1. Introduction

The Amundsen Sea Embayment (ASE), Antarctica, defines the drainage area of the Western Antarctic Ice Sheet, which has been losing mass, contributing to sea level rise through the Pine Island and Thwaites Glaciers (Scambos et al., 2017). These glaciers rest mainly on land that lies below sea level and often deepens toward the interior of the ice sheet, which together with increased mass loss, could lead to marine ice sheet instability and its eventual collapse (Alley et al., 2015) potentially rising sea level by up to 2.5 m within 500 years (Bamber et al., 2009; DeConto & Pollard, 2016).

This mass loss is mainly driven by warm water extending into the ice shelves’ cavities, transporting heat and consequently melting their undersides. This source of oceanic heat originates from the Antarctic Circumpolar Current north of the ASE. Warm (about 2°C) and saline (about 34.9 g/kg) Circumpolar Deep Water (CDW) can cross into the ASE forced by wind, through bottom Ekman layers and a topographically steered undercurrent, and becomes slightly cooler and fresher (Assmann et al., 2013; Thoma et al., 2008; Wåhlin et al., 2012, 2013; Walker et al., 2013). This modified CDW (mCDW) follows the bathymetry to the ice shelves while mixing with on-shelf water masses and is identified by a subsurface temperature maximum greater than 0°C between 350 and 650 dbar, with absolute salinities greater than 34.7 g/kg (Dutrieux et al., 2014; Jacobs et al., 2011; Nakayama et al., 2013). In the eastern ASE, mCDW follows the Central Troughs and Eastern Troughs (CT and ET, respectively), flowing from the ET over into the deeper CT north and south of Burke Island and then flows further south and splits up toward the Thwaites Glacier and into Pine Island Bay (PIB; Figure 1; Nakayama et al., 2013). The water column structure in the ASE varies on seasonal to interannual time scales, with large interannual variability in PIB (Dutrieux et al., 2014; Jacobs et al., 2011; Webber et al., 2017). The variability of the water mass structure in PIB can be driven by both...
local atmospheric forcing, changes in ocean circulation, heat advection and by meltwater outflow from the Pine Island (PIG) and Thwaites Glaciers (TG) (Thurnherr et al., 2014; Webber et al., 2017). Therefore, the basic processes and pathways transporting mCDW toward the glaciers are well documented, however, details are still not fully understood due to a lack of in situ observations.

Oceanographic observations in the ASE are traditionally very sparse due to ice cover limiting access and icebergs potentially catching and dragging moorings (Heywood et al., 2016). Therefore, our understanding of the spatio-temporal variability on the shelf has been severely restricted until recently. In the last 2 decades, animal-borne instruments have been proven invaluable to observe spatial variability in ice-covered oceans (Boehme et al., 2009; Charrassin et al., 2008; Roquet et al., 2013). This observational strategy was recently employed in the ASE, as part of the UK’s iSTAR Program (Heywood et al., 2016; Mallet et al., 2018) and International Thwaites Glacier Collaboration (ITGC, http://www.thwaitesglacier.org), which resulted in an increase of oceanographic profiles available from about one thousand to tens of thousands profiles, including winter data (Heywood et al., 2016). This large data set enables us to investigate the oceanographic structures of the ASE using nontraditional techniques.

Recently, machine learning classification techniques, such as Profile Classification Models (PCM; Maze, Mercier, & Cabanes, 2017), have been applied to large datasets of oceanographic profiles, revealing regimes with similar vertical structures (Jones et al., 2018; Maze, Mercier, & Cabanes, 2017; Rosso et al., 2020). PCMs have been applied to Argo float profiles in the North Atlantic (Maze, Mercier, Fablet, et al., 2017) and Southern Ocean (Jones et al., 2018; Rosso et al., 2020), capturing where regions are organized by similar properties and delimited by areas of transition and larger variability. This methodology not only helps to sort profiles into a given number of regimes, but it also provides the probability that a specific profile belongs to each one of the regimes. The probability can then be used to highlight areas with larger variability (in space and time), located between more uniform regimes (Rosso et al., 2020).

The aim of this study is to apply a PCM to temperature and salinity profiles from a large data set, in order to investigate oceanographic patterns and areas of larger spatial and temporal variability in the vertical oceanographic structure of the eastern ASE.

2. Materials and Methods

2.1. Oceanographic Data

Twelve Conductivity-Temperature-Depth-Satellite Relayed Data Loggers (CTD-SRDLs; Boehme et al., 2009) were deployed in early 2019 in the ASE, delivering 4,264 vertical profiles of temperature and salinity in our chosen domain of 110°W–99°W and 76°S–70°S. 5,881 additional profiles from 18 different CTD-SRDLs were downloaded from the Marine Animals Exploring the Oceans Pole to Pole (MEOP) project (Roquet et al., 2011, 2013; http://www.meop.net/), covering a time period from 2006 to 2014. All profiles underwent predeployment comparisons with ship-based CTD measurements and were quality controlled using methods similar to the ones used by the Argo float community (Roquet et al., 2011). Although MEOP data are provided as a quality-controlled data set, data were checked by comparison to recent ship-based measurements as well. Ship-based CTD profiles were collected during the cruises NBP19-02 (104 profiles) and NPB20-02 (40 profiles) in 2019 and 2020 respectively and are also included in the analyses. 343 CTD profiles from the World Ocean Database were extracted covering the years 1994, 2000, 2007, 2008 and 2009 (Boyer Figure 1. Overview map of the eastern Amundsen Sea Embayment from the shelf edge past Abbott Ice Shelf (AIS) and Burke Island (BI) toward the Pine Island (PIG) and Thwaites Glaciers (TG). CTD profiles collected by seals are shown as red dots, while ship-based CTD profiles are green diamonds. ET and CT indicate the Eastern and Central Troughs respectively. Blue arrows indicate the main pathways for mCDW into this area. Bathymetry is based on RTOPO-2 with contour lines at 2,000, 600 (both in black) and 500 m (blue) water depths. CTD, Conductivity-temperature-depth; CT, central trough; ET, eastern troughs; mCDW, modified circumpolar deep water.
et al., 2013). To keep information about the mCDW within the profiles, we retained only profiles with a maximum pressure of 500 dbar or deeper, and all profiles were interpolated onto 1 dbar levels.

To visualize the vertical structure across the ASE we used a mapping scheme to interpolate data onto linear sections. For each horizontal location \( m \) on a section, we selected all CTD profiles in the database within 10 km to focus on small-scale fields and gave each profile \( i \) a weight \( w_{im} \) based on the distance \( d_{im} \) between the profile and the section:

\[
    w_{im} = e^{-\frac{d_{im}}{\lambda}}
\]

The exponential decay scale is determined by the spatial scale \( \lambda \), which was set to 10 km because it encompasses the first baroclinic Rossby radius (Chelton et al., 1998). Throughout this work we use the RTOPO-2 bathymetry (Schaffer & Timmermann, 2016).

### 2.2. Profile Classification Model

Based on the previous results of Rosso et al. (2020), a Gaussian Mixture Model (GMM; Bishop, 2006) was chosen as classification algorithm for the PCM. This was applied to temperature and salinity CTD profiles, following several steps that are here briefly presented (see Rosso et al. [2020] for more details).

The GMM is a probabilistic model used to automatically classify data into regimes, based on the assumption that data are generated by a mixture of \( K \) multidimensional Gaussian distributions with unknown parameters (i.e., means and standard deviations). \( K \) indicates the number of regimes that the data can be organized in. The dimensions correspond to the data vertical levels, that is, pressure levels. We did not use all the available pressure levels, but reduced the vertical dimension by transforming the data using a Principal Component Analysis (PCA). The PCA decomposition allows to represent the data with a number of principal components, which capture the main patterns of vertical variability. We found that \( \sim 94\% (\sim 95\%) \) of the temperature (salinity) variance can be explained by the first three modes (not shown for brevity), comparable to the previous findings of Pauthenet et al. (2017, 2018) in the Southern Ocean and south Indian Ocean. The principal components of these modes are then the input for the classification algorithm.

To estimate the parameters of the GMM regimes, we trained the model using a subset of randomly selected profiles from 0.1 x 0.1° boxes to have a uniform spatial distribution in the entire domain. The model was then applied to the remaining data profiles, using the parameters found with the training data set. For each profile, the GMM computes the posterior probability to belong to each one of the regimes. The final profile classification presented below is then based on the regime with the highest probability.

### 3. Results

After quality control, we retained a total of 3,809 temperature and salinity profiles in our domain (Figure 1). The best spatial coverage is achieved in PIB and the southeastern part of the ASE in Cranton (\( \sim 74^\circ S, 103^\circ W \)) and Ferrero (\( \sim 73.5^\circ S, 104^\circ W \)) Bays. The ET has good coverage further north, up to 72.5°S. The CT has the best coverage along its slopes deeper than 500 m (Figure 1). The number of CTD profiles further north, close to the shelf edge, is limited. The best temporal coverage across all regimes is in autumn (March–May), as CTD-SRDLs make up a large number within the data set and are usually deployed at the end of the summer (Boehme et al., 2009). The Off-Shelf data are only present in the summer and autumn periods (Figure S1), mainly driven by seal behavior, as some seals arrive and leave the ASE in relation to their breeding and moulting requirements.

For the training set of the PCM, a total of 1,097 profiles (29% of the data) were randomly selected. The results are not sensitive to the choice of the profiles (not shown). To determine the optimal number of regimes \( K \), we used the Bayesian Information Criterion (BIC; Schwarz et al., 1978), calculated by running the algorithm using 10 different subsets of the data set. The BIC and visual inspection of the profile classification show an optimal range of \( K \) between 5 and 16 (Figure S2). Results with \( K = 3, 5, \) and 9 show that, while the different regimes are chosen by the PCM independently of location information, spatial patterns arise (Fig-
By limiting the PCM to three regimes, a clear distinction between on- and off-shelf data is not yet evident (Figure S3). Increasing the number of possible regimes to nine starts to explain the uncertainty rather than the pattern (Figures 2 and S3). Using five regimes is the best trade-off between distinguishing different spatial regimes on and off the shelf (Figure 2, left panel) and observing seasonal changes in the water column structure (i.e., warm and cold modes as explained below; Figure 3).

The distinction between regimes is most evident in the mean vertical profiles of conservative temperature and absolute salinity (Figure 3). Both show a change in the depth and value of the temperature maximum, as well as differences in the topmost values, especially for profiles that are similar at depth (i.e., ET-Warm and ET-Cold, CT-Warm and CT-Cold). North of the ASE (Off-Shelf), we find the highest temperatures (1.71°C) and salinities (34.86 g/kg) at 350 dbar, followed by a decrease in the value and depth of the maximum temperature and salinity from the ET (0.71°C, 34.77 g/kg at 500 dbar) into the CT (0.52°C, 34.70 g/kg at 610 dbar) (Figures 2 and 3). In each trough, a warm and a cold mode can be observed, with the warm mode having warmer temperatures, fresher salinities close to the surface and colder temperatures, lower salinities at depth and a deeper 27.70 kg/m³ isopycnal (Figure 3). The warm modes of the two on-shelf regimes (ET-Warm and CT-Warm) are mainly observed in summer and autumn, while the cold modes are mainly observed from autumn to spring (Figure S1).

The profile classification is robust, with a difference in the posterior probability between the two highest being bigger than 60% for more than 90% of the classified profiles (Figure S4). However, some areas feature profiles with a low difference between their two highest probabilities, meaning that there is an ambiguity in the classification of such profiles (Figure 2, right panel). This is not a result of missing regimes, as we find no differences when repeating this calculation for the $K = 9$ case: the ambiguity is still present in the same locations (not shown). This represents an uncertainty in the classification between two contiguous regimes, which indicates variability in the water column structure. Areas of such higher variability are found on the eastern side of the CT to the west of Burke Island and along the coast in PIB.

In order to investigate these areas of higher variability, temperature, and salinity were mapped onto two sections (Figure 4). One section (T1) runs north of Burke Island from the western side of the CT at 73.4°S, 108°W to the Abbot Ice Shelf. Here, the 27.70 kg/m³ and 27.76 kg/m³ isopycnals, which define the upper-

---

**Figure 2.** Map of regimes (left) and difference between two highest posterior probabilities (right) based on five regimes. Two sections (T1 and T2) are shown as dashed lines.
most boundary of the mCDW layer and the surface approximately 100 dbar above its core (27.80 kg/m³) respectively, slope toward shallower depths from west to east (Figure 4). The isopycnals then show a slight doming across the ET, from 104°W to 103°W. The second section (T2) is in the band of high variability in

Figure 3. Vertical mean (solid) and one standard deviation (dashed) of conservative temperature (left) and absolute salinity (right). Depths of potential density of 27.70 kg/m³ shown has horizontal lines (left). Metrics are colored by five different regimes as determined by the profile classification model.

Figure 4. Conservative temperature (top) and absolute salinity (bottom) sections T1 (left) and T2 (right) as highlighted in Figure 2. Potential density contours of 27.70 and 27.76 kg/m³ are shown in black. Locations of the Eastern and Central Troughs marked on top plot of section T1.
PIB (Figure 2, right panel). Here, is a clear doming in the isopycnals in the middle of the bay present, with the 27.70 kg/m³ isopycnal rising from about 600 dbar to 450 dbar.

4. Discussion

The PCM classifies the different water column structures in the eastern ASE, capturing their spatial patterns and potential variability. Two different regimes with warm and cold modes are identified (Figure 2). Areas with higher variability are highlighted, indicating pathways of warm waters toward the TG and PIG.

Previous models and observations have shown a thicker mCDW layer in winter than in summer (Mallet et al., 2018; Steig et al., 2012; Thoma et al., 2008), which we observe in the CT and ET modes as a shallower 27.70 kg/m³ isopycnal (Figure 3). The modes in the two troughs are clearly different in the depth at which the temperature maximum occurs, i.e. the core of the mCDW (Figures 2–4). Section T1 north of Burke Island shows that the spatial separation happens with a drop of the mCDW upper isopycnal from 400 dbar to deeper than 500 dbar at around 105.5°W (i.e., on the slope at the eastern side of the CT; Figures 2 and 4). This and the higher variability in the classification indicate the presence of a pathway of waters from the ET toward the CT. The mean profiles also show seasonal variability in the mCDW properties, with warmer and saltier conditions in winter in both troughs at about 600 dbar, similar to Mallet et al. (2018) (Figure 3). The more saline surface waters in the cold modes are the results of salinification during sea ice formation and less meltwater from the sea ice and glaciers in winter.

The southern end of the ET shows a range of profiles with higher variability in the deeper area linking toward the CT west of the Canisteo Peninsula (~73.7°S; Figure 2). This variability can be a result of mixing and convergence between the two spatial modes, as indicated by Heywood et al. (2016).

In the center of PIB, the vertical structures are linked to the ET (i.e. shallower mCDW), while toward the coast the regimes are linked to the CT (i.e., deeper mCDW; Figure 2), representative of a clockwise circulation, as in Thurnherr et al. (2014) and Heywood et al. (2016). Mallet et al. (2018) observed temporal changes in the upper limit of the mCDW arguing that this could be the result of a slowing of the gyre in autumn or a lateral movement. Section T2 shows a clear doming of the mCDW isopycnals, indicating a strong deep reaching clockwise gyre in the mean field (Figure 4). We also find high variability in the water column structure and in the classification of the profiles (Figure 2). This high variability is well known, but mainly based on summertime data (Mallet et al., 2018; Webber et al., 2017) and can therefore still be underrepresented. The variability can be a combined result of: interannual and seasonal variability of the transport of mCDW into PIB (Thoma et al., 2008); a change in the strength and location of the cyclonic gyre in PIB (Thurnherr et al., 2014) as observed by Heywood et al. (2016) and Figure 4; and, on much shorter scales, local fluxes between the ocean and the atmosphere (Webber et al., 2017). On these shorter timescales, the gyre is much more variable. The section in Figure 5 (from early March 2020) shows a very suppressed gyre only visible in the doming of the mCDW upper isopycnal below 600 dbar. This can be a result of several mechanisms, such as the slowing down of the gyre in autumn, a shift in the horizontal position of the gyre’s core, or the influence of local atmosphere-ocean-ice interactions. For example, a month before the occupation of the section, PIG calved and icebergs drifted to the west through PIB, potentially disturbing the upper 300 m of the water column and suppressing the expression of the gyre.

5. Conclusions

We showed that the PCM is a useful tool to analyze large datasets from shelf regions, enabling us to capture the pathways and variability of the mCDW in the ASE. Future work could include additional parameters for classification (e.g., dissolved oxygen), or use a different algorithm (e.g., the variational Bayesian Gaussian Mixture Model [Ghahramani & Beal, 2000]). By spatially expanding the analysis, the representation of sources, pathways and mixing products could be improved.

We note that the maximum depth of the analysis was limited by the data set, while deeper profiles would allow a better representation of the mCDW. We show that winter data are fundamental to capture the spatial-temporal scales of variability of the ASE; however, an improved spatio-temporal coverage of the data set would also help to capture the different mechanisms of variability in this region. This can be achieved by
developing an integrated observational strategy, which can be particularly important in PIB to capture the drivers and the variability of the observed gyre.

Acknowledgments
This work is from the TARSAN project, a component of the International Thwaites Glacier Collaboration (ITGC). Support from National Science Foundation (NSF: Grant 1738992) and Natural Environment Research Council (NERC: Grant NE/S006591/1). Logistics provided by NSF-U.S. Antarctic Program and NERC-British Antarctic Survey. ITGC Contribution No. ITGC-018. I. Rosso was supported by the NSF's SOCCOM project (NSF Award PLR-1425989). NOAA and NASA.

References


Schaffer, J., & Timmermann, R. (2016). Greenland and Antarctic ice sheet topography, cavity geometry, and global bathymetry (RTopo-2), links to NetCDF files. PANGAEA. https://doi.org/10.1594/PANGAEA.856844


