



Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean

Guillaume Maze ^{a,*}, Herlé Mercier ^b, Ronan Fablet ^c, Pierre Tandeo ^c, Manuel Lopez Radcenco ^c, Philippe Lenca ^c, Charlène Feucher ^a, Clément Le Goff ^a

^a Ifremer, Univ. Brest, CNRS, IRD, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, 29280, Plouzané, France

^b Univ. Brest, CNRS, IRD, Ifremer, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, 29280, Plouzané, France

^c Institut Mines-Telecom, Télécom Bretagne, UMR CNRS 6285 Lab-STICC, Brest, France

ARTICLE INFO

Article history:

Received 8 April 2016

Received in revised form 9 December 2016

Accepted 30 December 2016

Available online xxxx

Keywords:

Heat content

Classification

North Atlantic

Stratification

Water mass

Thermocline

Argo

Pattern

ABSTRACT

A quantitative understanding of the integrated ocean heat content depends on our ability to determine how heat is distributed in the ocean and identify the associated coherent patterns. This study demonstrates how this can be achieved using unsupervised classification of Argo temperature profiles. The classification method used is a Gaussian Mixture Model (GMM) that decomposes the Probability Density Function of a dataset into a weighted sum of Gaussian modes.

It is determined that the North Atlantic Argo dataset of temperature profiles contains 8 groups of vertically coherent heat patterns, or classes. Each of the temperature profile classes reveals unique and physically coherent heat distributions along the vertical axis. A key result of this study is that, when mapped in space, each of the 8 classes is found to define an oceanic region, even if no spatial information was used in the model determination. The classification result is independent of the location and time of the ARGO profiles.

Two classes show cold anomalies throughout the water column with amplitude decreasing with depth. They are found to be localized in the subpolar gyre and along the poleward flank of the Gulf Stream and North Atlantic Current (NAC). One class has nearly zero anomalies and a large spread throughout the water column. It is found mostly along the NAC. One class has warm anomalies near the surface (50 m) and cold ones below 200 m. It is found in the tropical/equatorial region. The remaining four classes have warm anomalies throughout the water column, one without depth dependence (in the southeastern part of the subtropical gyre), the other three with clear maximums at different depths (100 m, 400 m and 1000 m). These are found along the southern flank of the North Equatorial Current, the western part of the subtropical gyre and over the West European Basin. These results are robust to both the seasonal variability and to method parameters such as the size of the analyzed domain.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

As revealed by in situ and satellite observations, the ocean has undergone significant changes in recent decades. In particular, since the early 1970s, the ocean has stored 93% of the excess of heat added to the Earth's climatic system by the anthropogenically modified radiative balance at the top of the atmosphere (Stocker et al., 2013). The ocean has also been found to be more stratified (Levitus et al., 2012) and Western Boundary Currents have probably shifted poleward and intensified (Wu et al., 2012; Yang et al., 2016). To understand the drivers of these changes requires a quantitative understanding of the integrated ocean heat content.

The ocean temperature structure is very complex but a simple first-order description is possible. Near the surface, ocean temperature is primarily driven by air-sea heat fluxes and modulated by horizontal heat mean and eddy transports, especially in Western Boundary Current systems (Kwon et al., 2010). This causes the ocean surface temperature to decrease poleward. However, ocean currents are three-dimensional and redistribute heat at different depths. At mid-latitudes, a negative wind-stress curl forces a downward doming of isopycnal and isothermal surfaces (Vallis, 2006), which results in the temperature at depth, e.g. 500 m, to be higher in subtropical gyres than at the equator (Talley et al., 2011). This 3-dimensional redistribution of heat in the ocean makes our ability to identify remarkable heat patterns in the horizontal and vertical planes crucial to the understanding of the integrated ocean heat content.

* Corresponding author.

E-mail address: gmaze@ifremer.fr (G. Maze).

Along the vertical axis, remarkable patterns may be defined a priori using known water masses such as shallow, intermediate and deep waters or layers such as the mixed and Ekman layers and the permanent thermocline. These patterns can be used to partition horizontal or vertical heat transports (Talley, 2003; McCarthy et al., 2012; Buckley et al., 2014). However, they are not used to partition heat content variability despite efforts to formalize the use of reference surfaces in vertically integrating heat content (Palmer and Haines, 2009). The problematic is that the general lack of clear objective definition for vertical patterns, despite a recent effort with regard to the permanent pycnocline (Feucher et al., 2016), impedes their description, especially over long timescales during which their defining properties can change (e.g. Yang and Wang, 2009; Fiedler, 2010).

In the horizontal plan, remarkable large scale patterns are not defined per se. Simple geographical boxes of fixed size and shape are preferred. One is left with the difficult task of looking for relevant boxes to explain the large scale structure and variability of the heat content. Many studies define the subtropical and/or subpolar gyres as rectangular boxes from which box-averaged statistics are computed (e.g. Lozier et al., 2010; Bryden et al., 2014; Häkkinen et al., 2015; Grist et al., 2015). Due to limited availability of historical measurements, one can even find signals of entire regions to be approximated by a single location dataset (e.g. Curry and McCartney, 2001). Obviously, a serial issue with a rectangular box is that it does not take into account the complex structure of the ocean, which is not aligned along latitudes and longitudes. The problematic is that, although it is always possible to use more complex polygons than a rectangle to describe a region (e.g. Barrier et al., 2015), this will be difficult, if not possible, if a region has to be bounded by a dynamical structure such as a Western Boundary Current.

To identify remarkable heat patterns in the horizontal and vertical plans, their variability and climatology thus remains a challenge. In this study, we propose to tackle this problem with a method that belongs to the class of unsupervised classification methods. Classification, or clustering, is a statistical method that groups data into classes, or clusters, according to a given similarity metric.

Profile classification has already been used in oceanographic applications but for other purposes. Hjelmervik and Hjelmervik (2013) used a classification method on in situ profiles to predict the local vertical structure of temperature and salinity at a given location, without surface information. Indeed, it is rather common to predict the interior structure of the ocean based on surface data, such as sea surface height, and a model either based on physical principals (Ponte and Klein, 2013) or on historical local regressions (Guinehut et al., 2012). To do so without a surface information is much more complicated though. Hjelmervik and Hjelmervik (2013) grouped profiles according to their Euclidean distance in a reduced dimensional space for latitude/longitude/temperature/salinity and derived a prediction model of climatological profiles at a given latitude/longitude location. For the method to perform better than a classic box averaging method, they determined that 26 groups of profiles were necessary for the North Atlantic Ocean. They later adapted the method to real-time profile prediction using partial observations, and decreased the number of groups to 18 for the method to perform well (Hjelmervik and Hjelmervik, 2014). A classification based prediction method to fill in gaps in two-dimensional data has also been used for satellite measurements with clear success (Aretxabaleta and Smith, 2013). The work by Hjelmervik and Hjelmervik (2013, 2014) extends this idea to vertical profiles with strong promises.

Classification based prediction methods strive in dealing with non-Gaussian statistics, such as observed in frontal regions (Sura, 2010). However, for our goal, which is to characterize remarkable

heat patterns, they suffer from two limitations: (i) they take data latitude and longitude as parameters and (ii) they require a rather large number of classes to perform well. We understand that these requirements are imposed to ensure a satisfactory prediction performance, but here, we are interested in identifying remarkable patterns in vertical temperature profiles, their corresponding regional distributions (if any) and their climatology. Therefore, on the one hand there is no reason to impose data coordinates in the classification. Indeed, we should let the classification reveal the spatial coherence, or lack thereof, of temperature profiles rather than impose it. Tandeo et al. (2014) demonstrated how unsupervised mixture modeling can be used to classify sea surface temperature and height anomalies into small scale dynamical modes without using the latitude and longitude of the data. We shall demonstrate in this study that profile based classes are indeed coherent in space in the North Atlantic. On the other hand, we aim to understand, and therefore reduce, the information contained into a large collection of temperature profiles. Thus, it is crucial for the information to be contained into a limited number of classes. This however will depend on how many *remarkable patterns* are relevant for a given usage.

The paper is organized as follows: in Section 2, the dataset is presented; in Section 3, the classification method is introduced as well as the method we employed to apply it to the Argo temperature dataset; in Section 4, we apply classification to analyze the vertical integral of the heat content and in Section 5, the classification of temperature profiles is performed to reveal the ocean internal heat content structure in the North Atlantic. Last, discussion/conclusion are drawn in Section 6 while the appendices provide technical details about optimization (Appendix A) and sensitivity experiments (Appendix B).

2. Data

In this study, we used data from the Argo array. Argo is a real-time global ocean observation network. It consists of about 3000 autonomous profilers randomly distributed in all oceans to observe the large scale open ocean out of the high latitudes and marginal seas. Most of the profilers drift freely at a parking depth around 1000 m and every 10 days descend down to 2000 m to then rise up to the surface, measuring pressure, temperature and salinity. Once at the surface, profile data are transmitted to data assembly centers by satellite, after which profilers descend back to their parking depth and start another 10 day cycle. Argo data are now used routinely in physical oceanography and are key to the observation of the ocean climate (Riser et al., 2016).

The Argo database is a collection of more than 1.5 million temperature and salinity profiles going from the surface to 2000 m, evenly distributed throughout the seasonal cycle and with approximately 1 profile per month per $3^\circ \times 3^\circ$ cell between 2000 and 2014. We extracted the database in December 2014 from the Coriolis GDAC (Argo, 2014). We selected profiles located in the North Atlantic between the equator and 70N and between 90W and 0E. The collection was reduced to profiles and measurements with correct quality control flags (1, 2, 5 or 8, following the Argo reference Table 2 of the user manual, Carval et al., 2015) between the surface and 1400 m. The depth level of 1400 m was chosen as a compromise between the total number of profiles (the shallower the larger) in the analyzed dataset and the vertical extent of the analysis. We finally interpolated the data on a regular vertical grid with a 5 m resolution (the original resolution ranges from less than 10 m at the surface to 200 m at the bottom of the profile).

Fig. 1A shows the spatial density of the final collection of 100,684 profiles. The North Atlantic is a well observed basin and the spatial density is such that there are around 30 profiles per

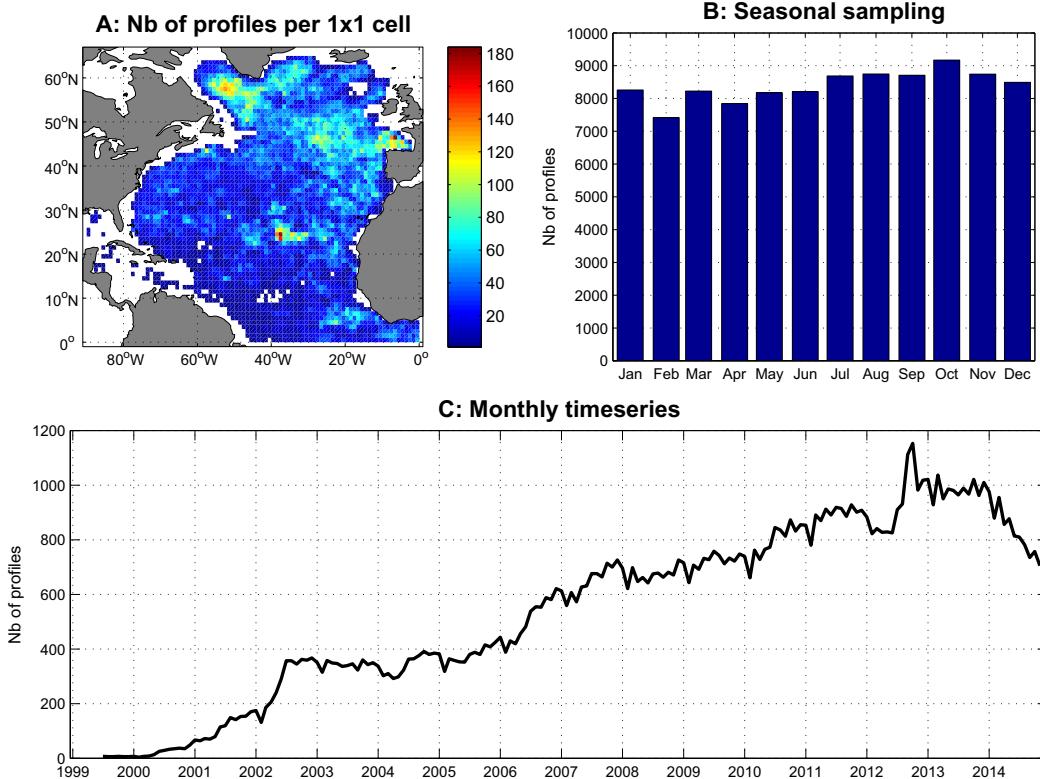


Fig. 1. Spatio/temporal sampling of the 100,684 temperature profiles in the Argo dataset used in this study.

$1^\circ \times 1^\circ$ cell over most of the area. For the period covered by the dataset (180 months), this is about 160% the initial target of the Argo sampling strategy. We note that the data density is not homogeneous though. The subpolar gyre and the North-East, off the Bay of Biscay, are more densely sampled than the Western tropical region. This will be taken into account in the analysis. Fig. 1B shows the temporal sampling of the collection of profiles: there are on average 8400 profiles per month with no significant seasonal bias. Finally, Fig. 1C shows the entire time series of number of profiles per month in the collection. It steadily increased from 2000 to the end of 2013, with a remarkable doubling in 2002, a 200 profile/month increase in 2006 and a peak at the end of 2012. The 2014 decrease is due to the fraction of profiles still waiting for delayed mode quality control and not incorporated in the dataset used here.

3. Method

Hereafter we justify and then present our method based on Gaussian Mixture Models (GMM). Readers already familiar with classification and GMM may want to move directly to Section 3.4. This presentation targets the physical oceanographic community and will therefore favor a pragmatic description to introduce the key elements required to manipulate this statistical tool. The reader is referred to Bilmes (1998) and Bishop (2006) for a detailed GMM description and determination method, and to Hannachi (2007) for an example of its usage in atmospheric dynamics.

Here we are interested in modeling the oceanic heat content structure, i.e. in understanding how, along the water column, heat reservoirs are organized. To do so, we propose to analyze the diversity of vertical temperature profiles by way of automatic identification of recurrent profile patterns throughout the collection of profiles. One can think about this as an extension of the water mass approach (whereby specific levels of the temperature profile are

considered and possibly associated with a water mass) toward the analysis of stacks of water masses.

Fig. 2A shows 50 superimposed profiles randomly drawn out of the Argo collection in order to illustrate the diversity of possible vertical structures: some profiles are almost linear from the surface to the bottom, others exhibit one or more layers with gradients (thermoclines) or homogeneities (mode waters).

The diversity of profiles within the complete collection can statistically be represented by a probability density function (PDF). The PDF captures the relative likelihood of a profile to taking a given pattern at a given depth. If a pattern is recurrent in the collection, its instances will accumulate and create a peak in the PDF. Similarly, if more than one pattern is recurrent, one will observe a PDF with several peaks.

It is important to note that such a PDF of the collection of profiles is not easy to visualize nor to compute because it requires enumeration of the number of profiles falling into small bins of profile types, something which cannot be determined at this point. We therefore adopted a simplified representation, noted PDFz, by enumerating the number of temperature data at a given depth falling into small bins of temperatures, and by scaling the resulting histogram so that the integral at each depth level went to 1. This is shown in Fig. 2B (at 4 sample depths) and C (at all depths). The observed PDFz has several peaks and we note that they are not necessarily connected through the vertical dimension, meaning that patterns at a given depth can be found in profiles with different patterns at other depths. Such complexity makes the collection of profiles strongly heteroscedastic, i.e. to have sub-collections with different statistical properties, in particular their variance. To tackle this we need: (i) to objectively identify recurrent patterns in the collection of profiles, i.e. PDF peaks and (ii) to describe such patterns with a minimum of information, i.e. to create a simple model of PDF peaks. It happens that GMM strives to do exactly that.

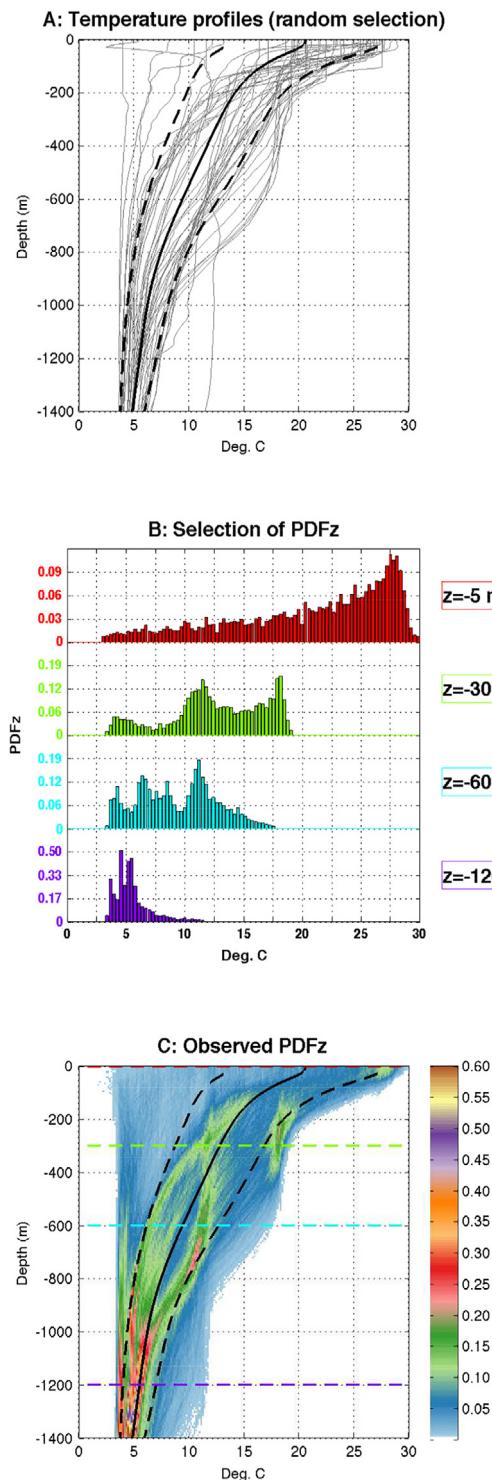


Fig. 2. Plot A: 50 profiles randomly drawn from the Argo collection. Plot B: Sample of PDFz at depth levels -5 , -300 , -600 and -1200 m. Plot C: All observed PDFz, i.e. the PDF of temperature at each vertical levels, here computed using the entire collection. The colorscale is the PDF value so that the integral of one PDFz goes to 1. On plots A and C the mean (plain) and mean \pm one standard deviation (dashed) at each depth are superimposed in black. On plot C, horizontal dashed lines indicate plot B selection of PDFz. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Thus we used an unsupervised classification method called GMM to decompose the PDF of profiles (for which one simplified representation, PDFz, was shown in Fig. 2B and C) into a weighted

sum of multi-dimensional Gaussian PDF. This will allow us to identify and model the typical vertical structures represented in the collection of profiles. The method is referred to as: (i) *classification* because it seeks to classify profiles into sub-collections, or classes, according to their similarity, (ii) *fuzzy* because it provides the probability for a profile to belong to each of the classes, (iii) *unsupervised* because no information about the properties of each class is known *a priori*, only their PDF family is imposed and (iv) *mixture modeling* because it provides a model for the PDF of the collection as a weighted sum (hence mixture) of Gaussian PDF.

One key point of our method that makes it fundamentally different from previous studies (Hjelmervik and Hjelmervik, 2013, 2014) is that we do not use the geographical locations of profiles to help identify groups of similar profiles. This choice was motivated by the idea that there is no reason for the vertical structure of a profile to be unique to a given region. We thus want to determine if profiles with similar vertical structures are also co-localized in space, hence defining physically coherent regions in the ocean by the sole virtue of their similarities. If this is the case, as we shall see, such a model could be used to determine the distribution in space and time of particular stack of water masses, very much like water mass properties (e.g. temperature, salinity and possibly biogeochemical tracers and stratification) can be used to localize water parcels of that water mass in space and time in order to study its distribution and variability (García-Ibáñez et al., 2015).

3.1. Probability density function of profiles

Let us start by introducing the key ingredient of GMM: a multi-dimensional Normal PDF with mean μ and covariance Σ :

$$\mathcal{N}(\mathbf{x}; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \quad (1)$$

where $|\cdot|$ is the determinant and \cdot^\top the transpose operators, $\mathbf{x} \in \mathbb{R}^{D \times 1}$ is a profile of the $\mathbf{x} \in \mathbb{R}^{D \times N}$ collection, $\mu \in \mathbb{R}^{D \times 1}$ and $\Sigma \in \mathbb{R}^{D \times D}$. The array \mathbf{x} is the dataset we want to analyze: it is made of N profiles (as columns) of D vertical levels (as rows). Note that Eq. (1) is a scalar (higher or equal to 0, but not bounded by 1) defined for \mathbf{x} whatever its number of dimensions. We shall refer to $\mathcal{N}(\mathbf{x}; \mu, \Sigma)$ as the $\mathbb{R}^{1 \times N}$ ensemble of PDF values obtained by computing Eq. (1) for all profiles of the dataset $\mathbf{x} \in \mathbb{R}^{D \times N}$.

Although in Section 4 we start by investigating a simpler unidimensional case, it is crucial to note here that we explicitly keep the vertical dimension (through the dimensions of \mathbf{x}, μ and Σ) in the formulation of the PDF Eq. (1) because it is along this dimension that we aim to identify coherent patterns.

Let us take an example using the collection of Argo profile statistics plotted in Fig. 2. The collection mean profile (black plain line in Fig. 2) is a column vector $\mu \in \mathbb{R}^{D \times 1}$ and the collection standard deviation profile (black dashed lines in Fig. 2 that are one standard deviation around the mean at a given depth) is the square root of the diagonal of a covariance matrix $\Sigma \in \mathbb{R}^{D \times D}$. We will keep in mind that the covariance matrix has no reason to be diagonal though, because of the vertically coherent patterns. The point here, is that a Gaussian PDF from D -dimensional data takes into account the structure and the scale of patterns exhibited by a population of profiles with D vertical levels.

In practice, and this was the motivation for this study, the collection of Argo profiles cannot be represented appropriately by a single Normal PDF.

3.2. Gaussian mixture modeling

The core foundation of a GMM is that any PDF can be described as closely as desired with a model of weighted sum of Gaussian PDF (Anderson and Moore, 1979):

$$p(\mathbf{x}) = \sum_{k=1}^K \lambda_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k) \quad (2)$$

where the PDF model has K components, each referred to as $\mathbf{c} = k$, from the same parametric Gaussian family given by $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)$ and being associated with weighting coefficients λ_k . One can show that the component weights must satisfy: $0 \leq \lambda_k \leq 1$ and $\sum_k \lambda_k = 1$ (see Bishop, 2006, p. 430). These weights are the prior probability densities $p(\mathbf{c} = k)$ of each component (Bishop, 2006, p. 110).

We aim to fit a PDF model of the form Eq. (2) to the observed PDF of the vertical profiles:

$$\hat{p}(\mathbf{x}) \simeq p(\mathbf{x}) = \sum_{k=1}^K \tilde{\lambda}_k \mathcal{N}(\mathbf{x}; \tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k) \quad (3)$$

where $\hat{p}(\mathbf{x})$ stands for the observed PDF and tildes on parameters ($\tilde{\lambda}_k, \tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k$) stand for their best estimates. Gaussian mixture modeling nails down to an optimization problem that can be tackled by maximizing the likelihood of observed profiles. This optimization is referred to as *model training*. It is solved with the Expectation-Maximization method (McLachlan and Krishnan, 2007) and was computed using the open source Netlab software (Nabney, 2002). For clarity in this method introduction, we provide the details of the model training in Appendix A.

The only input parameter to the model training procedure is the number of mixture components K . It is a rather difficult problem to determine automatically the most appropriate number of components. There exist different methods that are mostly based on estimating the most probable K , or minimizing a given metric such as the mixture entropy or misfit with the observed PDF (Fraley et al., 1998). A popular method is the Bayesian Information Criterion (BIC, Schwarz, 1978). The BIC is an empirical approach of the model probability computed as:

$$BIC(K) = -2\mathcal{L}(K) + N_f(K) \log(n) \quad (4)$$

where $\mathcal{L}(K)$ is the log likelihood of the trained model with K classes (see Eq. (A.1)), $N_f(K) = K - 1 + KD + KD(D - 1)/2$ is the number of independent parameters to be estimated (the sum of the component weights, Gaussian means and covariance matrix elements in the D -dimensional data space) and n is the number of profiles used to train the model. The BIC is empirical because the first r.h.s. term in Eq. (4) decreases as the number of classes K increases, while the second r.h.s. term is a penalty term that increases with K and thus prevents model overfitting the data. The sum of the two terms is expected to exhibit a minimum for the most appropriate K . For our dataset, we found the BIC method to provide insightful guidance only if computed using the minimal number of profiles expected to describe the collection. More details are provided in Section 5. Here, the user should keep in mind that the number K sets a compromise between compression (small K) and accuracy (large K) of the GMM and that it is meant to provide a number of classes that is physically meaningful.

At this point, we thus optimally decomposed the observed PDF into K Gaussian densities representing the K components of the mixture model. Each component is defined by a Gaussian density with parameters $(\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k)$ and by a $\tilde{\lambda}_k = p(\mathbf{c} = k)$. A component will gather a group, or class or cluster of profiles similar to each other while simultaneously being as different as possible from profiles

of other components (the similarity metric being the Mahalanobis distance from the Gaussian density).

How finally do we determine which component a profile x resembles most, i.e. how do we *classify* it? We compute the posterior probability density of a component $p(\mathbf{c} = k|x)$ which, given a prior density $p(\mathbf{c} = k)$, is obtained using Bayes' rule $p(\mathbf{c} = k|x)p(x) = p(x|\mathbf{c} = k)p(\mathbf{c} = k)$ as:

$$p(\mathbf{c} = k|x) = \frac{\tilde{\lambda}_k \mathcal{N}(x; \tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k)}{\sum_{k=1}^K \tilde{\lambda}_k \mathcal{N}(x; \tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k)} \quad (5)$$

where we recognize that: (i) the mixture PDF $p(x)$ is given by Eq. (3) for a profile x in the collection \mathbf{x} , (ii) the profile PDF conditioned to one component $p(x|\mathbf{c} = k)$ is given by Eq. (1) using the optimized set of parameters $(\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k)$, a probability sometimes referred to as the *activation*, and (iii) as already noted, the component prior density $p(\mathbf{c} = k)$ is the weighting coefficient $\tilde{\lambda}_k$.

The probability density $p(\mathbf{c} = k|x)$, defined Eq. (5), is the result to be used to *classify* a profile. It is interesting to note that there are K posterior values for each profile and that they add up to 1. This is why GMM is referred to as a probabilistic or *fuzzy* classification method, in opposition to *hard* ones like KMeans, which provide a binary classification (probability is reduced to either 0 or 1, Lloyd, 1982).

To ease visualization or discussion of the classification results, a profile is said to be *attributed to* or *labeled* with the class k for which its posterior is maximum:

$$\mathcal{C}(\mathbf{x}) = \arg \max_k p(\mathbf{c} = k|\mathbf{x}), k = 1 : K \quad (6)$$

One can interpret maximum posterior values as follows. As the posteriors $p(\mathbf{c} = k|\mathbf{x})$ range from 0 to 1, one can imagine the two extreme cases where: (1) the profile perfectly matches that of the class center and $p(\mathbf{c}|\mathbf{x}) \rightarrow [1, 0, \dots, 0]$ with $K - 1$ null values or (2) the profile does not match any of the classes and $p(\mathbf{c}|\mathbf{x}) = [1/K, 1/K, \dots, 1/K]$, i.e. classes are equi-probable. In that latter case one should note that all posteriors cannot converge toward zero, which might have been expected if none of the model classes corresponded to the profile, because their sum must come to one. Equi-probability is thus the indication of un-conclusive labeling. This simple line of arguments puts some bounds on the maximum posterior value: from $1/K$ for the worst case scenario of equi-probable classes up to 1 for the best case scenario if one of the class is a perfect fit for the profile (though never reached in practice). In this study, we introduce the following labeling metric as:

$$\mathcal{R}(\mathbf{x}) = \left(p_m - \frac{1}{K} \right) \frac{K}{K - 1} \quad (7)$$

where $p_m = \max(p(\mathbf{c}|\mathbf{x}))$ is the maximum posterior value used to attribute a class to a profile. This metric is a simple scaling of the likelihood of the class attributed to a profile. It ranges from 0 for an unlikely labeling (equi-probable label) up to 1 for a virtually certain labeling.

3.3. Dimensionality reduction

The main difficulty in training a GMM with oceanic profiles is that the dimension along which we want to extract hidden structures within the dataset, the vertical axis, is large. We have to deal in Argo data with at least 30–40 levels at which a parameter is defined. If, like in our case, vertical profiles are interpolated on a regular 5 m grid, we have no less than 280 levels. This large number of dimensions in the problem fundamentally translates into a large number of parameters to be determined in the Gaussian covariance matrices. The way to tackle such a problem is still a subject of

intense research (e.g. Krishnamurthy, 2011; Azizyan et al., 2014). Although this so-called “curse-of-dimensionality” problem typically occurs in datasets where the number of training data (profiles in our case) is small compared to the number of dimensions (vertical levels), here we consider a dimension reduction scheme to improve the computational efficiency and robustness of the GMM training.

To reduce the number of dimensions several methods are available. We used the most popular one: Principal Component Analysis (e.g. Thomson and Emery, 2014, p. 335) on which spatio/temporal EOFs are based (Bjornsson and Venegas, 1997). With PCA, we can decompose a dataset $\mathbf{x}(z)$ as:

$$\mathbf{x}(z) = \sum_{j=1}^d \mathbf{P}(z,j) \mathbf{y}(j) \quad (8)$$

where $\mathbf{P} \in \mathbb{R}^{D \times d}$ and $\mathbf{y} \in \mathbb{R}^{d \times N}$ with $d \leq D$. The first rows of \mathbf{P} contain profiles maximizing the structural variance throughout the collection of profiles. Thus, if we choose $d \ll D$, we can reduce the number of dimensions of the dataset \mathbf{x} while preserving most of its structure. This decomposition creates a new space where the N profiles are not defined with D vertical level values (the \mathbf{x} array) but with only d ones (the \mathbf{y} array). The transition between one space and the other is done through the matrix \mathbf{P} containing the definition of the new dimensions in the original ones (d vertical profiles of D levels, the eigenvectors of the covariance matrix $\mathbf{x}^\top \mathbf{x}$).

Dimensionality reduction comes at the expense of the amount of detail represented in the dataset. To investigate this loss of detail, one can reverse Eq. (8) with different values of d and compare the root mean squared difference (RMSD) between the reconstructed and original datasets. This is shown in Fig. 3 for d ranging

from 0 to 30. Not surprisingly, for all d values, the RMSD is larger near the surface. This is a typical behavior of the PCA decomposition of vertical profiles (see for instance Fig. 20 in Fukumori and Wunsch, 1991). One can also note small scale features appearing either at depth or with large $d = 20$ or $d = 30$. This noisy behavior arises from the interpolation of Argo profiles with an original 20–50 m vertical resolution onto standard depth levels with a 5 m resolution. This is not problematic here because PCA acts as a small scale filter. We choose $d = 11$ (black curve in Fig. 3) to obtain a RMSD smaller than 0.5°C at all depths and smaller than 0.1°C below 600 m (thus preserving 99.88% of the dataset variance). This level of RMSD is satisfactory to analyze the climatological structure of heat we are interested in. It would take twice that number of dimensions to decrease the RMSD below 0.25°C at all depths. Using PCA, we thus reduced the number of dimensions from 280 to 11, i.e. by one order of magnitude. This is a very significant data reduction that considerably improves the computational cost of the classification.

Note that to compute the decomposition Eq. (8), we centered and standardized the dataset \mathbf{x} along each dimension, i.e. along each vertical level, so that the variance of the upper or surface layers did not dominate over the deeper ones in the definition of the new space. One can ultimately note that in this study, PCA is used to compress the dataset (reduce its number of dimensions), rather than to investigate the structure of its covariance matrix (that, is left to GMM). So we purposely do not show or analyze profiles and maps of the $\mathbf{P}(z,j)$ and \mathbf{y} matrices for which a physical interpretation would be hard to derive.

3.4. Processing the irregular Argo dataset

Overall, we choose to train a GMM on a subset of the full collection of profiles and then to classify all the profiles using the trained model. Motivations for this choice are two fold: (i) computational efficiency and (ii) un-biased structural sampling. The later point is crucial. Indeed, we have shown in Section 2 and in Fig. 1 that the collection of Argo profiles is not evenly distributed in space, some regions having almost 100 more profiles than others. This is not an issue per se to train a GMM but we want to ensure that the GMM components will possibly reveal spatial information about the dataset, rather than its sampling spatial density, i.e. to answer the question of whether class, or group of profiles that resemble each other, have a geographical coherent signature or not.

We thus created a training subset by randomly selecting profiles within the full collection so that the training subset had a spatial density of 10 profiles per $2^\circ \times 2^\circ$ grid cell over at least 75% of the domain (the training subset has finally 7590 profiles). This value was determined by trials and errors as a compromise between the size of the training set and its domain coverage. If we imposed too high a density of profiles, we were excluding less populated regions (such as the western low-latitudes). If we imposed too low a density of profiles, we were dramatically reducing the size of the training set.

Last, note that this sub-setting is not related to a classic cross-validation analysis, which distinguishes training and test datasets. The sensitivity of the classification to sub-setting is discussed in Appendix B.

3.5. Procedure

Now that we have introduced all elements of the analysis, it is time to make a recap of the procedure used to perform the unsupervised classification of Argo temperature profiles with GMM:

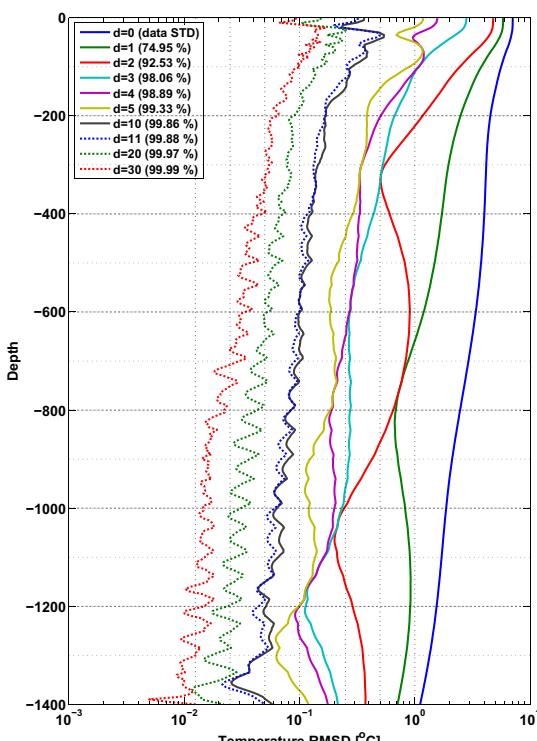


Fig. 3. Centered root mean squared difference between the reconstructed and original datasets of temperature for an increasing number of dimensions used during dimensionality reduction with PCA. The plot can be read as follows: it takes 11 reduced-dimensions (black curve) to compress the dataset in a way that the RMSD is lower than 0.5°C throughout the water column. Vertical tick marks are centered at 0.0125, 0.025, 0.05, 0.1, 0.25, 0.5 and 1°C .

- create a training subset from the dataset \mathbf{x} where profiles are homogeneously distributed in space,
- select the training subset and then:
 - center and standardize data at each vertical level,
 - reduce the number of dimensions, i.e. compute the $\mathbf{P}(z,j)$ and $\mathbf{y}(j)$ vectors,
 - set a number of classes K (using guidance from BIC Eq. (4)),
 - train a GMM using the reduced-dimension training set, i.e. compute the best set of parameters $\{\tilde{\lambda}_k, \tilde{\mu}_k, \tilde{\Sigma}_k, k = 1 \dots K\}$ with the EM algorithm and Eqs. (A.3)–(A.5)
- for all profiles of the dataset:
 - center and standardize data at each vertical level using training subset mean and standard deviation profiles,
 - reduce the number of dimensions, i.e. compute the $\mathbf{y}(j)$ vectors, given $\mathbf{P}(z,j)$,
 - classify the reduced-dimension profiles, i.e. compute the posteriors, class labels and robustness metric with Eqs. (5)–(7) given the $\{\tilde{\lambda}_k, \tilde{\mu}_k, \tilde{\Sigma}_k, k = 1 \dots K\}$.
- synthesize class patterns by computing weighted class quantile statistics (e.g. compute the median profile of class k as the median of all profiles weighted by their activation values for class k given by $\mathcal{N}(\mathbf{x}; \tilde{\mu}_k, \tilde{\Sigma}_k)$).

Table 1 summarizes all GMM important variables that we introduced in this section.

4. Structure of the vertical mean temperature

Here we analyze the PDF of the vertical mean temperature – a proxy for the heat content – with GMM. In the next section we will investigate the internal structure giving rise to that uni-dimensional synthetic representation of the temperature distribution. We used the 0–1400 m vertical mean temperature from the Argo dataset. This depth range is limited by the available dataset.

Fig. 4 illustrates one of the most important features of the oceanic temperature structure, namely that heat is not concentrated at low latitudes near the equator but in the mid-latitudes. More precisely, heat concentrates in the subtropical gyre where Fig. 4 shows temperature to be greater than 3 °C compared to the equatorial band. Note that the dataset is centered around the domain average, about 9 °C; thus, in the following, anomalies refer to that center.

The PDF of the vertical mean temperature dataset is shown in Fig. 5A as a gray bar plot. A striking feature that was not clearly visible in Fig. 4, is the modality of the water column mean temperature dataset. Indeed, values accumulate in 3 distinct regions to create PDF peaks: warm, cold and near-zero. Although we can distinguish other peaks, these three largest ones clearly capture the main structure of the dataset. One can also note that a significant fraction of the dataset values is located between the main peaks.

From a more methodological point of view, one can note that the warm and cold flanks of the PDF have different slopes: the cold flank is much steeper than the warm flank. This is due to the fact that ocean temperature cannot be negative (more precisely colder than the freezing point) and thus the dataset distribution on the negative range of possible values resemble a Rayleigh distribution. This lower bound effect is usually seen in PDF of wind speed time series (e.g. Pavia and O'Brien, 1986). For the sake of simplicity, we did not analyze this limitation of GMM applied to this dataset any further.

To train a GMM, i.e. to compute the optimized model parameters, on this dataset we need to choose a number of components K . Results for the straightforward choice of $K = 3$ are given Fig. 5B. The plot represents the most likely 3 Gaussian distributions, Eq. (1) for $k = 1 : 3$, weighted by their mixture probability (the λ_k) so that their sum is the model PDF (see Eq. (3)), also

Table 1

Summary of the different parameters and variables used in a GMM analysis. In practice, the optimized mean and covariance of the k th classes are computed in the reduced d -dimensional space.

Variable	Formulae	Dimension	Name
→ (ith) profile of D vertical levels			
x_i, x		$\mathbb{R}^{D \times 1}$	Profile, instance
→ Collection of N profiles of D vertical levels			
\mathbf{x}		$\mathbb{R}^{D \times N}$	Dataset
→ PDF obtained using data from a given depth and a range of r values			
PDF _z		$\mathbb{R}^{D \times r}$	PDF at depth levels
→ Observed PDF of the dataset \mathbf{x}			
$\hat{p}(\mathbf{x})$		$\mathbb{R}^{1 \times N}$	Observed PDF
→ Parametric PDF family used in the mixture model			
$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Eq. (1)	$\mathbb{R}^{1 \times N}$	Class PDF
→ Model PDF for the dataset \mathbf{x}			
$p(\mathbf{x})$	Eq. (2)	$\mathbb{R}^{1 \times N}$	Model PDF
→ Number of classes used to decompose the observed PDF			
K		\mathbb{I}	Number of class
→ k th class density (best estimate)			
$\tilde{\lambda}_k = p(\mathbf{c} = k)$	Eq. (A.3)	\mathbb{R}	Mixing weights, priors
→ Mean of the k th class			
$\tilde{\mu}_k$	Eq. (A.4)	$\mathbb{R}^{D \times 1}$	Center
→ Covariance matrix of the k th class			
$\tilde{\Sigma}_k$	Eq. (A.5)	$\mathbb{R}^{D \times D}$	Covariance
→ PDF of the dataset given the k th class			
$p(\mathbf{x} \mathbf{c} = k) = \mathcal{N}(\mathbf{x}; \tilde{\mu}_k, \tilde{\Sigma}_k)$	Eq. (1)	$\mathbb{R}^{N \times 1}$	Activation
→ PDF of the k th class given the dataset			
$p(\mathbf{c} = k \mathbf{x})$	Eq. (5)	$\mathbb{R}^{N \times 1}$	Posterior
→ Name of the most probable class for dataset \mathbf{x}			
$\mathcal{C}(\mathbf{x})$	Eq. (6)	$\mathbb{I}^{1 \times N}$	Attributed class
→ Labeling metric for the attributed class of dataset \mathbf{x}			
$\mathcal{R}(\mathbf{x})$	Eq. (7)	$\mathbb{R}^{1 \times N}$	Labeling metric
→ Vertical profiles defining the dimensions in the reduced space			
\mathbf{P}	Eq. (8)	$\mathbb{R}^{D \times d}$	Eigenvectors
→ The collection of N profiles projected in the reduced space			
\mathbf{y}	Eq. (8)	$\mathbb{R}^{d \times N}$	Eigenvalues

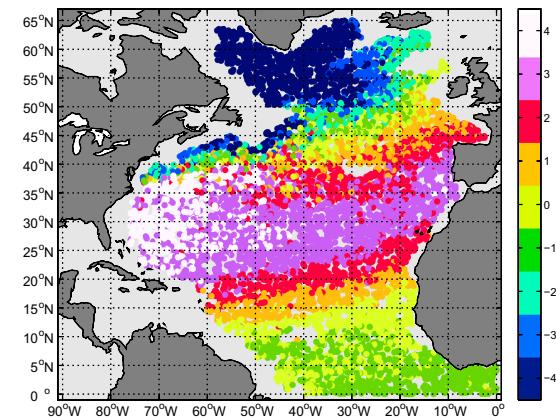


Fig. 4. Horizontal distribution of the 0–1400 m vertical mean temperature, plotted as anomalies with regard to the domain average (which is 9 °C). We want to investigate the vertical structure of heat anomalies leading to such pattern. Note that only 7.5% of the profiles were used to generate this plot (roughly 10 profiles per 2×2 grid cell) in order to limit overlapping dot pixels.

superimposed on the observed PDF panel A. Class 1 and 3 are distinct and capture the cold and warm peaks. However, the choice of $K = 3$ is clearly unsatisfactory to describe the near-zero peak. Class 2 is appropriately centered around the near-zero peak but has a large variance. It thus appears that GMM fills this class with all the data points that are neither in the warm nor in the cold peak regions and consequently loses the ability to properly describe

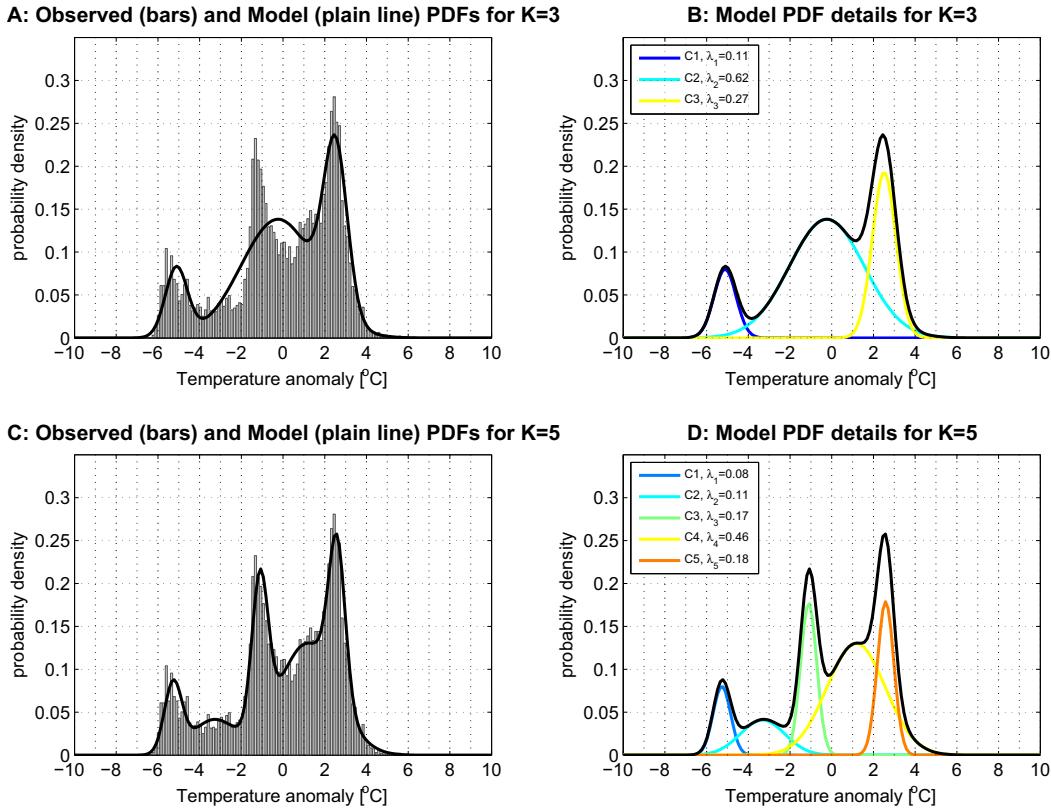


Fig. 5. Left plots: gray bars are the observed PDF of the uni-dimensional dataset represented in Fig. 4 and thick black lines are the GMM PDFs for top: $K = 3$ components and bottom: $K = 5$ components. Right plots: Decomposition of GMM PDFs into their 3 and 5 components. Colored PDFs are the prior weighted activations of each component. Three distinct modes are clearly visible in the dataset (at the extreme and center of the complete PDF) and are nicely captured by the GMM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the near-zero peak. This is fundamentally due to the fact that in GMM, data conditional component probabilities cannot be null, or in other words that a data point must be attributed to a class.

We consequently trained a GMM with $K = 5$ components. The model PDF is superimposed on the observed PDF in Fig. 5C, while details of the decomposition are shown in Fig. 5D. GMM has, in this case, enough degrees of liberty to characterize all peaks of the observed PDF. Class 1, 3 and 5 have small variance and capture the cold, near-zero and warm peaks. The two intermediate classes 2 and 4, have larger variances. These classes can be seen as *transition* classes ensuring that the significant amount of dataset values that are not within the peak regions are also classified. This decomposition is rather satisfactory but has yet some limitations. For instance the PDF structure near the 1°C anomaly range is not captured. One would need to further increase K to reproduce that structure (not shown). This illustrates that the choice for the number of components to impose in training a GMM is left to the operator, based on the observed PDF and critical analysis of the GMM results, and on the analysis goal. For instance, to capture the warm and cold peak regions, $K = 3$ is a good choice and $K = 5$ does not clearly add new information about these peaks.

Now that we have given a description of the dataset structure with GMM, we can visualize classes using other information in the database, like data latitude and longitude. Here we simply map data labels from Eq. (6). A map of the labels is shown in Fig. 6 for the $K = 5$ GMM. We see that the warm class (label 5) coincides with a large zonal band throughout the North-Atlantic subtropical gyre from 15°N to the Gulf Stream and 45°N to the East. The cold class (label 1) is confined to a much smaller region in the Labrador Sea and off the coast of the Newfoundland. Interestingly, the near-zero class (label 3) is found in two distinct geographical

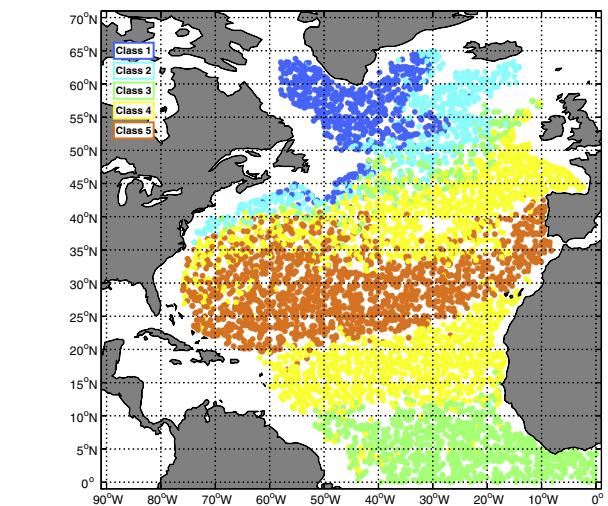


Fig. 6. Map of profiles color-coded with the class they were attributed by Eq. (6) for the uni-dimensional analysis of vertical mean temperature using a GMM with $K = 5$. Class color-codes are similar to those from Fig. 5D. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regions: along the North Atlantic Current and at low latitudes, south of 15°N .

With GMM components we detected a structure of the dataset that is far from obvious in Fig. 4. This analysis shows that the ocean vertical mean temperature distribution - or integral heat content - is much more complex than a simple equator-to-pole meridional

gradient. It clearly reveals how ocean dynamic organizes heat in reservoirs embedded between transition regions that may be associated with fronts. If heat would have been uniformly distributed, then the PDF would not have exhibited modality and would have mapped with similar probability from the equator to the pole. This would have resulted in a series of classes with large overlapping ranges of temperature anomalies.

5. Ocean internal heat content structure

The uni-dimensional analysis conducted in the previous section revealed key features in the distribution of the vertical mean temperature and its associated regional patterns. In this section we aim to determine the internal structure of temperature anomalies giving rise to that integrated point of view and to investigate the corresponding regional distribution.

5.1. Vertical distribution of heat

Fig. 7A presents the observed PDFz, i.e. the PDF at each level, of the temperature dataset where data have been centered and standardized at each depth level (which was not the case **Fig. 2C**). **Fig. 7C** additionally shows the PDFz at 5, 300, 600 and 1200 m depth levels. Like the top 50 m of the water column, the surface PDFz does not show clear modes and mostly reflects a meridional gradient of temperature. This is no longer true as we look deeper. From 100 m to 400 m depth, three peaks appear centered over cold, near-zero and warm temperature anomalies progressively shifting toward warmer classes. These peaks are very similar to

those found for the PDF of the vertical mean temperature (see **Fig. 5**). From 400 m to 800 m depth, three other peaks can be seen with centres progressively shifting from one anomalous temperature range to another. Below 800 m, a three-peak distribution also emerges with different centres converging toward a narrow range of negative temperature anomalies.

One purpose of this study was to identify coherent patterns within the collection of profiles that together lead to such a complex PDFz. To do so, we need to train a GMM and first, to determine the number of classes to be used. We used the BIC method presented in the methods section. For the method to perform appropriately, we found that a subset of independent profiles must be used to evaluate Eq. (4), not the full or only the training set. Using Argo data, [Ninove et al. \(2016\)](#) provided the most recent estimate of the horizontal correlation scale for temperature as a function of depth. These estimates are appropriate for our study focusing on the climatological temperature structure in the North Atlantic, for which the temporal correlation scale does not have to be taken into account. For the North Atlantic, they found that below 30 m depth, the horizontal temperature scale was smaller than 200 km and larger than 100 km, both zonally and meridionally. Thus, using a representative value of 150 km (see their **Fig. 9**), one can expect an approximate number of 900 independent profiles for a $4500 \times 4500 \text{ km}^2$ region such as the North Atlantic. As our collection of profiles is far more larger than this, we computed an ensemble of 50 realizations of Eq. (4) using members of 900 random profiles and K ranging from 1 to 30. The ensemble mean and spread is shown in **Fig. 8**. The BIC clearly exhibits a minimum between $K = 5$ and 10, near 7, 8 and 9. All values were tested

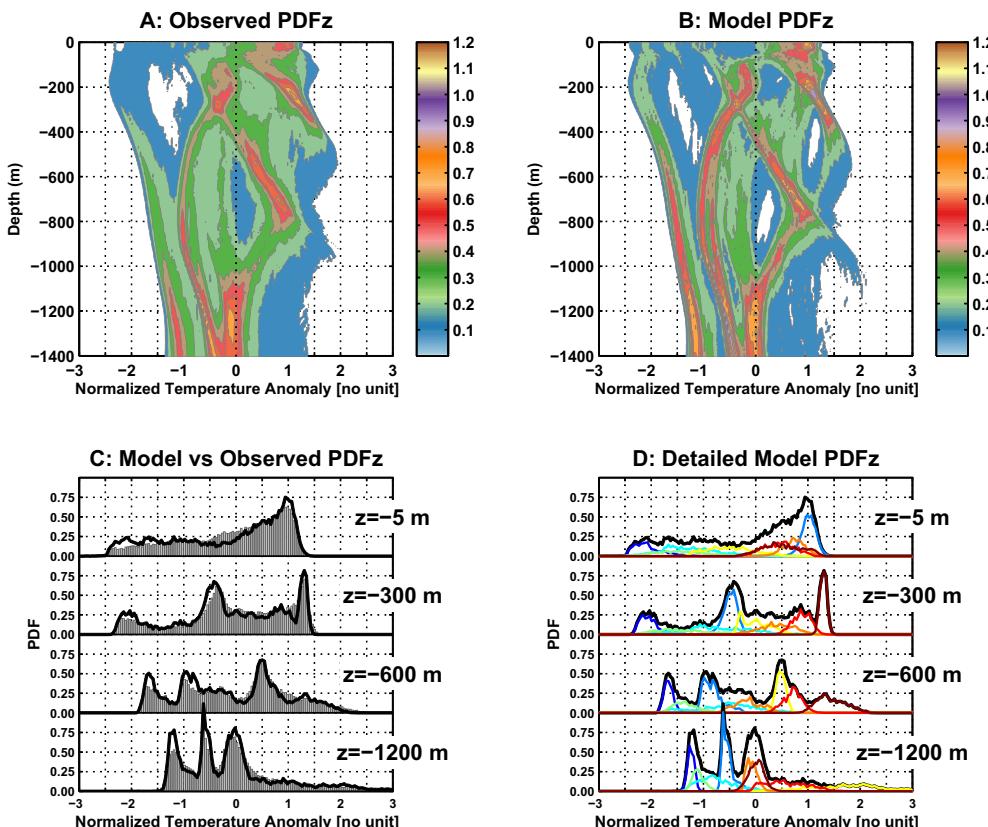


Fig. 7. Plot A: Observed PDFz (PDF at each vertical level) of profiles from the normalized (centered and standardized) temperature dataset. Plot B: Model PDFz from the prior weighted sum of class PDFz (shown in **Fig. 10**). Plot C: Observed (gray bars) and modelled (thick black line) PDFz for selected depth levels: 5, 300, 600 and 1200 m. Plot D: Decomposition of the model PDFz for the same selected depth levels. Colored lines are the prior weighted activations of each class, following the color convention used in **Fig. 9**. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

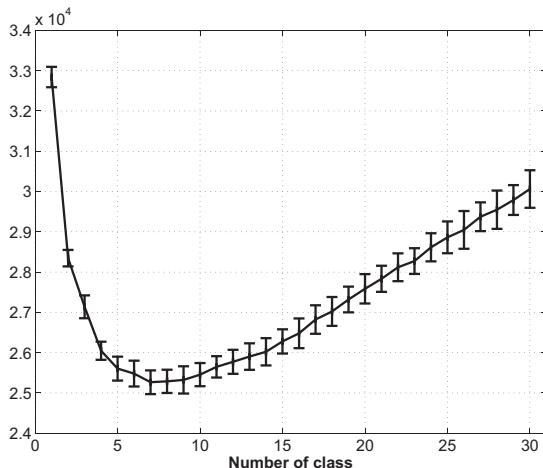


Fig. 8. Ensemble mean and spread (\pm one standard deviation) of the Bayesian information criteria (BIC) for an increasing number of GMM classes K. The ensemble was computed using 50 members of 900 independent profiles with Eq. (4).

and evaluated and we settled on $K = 8$ as it provided the most physically meaningful results. The sensitivity of the GMM classification to the number of classes is discussed in Appendix B.

The trained GMM defines classes as Gaussian distributions in the reduced dimensional space. To analyze the analytical class structures in that space would be quite complicated and we found much easier to present and discuss classes in the real space along

the vertical depth axis. Therefore, we used temperature profiles and their corresponding activation values to compute weighted quantile statistics at each depth level for each class.

Fig. 9 shows activation weighted median profiles for each of the classes obtained with $K = 8$ in the GMM. This decomposition is the key result of the study and can be described as follows. The largest temperature difference between median class profiles is 22°C at the surface, 12°C near -500 m , 6°C near -1000 m and 3°C near -1400 m (consistent with Fig. 2C). Class 1 is the coldest while four classes (2, 6, 8 and 5) are successively the warmer. Class 1 gathers profiles much colder throughout the water column than the dataset average. The median profile for this class has anomalies colder than -5°C above -600 m . Class 2 has a much more complex structure. It gathers profiles with cold anomalies below -100 m (with the largest amplitude of -3°C around -500 m), no anomalies between -100 and -200 m and the largest surface warm anomalies of all classes above -50 m (about 7°C). Class 3 on the other hand, resembles class 1 pattern, simply shifted toward the centre of the PDF at all levels. It is a class with decreasing cold anomalies from -9°C near the surface down to less than -3°C below -400 m . Class 4 is also colder throughout the water column than the dataset average. This class falls within the range of class 1 and 3 median profiles. Class 5 is the warmest class below -800 m depth. It is characterized by near neutral conditions above -400 m and warm anomalies below -600 m with a maximum centered at -1000 m of about 3°C . Class 6 has near neutral conditions below -400 m and a warm anomaly maximum of 7°C centered at -100 m , which makes it the warmest class at this depth. Class 7 gathers profiles warmer throughout the water column than

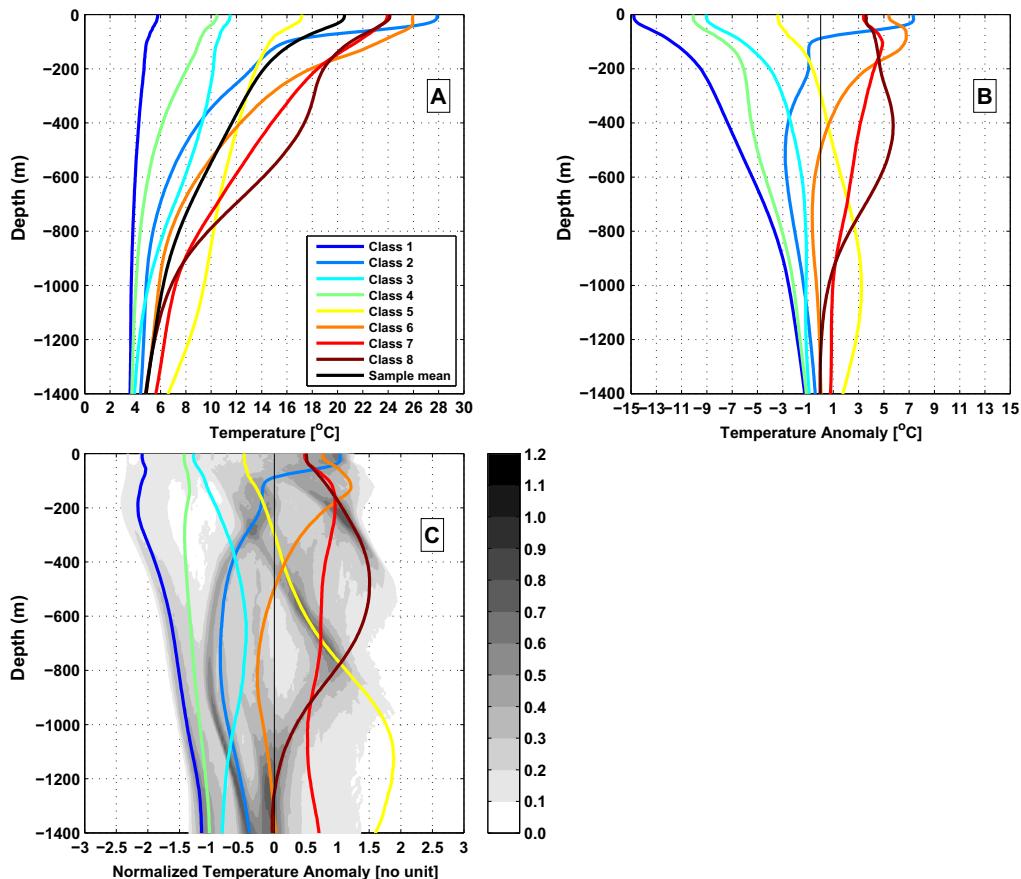


Fig. 9. Activation-weighted median profiles for each class. Plot A: with real temperature profiles (the sample mean reference profile is shown in black), plot B: with profiles of temperature anomalies (centered dataset) and plot C with profiles of normalized temperature anomalies (level centered/standardized dataset). Grey shading in plot C is the observed PDFs of the dataset (reproduced from Fig. 7A).

the dataset average, with anomalies higher than 1°C , but is never the warmest. Last, class 8 exhibits warm anomalies, higher than 3°C above -800 m , with a clear maximum at -400 m of about 5.5°C where this class is the warmest of all.

In Fig. 9C, class median profiles are superimposed on the observed PDFz reproduced from Fig. 7A. Fig. 7C and D shows the observed and modeled PDFz at 4 selected depth levels and the details of each of the GMM components (this can be seen as equivalent to the details of the uni-dimensional case analyzed in Section 4 and shown in Fig. 5). These plots illustrate how well the modeled PDFz reproduces the observed PDFz and how each of the $K = 8$ classes contribute to the peaks of the PDFz. For instance, one can note: how class 5 reproduces the peak of warm anomalies near -600 m , how class 6 and 8 contribute to the neutral peak at -1200 m , how class 8 controls the warm peak of the PDFz near -300 m and how class 1 shapes the colder flank of the PDFz with contribution from class 4 below -600 m .

More precisely, Fig. 10 represents the PDF at each depth of normalized profiles attributed to each of the $K = 8$ classes and each class subtitle indicates the prior value of classes (the λ_k , also reported Table 2). The prior weighted sum of class PDFz is shown in Fig. 7B to reproduce appropriately the observed one, shown in Fig. 7A. Priors are the densities of the class in the model. They can be seen as the fraction of the dataset that can be attributed to each class. This is confirmed by the explicit computation of that fraction, which is reported on the second line of Table 2. Priors and class fraction in the dataset are very close to each other and indicate, for instance, that class 3 accounts for 15% of the profile collection but class 4 only 8%. This, in fact, is not surprising because priors are computed as the average probabilities of classes within the dataset (Eq. (A.3)). Each of the classes has its own vertical structure, and its relative weight in the complete PDF is not related to its physical relevance as a vertical structure to interpret the integrated heat content. This, in fact, demonstrates one of the

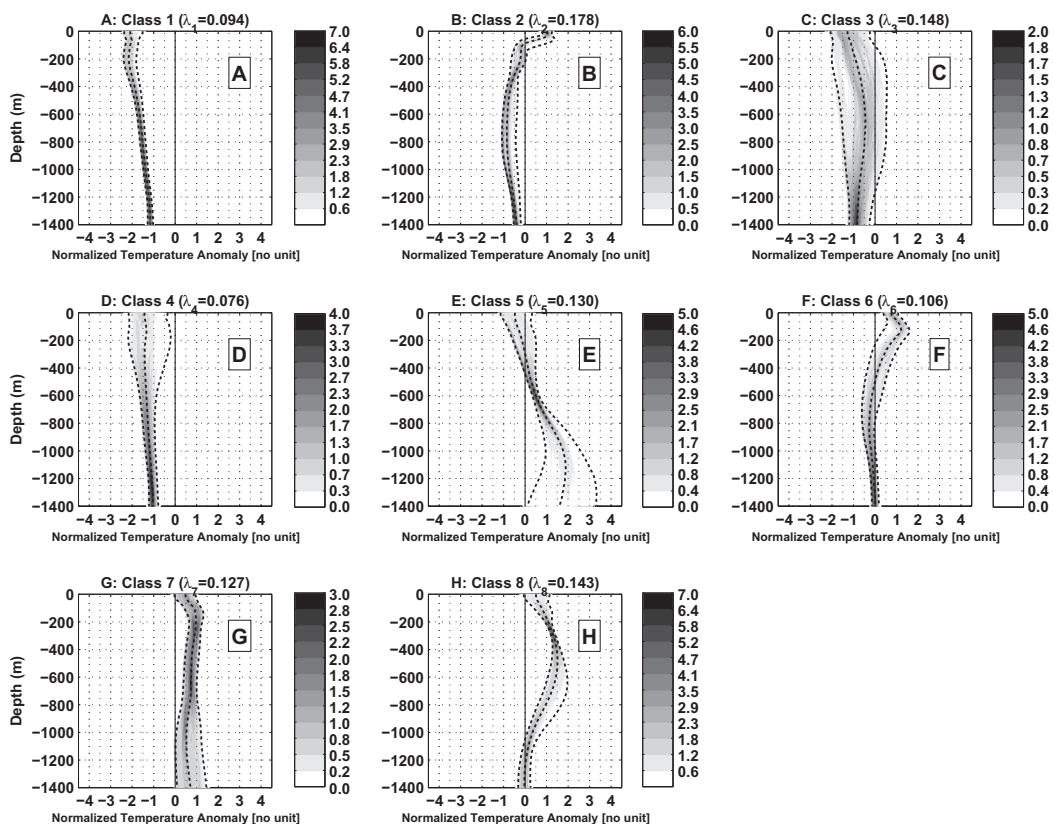


Fig. 10. Normalized temperature anomaly for each class as a function of depth: black dashed lines are the 5, 50 and 95% percentiles of the class (note that the 50% percentile is the median profile plotted in Fig. 9). Grey shading is the corresponding PDFz for each class temperature anomalies. Also given in the plot title are the prior values λ_k from Eq. (A.3) of the class. Statistics were computed using temperature data weighted by activations.

Table 2

For each of the $K = 8$ classes of the reference GMM analysis this table provides: the prior values and the dataset fraction of profiles attributed to each class (first and second lines). From the third to seventh lines are given the distribution by class of the labeling metric (Eq. (7)). The first table column provides similar statistics for the entire dataset. The table can be read as follows: 15% of the dataset can be attributed to class 3, 78% of the class 1 profiles are *virtually certain* to belong to class 1 given this GMM.

	All	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
λ_k Priors	100	9	18	15	8	13	11	13	14
Dataset fraction	100	10	14	15	8	14	9	14	15
Unlikely ($0.00 < \mathcal{R} < 0.33$)	0	0	0	0	0	0	0	0	0
About as likely as not ($0.33 < \mathcal{R} < 0.66$)	5	3	1	7	10	5	4	6	4
Likely ($0.66 < \mathcal{R} < 0.90$)	8	6	2	10	13	7	7	10	8
Very likely ($0.90 < \mathcal{R} < 0.99$)	13	13	4	13	17	10	14	16	17
Virtually certain ($0.99 < \mathcal{R} < 1.00$)	75	78	92	69	60	78	75	68	72

key strengths of the unsupervised classification method: it can distinguish different recurrent patterns without being biased by those having large amplitudes.

On the class PDFz shown in Fig. 10 are superimposed the 5, 50 and 95% percentile profiles. These allow us to investigate the vertical structure of the spread (the width of the class PDF at each level) for each class. Class 1 – which gathers profiles with cold anomalies throughout the water column – has a small spread and the 5–95% percentile envelope always corresponds to cold anomalies, which basically means that profiles attributed to class 1 are always colder than the dataset average at any level. Class 2 exhibits a moderate temperature spread below -200 m where it has negative anomalies for the 5–95% percentile. Near -100 m depth, the spread reaches a maximum and the PDF is positively skewed. Class 3 PDF shows a large spread throughout the water column. This is the class with the largest spread among the GMM. Class 4 has a large spread in the upper layers which decreases downward. Class 5 PDF is narrow around -600 m depth with an increasing and large spread downward. Class 6 spread is the largest at its maximum warm anomaly near -100 m and is relatively small otherwise. Class 7 spread makes it significantly warmer than the rest of the basin throughout the water column. Last, class 8 is extremely homogeneous near -300 m and has a maximum spread around -700 m.

Figs. 9 and 10 focus on statistics for a given class. We now examine statistics for a given profile, i.e. the posteriors (Eq. (5)). For a given profile, K posterior values are used to label it, i.e. to attribute it to its most probable class (Eq. (6)). In the method section we introduced a 0–1 label metric (Eq. (7)) from the maximum posterior value in order to determine if, given a model, a profile label can be trusted (maximum posterior close to 1) or should be taken with caution (equi-probable classes with maximum posterior close to $1/K$). Table 2 shows the class fractions of profiles having their label metric values in 5 pre-defined ranges (simply following the IPCC likelihood standards). Results are striking: only a very small fraction (no more than 5%) of the profiles are unlikely or about as likely as not to be correctly labeled. In other words, the vast majority of the profiles (88%) are very likely or virtually certain to be labeled correctly. This means that classes are appropriately defined and do not overlap significantly. We note that class 2 profiles have the maximum label likelihood (92%) while the minimum is found for the class 4 profiles.

5.2. Relating vertical to horizontal heat distribution

To generate Figs. 9 and 10, we didn't use any geographical information about profiles, nor did we use latitude or longitude in training the GMM. So the question remains of whether profiles attributed to each class are collocated in space. The answer is provided by Fig. 11 where we color-coded profiles with the class they were attributed to ($\mathcal{C}(\mathbf{x})$, see Eq. (6)). It is clear that all classes identified with the GMM method with Argo temperature profiles correspond to physically coherent regions. Class 1 covers the subpolar gyre from the Labrador and Irminger Sea to Eastern flank of the Reykjanes Ridge. Class 2 profiles are localized in the equatorial and tropical bands up to 20°N to the East and 10°N to the West. Class 3 spreads from 45°W to the Iceland Basin in the North Atlantic Current (NAC) region. Class 4 profiles are localized on the poleward flank of the Gulf Stream and NAC. Class 5 is centered over the West European Basin down to the Azores Islands. Class 6 covers the tropical band between 10 and 20°N on the Southern flank of the North Equatorial Current (NEC). Class 7 is in the South Eastern part of the subtropical gyre between 20 and 35°N , mostly to the East of 40°W . Last, class 8 profiles are localized in the Western part of the subtropical gyre, South of the Gulf Stream.

This coherent horizontal distribution of GMM classes provides key indications to interpret the class vertical patterns. The subpolar gyre gathers profiles with relatively uniform cold temperature anomalies throughout the water column (class 1). Low latitudes are regions where heat (positive anomalies in temperature classes 2 and 6), is concentrated in the upper layers while negative anomalies are found at intermediate depth (class 2). These contrasted layers are separated by the equatorial thermocline, which explains the spread in the class 2 PDF near -100 m depth (Fig. 10B). The fact that class 3 and 4 profiles are observed in frontal regions explains the large spread of possible temperature anomalies because front meanders and meso-scale eddies generated by their instability imply a diversity of possible profiles at the same location. Moreover, class 3 profiles have nearly no heat content anomalies because they are located between the warm and cold reservoirs of the subtropical vs subpolar regions (as illustrated by the vertical heat content integral in Fig. 5). Class 5 profiles are clearly located within the sphere of influence of the Mediterranean Outflow. This warm water mass spreads in the North Atlantic near -1000 m, which explains the deep structure of class 5 profiles. Small spread near -600 m depth for this class corresponds to the lower Central $11\text{--}12^{\circ}\text{C}$ Water (Paillet and Arhan, 1996).

The last two regions are located on the eastern and western parts of the southern subtropical gyre, where the integral heat content is the largest. The horizontal distribution of classes allows us to attribute the class 8 large heat reservoir between -200 and -800 m depth to the subtropical thermocline and Eighteen Degree Mode Water (EDW, Maze et al., 2009). The EDW thermostad signature can easily be seen Fig. 9A in class 8 between -200 and -400 m while the thick permanent thermocline ranges from -400 to -800 m (Feucher et al., 2016). Class 7 localizes a region where heat is almost uniformly distributed through the water column, similar to class 1 for cold anomalies.

Last, we examine the geographical distribution of the label metric defined by Eq. (7). This is shown in Fig. 12. When compared with Fig. 11, the distribution of the label metric seems to primarily indicate that profiles located on the edge of regions delimiting classes are those close to equi-probable labels, i.e. with maximum posterior values near $1/K$ (see Fig. 12A). However, similar maps produced using only profiles attributed to each class (Fig. 12B–I)

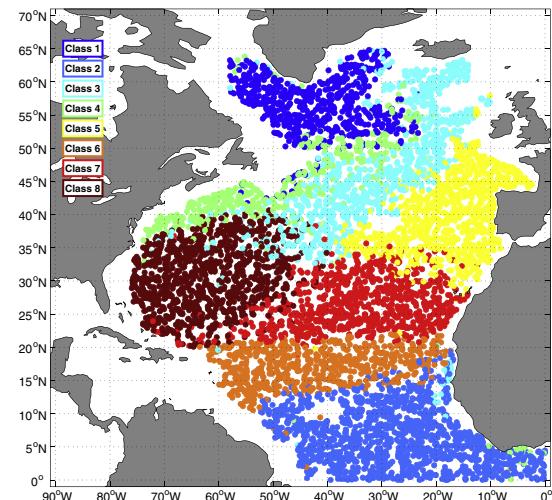


Fig. 11. Map of profiles color-coded with the class they have been attributed to with Eq. (6) by the classification of temperature profiles using a GMM with $K = 8$. Only profiles with a label metric higher than 90% (Eq. (7)) with a limit of 10 profiles per 2×2 grid cell are used in order to limit overlapping dot pixels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

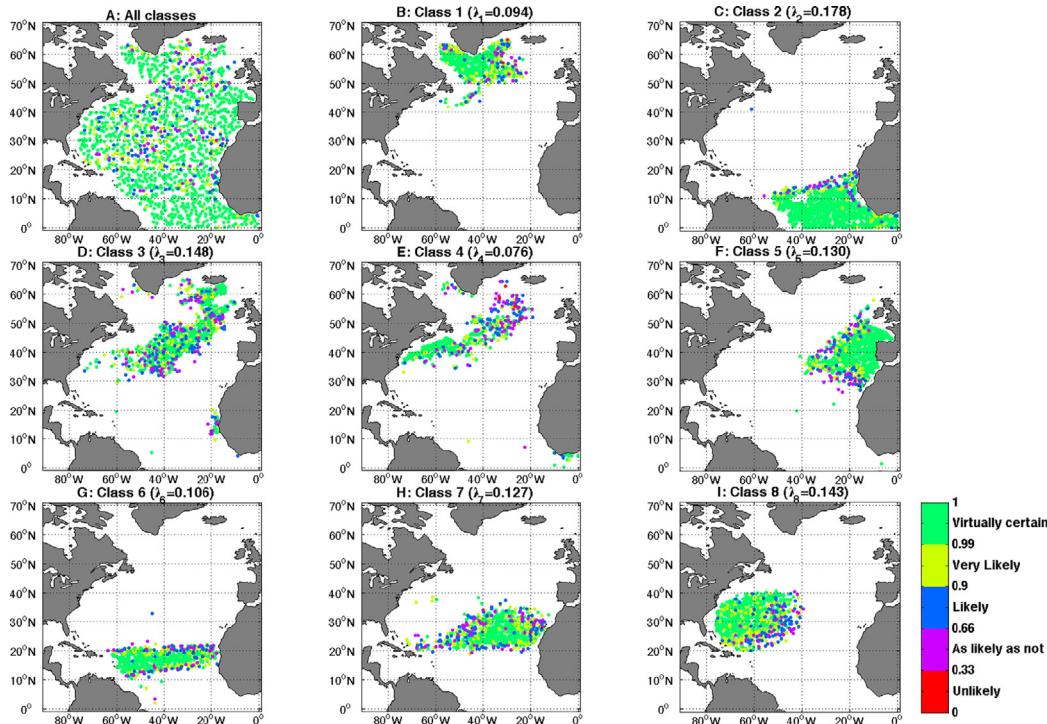


Fig. 12. Map of profiles color-coded with their label metric from Eq. (7) for the classification of temperature profiles using a GMM with $K = 8$. See how labels are less robust along the edges of classes shown in Fig. 11. Map A is for all profiles while maps B to I are for profiles attributed to classes 1 to 8 respectively. No more than 10 profiles per 2×2 grid cell are plotted in order to limit overlapping dot pixels.

reveal a more complex information where robust labels can be found at the edge of class regions when a frontal dynamical structure is involved. For instance, the class 8 northern edge shows a rather robust labeling (greenish points Fig. 12I) where the Gulf Stream is located (similarly with class 4 south-western profiles). To the opposite, class 5 edges do not correspond clearly with a front and have profile labels with a smaller likelihood all around.

6. Discussion and conclusion

In this study we used a Gaussian Mixture Model (GMM) that is an unsupervised classification method (Bishop, 2006) on Argo temperature profiles. Unlike previous studies using such a method (Hjelmervik and Hjelmervik, 2013, 2014) our goal was not to extrapolate a dataset: our goal was to investigate the information contained into a limited number of classes of profiles identified with classification. We have shown that this information was physically meaningful and brought some insight to the oceanic temperature structure.

The GMM was applied to the observed PDF of normalized and compressed temperature profiles. GMM decomposes the observed PDF into a weighted sum of K multi-dimensional Gaussians, the dimension being the vertical axis, using the Expectation-Maximization algorithm. Each Gaussian corresponds to a group of profiles that resembles each other while being as different as possible from profiles in other groups.

We started the analysis with the classification of 0–1400 m vertical mean temperatures, a simple uni-dimensional dataset, which scales as the vertical integral of heat content. In this case, we found that $K = 5$ classes was an appropriate choice to capture the structure and modality of the PDF (Fig. 5C). Three of the classes correspond to clear peaks of the PDF for cold, near-zero and warm temperature anomalies with regard to the dataset average. We showed that the cold class is located in the subpolar gyre, the near-zero class mostly at latitudes lower than 10°N and the warm

class in the western and southern part of the subtropical gyre (Fig. 6). The remaining 2 classes identify transition regions between each of the 3 peaks. It is interesting to note that the simple choice of $K = 3$ classes to fit this 3-peak PDF was not the most appropriate. Indeed, as a significant amount of data values were in the transition regions between the peaks, additional Gaussian classes were necessary to fulfill the GMM requirement for each data to be classified.

The uni-dimensional description of the heat content thus reveals a classic meridional distribution. But to non-experts it might be surprising because it is different from a trivial south-to-north gradient of temperature, such as that observed near the surface. There is indeed more heat stored at mid-latitudes than along the equator because of the internal structure of the oceanic stratification. However this internal structure (from which the PDF at each depth is shown in Fig. 7A) cannot be revealed by the simple uni-dimensional analysis.

Therefore, as a second step, we classified the full temperature profiles using multi-dimensional Gaussians. Given the high dimensionality of the problem, with 280 vertical levels, we had to compress the vertical structural information contained in the profiles. To do so, we used a standard Principal Component Analysis and reduced that information from 280 temperature values to only 11 eigenvalues for each profile. This reduction preserved 99.88% of the profile collection structural variance, so that the RMSD between the original and compressed dataset is smaller than 0.5 °C throughout the water column and smaller than 0.1 °C below 600 m (Fig. 3).

Within this reduced-dimensional space, we trained a series of GMM and used the BIC for model selection, i.e. to determine the most appropriate number of classes to decompose the PDF. We should point here to the fact that the initial application of this popular method was not satisfactory because the size of the dataset (more than 7000 profiles in the training set) implied a log likelihood much larger than the penalty term, which was thus ineffective (see Eq. (4)). The solution to this issue was to limit

the size of the dataset used to compute the BIC to the expected number of independent profiles, 900 for the North Atlantic Ocean climatology. It is thus key to note here that for an application to a dataset with spatio/temporal correlations, one should use data mining methods with caution.

We then detailed results obtained with $K = 8$ classes. First, we computed the statistical properties of each class (activation-weighted profile statistics) and investigated their median and spread vertical structures (Figs. 9 and 10). We were able to attribute features and known water masses or thermoclines to each of the classes in a simple way. This part of our study helps answer the problematic of the identification of remarkable temperature patterns along the vertical axis. One can also note that two types of classes were found: classes with almost no zero-crossing (classes 1, 3, 4 and 7) and classes with distinct heat anomaly reservoirs (classes 2, 5, 6 and 8).

Furthermore, it is striking to note that the diversity of structures is found mostly in the warm water sphere. Indeed, cold reservoirs from class 1 and 4 have no clear peaks. On the other hand, warm reservoirs from classes 5, 6 and 8 exhibit clear maxima in the sub-surface (class 2), intermediate depths (class 8) and at depths (class 5). Class 2 is the only group of profiles with a maximum and a minimum of heat clearly identifiable near the surface and at intermediate depths respectively. On the other hand, class 7 is the only warm group of profiles without a clear vertical structure. The geographic comparison of the vertical mean $K = 5$ class (Fig. 6) distribution with the full profile $K = 8$ class one (Fig. 11) further illustrates this point. It can be seen that warm classes 4–5 of the vertical mean $K = 5$ case, decompose into classes 3–5–6–7–8 in the full profile analysis, while cold classes 1–2 of the vertical mean $K = 5$ case remain colocated with classes 1–4 without vertical structure in the full profile analysis. This lack of structure in the cold water sphere is simply due to the fact that ocean loses heat at the surface so that loss of heat, triggering vertical mixing, is similar to a loss of vertical structure. Inversely, gaining heat at the surface further increases the stratification and creates local maxima. The fact that these maxima are not always found at the same depth is due to the horizontal redistribution of heat by currents.

Secondly, we answered the question of whether classes of profiles are co-localized in space. The answer is yes. When the horizontal distribution of classes was drawn (Fig. 11) we found that, although no geographical information was included in the GMM training, each class corresponded to a specific region in the North Atlantic. This is one of the key results of our study: it fundamentally means that even if heat circulates and anomalies may be found far from the region where they are formed, the possible vertical stacks of anomalies are specific to a region. In other words, there are no distinct regions in the North Atlantic where the vertical distribution of heat is similar. Note that the vertical structure of the class spread allows us to be more precise: the vertical structure of temperature profiles combines information from water masses (where the spread is small, see for instance class 8 around -300 m where the signature of the EDW is found) but also thermoclines or transition layers (where the spread is larger, see for instance in class 2, around -100 m, the signature of the equatorial thermocline). So, as the dynamic of the ocean uniquely distributes heat in space, both vertically and horizontally, unsupervised classification of profiles is able to identify the coherent heat patterns arising from that distribution. An analysis of the profile label metric (Fig. 12) additionally suggests that horizontal fronts influence the classification probability of a profile. In other words, the probabilistic transition in the horizontal plane from one class to the other is an indication of the presence of a dynamical front.

We tested the sensitivity of the results to the number of classes and to the spatio/temporal sampling of the training set (Appendix B). The number of classes reveals a non-linear, or non-hierarchical,

behavior of the classification: as the number of classes increases, new classes were not simple subsets of *parent* classes. Using BIC method guidance, the choice of the number of classes will ultimately depend on the user's scientific question and the degree of details necessary. We also found that as long as modes are in the PDF, the classification identifies classes in a coherent way, whether they are from a small or a large region. This is quite an advantage compared to other analysis tools such as EOFs, which are domain dependent. Last, we found that our classification of temperature profiles within the North Atlantic Ocean does not depend on the season of the training set. This result, which may be surprising, relates to the fact that classes were found to identify natural oceanographic domains. It means that even if summer and wintertime profiles from the subpolar gyre are different within the thin layer above the seasonal thermocline, they remain more similar to each other than to summer or wintertime profiles from another region, such as the subtropical gyre. This result also highlights the fact that the classification outcome is related to what is used as the reference of the dataset, in our case the North Atlantic mean temperature profile. A classification of local temperature anomalies, for instance, would undoubtedly provide different classification results.

All these results suggest a wide range of possible applications for unsupervised classification of profiles. For example:

1. Figs. 5 and 6 and 9–11 are highly synthetic benchmarks to validate Ocean General Circulation Models and to analyze climatic projections with regard to horizontal and vertical heat content changes.
2. The geographic distribution of classes reveals, in a coherent and elegant way, the horizontal extent of natural oceanographic *domains* such as the subpolar gyre or the western subtropical gyre. This information can be used: (i) to define such domains with interior data only, without using un-natural rectangular boxes as is usually done, nor surface variables such as Sea Surface Height, (ii) to characterize the horizontal extent variability of a region and (iii) to objectively compare regions between different datasets, in particular those from numerical models.
3. The temporal evolution of the class representative profiles (median or mean) may be used to characterize the interior variability of a region, as an intermediate representation between domain average and local Eulerian anomalies.
4. Separating profiles from one side of a front from the other may prove useful for data validation and fronts detection.
5. Classifying a profile may help in selecting appropriate parameters for complex profile-based diagnostics such as the characterization of the permanent pycnocline (Feucher et al., 2016) or simpler ones, such as the density threshold used in the determination of the mixed layer depth.
6. The vertical structure of the class spread may be used to detect water masses and thermoclines.

To encourage these applications, the classification model parameters and results have been made publicly available (<http://dx.doi.org/10.17882/47106>, Maze, 2017) and an open source software to handle Profile Classification Model is also distributed (<http://forge.ifremer.fr/projects/pcm>).

Last, additional improvements to the method used here could be explored. There is in fact no reason to limit the classification of profiles to temperature data. Salinity could easily be added to investigate the ocean density structure and the role of temperature vs salinity in controlling stratification patterns. Moreover, rather than a classic PCA, more modern methods for dimensionality reduction, such as deep autoencoders (Hinton and Salakhutdinov, 2006) could also be used to feed the GMM classification.

To conclude, we have shown that unsupervised classification of temperature profiles, when not performed in order to complete or fill gaps within a dataset and thus using a limited number of classes and spatial information, can be used to reveal in an intuitive and physically coherent way remarkable heat patterns, their regional distributions and their climatology. It was shown that the possible stack of water masses and thermoclines are unique to a given region in the North Atlantic Ocean. This result can be used to objectively and coherently define the climatology of such regions, and possibly their variability. As an intermediate between domain averaging and local Eulerian analysis, classes of temperature profiles can provide a new framework for the analysis of the heat content climatology and variability and a synthetic benchmark to validate Ocean General Circulation Models and to analyze climatic projections with regard to horizontal and vertical heat content changes. Unsupervised classification of profiles is not widely used in physical oceanography and we hope with this study to contribute to demonstrate its potential. More generally, we believe that the increasing amount of data from in situ observations, typically from Argo and its ever increasing number of sensors, will be pivotal in the development of a data-driven model of the structure and variability of the ocean interior.

Acknowledgments

Argo data used in this study were collected and made freely available by the International Argo Program and the national programs that contribute to it (<http://www.argo.ucsd.edu>). The Argo Program is part of the Global Ocean Observing System. This study was supported by the French institute IFREMER under the Brittany regional collaboration enhancement program. We thank the editor and anonymous reviewers for their comments, which helped to improve the manuscript.

Appendix A. GMM training

GMM is an optimization problem that consists in determining the best estimate for the set of parameters $\theta = \{\lambda, \mu, \Sigma\}$ to minimize the misfit between the PDF of the model and the PDF of observations. Formally, this is stated as the maximization of the log-likelihood of observed profiles conditioned to the model parameters. The log-likelihood of the dataset, assuming independent observations, is:

$$\mathcal{L}(\theta) = \sum_{i=1}^N \log p(x_i; \theta) \quad (\text{A.1})$$

where it is explicit that the log-likelihood is a function of the set of parameters θ , not the dataset, and where $p(x_i; \theta)$ is the probability given Eq. (2) for the dataset instance x_i using parameters θ .

To train a GMM, i.e. to maximize $\mathcal{L}(\theta)$ with regard to θ , we need a dataset \mathbf{x} and a given number of components K . Several methods exist to determine the optimum parameters but the most popular is certainly the Expectation-Maximization (EM) algorithm (McLachlan and Krishnan, 2007). This is an iterative procedure that consists in looping through a two-step procedure that is guaranteed to increase the likelihood after each iteration (Dempster et al., 1977). The procedure is stopped once convergence in the parameters or likelihood values is reached. At iteration n , the two steps of the procedure are:

- the Expectation-step where one computes the posteriors $p(\mathbf{c} = k|\mathbf{x}; \theta^n)$ for all the profiles of the dataset using the n th iteration parameter values θ^n . Note that the dependence on model parameters is made explicit compared with Eq. (5).

- the Maximization-step which update model parameters θ^n so that they maximize the log-likelihood of the dataset while considering the imposed constraints on the λ_k .

The Maximization-step maximizes with regard to each of the model parameters the following function:

$$J(\lambda, \mu, \Sigma) = \sum_{i=1}^N \log \left(\sum_{j=1}^K \lambda_j p(x_i; \mu_j, \Sigma_j) \right) - \omega \left(1 - \sum_{j=1}^K \lambda_j \right) \quad (\text{A.2})$$

where ω is a Lagrange multiplier introduced to consider constraints on mixing coefficients. The detailed derivation of the optimum solution to Eq. (A.2) is beyond the scope of this introduction to GMM. It comes to the following updates, where one may recognize weighted versions of classical Maximum Likelihood estimates:

$$\lambda_k^{n+1} = \frac{1}{N} \sum_{i=1}^N p(\mathbf{c} = k|x_i; \theta^n) \quad (\text{A.3})$$

$$\mu_k^{n+1} = \frac{\sum_{i=1}^N p(\mathbf{c} = k|x_i; \theta^n) x_i}{\sum_{i=1}^N p(\mathbf{c} = k|x_i; \theta^n)} \quad (\text{A.4})$$

$$\Sigma_k^{n+1} = \frac{\sum_{i=1}^N p(\mathbf{c} = k|x_i; \theta^n) (x_i - \mu_k^{n+1})(x_i - \mu_k^{n+1})^\top}{\sum_{i=1}^N p(\mathbf{c} = k|x_i; \theta^n)} \quad (\text{A.5})$$

The model probability of the k th component λ_k^{n+1} is the dataset averaged probability of that component (given by the $p(\mathbf{c} = k|x_i; \theta^n)$ posterior values). The center μ_k^{n+1} and covariance Σ_k^{n+1} of the k th component are given by the data mean and covariance weighted by the posteriors.

All computations were conducted on a desktop computer with Matlab software and the open source Netlab library (Nabney, 2002).

Appendix B. Sensitivity analysis

B.1. Sensitivity to the number of classes

We have shown in Fig. 8 that a classic method such as the BIC can be used to determine the number of classes to train a GMM of profiles. However, the spread of the estimate does not provide a distinct single value for the best K choice, only useful guidance within a small range of possible values between 5 and 10. Therefore, it is in fact the user who will ultimately decide what is the most appropriate K for a given application.

As an indication, Fig. B13 shows the map of labels for an increasing number of components (from 3 to 9, case 8 was shown in Fig. 11). For $K = 3$ (and 2, not shown), the classification is mostly meridional and somehow captures the uni-dimensional analysis information discussed section in Section 4. For larger K , Eastern and Western regions appear distinctly. It may be noted that the classification is not hierarchical in a sense that for K larger than 4, one class is not necessarily a subset of a larger parent class. For instance, class 1 for $K = 6$ (Fig. B13D) is a mixture of some of the profiles from class 2, 3 and 4 for $K = 5$ (Fig. B13C).

In the previous section, we choose $K = 8$ because it refined the $K = 7$ description of the NAC and subpolar gyre while avoiding the $K = 9$ split of the equatorial region and thermocline vertical structure into 2 distinct eastern and western classes.

Last, one may note that even if the number of classes increases, some regions remain unchanged and identified as distinct classes. This is particularly true for 4 regions: the subpolar gyre, the western subtropical gyre south of the Gulf Stream, the Mediterranean outflow region and the lowest latitude band. This is already

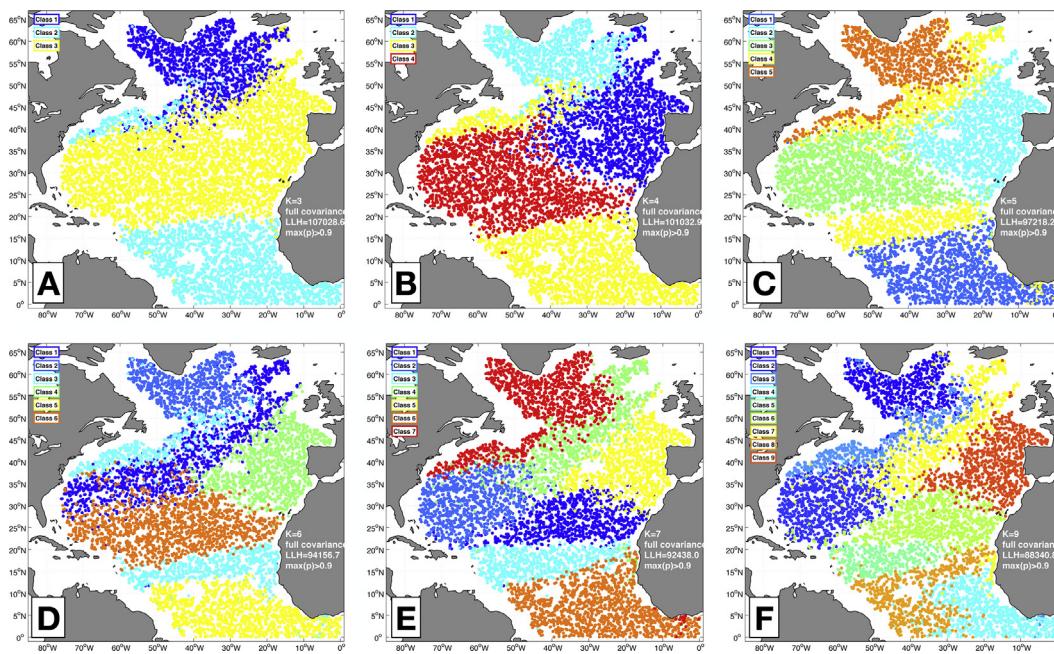


Fig. B13. Map of labels obtained with different number of classes K . Map A corresponds to $K = 3$, map B to $K = 4$, map C to $K = 5$, map D to $K = 6$, map E to $K = 7$ and map F to $K = 9$ ($K = 8$ is shown in Fig. 11).

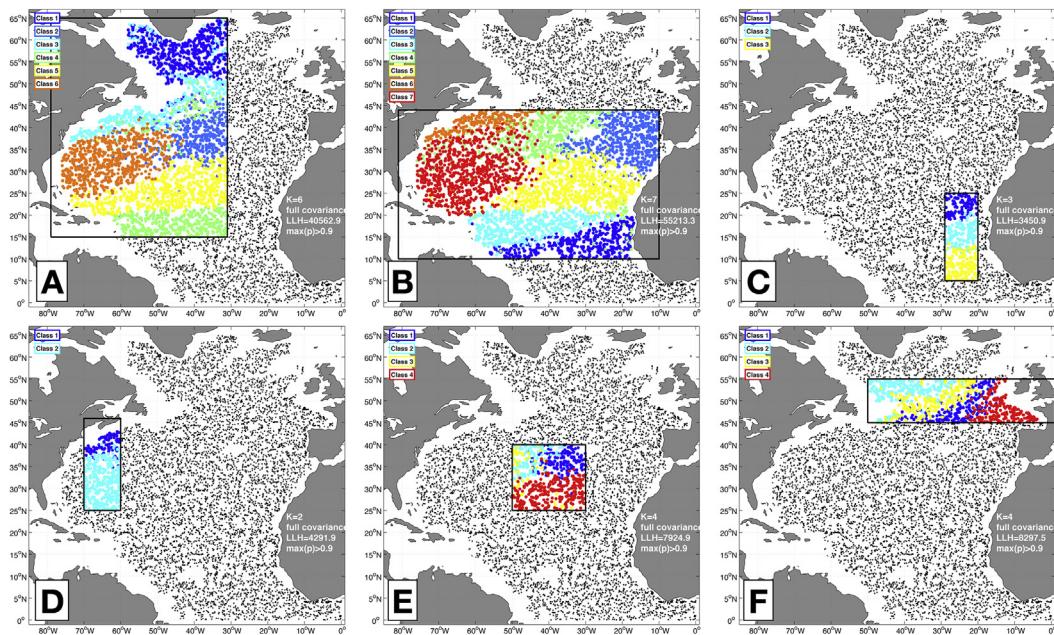


Fig. B14. Sensitivity of the classification to the choice of domain. On each maps, the black box indicates the domain used to train a new GMM, and profiles are color-coded with the class they have been attributed to. Profiles not used in the GMM training are simply shown as black dots. Note that class colors are arbitrary and only the relative distribution of classes is to be compared with the reference case shown in Fig. 11. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

apparent with $K = 4$ classes. This basically means that the core information of the climatological distribution of heat in the North Atlantic is: cold anomalies throughout the vertical in the subpolar gyre, warm anomalies near the surface and cold anomalies at depth at low latitudes and warm anomalies in the upper-mid depth in the western and lower mid-depth in the eastern side of the subtropical gyre. This may sound trivial in terms of descriptive oceanography, except that here the structure was elegantly and objectively revealed by unsupervised classification of profiles.

The detailed analysis for $K = 8$ conducted in the previous section provided additional details to refine this big picture.

B.2. Sensitivity to the domain and seasonal sampling

One can also wonder how sensitive is the classification to the geographical extent of the training set. To investigate this issue, we defined a series of rectangular sub-domains and, using the reference GMM classification map (Fig. 11), we determined the

expected number of classes to be identified in each sub-domain. We then used these number of classes to train different GMMs on each of the sub-domains and compared the classification map to the reference case. Results are shown in Fig. B14 for 6 sub-domains covering different regions and having different spatial scales, both meridionally and zonally. We found that for each sub-domain, the classification map reproduces the expected geographical distribution of labels when compared to the reference case. We conclude that for an appropriate choice of the number of classes, the GMM classification results are not particularly sensitive to the geographical extent of the training set. But one should not forget, that this will be verified as long as the sub-domain PDF will contain the modes to be identified by GMM.

We furthermore investigated the sensitivity of the classification results to the seasonal sampling of the training set. Indeed, as we used the entire collection of Argo profiles without distinction on the time of sampling, one can wonder if the classification results are not biased toward a particular season or if, for instance, deeper mixed layers from the winter time would lead to another classification outcome. We trained 4 new GMMs using the 4 subsets of profiles sampled in the four seasons: December/January/February, etc... Using $K = 8$, the 4 seasonal classifications (not shown) produce classes and maps of labels very similar to those from the complete dataset (Figs. 9 and 11). It is probable that using a higher number of classes would give GMM the ability to distinguish seasonal signals in the classification outcome.

References

- Anderson, B.D., Moore, J.B., 1979. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.
- Aretxabaleta, A.L., Smith, K.W., 2013. Multi-regime non-gaussian data filling for incomplete ocean datasets. *J. Mar. Syst.* 119–120 (0), 11–18. <<http://www.sciencedirect.com/science/article/pii/S0924796313000493>>.
- Argo, 2014. Argo Float Data and Metadata from Global Data Assembly Centre (Argo GDAC) – Snapshot of Argo GDAC as of December, 8th 2014. <http://sextant.ifremer.fr/record/7693bfc1-a695-484d-a55c-7453d54719ec>.
- Azizyan, M., Singh, A., Wasserman, L., 2014. Efficient Sparse Clustering of High-dimensional Non-spherical Gaussian Mixtures. Available from: <[arXiv:1406.2206](https://arxiv.org/abs/1406.2206)>.
- Barrier, N., Deshayes, J., Treguier, A.-M., Cassou, C., 2015. Heat budget in the North Atlantic subpolar gyre: impacts of atmospheric weather regimes on the 1995 warming event. *Prog. Oceanogr.* 130, 75–90. <<http://www.sciencedirect.com/science/article/pii/S0079661114001645>>.
- Bilmes, J.A., 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.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. Springer, New York.
- Björnsson, H., Venegas, S.A., 1997. *A Manual for EOF and SVD Analyses of Climatic Data*. Center for Climate and Global Change Research.
- Bryden, H.L., King, B.A., McCarthy, G.D., McDonagh, E.L., 2014. Impact of a 30% reduction in atlantic meridional overturning during 2009–2010. *Ocean Sci. Discuss.* 11 (2), 789–810. <<http://www.ocean-sci-discuss.net/11/789/2014/>>.
- Buckley, M.W., Ponte, R.M., Forget, G., Heimbach, P., 2014. Low-frequency SST and upper-ocean heat content variability in the North Atlantic. *J. Clim.* <http://dx.doi.org/10.1175/JCLI-D-13-00316.1>.
- Carval, T., Keeley, R., Takatsuki, Y., Yoshida, T., Schmid, C., Goldsmith, R., Wong, A., Thresher, A., Tran, A., Loch, S., McCREADIE, R., Argo Data Management Team, 12 2015. Argo User's Manual v3.2. Tech. Rep., Ifremer. <<http://archimer.ifremer.fr/doc/00187/29825/>>.
- Curry, R.G., McCartney, M.S., 2001. Ocean gyre circulation changes associated with the North Atlantic oscillation. *J. Phys. Oceanogr.* 31 (12), 3374–3400.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. *J. Roy. Stat. Soc. Ser. B (Methodol.)* 39 (1), 1–38. <<http://www.jstor.org/stable/2984875>>.
- Feucher, C., Maze, G., Mercier, H., 2016. Mean structure of the North Atlantic subtropical permanent pycnocline from in-situ observations. *J. Atmos. Ocean. Technol.* <http://dx.doi.org/10.1175/JTECH-D-15-0192.1>.
- Fiedler, P.C., 2010. Comparison of objective descriptions of the thermocline. *Limnol. Oceanogr. Methods* 8, 313–325.
- Fraley, C., Raftery, A.E., 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput. J.* 41 (8), 578–588. <<http://comjnl.oxfordjournals.org/content/41/8/578>>.
- Fukumori, I., Wunsch, C., 1991. Efficient representation of the North Atlantic hydrographic and chemical distributions. *Prog. Oceanogr.* 27 (1–2), 111–195. <<http://www.sciencedirect.com/science/article/pii/007966119190015E>>.
- García-Ibáñez, M.I., Pardo, P.C., Carracedo, L.I., Mercier, H., Lherminier, P., Ríos, A.F., Pérez, F.F., 2015. Structure, transports and transformations of the water masses in the Atlantic subpolar gyre. *Prog. Oceanogr.* 135, 18–36. <<http://www.sciencedirect.com/science/article/pii/S0079661115000506>>.
- Grist, J., Josey, S., Jacobs, Z., Marsh, R., Sinha, B., Van Sebille, E., 2015. Extreme air-sea interaction over the North Atlantic subpolar gyre during the winter of 2013–2014 and its sub-surface legacy, 1–19. <http://dx.doi.org/10.1007/s00382-015-2819-3>.
- Guinehut, S., Dhomps, A.-L., Larnicol, G., Le Traon, P.-Y., 2012. High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. *Ocean Sci.* 8 (5), 845–857. <<http://www.ocean-sci.net/8/845/2012/>>.
- Häkkinen, S., Rhines, P.B., Worthen, D.L., 2015. Heat content variability in the North Atlantic ocean in ocean reanalyses. *Geophys. Res. Lett.* <http://dx.doi.org/10.1002/2015GL063299>. 2015GL063299.
- Hannachi, A., 2007. Tropospheric planetary wave dynamics and mixture modeling: two preferred regimes and a regime shift. *J. Atmos. Sci.* 64 (10), 3521–3541. <http://dx.doi.org/10.1175/JAS4045.1>.
- Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. *Science* 313 (5786), 504–507.
- Hjelmervik, K., Hjelmervik, K., 2013. Estimating temperature and salinity profiles using empirical orthogonal functions and clustering on historical measurements. *Ocean Dyn.* 63 (7), 809–821. <http://dx.doi.org/10.1007/s10236-013-0623-3>.
- Hjelmervik, K., Hjelmervik, K., 2014. Time-calibrated estimates of oceanographic profiles using empirical orthogonal functions and clustering. *Ocean Dyn.* 64 (5), 655–665. <http://dx.doi.org/10.1007/s10236-014-0704-y>.
- Krishnamurthy, A., 2011. High-dimensional clustering with sparse gaussian mixture models.
- Kwon, Y.-O., Alexander, M.A., Bond, N.A., Frankignoul, C., Nakamura, H., Qiu, B., Thompson, L.A., 2010. Role of the gulf stream and Kuroshio–Oyashio systems in large-scale atmosphere–ocean interaction: a review. *J. Clim.* 23 (12), 3249–3281. <http://dx.doi.org/10.1175/2010JCLI3343.1>.
- Levitus, S., Antonov, J.I., Boyer, T.P., Baranova, O.K., Garcia, H.E., Locarnini, R.A., Mishonov, A.V., Reagan, J.R., Seidov, D., Yarosh, E.S., Zweng, M.M., 2012. World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.* 39 (10). <http://dx.doi.org/10.1029/2012GL051106>. n/a-n/a.
- Lloyd, S., 1982. Least squares quantization in PCM. *IEEE Trans. Inform. Theory* 28 (2), 129–137.
- Lozier, M.S., Rousseenov, V., Reed, M.S.C., Williams, R.G., 2010. Opposing decadal changes for the North Atlantic meridional overturning circulation. *Nat. Geosci.* 3 (10), 728–734. <http://dx.doi.org/10.1038/ngeo947>.
- Maze, G., 2017. A Profile Classification Model from North-Atlantic Argo Temperature Data. SEANOE. <http://dx.doi.org/10.1175/2009PO3985.1>.
- Maze, G., Forget, G., Buckley, M., Marshall, J., Cerovecki, I., 2009. Using transformation and formation maps to study the role of air-sea heat fluxes in North Atlantic eighteen degree water formation. *J. Phys. Oceanogr.* 39 (8), 1818–1835. doi.org/10.1175/2F2009JPO3985.1.
- McCarthy, G., Frajka-Williams, E., Johns, W.E., Baringer, M.O., Meinen, C.S., Bryden, H.L., Rayner, D., Duchez, A., Roberts, C., Cunningham, S.A., 2012. Observed interannual variability of the atlantic meridional overturning circulation at 26.5°N. *Geophys. Res. Lett.* 39 (19). <http://dx.doi.org/10.1029/2012GL052933>.
- McLachlan, G., Krishnan, T., 2007. *The EM Algorithm and Extensions*, vol. 382. John Wiley & Sons.
- Nabney, I., 2002. *NETLAB: Algorithms for Pattern Recognition*. Springer Science & Business Media <<http://www.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/>>.
- Ninove, F., Le Traon, P.-Y., Remy, E., Guinehut, S., 2016. Spatial scales of temperature and salinity variability estimated from argo observations. *Ocean Sci.* 12 (1), 1–7. <<http://archimer.ifremer.fr/doc/00317/42845/>>.
- Paillet, J., Arhan, M., 1996. Shallow pycnoclines and mode water subduction in the Eastern North Atlantic. *J. Phys. Oceanogr.* 26 (1), 96–114.
- Palmer, M.D., Haines, K., 2009. Estimating oceanic heat content change using isotherms. *J. Clim.* 22 (19), 4953–4969. <http://dx.doi.org/10.1175/2009JCLI2823.1>.
- Pavia, E.G., O'Brien, J.J., 1986. Weibull statistics of wind speed over the ocean. *J. Clim. Appl. Meteorol.* 25 (10), 1324–1332. [http://dx.doi.org/10.1175/1520-0450\(1986\)025<1324:WSOS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1986)025<1324:WSOS>2.0.CO;2).
- Ponte, A., Klein, P., 2013. Reconstruction of the upper ocean 3D dynamics from high-resolution sea surface height. *Ocean Dyn.* 63 (7), 777–791. <http://dx.doi.org/10.1007/s10236-013-0611-7>.
- Riser, S.C., Freeland, H.J., Roemmich, D., Wijffels, S., Troisi, A., Belbeoch, M., Gilbert, D., Xu, J., Pouliquen, S., Thresher, A., Le Traon, P.-Y., Maze, G., Klein, B., Ravichandran, M., Grant, F., Poulain, P.-M., Suga, T., Lim, B., Sterl, A., Sutton, P., Mork, K.-A., Velez-Belchi, P.J., Ansorge, I., King, B., Turton, J., Baringer, M., Jayne, S.R., 2016. Fifteen years of ocean observations with the global argo array. *Nat. Clim. Change* 6 (2), 145–153. <http://dx.doi.org/10.1038/nclimate2872>.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Stat.* 6 (2), 461–464. <http://dx.doi.org/10.1214/aoas/1176344136>.
- Stocker, T., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, A., Nauels, A., Xia, Y., Bex, V., Midgley, M., 2013. *The physical science basis*. In: *Climate Change 2013. Working Group I Contribution to the Fifth Assessment Report of the IPCC*. Cambridge University Press.
- Sura, P., 2010. On non-gaussian SST variability in the gulf stream and other strong currents 60 (1), 155–170. <http://dx.doi.org/10.1007/s10236-009-0255-9>.
- Talley, L.D., 2003. Shallow, intermediate and deep overturning components of the global heat budget. *J. Phys. Oceanogr.* 33, 530–560.

- Talley, L.D., Pickard, G.L., Emery, W.J., Swift, J.H., 2011. Chapter 4 – typical distributions of water characteristics. In: Swift, L.D.T.L.P.J.E.H. (Ed.), *Descriptive Physical Oceanography*. 6th ed. Academic Press, Boston, pp. 67–110 <<http://www.sciencedirect.com/science/article/pii/B9780750645522100046>>.
- Tandeo, P., Chapron, B., Ba, S., Autret, E., Fablet, R., 2014. Segmentation of mesoscale ocean surface dynamics using satellite SST and SSH observations. *IEEE Trans. Geosci. Rem. Sens.* 52 (7), 4227–4235.
- Thomson, R.E., Emery, W., 2014. *Data Analysis Methods in Physical Oceanography*. Newnes.
- Vallis, G.K., 2006. *Atmospheric and Oceanic Fluid Dynamics, Fundamentals and Large-scale Circulation*. Cambridge University Press.
- Wu, L., Cai, W., Zhang, L., Nakamura, H., Timmermann, A., Joyce, T., McPhaden, M.J., Alexander, M., Qiu, B., Visbeck, M., Chang, P., Giese, B., 2012. Enhanced warming over the global subtropical western boundary currents. *Nat. Clim. Change* 2 (3), 161–166. <http://dx.doi.org/10.1038/nclimate1353>.
- Yang, H., Lohmann, G., Wei, W., Dima, M., Ionita, M., Liu, J., 2016. Intensification and poleward shift of subtropical western boundary currents in a warming climate. *J. Geophys. Res.: Oceans*. <http://dx.doi.org/10.1002/2015JC011513>. n/a–n/a.
- Yang, H., Wang, F., 2009. Revisiting the thermocline depth in the equatorial pacific. *J. Clim.* 22 (13), 3856–3863. <http://dx.doi.org/10.1175/2009JCLI2836.1>.