atmospheric variability. *J. Climate*, **21**, 3020–3039, doi:10.1175/2007JCLI1702.1.

Smith, R. O., R. Vennell, H. C. Bostock, and M. J. Williams, 2013: Interaction of the subtropical front with topography around southern New Zealand. *Deep-Sea Res. I*, **76**, 13–26, doi:10.1016/j.dsr.2013.02.007.

Sokolov, S., and S. R. Rintoul, 2002: Structure of Southern Ocean fronts at 140°E. *J. Mar. Syst.*, **37**, 151–184, doi:10.1016/S0924-7963(02)00200-2.

——, and ——, 2007: Multiple jets of the Antarctic Circumpolar Current south of Australia. *J. Phys. Oceanogr.*, **37**, 1394–1412, doi:10.1175/JPO3111.1.

——, and ——, 2009: Circumpolar structure and distribution of the Antarctic Circumpolar Current fronts: 1. Mean circumpolar paths. *J. Geophys. Res.*, **114**, C11018, doi:10.1029/2008JC005108.

Stewart, K. D., and T. W. Haine, 2016: Thermobaricity in the transition zones between alpha and beta oceans. *J. Phys. Oceanogr.*, **46**, 1805–1821, doi:10.1175/JPO-D-16-0017.1.

Talley, L. D., G. L. Pickard, W. J. Emery, and J. H. Swift, 2011: *Descriptive Physical Oceanography: An Introduction.* 6th ed. Academic Press, 555 pp.

Viviani, R., G. Grön, and M. Spitzer, 2005: Functional principal component analysis of fMRI data. *Hum. Brain Mapp.*, **24**, 109–129, doi:10.1002/hbm.20074.

Wong, A. P. S., G. C. Johnson, and W. B. Owens, 2003: Delayed-mode calibration of autonomous CTD profiling float salinity data by θ–S climatology. *J. Atmos. Oceanic Technol.*, **20**, 308–318, doi:10.1175/1520-0426(2003)020<0308:DMCOAC>2.0.CO;2.

Wunsch, C., and P. Heimbach, 2007: Practical global oceanic state estimation. *Physica D*, **230**, 197–208, doi:10.1016/j.physd.2006.09.040.

Wyrtki, K., 1960: The Antarctic Circumpolar Current and the Antarctic polar front. *Dtsch. Hydrogr. Z.*, **13**, 153–174, doi:10.1007/BF02226197.