Title
Water Mass and Biogeochemical Variability in the Kerguelen Sector of the Southern Ocean: A Machine Learning Approach for a Mixing Hot Spot

Journal
Journal of Geophysical Research - Oceans, 125(3)

ISSN
2169-9275

Authors
Rosso, Isabella
Mazloff, Matthew R
Talley, Lynne D
et al.

Publication Date
2020-03-01

DOI
10.1029/2019jc015877

Peer reviewed

Water Mass and Biogeochemical Variability in the Kerguelen Sector of the Southern Ocean: A Machine Learning Approach for a Mixing Hotspot

Isabella Rosso[1], Matthew R. Mazloff[1], Lynne D. Talley[1], Sarah G. Purkey[1], Natalie M. Freeman[1], and Guillaume Maze[2]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA.

[2]Ifremer, University of Brest, CNRS, IRD, Laboratoire d'Océanographie Physique et Spatiale, IUEM, 29280, Plouzané, France

Key Points:

- An unsupervised classification technique, applied to temperature and salinity float data, is used to sort the profiles into frontal zones.
- In eddy fields the variability of physical and biogeochemical properties is more than twice as large as the mean zonal variability.
- The intense eddy variability drives lateral physical processes that cause the large property variance.

Corresponding author: I. Rosso, `irosso@ucsd.edu`

Abstract

The Southern Ocean (SO) is one of the most energetic regions in the world, where strong air-sea fluxes, oceanic instabilities, and flow-topography interactions yield complex dynamics. The Kerguelen Plateau (KP) region in the Indian sector of the SO is a hotspot for these energetic dynamics, which result in large spatio-temporal variability of physical and biogeochemical (BGC) properties throughout the water column.

Data from Argo floats (including biogeochemical) are used to investigate the spatial variability of intermediate and deep water physical and BGC properties. An unsupervised machine learning classification approach is used to organize the float profiles into five SO frontal zones based on their temperature and salinity structure between 300 and 900 m, revealing not only the location of frontal zones and their boundaries, but also the variability of water mass properties relative to the zonal mean state. We find that the variability is property-dependent and can be more than twice as large as the mean zonal variability in intense eddy fields. In particular, we observe this intense variability in the intermediate and deep waters of the Subtropical Zone; in the Subantarctic Zone just west of and at KP; east of KP in the Polar Frontal Zone, associated with intense eddy variability that enhances deep waters convergence and mixing; and, as the deep waters upwell to the upper 500 m and mix with the surface waters in the southernmost regimes, each property shows a large variability.

Plain Language Summary

The Southern Ocean strongly influences the global climate system, by absorbing, storing and redistributing heat and carbon across the different ocean basins. Thanks to an increasing number of observations from autonomous instruments, called Argo floats, our understanding of this harsh environment has deepened in the last two decades. Here we use a machine learning technique to automatically classify the float measurements and sort them in regimes with similar properties based on their temperature and salinity vertical structure. The classification results are consistent with previous studies, but are here used to reveal regions where mixing between different types of waters is likely to be occurring. By sorting the float profiles into regimes, we can diagnose regions with larger variation of properties and highlight the transition of the properties across regimes. Given the increasing volume of observations that instruments like the Argo floats are building, a method such as the technique implemented in this study represents a valuable tool that can help to automatically reveal similarities in dynamical regimes.

1 Introduction

The Kerguelen Plateau (KP) is a prominent shallow topographic barrier to the Antarctic Circumpolar Current (ACC) in the Indian sector of the Southern Ocean, spanning 2000 km of latitude and reaching 3000 m in depth. The surface, intermediate, deep and abyssal circulation around the plateau is complex and acts to mix waters from different sources [e.g., *Aoki et al.*, 2008; *Park and Gamberoni*, 1997; *Tamsitt et al.*, 2017; *Llort et al.*, 2018], enhance phytoplankton productivity [e.g., *Maraldi et al.*, 2009; *Park et al.*, 2008a; *Van Beek et al.*, 2008], and connect Antarctic–sourced bottom waters with the lower latitudes [*Donohue et al.*, 1999; *Fukamachi et al.*, 2010], with implications for carbon and heat budgets [*Tamsitt et al.*, 2016; *Rosso et al.*, 2017]. Upon interaction with the plateau, the ACC is deflected, with most of the transport occurring north of the plateau [*Park et al.*, 1993] and through the Fawn Trough [*Park and Gamberoni*, 1997; *McCartney and Donohue*, 2007; *Park et al.*, 2008a] that divides KP into northern and southern parts (Fig. 1a). However, much of these complex dynamics are still poorly understood.

In recent decades, core (i.e., temperature and salinity only data) and biogoechemical-Argo (BGC-Argo) profiling floats have greatly augmented the spatial and temporal coverage of the top 2000 m of the Southern Ocean, a region that, because of the extreme conditions, is only marginally observed by ship-based platforms. In the present study, we use a set of core and BGC-Argo floats (the latter as part of the Southern Ocean Carbon and Climate Observations and Modeling project; SOCCOM) to explore the variability of physical and biogeochemical parameters (Section 4) within the intermediate and deep waters of the south Indian Ocean, in relation to Southern Ocean regimes and to the area around KP.

In order to classify individual Argo and BGC-Argo profiles into unique Southern Ocean regimes, we use a Profile Classification Model (PCM) approach based on machine learning unsupervised classification techniques [*Maze et al.*, 2017a]. The PCM is applied to individual temperature and salinity profiles, which are organized into groups with similar properties. This approach has shown skills in systematically classifying vertical profiles of the North Atlantic [*Maze et al.*, 2017b] and Southern Oceans [*Jones et al.*, 2018], without relying on user specified ad–hoc criteria for each profile. The PCM automatically identifies: multiple Southern Ocean (SO) zones, the Agulhas Current, subtropical and subantarctic mode water formation regions, and the system of currents around Australia.

A short review of the circulation, dynamics and water masses examined in this study is presented in Section 2. The dataset is presented in Section 3. We introduce the use of the PCM technique to classify float profiles in Section 4 and present the resulting Southern Ocean frontal zones in Section 5.1. We then highlight the variability of the physical and biogeochemical properties of the Antarctic Intermediate Water (AAIW) and Upper Circumpolar Deep Water (UCDW) across the different regimes depicted by the PCM classification, and in relation to topographic features (Section 5.3 and 5.4, respectively). Section 6 is a discussion with final remarks.

2 Background

In this section we describe some of the fundamental features of the South Indian Ocean circulation and dynamics, and highlight key questions for this study.

As the ACC encounters KP, its transport is divided into three different pathways, each with intensified, narrow currents guided by topography. The portion of the ACC that flows north of KP interacts with the southern limb of the South Indian subtropical gyre, which reaches the plateau via the Agulhas Return Current [ARC; *Park et al.*, 1993]. The ARC enters from the Crozet Basin, just west of KP [*Park et al.*, 1993], carrying salty and warm waters to the fresher and colder waters of the ACC. Here, eddy–induced transport convergence and subduction at both the mode and intermediate classes occurs [*Sallée et al.*, 2010]. A narrow and deep passage, the Fawn Trough (sill depth: 2650 m), divides KP around 56°S, 78°E and channels Antarctic waters into a strong, northeastward–flowing current (Fawn Trough Current) towards the Australian–Antarctic Basin (east of KP) [e.g., *Roquet et al.*, 2009; *Park et al.*, 2009, 2014].

Just south of KP, the eastward flow of the ACC navigates a narrow opening through the Princess Elizabeth Trough (PET; as deep as ~3700 m at 64°S). In the southern part of PET, the westward flow of the Antarctic Slope Front carries waters from the Australian-Antarctic basin [*Donohue et al.*, 1999; *Aoki et al.*, 2008]. These flows in the PET mingle the waters from both the Weddell Basin and the Adélie coast, turn northward around KP, and form a northward deep western boundary current that hugs the eastern edge of the southern plateau [e.g., *Donohue et al.*, 1999; *Aoki et al.*, 2008; *Fukamachi et al.*, 2010]. Can float-based temperature and salinity profiles be used by the PCM to automatically identify these pathways?

Downstream (East) of KP, the surface, deep and bottom waters all converge into a system of highly energetic mesoscale and submesoscale eddies and fronts [e.g., *Rosso et al.*, 2014; *Llort et al.*, 2018], which can facilitate the exchange of different water masses by injection and intrusion mechanisms [*Llort et al.*, 2018]. Their pathways, then, continue northwards across the mean flow, modified by mixing with ACC waters. Frontal positions are complex and highly variable in space and intensity [*Sokolov and Rintoul*, 2009a; *Freeman et al.*, 2016], especially upstream and downstream of topographic features that play a major role in controlling their position. Specifically, near and at Crozet Plateau and KP, the Subtropical (STF), Polar (PF), and Subantarctic (SAF) fronts can merge and divide into multiple branches [*Park et al.*, 2008b; *Sokolov and Rintoul*, 2009b; *Freeman et al.*, 2016]. Facilitated by strong cross–frontal injections and high eddy kinetic energy downstream of major topographic features, saltier subtropical waters exchange with subantarctic waters in Crozet Basin and pulses of AAIW are injected into the Subtropical Zone [*Park and Gamberoni*, 1997]. In addition, enhanced vertical velocities associated with increased meso- and sub-mesoscale energy downstream of topography can trigger subduction events from the surface to the mixed layer. These subduction events can facilitate the (1) export of carbon below the mixed layer [*Llort et al.*, 2018] or (2) flux of dissolved inorganic iron into the surface waters and are fundamental in triggering enhanced phytoplankton productivity in this region [*Rosso et al.*, 2016]. Can the PCM be used to identify frontal positions? How do injection events impact the classification of a float profile?

High mesoscale energy downstream of the plateau favors not only the vertical flux of iron [*Park et al.*, 2014; *Rosso et al.*, 2014], but also the upwelling of carbon and macronutrient rich deep water masses to the surface [*Tamsitt et al.*, 2017]. Localized upwelling into the mixed layer also occurs at the mode and intermediate water classes near KP [*Sallée et al.*, 2010]. Here, turbulent diapycnal mixing is enhanced throughout the full water column, driven by processes associated with local wind and tides (near-surface), internal wave shear and strain variance (interior), or generated by geostrophic flow over rough topography [near-bottom; *Meyer et al.*, 2015; *Whalen et al.*, 2015]. *Meyer et al.* [2015] found that diapycnal mixing is particularly enhanced in this ACC frontal region, driven by the dissipation of internal waves generated by the ACC's interaction with KP. This mixing is particularly important at the boundary between the AAIW and the denser Upper Circumpolar Deep Water (UCDW), which drives water mass transformation and consequently contributes to the overturning circulation [*Meyer et al.*, 2015]. Are there areas of major mixing that can be identified by the PCM classification method?

Following *Orsi et al.* [1995], the Southern Ocean regimes can be classified into the following zones, from north to south: Subtropical Zone (STZ; north of the STF), Subantarctic Zone (SAZ; between the STF and SAF), Polar Frontal Zone (PFZ; between the SAF and PF), and Antarctic-Southern Zone (ASZ; south of the PF and north of the southern boundary of the ACC, and including both of Orsi et al.'s Antarctic and Southern Zones). In this study, we focus on the spatial variability of the intermediate and deep water masses in the Indian sector of the Southern Ocean in relation to KP and these Southern Ocean frontal zones, which are in fact uniquely represented using the PCM technique applied here. AAIW represents one of the major water masses originating in the Southern Ocean and allows for the ventilation and transport of surface signals through much of the world's oceans [*Talley*, 1996]. AAIW is a cold and low-salinity water mass, defined by a salinity minimum at an intermediate depth of ~600–1000 m in the waters north of the SAF [*McCartney*, 1977; *Orsi et al.*, 1995; *Talley*, 2013], forming in the SAZ from the sinking waters south of the SAF at specific sites, such as the southeast Pacific [e.g., *McCartney*, 1977; *Sloyan et al.*, 2010] and the southwest Atlantic Ocean [e.g., *Piola and Gordon*, 1989]. AAIW is modified by mixing and intrusion of waters with different source origins throughout the Southern Ocean, such as in the southeast Pacific, southwest Atlantic [e.g., *McCartney*, 1977; *Piola and Georgi*, 1982], or central south Indian Ocean [e.g., *Park and Gamberoni*, 1997]. As a water mass sourced from surface waters, AAIW oxygen content is relatively high, but varies spatially across the Southern Ocean due to localization of its source and subsequent modification [e.g., *Talley*, 1996]. In the south Indian Ocean, the AAIW's core sits at an isopycnal with potential density $\sigma_0 \sim 27.3$ kg m^{-3} [*Talley et al.*, 2011], and with oxygen concentration as high as ~270 μmol kg^{-1} [*Park and Gamberoni*, 1997]. Where are the regions of larger variability associated with the different physical and biogeochemical properties? Can these regions be described by a single property, or are they property–dependent?

Above AAIW and north of the SAF, at a potential density of $\sigma_0 \sim 26.8$ kg m^{-3}, lies Subantarctic Mode Water [SAMW; e.g., *McCartney*, 1977; *Hanawa and Talley*, 2001; *Sloyan and Rintoul*, 2001; *Aoki et al.*, 2007]. This thick homogeneous layer ventilates the thermocline and originates from a combination of different processes, such as air–sea exchange, deep wintertime mixed layers, diapycnal mixing, advection, and eddy mixing [*Hanawa and Talley*, 2001; *Sallée et al.*, 2006; *Sloyan et al.*, 2010; *Cerovečki and Mazloff*, 2016] and is characterized by a minimum in potential vorticity [*Hanawa and Talley*, 2001]. The southeast Indian Ocean east of KP is a major source of SAMW [*McCarthy and Talley*, 1999], with a pool of low potential vorticity centered around 90°E,40°S and extending toward Australia. Here, the region's unique bathymetry controls the location of fronts [e.g., *Sallée et al.*, 2006; *Sokolov and Rintoul*, 2009b] and SAMW formation [located at the divergence of the STF and SAF; *Sallée et al.*, 2006], as well as the flavors of different types of SAMW. Both AAIW and SAMW play a fundamental role in regulating fluxes, storage, and transport of carbon, freshwater, heat, and nutrients [*Sabine et al.*, 2004; *Ito et al.*, 2010; *Sarmiento et al.*, 2004], and thus play a major role in controlling Earth's climate. Based on temperature and salinity profiles only, can the PCM classification method identify the south Indian Ocean SAMW region?

Below AAIW lies UCDW, which can be identified by its core at σ_0 ~27.6 kg m^{-3} [*Talley et al.*, 2011]. UCDW is a large volume of water which originates from the deep waters of the Pacific and Indian Oceans, with modifications in the Southern Ocean [*Talley et al.*, 2011]. As an old water mass, UCDW is characterized by an oxygen minimum and high nutrient content. In the southernmost zones of the Southern Ocean (i.e., south of the PF), upwelling UCDW brings very old and nutrient–rich waters to the surface, stimulating carbon outgassing [e.g., *Gruber et al.*, 2009; *Takahashi et al.*, 2009; *Gray et al.*, 2018] and local [*Prézelin et al.*, 2000] and remote biological productivity [*Sarmiento et al.*, 2004]. Biogeochemical profiling floats have recently been used to identify a much stronger outgassing of natural carbon in these regions than previously understood [*Gray et al.*, 2018]. Where are the regions of major variability associated with the different properties in the UCDW, and can these regions be described by a single property, or are they property–dependent?

3 Data: autonomous profiling floats

The present study is focused on variability of physical and biogeochemical (BGC) properties in the Indian sector of the Southern Ocean, using core Argo floats that measure temperature and salinity, and BGC–Argo floats that additionally measure oxygen (O_2), nitrate (NO_3^-), and pH. Our focus is on Southern Ocean regimes up- and downstream of Kerguelen Plateau.

Core and BGC-Argo profilers drift freely at 1000 m, descend to 2000 m after ~10 days, and ascend to the surface, profiling the water column. At the surface, they transmit the measurements via satellite. In this work, we use only core and BGC-Argo float data with quality control equal to 1, 5 or 8 [i.e. flagged as “good”, “value changed” or “estimated value”, respectively, as per Table 2 of the Argo manual, *Argo Data Management Team*, 2017; *Wong et al.*, 2012; *Johnson et al.*, 2017]. Argo float vertical sampling varies across float and sensor types (see Table 16 in *Argo Data Management Team* [2017] for a description). The vertical resolution of BGC-Argo data is higher in the upper 100 m, and decreases with increasing depth [*Johnson et al.*, 2017]. We therefore linearly interpolate each Argo and BGC-Argo vertical profile onto regular 1 dbar vertical spacing. The accuracy of oxygen data is 1±1%, nitrate is 0.5±0.5 μmol kg^{-1}, and pH is 0.005±0.007 [*Johnson et al.*, 2017]. For more technical details about BGC sensors, see *Johnson et al.* [2017] and *Riser et al.* [2018].

This study uses Argo and BGC-Argo profiles south of 30°S and between 0° and 180°E. The quality-controlled September 2018 Argo snapshot was extracted from the Global Data Assembly Center [*Argo Data Management Team*, 2018]; 822 Argo floats were selected in the area of study between December 2010 and September 2018, with a total of 103718 profiles, not including the BGC-Argo profiles. The quality-controlled SOCCOM September 2018 snapshot used in this study can be found in *Johnson et al.* [2018]. Between December 2014 and September 2018, 36 BGC-Argo profiling floats (with more than 5 profiles) were present in the study area (Table S1 in the Supporting Material and Fig. 1), for a total of 1847 profiles. The SOCCOM floats, mostly fabricated at the University of Washington from components purchased from Teledyne/Webb Research (Apex floats), but with some BGC Navis floats purchased from SeaBird Electronics, are listed in Table S1. They were deployed during the course of several US and international oceanographic campaigns (Table S1): three GO–SHIP (A12, I08S and SR03; `https://usgoship.ucsd.edu`) and four non GO–SHIP cruises (IN2016_V01, AU1603, SOE10, and ACE).

4 Method

4.1 Classification of profiles into regimes using machine learning

The waters in the Southern Ocean are often classified into zones divided by fronts, according to their properties (e.g., temperature, salinity, nitrate, oxygen) [*Orsi et al.*, 1995; *Gray et al.*, 2018]: the Subtropical Zone (STZ), with the warmest and most saline waters; the Subantarctic Zone (SAZ), with cooler and fresher waters relative to the STZ; the Polar Frontal Zone (PFZ), with a characteristic subsurface temperature minimum that tracks the

Antarctic surface water; the Antarctic-Southern Zone (ASZ) within the southern ACC, which is characterized by carbon- and nutrient-rich waters shoaling towards the surface; and the Sea Ice Zone (SIZ), which includes waters covered with sea ice during colder months. Because these fronts meander in space and time [*Sokolov and Rintoul*, 2009a; *Freeman et al.*, 2016], and the Argo data set is large, investigating changes of water properties within each regime can be complicated. A straightforward but limited approach is to separate float profiles according to the mean position of fronts [e.g., as in *Gray et al.*, 2018]. A more data–intensive approach, which takes frontal meandering into account, is to classify each profile based on specified characteristics of each frontal zone [e.g., as in *Williams et al.*, 2018], using *Orsi et al.* [1995] for the zone definitions.

Here, we use an unsupervised machine learning approach that groups profiles with similar vertical distributions of properties [*Maze et al.*, 2017a, called this a Profile Classification Model, or PCM]. If the historical choices of frontal zone properties are reasonable, then the machine learning approach should result in groupings that closely resemble the frontal zone groupings, and in fact this is what we find. In this method, the time-variable position of the fronts, or regimes, naturally arises from the spatial distribution of the groups defined by the data, and the statistics of the groups can account for the variability of front meanders. Given the large number and sparse coverage of profiles, the use of a PCM is particularly suited to investigate the variability of water properties in relation to the presence of the Kerguelen Plateau and Southern Ocean regimes.

A Profile Classification Model determines, without supervision, categories for a collection of ocean profiles. For each individual profile, the model gives the probability that it belongs to one of the determined categories. Our PCM methodology is based on *Maze et al.* [2017b], where the authors successfully used Argo temperature profiles in the North Atlantic Ocean to characterize the different regimes in the region. The PCM method was applied recently to Southern Ocean Argo floats by *Jones et al.* [2018], who classified the profiles into the major current systems and regimes using temperature data alone. In the present work, we significantly extend the PCM procedure using both temperature and salinity data, an implementation that gives a more robust identification of Southern Ocean zones, which are commonly defined by both temperature and salinity [*Orsi et al.*, 1995]. The advantage of using this unsupervised technique is that it treats the numerous profiles systematically, without relying on ad–hoc criteria for each profile, which can change over time [*Jones et al.*, 2018]. The method identifies the different regimes, allowing one to then diagnose hotspots of larger property variability within the identified regime. Frontal locations become apparent as connected quasi-zonal bands of variability arising from frontal meandering.

4.2 The PCA methodology

The PCM requires float data preparation and a classifier algorithm. For the classifier algorithm we chose a Gaussian Mixture Model [*Bilmes et al.*, 1998; *Bishop*, 2006]. This algorithm is based on the assumption that the data are generated by a mixture of a number of Gaussian distributions in D dimensions (defined by the number of principal components, as explained below), and takes into account the covariance of the data set as in *Maze et al.* [2017b]. We modified their data preparation procedure to include both temperature and salinity. Our procedure is as follows:

1. Float in situ temperature was transformed into potential temperature (θ), and practical salinity into absolute salinity (SA), using TEOS-10 [*IOC, SCOR, IAPSO*, 2010]. We have tested the method using practical salinity and θ, which would be more consistent with TEOS-10, and found no differences in the results (not shown for brevity). Thereby, we decided to use the combination of θ and SA for an easier comparison with previous works.
2. Profiles with valid QCed data between 300 m and 900 m were selected. The upper 300 m limit is below most of the deepest mixed layer depths in this area (note that here the number of profiles with deep winter mixed layers, down to ~500 m, are only a few tens, compared to thousands of profiles), which avoids influencing the classification algorithm by large seasonal variations. We have tested the algorithm using profiles with QCed data up to 50 m, but found that 300 m resulted in a more correct profile classification (not shown for brevity). The lower 900 m limit is selected for the practical reason that most floats have continuous quality data above this depth: extending the lower limit to a deeper depth would reduce the number of profiles that we can use for the study. Only 4% of the Argo profiles and none of the BGC-Argo profiles were rejected. It is possible to use the profiles to their individual maximum depths, but the missing values would need to be filled with, for example, the median of the data set or the most frequent value; this would create an unrealistic portion of the data set, thus it is not an acceptable solution, nor is it necessary.

3. We normalized each property measurement, i, by its standard deviation calculated at each 1-dbar level: $\theta_{i,n} = (\theta_i - \mu_\theta)/std(\theta)$ and $SA_{(i,n)} = (SA_i - \mu_{SA})/std(SA)$, where μ_θ and μ_{SA} are the depth-dependent averages of θ and SA, respectively, across the Argo float profiles, and $std(\theta)$ and $std(SA)$ are their depth-dependent standard deviations.
4. Following *Maze et al.* [2017b] and *Jones et al.* [2018], we reduced the vertical dimensionality of the problem (1 dbar data creates 600 vertical dimension points) by decomposing the Argo and BGC-Argo data set using Principal Component Analysis (PCA) applied to the 300–900 m layer. A PCA decomposition is a common method used in climate science and machine learning to detect the main covariance patterns in the data and reduce the number of dimensions. We found that ~99% of the property variance of θ and SA can be explained by the first 2 PCAs (Fig. S2 in the Supplementary Material), which are then used to reduce the profile dimension from 1,200 points (i.e. 600 depth levels for θ and 600 for SA) to 4 points (i.e. the 2 modal amplitudes for θ and 2 for SA). These 4 points are the inputs for the classifier algorithm. Note that, compared to *Maze et al.* [2017b] and *Jones et al.* [2018] we found a smaller number of PCAs, which explains the variability of our profiles. This is mainly due to the fact that we apply the algorithm to a only a portion of the full vertical range of the profiles: i.e., this range does not capture the deeper water masses, nor the surface ones which would increase the number of modes.
5. To assure an optimal and unbiased coverage of the analysis domain we selected a random profile in every $0.1° \times 0.1°$ box, similar to *Jones et al.* [2018]. 87,032 profiles were randomly selected this way and used as a training set for the classifier algorithm. This corresponds to ~ 89% of the Argo data set for this region.
6. The standardized and reduced θ and SA data were combined together in the same array $\vec{X}$ of dimensions 4×87032. The Gaussian Mixture Model algorithm then computes the optimal[1] Gaussian weights λ_k, mean μ_k and covariance Σ_k allowing one to compute the probability of a profile $x \in \vec{X}$ belonging to each "component" k of the Gaussian mixture:

$$p(k|x) = \frac{\lambda_k \mathcal{N}(x; \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \lambda_k \mathcal{N}(x; \mu_k, \Sigma_k)} \quad (1)$$

[1] i.e. the set of parameters maximizing the likelihood of all the data belonging to one of the clusters. This is computed using an Expectation-Maximization algorithm (see *Maze et al.* [2017b] for more details.)

where a Gaussian distribution $\mathcal{N}$ is given by:

$$\mathcal{N}(x; \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_k|}} \exp\left(-\frac{1}{2}(x - \mu_{\mathbf{k}})^\top \Sigma_k^{-1} (x - \mu_{\mathbf{k}})\right), \tag{2}$$

with $|\cdot|$ the determinant and T the transpose operators.

7. The sum over all components of the $p(k|x)$ is 1. The Gaussian mixture model is thus a probabilistic classifier, but note that each profile can be attributed to the component k for which the $p(k|x)$ is maximum. The relative amplitudes of the $p(k|x)$ are then used to assess the robustness of the classification.
8. Assessment of the classification is a fundamental step and can require subjective adjustments of the results (see an example in Section 5.1).

We performed several tests, using only one property (either θ or SA), or a combination of the different properties (θ, SA, nitrate, pressure), or single depth data (e.g., ~50 m and ~200 m), to reduce dimensionality instead of eigenvectors; but the combination of the first 2 PCAs of θ and SA was found to be the best choice, as this allows a definition of clusters that automatically capture most of the Southern Ocean regimes. In particular, in the subtropical and the southernmost zones, where salinity plays a fundamental role in setting the stratification, using temperature–only data would not correctly classify these areas. We also tested the PCM approach using an alternative classifier: the k-means algorithm, which assigns each profile to only one cluster k, based on the Euclidean distance of the profile to the nearest cluster mean [*Hartigan and Wong*, 1979], but found that k-means poorly separates the data in the southernmost regions.

Both PCA analysis and the Gaussian Mixture Model were performed using the Python scikit-learn version 0.20 machine learning package [*Pedregosa et al.*, 2011]. Our code was adapted from the pyXpcm software (`https://pyxpcm.readthedocs.io`), a Python implementation of Profile Classification Modelling [*Maze et al.*, 2017a]. The Gaussian Mixture Model used a "full" covariance matrix and 9 clusters (or components k). We tested the number of clusters between 5 and 15, and ultimately chose 9 as it allowed for a meaningful separation of the profiles into the desired Southern Ocean regimes. While there is no perfect way to choose between different numbers of clusters, we have validated this choice by looking at the Bayesian Information Criterion [BIC, *Schwarz et al.*, 1978; *Konishi and Kitagawa*, 2008], computed using 10 sets of randomly selected profiles, with a total number of 2166 profiles (~2.2% of the dataset). Although a clear minimum does not appear (as already found by *Jones et al.* [2018]), the method suggests an optimum value of between 9 and 15 clusters (see Fig. S3 in the Supporting Material).

4.3 Spatial variability of water masses

To describe how properties change with longitude, with respect to the location of KP and across the five regimes, we define four regions: West (0°–40° E), Upstream (40° E–68° E), Downstream (68° E–120° E), and East (120° E–180° E) of KP. These four regions identify different regimes of eddy kinetic energy both at surface [*Sallee et al.*, 2011] and at 1000 m depth [*Roach et al.*, 2018]. The properties are investigated in the intermediate (5.3) and deep water masses (5.4). In order to focus on the variability associated with specific water masses and remove the effect of isopycnal heave, we analyze the profiles in σ_0 coordinates rather than depth coordinates. We linearly interpolate each property profile in σ_0 density anomaly space with respect to a reference pressure of 0 dbar. We use a resolution of 0.03 kg m^{-3} between the 25.4 kg m^{-3} and 27.5 kg m^{-3} isopycnals, and a 0.01 kg m^{-3} step for denser classes, which resolves the density variations in both the upper and deep ocean.

In order to investigate the major hotspots of the variability associated with each property, we also compute the ratio of 1) the AAIW (or UCDW) property variance in 2° longitude bins and 2) the total AAIW (UCDW) property variance (computed for the whole domain, from 0° to 180° longitude): $\frac{var(C)_{2°}}{var(C)_{tot}}$, where $C = \left\{SA, \theta, O_2, NO_3^-, pH\right\}$.

5 Results

5.1 Resulting Argo profile clusters

The resulting 9 Argo clusters are shown in Fig. 2 ordered from north to south, where climatological fronts [*Orsi et al.*, 1995] are plotted in black for reference. The classification captures a roughly meridional structure from south of the Subtropical Front (STF) through the ACC that resembles the *Orsi et al.* [1995] frontal zones. North of the STF, the subtropical waters are classified into 5 distinct, quasi-zonal groups: the Agulhas Current region (k=2), the SAMW pool in the central subtropical gyre (k=3), the Australian currents system including the Tasman Sea, Great Australian Bight, and Leeuwin Current region (k=4), and the Subtropical Front, together with the Benguela Current and waters around Australia, at k=5. Clusters 6–8 identify the ACC waters, while cluster 9 depicts subpolar waters that comprise also the seasonal sea ice zone. Our classification compares well with the 8 clusters found by *Jones et al.* [2018] for the entire Southern Ocean. The biggest difference is north of the Subtropical Front, where our approach separates the Agulhas waters from the central and eastern Indian Ocean (SAMW and Australian waters).

The posterior probability, given as % value, is mostly large (more than 80%) for each cluster (Fig. 3 and 4). However, some profiles (more than 20% of the total number) in each classification component have a probability ≤70% which corresponds to a probability of ≥30% in at least a contiguous cluster, in particular in $k = 1$, 3, 4 and 5 (Fig. 5); these profiles tend to be concentrated in areas of strong currents (such as the Agulhas Return Current in cluster 4; the East Australian Current in clusters 1 and 5; the large air–sea exchange and deep mixed layers in the SAMW formation sites of clusters 3 and 4; the Southern Ocean fronts in clusters 7, 8 and 9), and in the Subantarctic Zone (cluster 6). The ambiguity in the classification comes from adjacent clusters, which is likely not due to a missing cluster to define these points. In order to check this possibility, we have calculated the same metric using a larger number of clusters (i.e., 15) and found no discernible difference with Fig. 5 (not shown). Thus, we do not discard any point with low probability, as this may be indicative of strong eddy and frontal dynamics, or of seasonal and interannual variability [*Jones et al.*, 2018].

5.1.1 Southern Ocean Zones

The variability of the waters in the Indian sector of the Southern Ocean is examined in terms of potential temperature, salinity, dissolved oxygen, nitrate and pH. In particular, we identify the variability for specific water masses 1) across fronts and 2) driven by the presence of the Kerguelen Plateau, as this large topographic feature is a site of convergence of upper, intermediate, deep and bottom waters [*Donohue et al.*, 1999; *Fukamachi et al.*, 2010; *Park et al.*, 2008a; *Tamsitt et al.*, 2017]. In order to identify specific water bodies, such as the Antarctic Intermediate Water (AAIW) or the Upper Circumpolar Deep Water(UCDW), we first classify each profile of the Argo and BGC-Argo data set by its Southern Ocean regime and then select the associated density class.

We define four Southern Ocean zones by grouping together some of the 9 clusters identified in Fig. 2, based on the $\theta - SA$ of each cluster (not shown): the Subtropical Zone (STZ; green profiles in Fig. 7, the set of all profiles with k in the range 1 to 5), Subantarctic Zone (SAZ; red, $k = 6$), Polar Frontal Zone (PFZ; blue, $k = 7$), Antarctic-Southern Zone (ASZ; orange, $k = 8$) and Sea Ice Zone (SIZ; magenta, $k = 9$). The resulting classification of the Southern Ocean regimes south of the PF (PFZ, ASZ and SIZ) is not affected by the number of clusters (k) from 5 to 15 (not shown). However, the classification is sensitive to the choice of k in the SAZ and especially in the STZ (not shown), where for $k < 9$ the algorithm grouped the southernmost STZ profiles in the SAZ.

The Profile Classification Model highlights areas of mixing between regions, where profiles show marked intrusions of waters in the upper ocean. An example of these intrusions is shown by the intense interleaving of temperature and salinity layers found in the profiles of BGC-Argo float with WMO ID #5904676 (Fig. 9). These interleaving structures indicate the occurrence of mixing and intrusions due to cross–frontal exchange, facilitated by vigorous eddy activity at the location of the front [*Park et al.*, 1993; *Llort et al.*, 2018]. These features are found both in the upper ocean (dashed lines in panel b, corresponding to the 2 markers in panel a and to the warmer and saltier intrusion of waters at 100–400 dbar in panels c and d) and at the salinity minimum of the AAIW (values at potential density anomalies σ_0 between 27–27.2 kg m^{-3} in panel b). The interleaving here occurs when the fresher and colder Subantarctic Surface Water comes into contact with the warmer and more saline water of the Agulhas Return Current, originating from the Agulhas Current and encountering ACC waters north of Crozet Plateau (~53°E) first, and then at Kerguelen Plateau [*Park et al.*, 1993; *Sallée et al.*, 2010].

The PCM method captures a feature in the area west and south–west of South Africa (in the Agulhas rings and the Benguela current, Fig. 2), in cluster 6 at ~30°S–35°S, ~10°E–20°E, which should be grouped within STZ waters, according to its θ and SA properties (Fig. 6). Most of the profiles of this cluster shows high posterior probability (Fig. 6a) and thus cannot be discarded. The $\theta - SA$ diagram in Fig. 6 shows that at latitudes close to 37°S (panel b), the surface waters (c) have temperatures warmer than 14°C, even in winter months (yellow colors in panel d), typical of subtropical waters in this region [*Park et al.*, 1993]. Thus, because of their upper ocean structure, we rely on a subjective definition and manually place these profiles within the STZ. We find that this feature is independent of the choice of k (not shown) and highlights an interesting connection at levels below 300 m (i.e., an undercurrent), between ACC waters and the region south and south-west of South Africa, which may indicate a pathway of intermediate and deep waters. This is not an error of the PCM method, as the PCM here captures similarities across data that connect intermediate waters from the ACC, but we decided to manually separate these points from the subantarctic zone, because of the usual classification of Southern Ocean zones.

Profiles with interleaving in the upper ocean are also found at warmer temperatures (not shown), likely due to either the passage of meanders and eddies of the SAF, by the proximity (and in some cases the merging) of the SAF and STF in some locations of the South Indian Ocean: close to ~30°E, near the Crozet Islands (~53°E), north of KP, east of ~125°E on the northern flank of the Southeast Indian Ridge [*Read and Pollard*, 1993; *Park et al.*, 1993; *Moore et al.*, 1997; *Freeman and Lovenduski*, 2016]; or due to drifting floats crossing a front. Of this type of interleaving, we find a group of profiles (Fig. 7a) with an upper thermohaline structure with temperatures warmer than 14°C and salinity larger than 34.7, which are typical of Subtropical Surface Water, rather than Subantarctic Surface Water of the upper SAZ [*Orsi et al.*, 1995; *Talley et al.*, 2011]. Therefore, we manually group these waters into the STZ. Yet, the advantage of the PCM approach is to use an algorithm that efficiently and automatically classifies a large number of data, without defining specific rules for each case. It is therefore beyond the scope of the present work to check every profile that could fall into this case, so we rely instead on a probabilistic approach.

The final Argo and BGC-Argo profile classification in Southern Ocean regimes used in this study is shown in Fig. 8. The classification shows generally a good comparison between the instantaneous margins of the STZ, SAZ, PFZ, ASZ and SIZ and the climatological STF, SAF, PF, and the southern boundary of the ACC [black contours; *Orsi et al.*, 1995], except for the STF east of 80°E. Qualitatively, the location of the different zones is also aligned with the areas delimited by the fronts identified by *Sokolov and Rintoul* [2009b] (not shown), where front locations are based on sea surface height. However, the temporal variability of the fronts is also large, as discussed in *Sokolov and Rintoul* [2009a] and *Freeman et al.* [2016], which could in part explain the largest misfit between the PCM zones and the climatological zones defined by *Orsi et al.* [1995]. Other regime-front mismatches may be due to the methodology for detecting a front's location, especially in proximity of complex topography [*Sparrow et al.*, 1996; *Sokolov and Rintoul*, 2009b]. In particular, the SAF and the PF have been shown to differ significantly in the Indian sector of the Southern Ocean, especially near Crozet Plateau and KP [e.g., *Orsi et al.*, 1995; *Park et al.*, 2009; *Sokolov and Rintoul*, 2009a; *Freeman and Lovenduski*, 2016].

The distinction between the different regimes is also evident in the mean float-based vertical profiles of θ and SA (Fig. 10). Here, we can see a net gradient in all the properties, both in the upper and deep ocean (except for salinity, as the ACC is the freshest), with a transition from the warmest and saltiest STZ waters to the coldest and fresher SIZ waters. South of the PF, in the PFZ, ASZ and SIZ waters (panel a, blue, orange and magenta lines), the upper ocean waters are very cold and fresh (Antarctic Surface Water) and show the typical subsurface temperature minimum of the Winter Water south of the PF, which is the remnant of the cold winter waters [*Talley et al.*, 2011]. At depth, the salinity minimum in the SAZ and PFZ is well captured, around 500 m and 1000 m, respectively (Fig. 10b). Furthermore, property variability, shown as the variance computed over each regime, is largest in the upper water column and at intermediate depths (~1000 m). This variability is due to the seasonal and interannual variability of the profiles, but may also reflect some mixing and interleaving.

Selecting only the BGC-Argo profiles, the distinction between the different zones is evident in the $\theta - SA$ and $\theta - O_2$ property diagrams (Fig. 7), with the oxygen increasing in the surface waters from the STZ towards Antarctica and with a minimum in each regime, which characterizes the UCDW core.

To highlight the effect of KP, in the following sections, we will describe how the properties associated with the four Southern Ocean regimes (STZ, SAZ, PFZ, ASZ and SIZ) vary as a function of longitude.

5.2 Subantarctic Mode Water floats

Compared to the full Argo data set used in this study, the BGC-Argo floats only marginally captures the SAMW formation pool in the southeast Indian Ocean region [see *Sallée et al.*, 2006; *Aoki et al.*, 2007, for maps of SAMW pool distribution in this region]. Hence, a complete discussion of the BGC property variability associated with this water mass and the influence of the ocean circulation on the SAMW is not possible. Instead, a description of the local properties captured by the BGC-Argo floats is given here.

BGC-Argo floats #5904688 (UW 9600), #5904683 (UW 9650), #5904682 (UW 9637) and #5904675 (UW 9749) (see Table S1 and Fig. 1) show the presence of SAMW formation in the wintertime deep mixed layers, in the σ_0 range between 26.65–26.85 kg m^{-3}. For example, Fig. 11 shows the vertical sections of θ (a), SA (b), O_2 (c) and potential vorticity (d) ($PV = -\frac{f}{\rho}\frac{\partial \rho}{\partial z}$, where f is the Coriolis parameter and ρ is the density) for float #5904688 (UW 9600). Between mid June and mid September 2016 and 2017, the winter mixed layer, computed using a density criterion of $\Delta\sigma_0$=0.03 kg m^{-3}, develops typical SAMW deep values ranging $\sim 400 - 700$ m. The temperature and salinity are well mixed in this volume of water (panels a and b), the oxygen concentration is high, with values around 270 μmol kg^{-1} (c), and the *PV* (d) is, as expected, very low ($\sim 20 \times 10^{-12}$ m^{-1} s^{-1}).

5.3 Property variability of the intermediate waters

We identify AAIW using σ_0 between 27.1 and 27.3 kg m^{-3} [Fig. 7; *Talley et al.*, 2011], and investigate its zonal variability in the STZ and SAZ, and at depth (Fig. 12). AAIW properties in figure are averaged across the AAIW density class and have a distinct meridional gradient, with saltier, warmer and lower oxygen waters in STZ than in the SAZ. The core Argo and BGC-Argo results are comparable as they should be since BGC-Argo is a subset of core Argo. Both show larger variability in the SAZ than in the STZ. Between 20°–40°E and 145–180°, the core Argo float data show high salinity waters that come from the Agulhas Current and the subtropical waters east of Australia, regions where BGC-Argo floats are not present. In the STZ, the largest difference in θ between the core and the BGC-Argo data (20°–180°E) is due to the warmer subtropical waters present in the northernmost region where there are no BGC-Argo floats (Fig. 8).

The largest oxygen concentration in the STZ ($\sim$230 μmol kg^{-3}) is found in the Upstream region, west of 20°E. East of this location, the oxygen concentration in the STZ is overall lower. The SAZ shows larger values (more than 260 μmol kg^{-3}) and spread for the oxygen concentration in the Downstream region, up to longitudes $\sim$100°E, east of which the oxygen concentration first decreases and then rises again east of approximately 150°E (nearly 260 μmol kg^{-3}). Larger oxygen concentration in the SAZ is consistent with the vicinity to source waters. The largest changes in the SAZ oxygen concentration are $\sim$52 μmol kg^{-3}, found in the Downstream region. Since the seasonal cycle is well captured in both the core and the BGC-Argo data sets, the difference might be due to local mixing and interannual variability.

Nitrate concentration in the STZ has larger values in the East region, with an overall range between 29–31.5 μmol kg^{-3}, with the maximum increasing towards east. We notice a large nitrate concentration ($\sim$33.5 μmol kg^{-3}) at approximately 110°E (Fig. 12), more similar to values found in the SAZ. This is an indication of the mixing across the subtropical front and intrusion of STZ waters, which create some ambiguity in the PCM classification of temperature and salinity profiles. In the SAZ, again we notice the largest concentration and spread in the Downstream region, with a maximum nitrate concentration of 32.2 μmol kg^{-3}

and a change in nitrate concentration of approximately 3.5μmol kg^{-3}. East of the Downstream region, the nitrate concentration decreases again, to values similar to those found in the West and Upstream regions (about 31 μmol kg^{-3}). We notice a group of data points in the East region (east of 125°E), with values less than 29 μmol kg^{-3}, which could be indicative of, as in the previous case, ambiguity in the classification or intrusion of low nitrate waters. This location corresponds to the Australian–Antarctic Discordance, a deep and rough section of the South East Indian Ridge that allows the passage of the Antarctic Bottom Water [e.t. *McCartney and Donohue*, 2007]. Lower values of nitrate here can be indicative of intrusion of low–nitrate waters through the Discordance.

The pH sensors mounted on some floats failed (`http://soccom.ucsd.edu/floats/SOCCOM\_sensor\_stats.html`), so the spatial coverage of pH data is not as dense as for the other parameters, neither in the STZ nor in the SAZ waters. However, we notice that the STZ pH is overall lower than in the SAZ, except around 100°E in the Downstream region and further in the East region, east of 140°E. At 100°E two distinct regimes appear: this could be due to the more limited sampling coverage of pH compared to other properties, due to sensor failure. However, BGC properties are found to have different gradients and fronts, compared to temperature and salinity (see for example the difference between the nitrate and the physical fronts found by *Freeman et al.* [2019]). Investigating what drives the difference between pH and nitrate is left for future work, which would require to consider the frontal structure of the BGC fields, together with the temperature and salinity fronts. As mentioned above, the property variability in the STZ waters, depicted by the spread of the values in Fig. 12, is smaller than the variability in the SAZ. Furthermore, the variability is predominantly associated to the two frontal zones and is zonally dependent, with larger spread in the Downstream region for each property and no evident impact due to the seasonality (not shown). As previously stated, we note that for those profiles at the edge of two separate zones or with marked interleaving (associated to profiles with water characteristics that are intermediate between two adjacent regimes), the PCM method could show some ambiguity and the classifier algorithm might not robustly distinguish the profiles' regime (Fig. 4). This can be indicative of leaking of fronts, where mixing is not instantaneous, and can explain the similarity of some values in the STZ and SAZ noted in Fig. 12.

The variability of AAIW in the STZ (Fig. 13 red markers; the black line corresponds to the ratio of 1) shows a large hotspot in the West region, around 10–30°E, for SA (the core Argo $var(SA)_{2^\circ}$ is more than twice as large than $var(SA)_{tot}$), θ (more than 1.5 times larger for both the Argo and BGC-Argo) and O_2 (more than 1.5 times larger). The core Argo SA and θ variance is large also in the East region, at about 170°E. The BGC-Argo does not capture the same magnitude of the Argo variance in the West region, likely due to the Agulhas current variability, which is not captured by the BGC-Argo profiles, or by interannual variability. The variability in SA and θ decreases with the longitude east of ~30°E (i.e. away from the Agulhas current), in both the core and BGC-Argo ensembles. Finally, in the STZ both oxygen, nitrate and pH (despite the coverage in space being less dense than the other properties) show hotspots of binned variance larger than the total variance east of 140°E. Additionally, nitrate variance is larger in the Downstream region, around 110°E.

The regional transition in the SAZ (Fig. 13 cyan markers) highlights hotspots of larger variability (i.e. > 1) for salinity and temperature around 70°E between the Up- and Downstream regions, then around 80°E and 160°E for both salinity, temperature and oxygen. NO_3^- shows a slightly larger variability at 60°E in the Upstream region, and between ~125–145°E in the East region. Finally, pH shows large variance in the eastern side of the Downstream region (~ 110°E and in the East region, at approximately 125°E).

5.4 Property variability of the deep waters

The site of Circumpolar Deep Water (CDW) is found below AAIW in the STZ and SAZ and below Antarctic Surface Water south of the SAF. CDW is made up of an upper (UCDW), characterized by an oxygen minimum centered around $\sigma_0 = 27.6$ kg m^{-3} (from *Talley et al.* [2011]), and a lower layer (LCDW), with a typical salinity maximum ($\sigma_0 =$ 27.8–28.27 kg m^{-3}, from *Talley et al.* [2011]) originating from North Atlantic Deep Water. Because LCDW is mostly found below 2000 m and in the upper water only in the southernmost regimes, we here only analyze UCDW.

The average of the properties at densities $27.6 \leq \sigma_0 < 27.8$ kg m^{-3} is given in Fig. 14. The values for the ASZ and the SIZ have been separated from the other zones for visualization purposes, and are presented in the right panels of the figure. SA and θ data are shown for both core (small markers) and BGC-Argo data (large markers), for comparison and statistics.

The meridional gradients of each property, from the STZ to the PFZ (left panels in Fig. 14), are well defined, with increasingly colder UCDW waters poleward. In the Downstream region, the meridional gradients switch sign: from the saltier, oxygen–rich, nitrate–

poor and higher pH waters of the West region (particularly in the STZ south west of South Africa), to fresher, oxygen–poor, nitrate–rich and pH low UCDW waters of the East region. We notice that the change is larger in the STZ than in the SAZ and PFZ. Very interesting is the larger spread associated with each of the properties in the PFZ Downstream region, specifically in the area east of the Kerguelen Plateau, which is indicative of mixing processes. We also notice that in between the Upstream and Downstream regions the nitrate concentration is similar across the different Southern Ocean zones, with a difference in magnitude of approximately 3 μmol kg^{-3}.

The southernmost zones (ASZ and SIZ) have the coldest and, in the SIZ, the most oxygenated UCDW waters. The range in each property values surpasses the northern zones, especially in the SIZ, because of the interaction of waters with the atmosphere and the formation/melting of sea ice. The largest ranges of the BGC-Argo data in the SIZ are found at 0–20°E, 80–95°E and 140–155°E, which are well comparable to the core Argo data spread. The larger spread in the SIZ is indicative of the proximity with surface waters, as in these southernmost regions the UCDW upwells to the surface and interacts with the mixed layer (not shown for brevity). In particular, the higher oxygen in these zones can be explained by the temperature–driven higher solubility of these waters and by the springtime primary productivity in the sea ice zone, as a consequence of melting of sea ice [*Briggs et al.*, 2018]. Note that a caveat of our analysis resides in the fact that the outcropping of UCDW can largely impact the property variance from surface processes rather than the interior mixing of water masses. To discern between surface processes and mixing events, one should remove the layers that interact with the mixed layers.

The large spread in the PFZ properties (left panels in Fig. 14, blue markers) is likely a signature of mixing with ASZ and SIZ waters that are interacting with the surface. Whether any longitudinal property evolution or meridional property gradient between zones is the signature of strictly irreversible isopycnal transformations (isopycnal mixing) or results from a combination of isopycnal and diabatic processes (e.g., from the surface turbulent layer, topographic induced turbulence or other irreversible processes) can be explored using potential vorticity PV. Averaged over the depths of the UCDW density classes, PV shows three striking features (Fig. 15): first, in the SAZ and the SIZ PV is much larger due to the interaction of UCDW with the atmosphere; second, PV shows hotspots of larger variability in the Downstream region and around 150°E in the PFZ, ASZ and SIZ; and third, in the STZ, SAZ and PFZ PV has a positive trend toward the east. The various hotspots of larger PV variability and the *PV* trend are indicative of the presence of diabatic processes [*Whalen et al.*, 2015] or air-sea interactions changing the UCDW density class stratification. Diabatic mixing of UCDW with the overlaying waters, which have greater *PV* (not shown), is facilitated by enhanced turbulence at topographic features such as KP.

We find several hotspots of larger property variability (computed as $\frac{var(C)_{2^\circ}}{var(C)_{tot}}$), associated with the different zones (Fig. 16). The BGC-Argo temperature and salinity variability compares generally well with core Argo, indicating that the spatial and temporal distribution of BGC-Argo floats in this region captures the overall variability in the time and space covered by core Argo. Nevertheless, we find some differences between the core and BGC-Argo temperature and salinity variability in the some locations, likely due to the poor coverage in time and space of BGC-Argo (where core Argo variance is larger than BGC-Argo) or to some local variability close to the surface waters (e.g. in the SIZ where BGC-Argo variance is larger than the core Argo variability): 1) in the STZ west of 20°E; 2) in the SAZ Upstream region; 3) between ~20–80°E in the PFZ; 4) east of 60°E in the ASZ; and almost everywhere in the SIZ.

In the STZ, the variability of BGC-Argo properties is larger in the West region. The variability in the SAZ shows hotspots in the West region in both salinity, oxygen, nitrate and pH, in the Downstream region around ~70°E (salinity, temperature, oxygen and nitrate) and around 100–110°E. We find hotspots of variability in the PFZ between ~55–100°E (Upstream to Downstream), ~110°E (Downstream, only for temperature), and around 140–160°E (East, for temperature and pH). There are several hotspots of larger variance for the ASZ profiles, which vary across the properties: the BGC-Argo variance is larger around 20°E (West region) for temperature and oxygen, around 60°E (Upstream) for nitrate and pH and around 80°E (Downstream) for each of the properties. Finally, the SIZ shows larger variability west of 10°E (West) for temperature, oxygen, nitrate and pH, and a hotspot in variance at approximately 140°E (Upstream) for salinity, temperature and oxygen.

The structure of the variability is complex and hotspots of temperature and salinity normalized variance do not necessarily correspond to locations of highest variability in BGC properties. For example, this can be due to: 1) BGC processes (e.g., carbonate production/dissolution), which modify the concentration of oxygen and nitrate and the pH of the waters; 2) the differing gradients of each property between fronts; 3) and the reduced spatial coverage of BGC data in some locations.

6 Discussion and Conclusions

The Profile Classification Modelling (PCM) approach [*Maze et al.*, 2017a] based on an unsupervised classification algorithm [here a Gaussian Mixture Model; *Bilmes et al.*, 1998; *Bishop*, 2006] is applied to classify core Argo and BGC-Argo profiles into water mass regimes in the Indian sector of the Southern Ocean. To define SO frontal zones, the PCM method can be skillfully applied to vertical profiles of temperature and salinity between 300–900 m, below the portion of the water column most sensitive to air-sea exchange for most of the Argo profiles. We build upon recent studies that used temperature alone for the PCM [e.g., *Maze et al.*, 2017b; *Jones et al.*, 2018] and we included salinity in addition to temperature, measured by autonomous profiling floats, as this property is especially important in setting the stratification of the upper ocean in the subtropical and sea ice zone. The PCM identifies boundaries between the frontal zones, which are not continuous lines, but rather regions of sharp changes, gradients as in *Chapman* [2017]. In addition to automatically classify each hydrographic profile into a unique frontal regime, this method allows us to use posterior probability as a metric to highlight/identify possible regions of strong mixing and temporal variability, particularly in regions of strong currents (Fig. 3 and 4). A region of larger mixing can be identified by lower probability (< 70%, Fig. 5). Mixing, such defined, is larger 1) at Southern Ocean fronts, 2) in the SAZ, and 3) in the major current systems of the STZ (compare Fig. 5 with Fig. 8). Due to flow–topography interaction, these areas are hotspots of eddy kinetic energy (see Fig. 5 in *Llort et al.* [2018] and Fig. 3 in *Roach et al.* [2018]), which can sustain intense events of vertical property exchange [*Llort et al.*, 2018; *Rosso et al.*, 2014], both within the AAIW density class (between the STF and PF) and within the UCDW (south of PF).

In this study, we find that the variability of the intermediate and deep waters is enhanced at topographic features (e.g., Crozet Plateau and Kerguelen Plateau) and in strong currents (e.g., at the subtropical Agulhas rings and Agulhas Return Current), but that the degree of variability differs for individual properties. AAIW temperature and salinity in the West region of the STZ are lowest west of 40°E, where colder, fresher and oxygen–rich waters from the AAIW source in the Atlantic sector (west) and from the SAZ mix with the warmer, more saline and oxygen–poorer Indian subtropical waters (east; Fig. 12) [e.g., *Talley*, 1996]. In the SAZ, AAIW mixes with the colder and fresher surface waters from the PFZ at KP, as evidenced by its minimum θ of 2°C and SA of 34.25 in the Downstream region (Fig. 12). This mechanism is likely associated with cross–frontal intrusions, as suggested by *Park et al.* [1993]; *Park and Gamberoni* [1997]; *Sloyan and Rintoul* [2001], where the convergence of fronts and highly energetic eddies can facilitate this injection of different waters and the subsequent modification of water masses [*Park et al.*, 1993]. Upon traversing KP, the averaged SAZ AAIW temperature and salinity increase again, while the oxygen concentration decreases, suggesting mixing with subtropical waters. Below the AAIW, the variation in the properties of the UCDW is not only marked by a strong meridional gradient across the different regimes, but also by a large transition in the properties from west to east, where KP acts to homogenize the water mass (Fig. 14). Hotspots of larger *PV* variability (Fig. 15) suggest that the large change in each property is due not only to isopycnal, but also diabatic transformations, as expected from the stronger vertical mixing that *Whalen et al.* [2012] and *Whalen et al.* [2015] show in these locations.

Argo has enabled us to study temperature, salinity and, in some cases, oxygen properties across the vast Southern Ocean. Here, we demonstrate how the complementary array of BGC-Argo floats enables the assessment of the spatial variability of BGC properties. As physical and BGC states are influenced by diverse dynamics and gradients, we cannot fully infer where BGC properties might show larger variability by looking at only temperature and salinity variability. Given the rapidly increasing amount of Argo data (both core and BGC) and model output, the PCM method used here can serve as an important tool in future studies aiming to identify similarities in dynamical regimes [e.g., *Maze et al.*, 2017b; *Ardyna et al.*, 2017; *Jones et al.*, 2018; *Liang et al.*, 2018] and to reveal regions of strong mixing. Furthermore, BGC-Argo floats, strategically planned to target this KP region [*Talley et al.*, 2019], can provide great insight regarding the distribution and the modification of water mass properties, highlighting the importance of targeting mixing hotspots in future observing arrays. Future work should explore these statistical methods to assess property changes and water mass evolution (e.g., in UCDW upwelling and its contribution to air–sea fluxes of carbon, oxygen, and heat) over the entire Southern Ocean.

Finally, we have only qualitatively compared our classification to fronts defined by *Orsi et al.* [1995]. We note that a future work should focus on a quantitative comparison of our results with the existing fronts definition based on hydrography [e.g., *Orsi et al.*, 1995], gradient of sea surface height [*Sokolov and Rintoul*, 2009b] or the new method by *Chapman* [2017] applied to absolute dynamic topography. Furthermore, the scientific community would benefit for a thorough comparison of our methodology with other existing algorithms, especially those methods that does not rely on a fixed number of clusters, such as the variational Bayesian GMM [*Ghahramani and Beal*, 2000].

Acknowledgments

SOCCOM data were collected and made freely available by the Southern Ocean Carbon and Climate Observations and Modeling Project funded by the National Science Foundation, Division of Polar Programs (NSF PLR -1425989), supplemented by NASA, and by the International Argo Program and the NOAA programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it. The Argo Program is part of the Global Ocean Observing System. IR thanks L. Böhme (University of St Andrews), J. Bowman (Scripps Institution of Oceanography), Y. Liang (Woods Hole Oceanographic Institution) and M. Kuusela (Carnegie Mellon University) for all the valuable discussions on fronts and classification methods. This work is a contribution of the Southern Ocean Carbon and Climate Observations and Modeling project (SOCCOM). SOCCOM is supported by the National Science Foundation under NSF Award PLR-1425989. GM was supported by the French national programme LEFE/INSU, project SOMOVAR. We thank the anonymous reviewers who helped to improve this manuscript.

References

Amante, C., and B. W. Eakins (2009), *ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis*, US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center, Marine Geology and Geophysics Division Colorado.

Aoki, S., M. Hariyama, H. Mitsudera, H. Sasaki, and Y. Sasai (2007), Formation regions of Subantarctic Mode Water detected by OFES and Argo profiling floats, *Geophysical Research Letters*, *34*(10).

Aoki, S., N. Fujii, S. Ushio, Y. Yoshikawa, S. Watanabe, G. Mizuta, Y. Fukamachi, and M. Wakatsuchi (2008), Deep western boundary current and southern frontal systems of the Antarctic Circumpolar Current southeast of the Kerguelen Plateau, *Journal of Geophysical Research: Oceans*, *113*(C8).

Ardyna, M., H. Claustre, J.-B. Sallée, F. D'Ovidio, B. Gentili, G. Van Dijken, F. D'Ortenzio, and K. R. Arrigo (2017), Delineating environmental control of phytoplankton biomass and phenology in the Southern Ocean, *Geophysical Research Letters*, *44*(10), 5016–5024.

Argo Data Management Team (2017), Argo user's manual v3.2, doi:10.13155/29825.

Argo Data Management Team (2018), Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) - Snapshot of Argo GDAC of September 8st 2018, doi: 10.17882/42182.

Bilmes, J. A., et al. (1998), A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, *International Computer Science Institute*, *4*(510), 126.

Bishop, C. M. (2006), *Pattern recognition and machine learning*, springer.

Briggs, E. M., T. R. Martz, L. D. Talley, M. R. Mazloff, and K. S. Johnson (2018), Physical and biological drivers of biogeochemical tracers within the seasonal sea ice zone of the southern ocean from profiling floats, *Journal of Geophysical Research: Oceans*, *123*(2), 746–758.

Cerovečki, I., and M. R. Mazloff (2016), The spatiotemporal structure of diabatic processes governing the evolution of subantarctic mode water in the southern ocean, *Journal of Physical Oceanography*, *46*(2), 683–710.

Chapman, C. C. (2017), New perspectives on frontal variability in the Southern Ocean, *Journal of Physical Oceanography*, *47*(5), 1151–1168.

Donohue, K. A., G. E. Hufford, and M. S. McCartney (1999), Sources and transport of the deep western boundary current east of the kerguelen plateau, *Geophysical Research Letters*, *26*(7), 851–854.

Freeman, N. M., and N. S. Lovenduski (2016), Mapping the Antarctic Polar Front: weekly realizations from 2002 to 2014, *Earth System Science Data*, *8*(1), 191–198.

Freeman, N. M., N. S. Lovenduski, and P. R. Gent (2016), Temporal variability in the Antarctic Polar Front (2002–2014), *J. Geophys. Res.: Oceans*, *121*(10), 7263–7276.

Freeman, N. M., D. R. Munro, J. Sprintall, M. R. Mazloff, S. Purkey, I. Rosso, C. A. DeRanek, and C. Sweeney (2019), The observed seasonal cycle of macronutrients in Drake Passage: relationship to fronts and utility as a model metric, *Journal of Geophysical Research: Oceans*, *124*, 4763–4783, doi:10.1029/2019JC015052.

Fukamachi, Y., S. Rintoul, J. Church, S. Aoki, S. Sokolov, M. Rosenberg, and M. Wakatsuchi (2010), Strong export of antarctic bottom water east of the kerguelen plateau, *Nature*

Geosc., *3*(5), 327.

Ghahramani, Z., and M. J. Beal (2000), Variational inference for Bayesian mixtures of factor analysers, in *Advances in neural information processing systems*, pp. 449–455.

Gray, A. R., K. S. Johnson, S. M. Bushinsky, S. C. Riser, J. L. Russell, L. D. Talley, R. Wanninkhof, N. L. Williams, and J. L. Sarmiento (2018), Autonomous Biogeochemical Floats Detect Significant Carbon Dioxide Outgassing in the High-Latitude Southern Ocean, *Geophysical Research Letters*, *45*(17), 9049–9057.

Gruber, N., M. Gloor, S. E. Mikaloff Fletcher, S. C. Doney, S. Dutkiewicz, M. J. Follows, M. Gerber, A. R. Jacobson, F. Joos, K. Lindsay, et al. (2009), Oceanic sources, sinks, and transport of atmospheric CO_2, *Global Biogeochemical Cycles*, *23*(1).

Hanawa, K., and L. D. Talley (2001), Mode waters. Ocean Circulation and Climate, G. Siedler et al., Eds, *International Geophysics Series*, pp. 373–386.

Hartigan, J. A., and M. A. Wong (1979), Algorithm AS 136: A k-means clustering algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108.

IOC, SCOR, IAPSO (2010), The international thermodynamic equation of seawater–2010: calculation and use of thermodynamic properties.[includes corrections up to 31st october 2015].

Ito, T., M. Woloszyn, and M. Mazloff (2010), Anthropogenic carbon dioxide transport in the Southern Ocean driven by Ekman flow, *Nature*, *463*(7277), 80–83.

Johnson, K. S., J. N. Plant, L. J. Coletti, H. W. Jannasch, C. M. Sakamoto, S. C. Riser, D. D. Swift, N. L. Williams, E. Boss, N. Haëntjens, et al. (2017), Biogeochemical sensor performance in the SOCCOM profiling float array, *Journal of Geophysical Research: Oceans*, *122*(8), 6416–6436.

Johnson, K. S., S. C. Riser, E. S. Boss, L. D. Talley, J. L. Sarmiento, D. D. Swift, J. N. Plant, T. L. Maurer, R. M. Key, N. L. Williams, R. H. Wanninkhof, A. G. Dickson, A. Feely Richard, and J. L. Russell (2018), SOCCOM float data - Snapshot 2018-09-12. In Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Float Data Archive, doi:doi.org/10.6075/J0QJ7FJP.

Jones, D. C., H. J. Holt, A. J. Meijers, and E. F. Shuckburgh (2018), Unsupervised clustering of Southern Ocean Argo float temperature profiles, *Journal of Geophysical Research: Oceans*.

Konishi, S., and G. Kitagawa (2008), *Bayesian Information Criteria*, pp. 211–237, Springer New York, New York, NY.

Liang, Y.-C., M. R. Mazloff, I. Rosso, S.-W. Fang, and J.-Y. Yu (2018), A Multivariate Empirical Orthogonal Function Method to Construct Nitrate Maps in the Southern Ocean, *Journal of Atmospheric and Oceanic Technology*, *35*(7), 1505–1519.

Llort, J., C. Langlais, R. Matear, S. Moreau, A. Lenton, and P. G. Strutton (2018), Evaluating Southern Ocean Carbon Eddy-Pump From Biogeochemical-Argo Floats, *Journal of Geophysical Research: Oceans*, *123*(2), 971–984.

Maraldi, C., M. Mongin, R. Coleman, and L. Testut (2009), The influence of lateral mixing on a phytoplankton bloom: Distribution in the Kerguelen Plateau region, *Deep Sea Res. Part I*, *56*, 963–973.

Maze, G., H. Mercier, and C. Cabanes (2017a), Profile Classification Models, *Mercator Ocean Journal*, *55*, 48–56.

Maze, G., H. Mercier, R. Fablet, P. Tandeo, M. L. Radcenco, P. Lenca, C. Feucher, and C. Le Goff (2017b), Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean, *Progress in Oceanography*, *151*, 275–292.

McCarthy, M. C., and L. D. Talley (1999), Three-dimensional isoneutral potential vorticity structure in the indian ocean, *Journal of Geophysical Research: Oceans*, *104*(C6), 13,251–13,267.

McCartney, M. (1977), Subantarctic Mode Water A Voyage of Discovery, *George Deacon 70th Anniversary Volume*, pp. 103–119.

McCartney, M. S., and K. A. Donohue (2007), A deep cyclonic gyre in the Australian-Antarctic Basin, *Progr. Oceanogr.*, *75*(4), 675–750, doi:10.1016/j.pocean.2007.02.008.

Meyer, A., B. M. Sloyan, K. L. Polzin, H. E. Phillips, and N. L. Bindoff (2015), Mixing variability in the southern ocean, *Journal of Physical Oceanography*, *45*(4), 966–987.

Moore, J. K., M. R. Abbott, and J. G. Richman (1997), Variability in the location of the antarctic polar front (90–20°w) from satellite sea surface temperature data, *Journal of Geophysical Research: Oceans*, *102*(C13), 27,825–27,833.

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. Part I: Oceanographic Research Papers*, *42*(5), 641–673.

Park, Y.-H., and L. Gamberoni (1997), Cross-frontal exchange of antarctic intermediate water and antarctic bottom water in the crozet basin, *Deep Sea Research Part II: Topical Studies in Oceanography*, *44*(5), 963–986.

Park, Y.-H., L. Gamberoni, and E. Charriaud (1993), Frontal structure, water masses, and circulation in the crozet basin, *Journal of Geophysical Research: Oceans*, *98*(C7), 12,361–12,385.

Park, Y.-H., F. Roquet, I. Durand, and J.-L. Fuda (2008a), Large-scale circulation over and around the Northern Kerguelen Plateau, *Deep Sea Res. Part II*, *55*, 566–581.

Park, Y.-H., J.-L. Fuda, I. Durand, and A. C. Naveira Garabato (2008b), Internal tides and vertical mixing over the Kerguelen Plateau, *Deep Sea Res. Part II*, *55*(5-7), 582–593.

Park, Y.-H., F. Vivier, F. Roquet, and E. Kestenare (2009), Direct observations of the acc transport across the kerguelen plateau, *Geophysical Research Letters*, *36*(18).

Park, Y.-H., I. Durand, E. Kestenare, G. Rougier, M. Zhou, F. d'Ovidio, C. Cotté, and J.-H. Lee (2014), Polar Front around the Kerguelen Islands: An up-to-date determination and associated circulation of surface/subsurface waters, *J. Geophys. Res.*, *119*(10), 6575–6592.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. (2011), Scikit-learn: Machine learning in Python, *Journal of machine learning research*, *12*(Oct), 2825–2830.

Piola, A. R., and D. T. Georgi (1982), Circumpolar properties of Antarctic intermediate water and Subantarctic Mode Water, *Deep Sea Research Part A. Oceanographic Research Papers*, *29*(6), 687–711.

Piola, A. R., and A. L. Gordon (1989), Intermediate waters in the southwest South Atlantic, *Deep Sea Research Part A. Oceanographic Research Papers*, *36*(1), 1–16.

Prézelin, B. B., E. E. Hofmann, C. Mengelt, and J. M. Klinck (2000), The linkage between upper circumpolar deep water (ucdw) and phytoplankton assemblages on the west antarctic peninsula continental shelf, *Journal of Marine Research*, *58*(2), 165–202.

Read, J., and R. Pollard (1993), Structure and transport of the antarctic circumpolar current and agulhas return current at 40°e, *Journal of Geophysical Research: Oceans*, *98*(C7), 12,281–12,295.

Riser, S. C., D. Swift, and R. Drucker (2018), Profiling Floats in SOCCOM: Technical Capabilities for Studying the Southern Ocean, *Journal of Geophysical Research: Oceans*.

Roach, C. J., D. Balwada, and K. Speer (2018), Global Observations of Horizontal Mixing from Argo Float and Surface Drifter Trajectories, *Journal of Geophysical Research:*

Oceans, *123*(7), 4560–4575, doi:doi.org/10.1029/2018JC013750.

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. Sys.*, *78*(3), 377–393.

Rosso, I., M. A. Hogg, G. P. Strutton, E. A. Kiss, R. Matear, A. Klocker, and E. van Sebille (2014), Vertical transport in the ocean due to sub-mesoscale structures: Impacts in the Kerguelen region, *Ocean Modell.*, *80*, 10–23.

Rosso, I., A. M. Hogg, R. Matear, and P. G. Strutton (2016), Quantifying the influence of sub-mesoscale dynamics on the supply of iron to Southern Ocean phytoplankton blooms, *Deep Sea Research Part I: Oceanographic Research Papers*, *115*, 199–209.

Rosso, I., M. R. Mazloff, A. Verdy, and L. D. Talley (2017), Space and time variability of the s outhern o cean carbon budget, *Journal of Geophysical Research: Oceans*, *122*(9), 7407–7432.

Sabine, C. L., R. A. Feely, G. C. Johnson, P. G. Strutton, M. F. Lamb, and K. E. McTaggart (2004), A mixed layer carbon budget for the GasEx-2001 experiment, *Journal of Geophysical Research: Oceans (1978–2012)*, *109*(C8).

Sallée, J.-B., N. Wienders, K. Speer, and R. Morrow (2006), Formation of subantarctic mode water in the southeastern Indian Ocean, *Ocean Dynamics*, *56*(5-6), 525–542.

Sallée, J.-B., K. Speer, S. Rintoul, and S. Wijffels (2010), Southern Ocean thermocline ventilation, *Journal of Physical Oceanography*, *40*(3), 509–529.

Sallee, J.-B., K. Speer, and S. Rintoul (2011), Mean-flow and topographic control on surface eddy-mixing in the Southern Ocean, *Journal of Marine Research*, *69*(4-5), 753–777, doi:doi.org/10.1357/002224011799849408.

Sarmiento, J. á., N. Gruber, M. Brzezinski, and J. Dunne (2004), High-latitude controls of thermocline nutrients and low latitude biological productivity, *Nature*, *427*(6969), 56.

Schwarz, G., et al. (1978), Estimating the dimension of a model, *The annals of statistics*, *6*(2), 461–464.

Sloyan, B. M., and S. R. Rintoul (2001), Circulation, renewal, and modification of Antarctic Mode and Intermediate Water, *Journal of Physical Oceanography*, *31*(4), 1005–1030.

Sloyan, B. M., L. D. Talley, T. K. Chereskin, R. Fine, and J. Holte (2010), Antarctic Intermediate Water and Subantarctic Mode Water formation in the southeast Pacific: The role of turbulent mixing, *Journal of Physical Oceanography*, *40*(7), 1558–1574.

Sokolov, S., and S. R. Rintoul (2009a), Circumpolar structure and distribution of the Antarctic Circumpolar Current fronts: 2. Variability and relationship to sea surface height, *J. Geophys. Res.*, *114*(C11), C11,019.

Sokolov, S., and S. R. Rintoul (2009b), Circumpolar structure and distribution of the Antarctic Circumpolar Current fronts: 1. Mean circumpolar paths, *J. Geophys. Res.*, *114*(C11), C11,018.

Sparrow, M. D., K. J. Heywood, J. Brown, and D. P. Stevens (1996), Current structure of the south indian ocean, *Journal of Geophysical Research: Oceans*, *101*(C3), 6377–6391.

Takahashi, T., S. C. Sutherland, R. Wanninkhof, C. Sweeney, R. A. Feely, D. W. Chipman, B. Hales, G. Friederich, F. Chavez, C. Sabine, et al. (2009), Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans, *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(8), 554–577.

Talley, L. (1996), Antarctic intermediate water in the South Atlantic, in *The South Atlantic*, pp. 219–238, Springer.

Talley, L., I. Rosso, I. Kamenkovich, M. Mazloff, J. Wang, E. Boss, A. Gray, K. Johnson, R. Key, S. Riser, et al. (2019), Southern Ocean biogeochemical float deployment strategy, with example from the Greenwich Meridian line (GO-SHIP A12), *Journal of Geophysical Research: Oceans*.

Talley, L. D. (2013), Closure of the global overturning circulation through the Indian, Pacific, and Southern Oceans: Schematics and transports, *Oceanogr.*, *26*(1), 80–97.

Talley, L. D., G. L. Pickard, W. J. Emery, and J. H. Swift (2011), *Descriptive Physical Oceanography: An Introduction*, Academic press.

Tamsitt, V., L. D. Talley, M. R. Mazloff, and I. Cerovečki (2016), Zonal variations in the Southern Ocean heat budget, *Journal of Climate*, *29*(18), 6563–6579.

Tamsitt, V., H. F. Drake, A. K. Morrison, L. D. Talley, C. O. Dufour, A. R. Gray, S. M. Griffies, M. R. Mazloff, J. L. Sarmiento, J. Wang, et al. (2017), Spiraling pathways of global deep waters to the surface of the southern ocean, *Nature communications*, *8*(1), 172.

Van Beek, P., M. Bourquin, J.-L. Reyss, M. Souhaut, M. Charette, and C. Jeandel (2008), Radium isotopes to investigate the water mass pathways on the Kerguelen Plateau (Southern Ocean), *Deep Sea Res. Part II*, *55*(5), 622–637.

Whalen, C., L. Talley, and J. MacKinnon (2012), Spatial and temporal variability of global ocean mixing inferred from Argo profiles, *Geophysical Research Letters*, *39*(18).

Whalen, C. B., J. A. MacKinnon, L. D. Talley, and A. F. Waterhouse (2015), Estimating the mean diapycnal mixing using a finescale strain parameterization, *Journal of Physical Oceanography*, *45*(4), 1174–1188.

Williams, N., L. Juranek, R. Feely, J. Russell, K. Johnson, and B. Hales (2018), Assessment of the carbonate chemistry seasonal cycles in the southern ocean from persistent observational platforms, *Journal of Geophysical Research: Oceans*, *123*(7), 4833–4852.

Wong, A., R. Keeley, T. Carval, and the Argo Data Management Team (2012), Argo quality control manual Version 2.7.

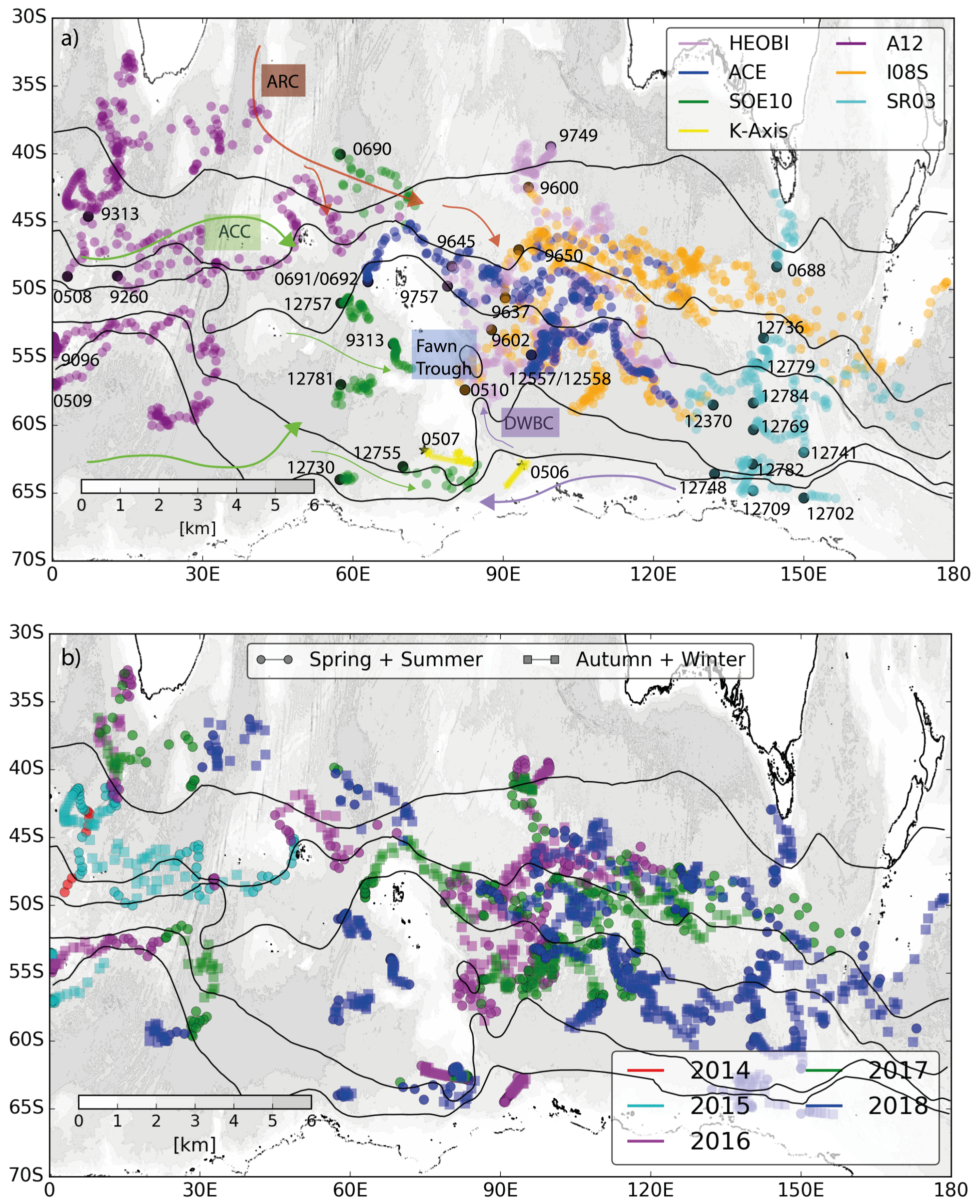

Figure 1. Trajectories of the biogeochemical-Argo floats from the Southern Ocean Carbon and Climate Observations and Modeling project in the Indian sector of the Southern Ocean (September 2018 snapshot), colored by (a) their deployment cruise (Table S1) and (b) profile year [markers indicate sample season: warmer austral spring and summer months (circles) and colder austral autumn and winter months (squares)]. The *Orsi et al.* [1995] fronts (black contours), from south to north, the Southern Boundary of the ACC, Southern Antarctic Circumpolar Current Front, Polar Front, Subantarctic Front, and Subtropical Front), are overlain on bathymetry [grayscale map; ETOPO1; *Amante and Eakins*, 2009]. Trajectories of southernmost floats sampling under sea ice are estimated using linear interpolation (appear as near-straight pathways; e.g., float #12702). Major currents in (a) are labeled, with flow direction indicated by arrows: Agulhas Return Current (ARC), Antarctic Circumpolar Current (ACC), and Deep Western Boundary Current (DWBC). The Fawn Trough is also indicated.

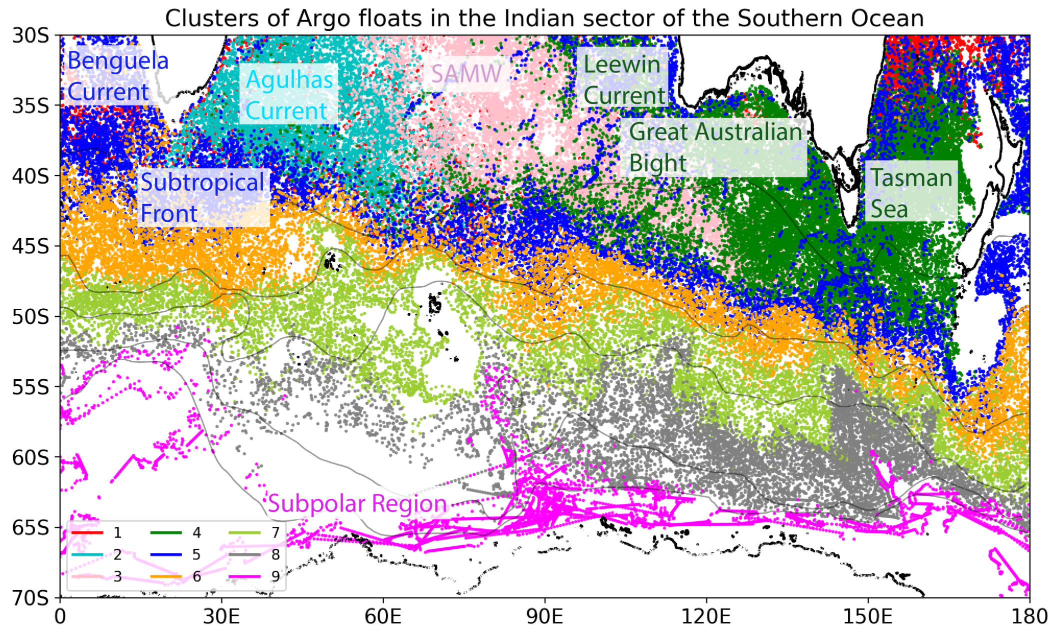

Figure 2. Spatial distribution of the 9 clusters (colors) overlaid on bathymetry [grayscale map; ETOPO1; *Amante and Eakins*, 2009] in the Indian sector of the Southern Ocean. Clusters were identified by the Profile Classification Model method using the full Argo data set. The five *Orsi et al.* [1995] fronts are indicated by black contours, as in Figure 1.

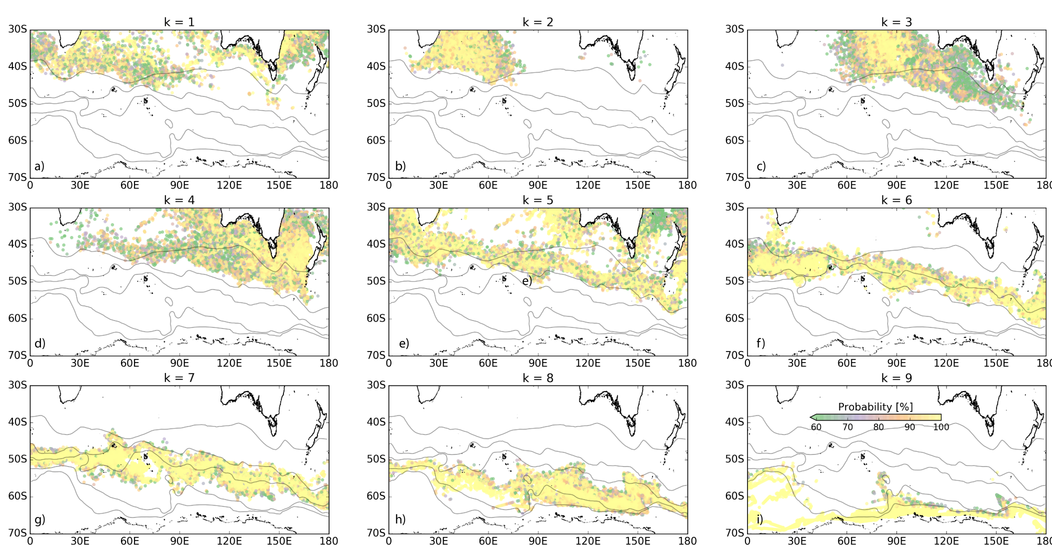

Figure 3. Posterior probability (%; colors) associated with each of the 9 clusters (k) shown in Fig. 2. A colorbar is shown in panel (i). The *Orsi et al.* [1995] fronts (black contours) are shown for reference.

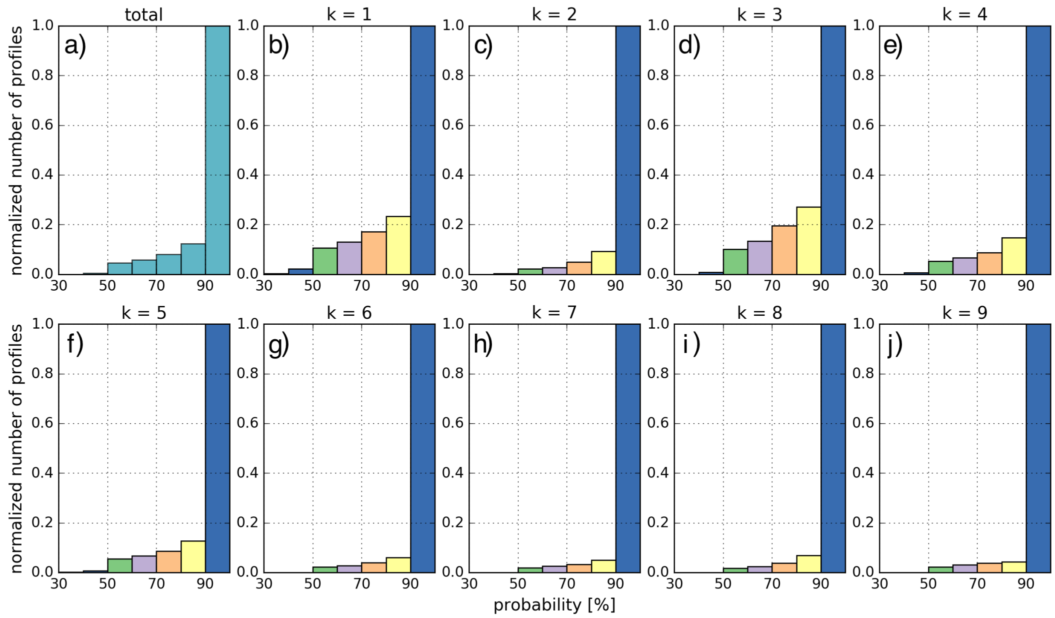

Figure 4. Histograms of the percent posterior probability associated with (a) the full Argo data set and (b–j) each of the 9 clusters (k) presented in Figure 2. The number of profiles for each 10% bin is normalized by the total number of profiles in the 90–100% bin, and are color coded in panels (b–j).

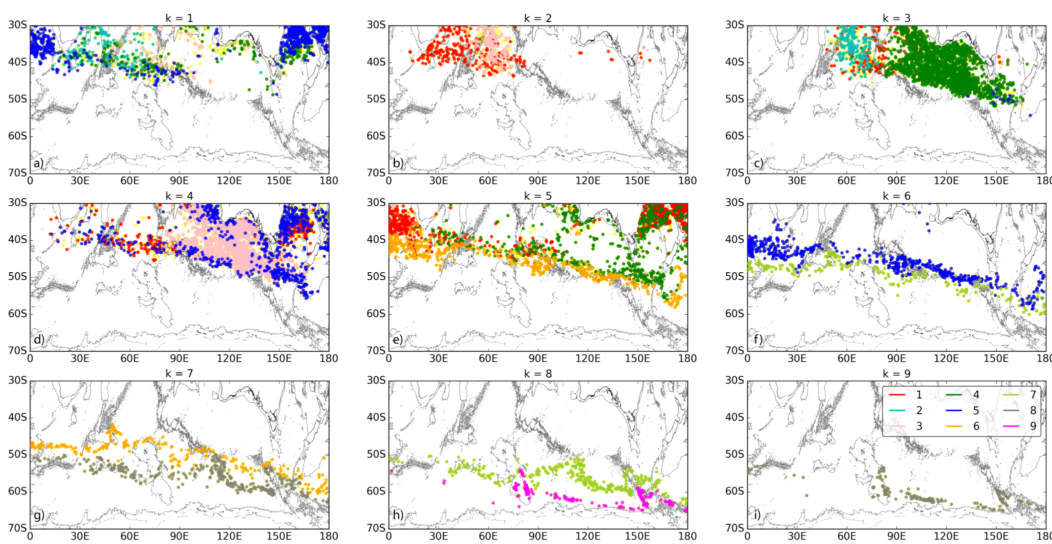

Figure 5. Profiles with a posterior probability less than 70%. Those profile that have a probability of more than 30% to belong to the remaining clusters are color coded as in legend in panel (i). In yellow are those profile that do not have a probability of ≥30% for the remaining clusters. Black contours are coastlines and the 3000 m isobath (bathymetry from from ETOPO1 [*Amante and Eakins*, 2009]).

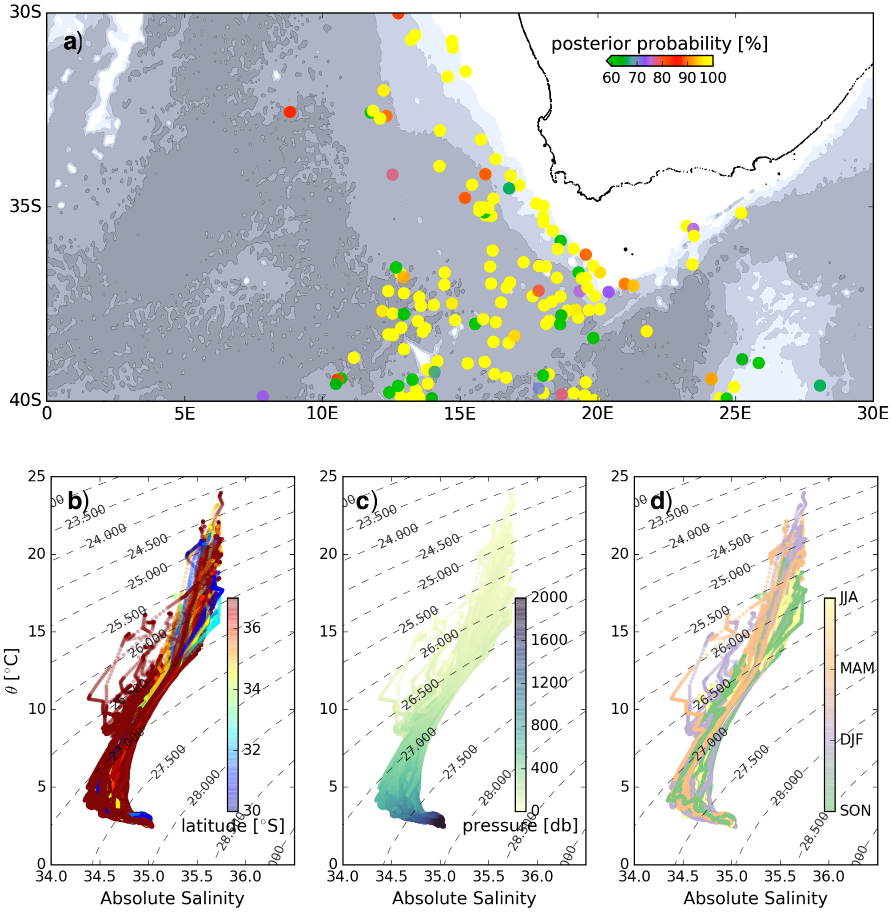

Figure 6. (a) Percent posterior probability (colors) and (b–d) θ and SA properties of subtropical profiles classified within the Subantarctic Zone (cluster $k = 6$) off the coast of Africa. θ and SA diagrams are colored by latitude in (b), pressure in (c), and season in (d).

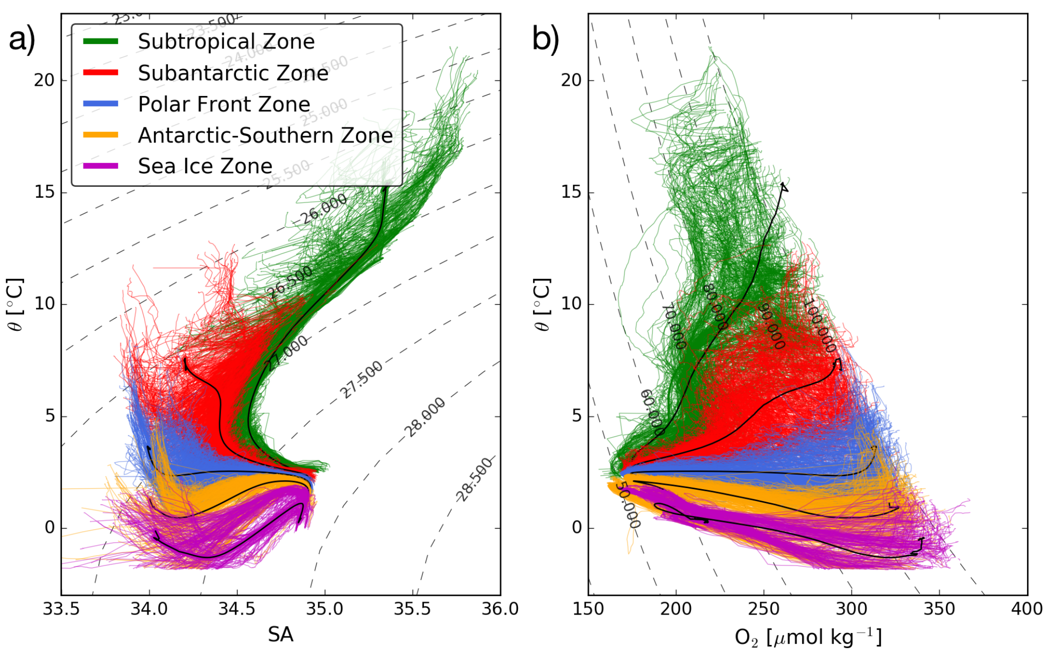

Figure 7. BGC-Argo profiles colored by Southern Ocean regime as a function of (a) $\theta - SA$ and (b) θ–O_2: Subtropical Zone (STZ; green), Subantarctic Zone (SAZ; red), Polar Frontal Zone (PFZ; blue), Antarctic-Southern Zone (ASZ; orange) and Sea Ice Zone (SIZ; magenta). The black contours indicate the mean Argo profile by frontal zone. Dashed lines in (a) show σ_0 contours, while in (b) % of oxygen saturation.

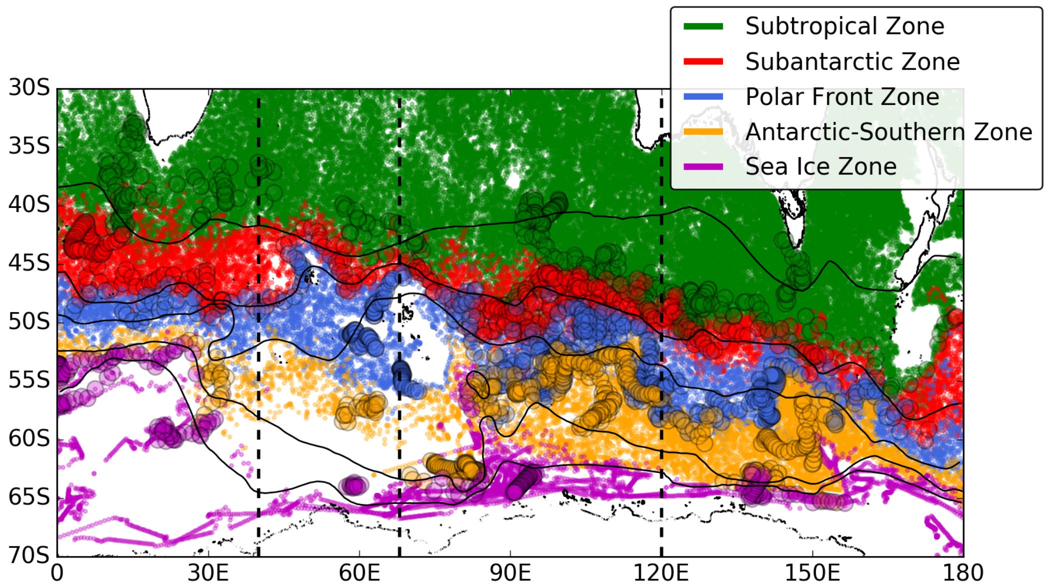

Figure 8. Argo (colored dots) and biogeochemical-Argo (larger, outlined colored circles) float profile locations colored by Southern Ocean frontal zone: Subtropical, Subantarctic, Polar Frontal, Antarctic-Southern and Sea Ice zones. Profiles are classified using the Gaussian Mixed Models algorithm applied to potential temperature and absolute salinity measurements (see Section 4). We define four regions relative to the Kerguelen Plateau by longitude (delineated by vertical dashed black lines): West (0°–40° E), Upstream (40° E–68° E), Downstream (68° E–120° E), and East (120° E–180° E). For reference, gray contours indicate the location of the five *Orsi et al.* [1995] fronts.

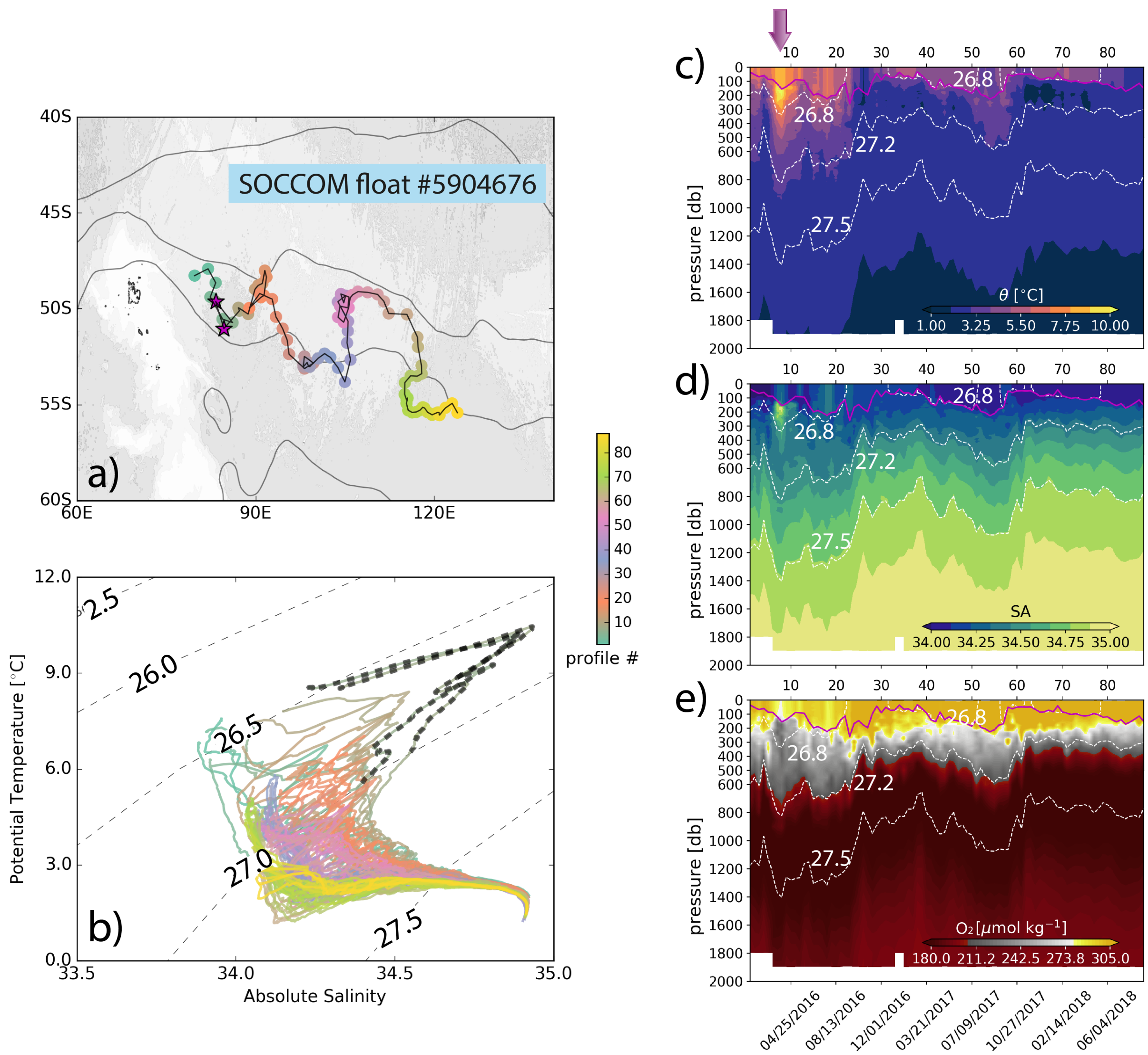

Figure 9. The (a) sampling trajectory and (b) θ and SA properties of biogeochemical-Argo float #5904676, colored by profile number (out of total 88), which collected (c) potential temperature, (d) absolute salinity and (e) dissolved oxygen of the upper 2000 m, between January 2016 and June 2018. Note, two intrusion events (red stars in (a) and dashed black lines in (b) are identified in profile #8 and #9 (magenta arrow in (c–e).) Bathymetry [grayscale map; ETOPO1; *Amante and Eakins*, 2009] and *Orsi et al.* [1995] fronts are indicated in (a) as in Figure 1. Dashed white contours in panels (c–e) are σ_0 isolines, with values indicated in white.

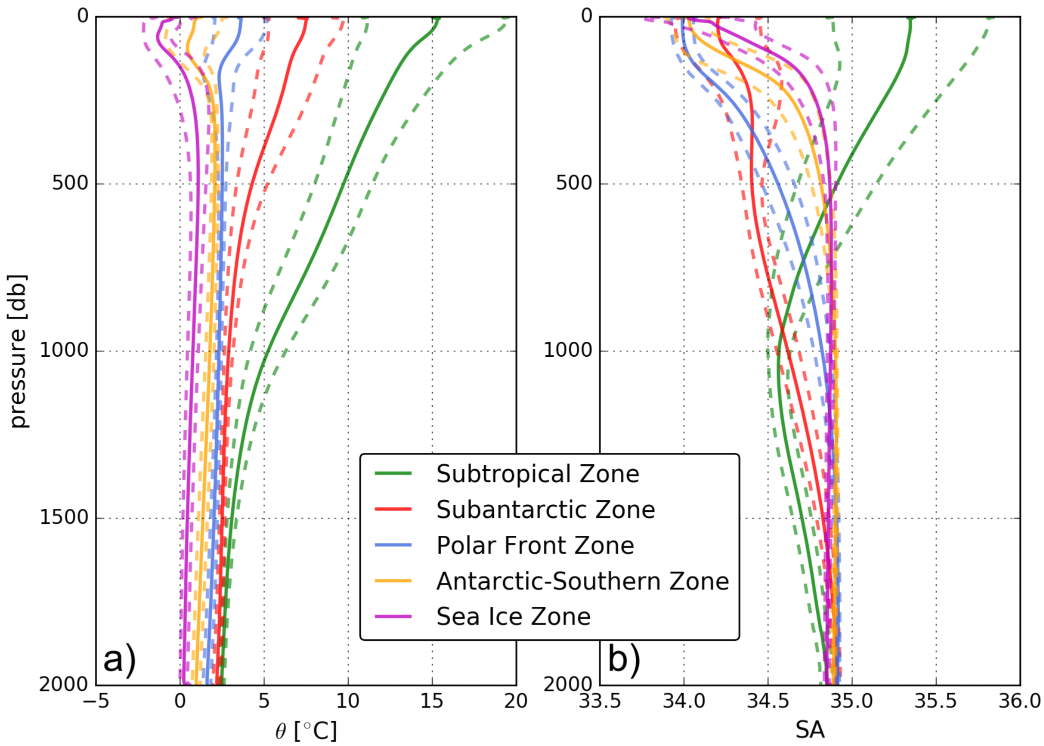

Figure 10. Vertical mean (solid) and one standard deviation (dashed) of (a) potential temperature and (b) absolute salinity of all Argo profiles from the Indian sector of the Southern Ocean used in this study. Metrics are colored by Southern Ocean frontal zone: Subtropical Zone (green), Subantarctic Zone (red), Polar Frontal Zone (blue), Antarctic-Southern Zone (orange) and Sea Ice Zone (magenta).

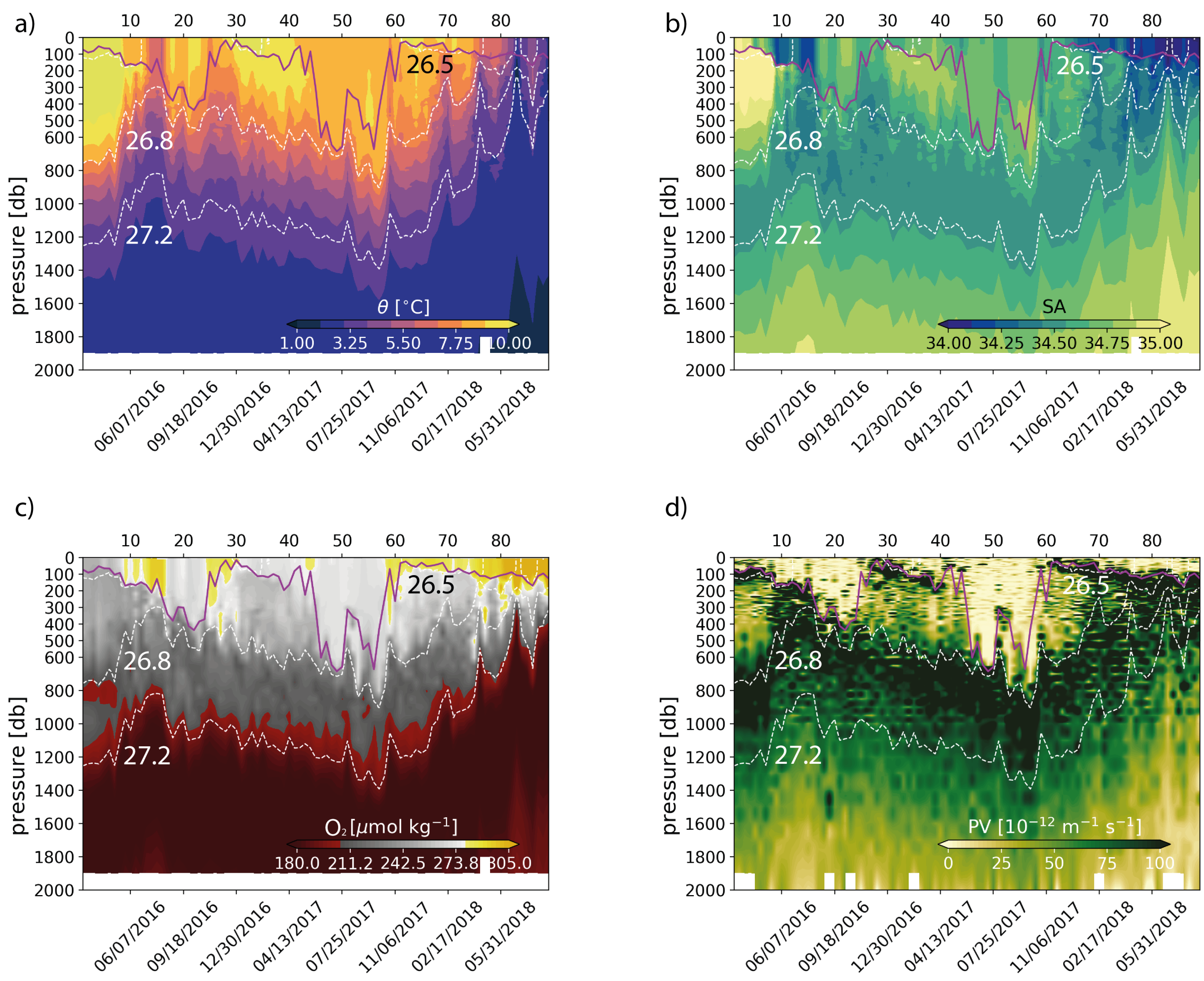

Figure 11. Upper 2000 m (a) θ, (b) SA, (c) dissolved oxygen and (d) potential vorticity of biogeochemical-Argo float #5904688. Dashed white lines are σ_0 isolines, with values indicated in white.

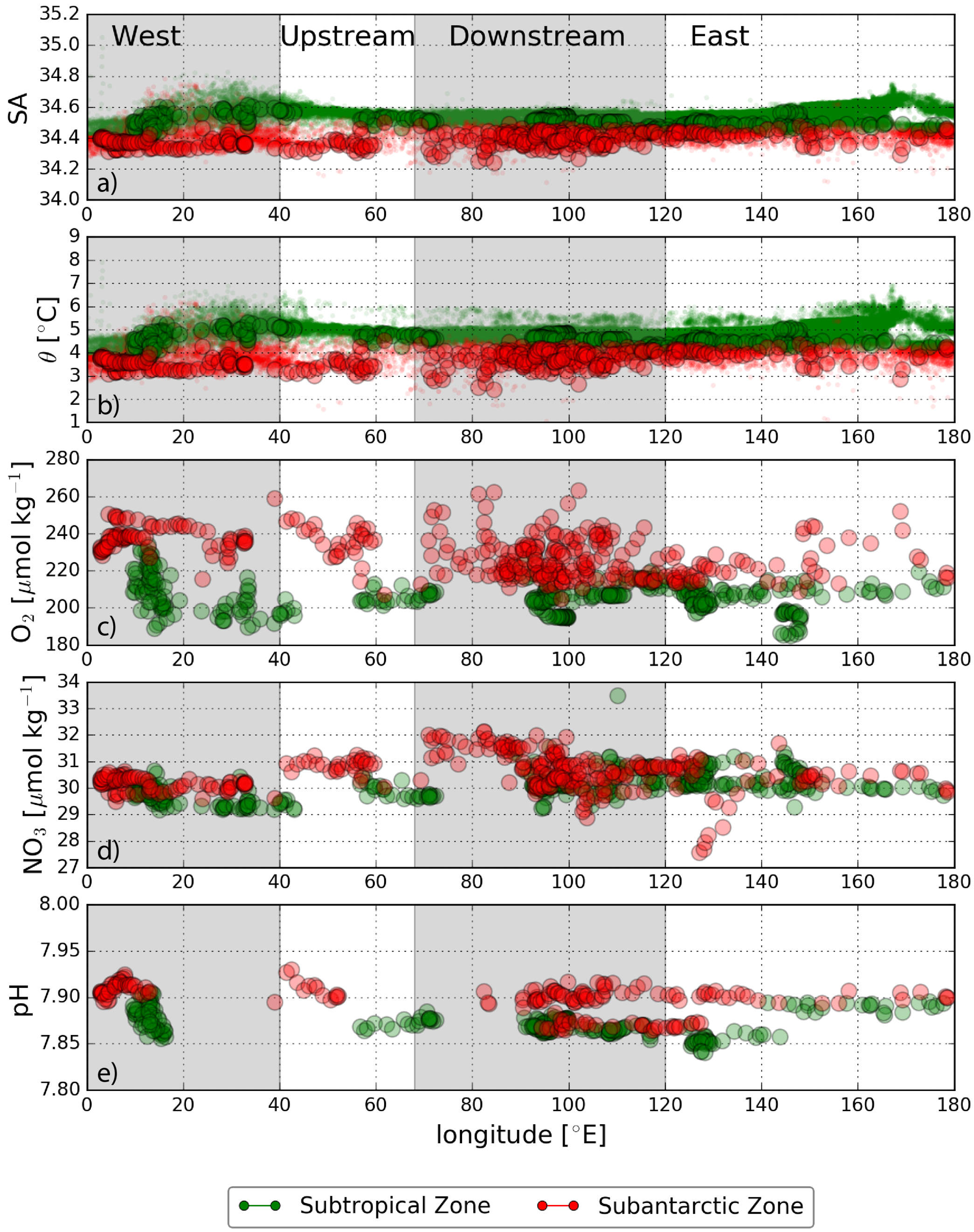

Figure 12. (a) Absolute salinity, (b) potential temperature, (c) dissolved oxygen, (d) nitrate, and (e) dissolved inorganic carbon, as measured by core Argo (colored dots) and biogeochemical-Argo (outlined colored circles) floats across the northernmost frontal zones, the (green) Subtropical Zone and (red) Subantarctic Zone. Properties are averaged over the Antarctic Intermediate Water layer (with density between σ_0 = 27 kg m^{-3} and 27.2 kg m^{-3}).

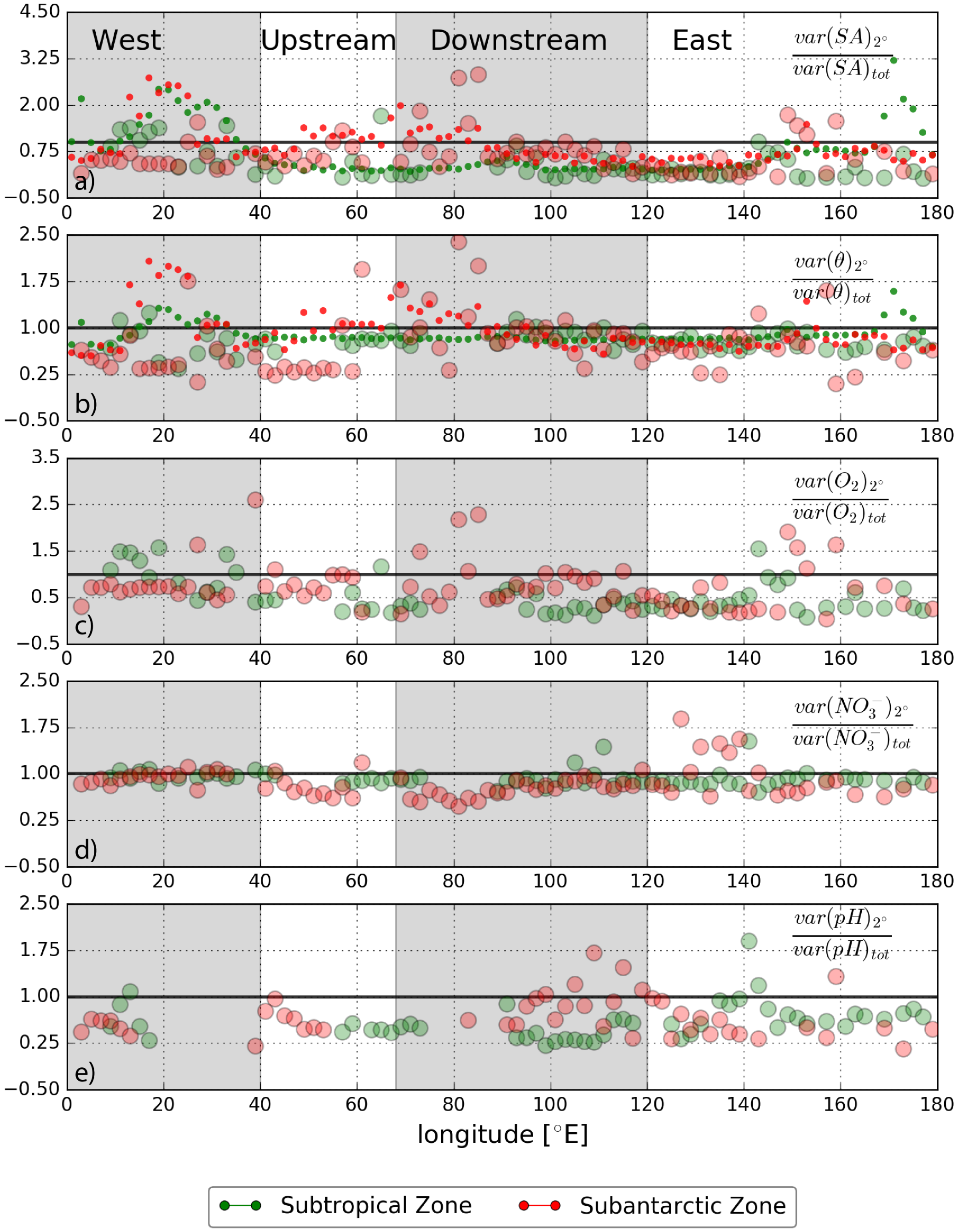

Figure 13. Regional variability of the ratio of binned (2° in longitude) over the total variance of (a) absolute salinity (SA), (b) potential temperature (θ), (c) oxygen (O_2), (d) nitrate (NO_3^-), and (e) pH, along σ_0, computed across the density classes around the Antarctic Intermediate Water. Markers are color coded as in Fig. 12.

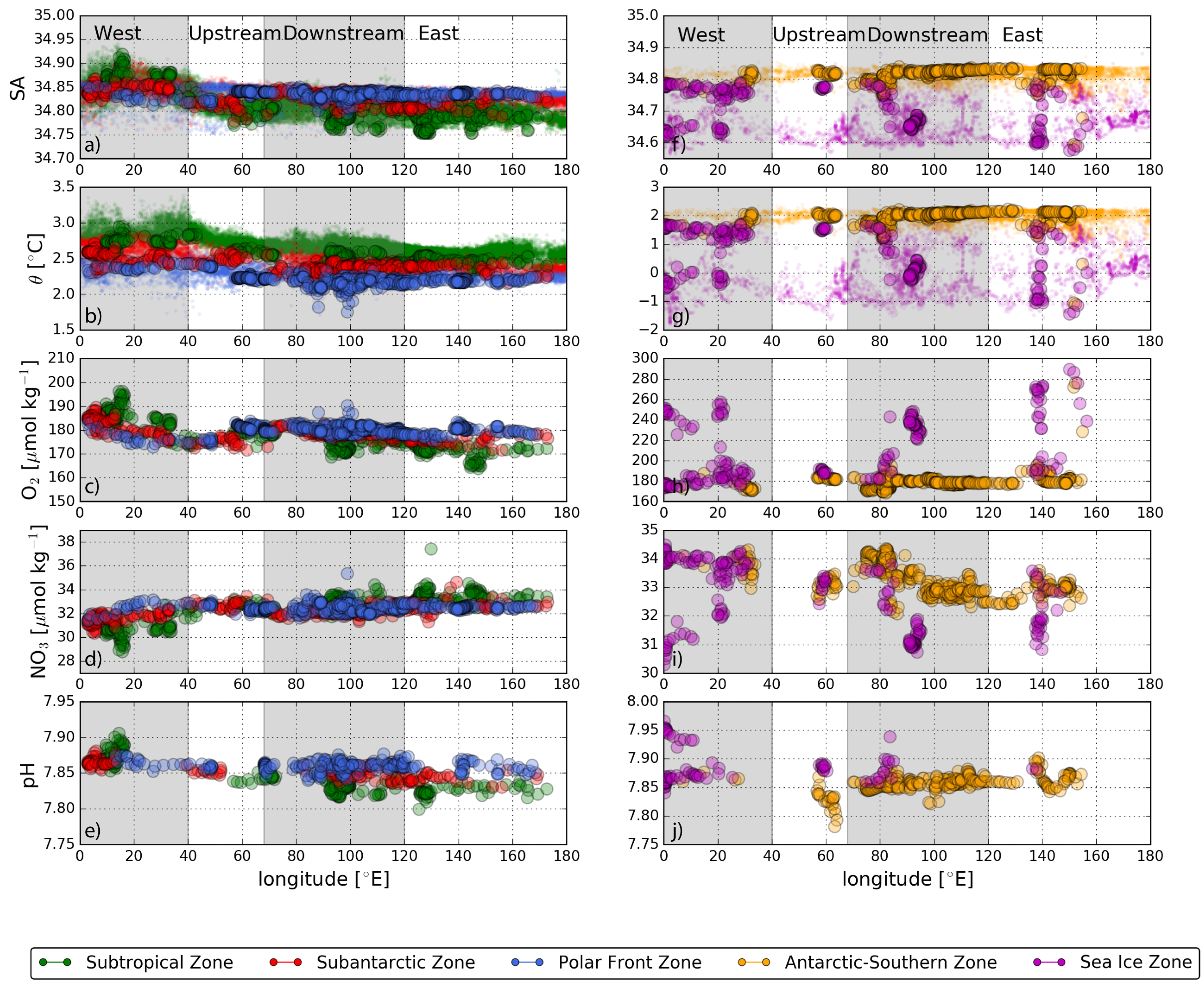

Figure 14. (a) Absolute salinity, (b) potential temperature, (c) dissolved oxygen, (d) nitrate, and (e) dissolved inorganic carbon, as measured by core Argo (colored dots) and biogeochemical-Argo (outlined colored circles) floats across the northernmost frontal zones, the (red) Subtropical Zone and (cyan) Subantarctic Zone. Properties are averaged over the Upper Circumpolar Deep Water (with density σ_0 larger than 27.3 kg m^{-3}).

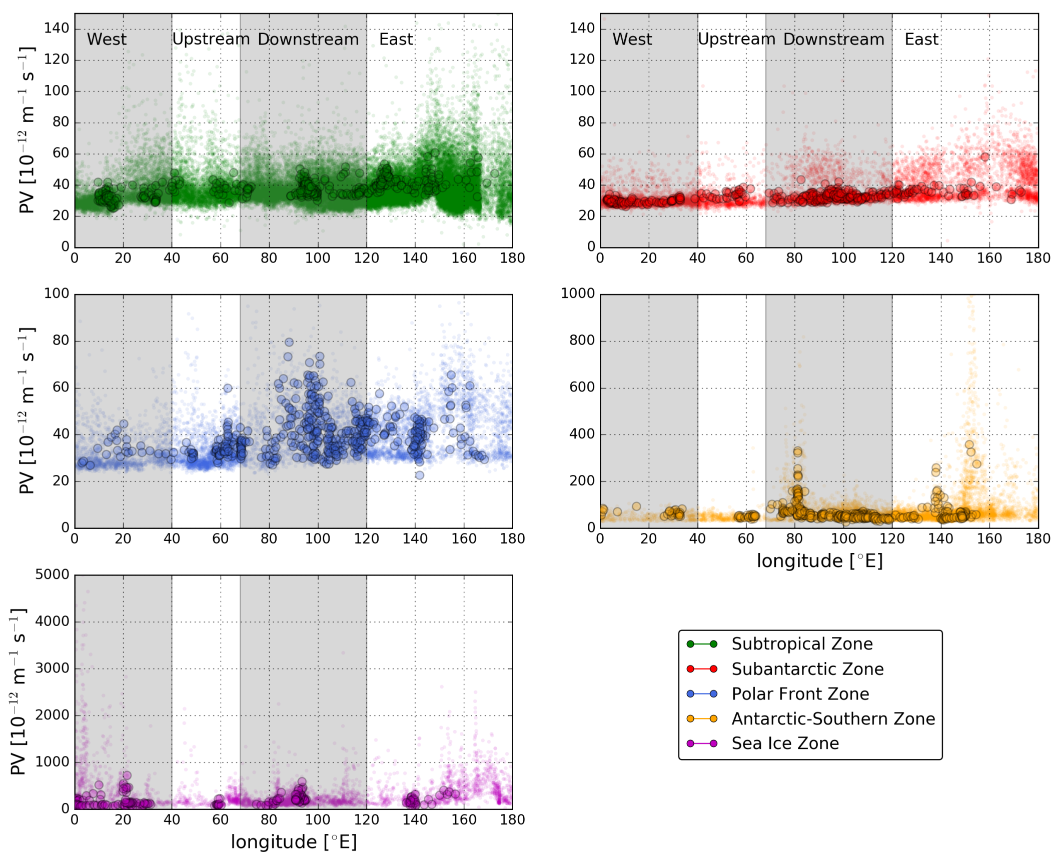

Figure 15. Potential vorticity along longitude, color coded by zone as in Fig. 14 and averaged over the Upper Circumpolar Deep Water (with density σ_0 larger than 27.3 kg m^{-3}).

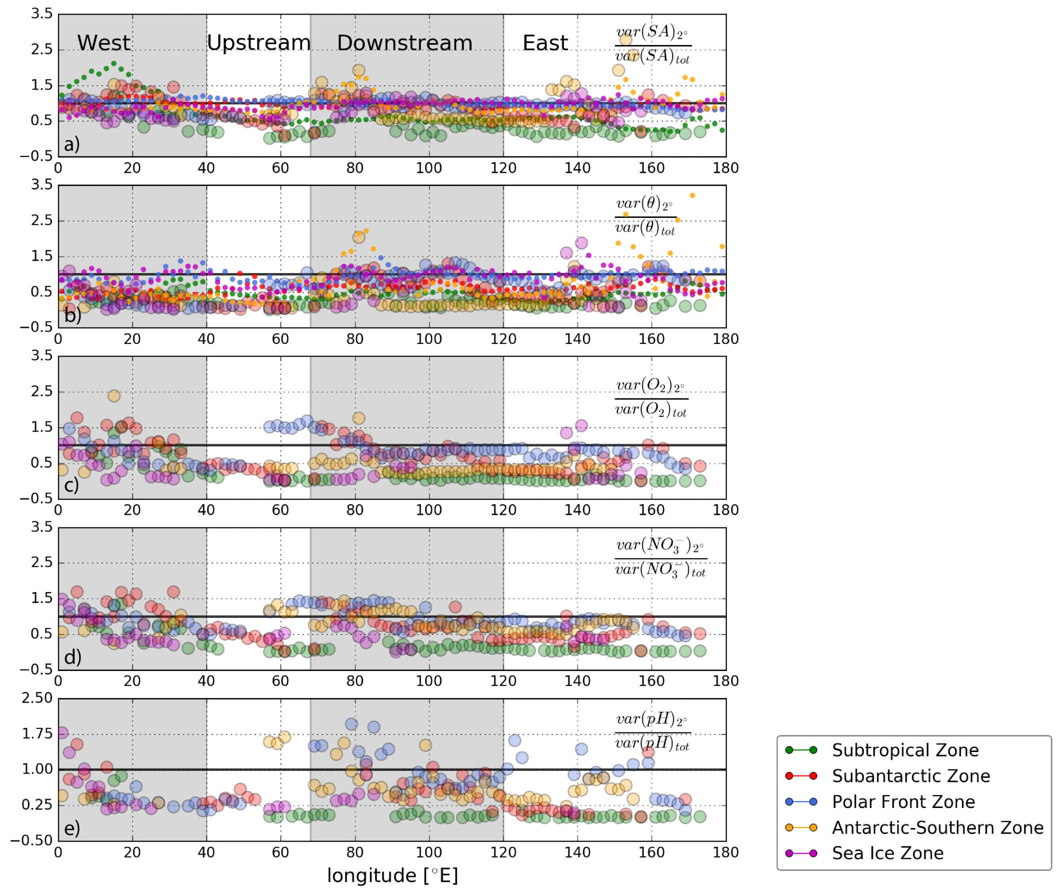

Figure 16. Regional variability of the ratio of binned (2° in longitude) over the total variance of (a) absolute salinity (SA), (b) potential temperature (θ), (c) oxygen (O_2), (d) nitrate (NO_3^-), and (e) pH, along σ_0, computed across the density classes around the Upper Circumpolar Deep Water. Markers are color coded as in Fig. 14.