DEPARTMENT OF MARINE SCIENCES

DETECTABILITY OF GLOBAL TRENDS IN THE MIXED-LAYER DEPTH SINCE 1970

Lovisa Holmquist

Degree project for Master of Science (120 hp) with a major in Physical Oceanography
[MAR701, 60hp]
Semester/year: 2024
Supervisor: Dr. Fabien Roquet
Examiner: Dr. Sebastiaan Swart

Acknowledgements

Firstly, I would like to thank my supervisor, Fabien, for guiding me throughout the entire project. It is easy to get lost in the coding and lose sight of the objective, but you encouraged me to learn new skills while steering me through each step, always reminding me of the bigger picture. Our meetings ranged from troubleshooting Python issues and discussing statistical methods to watching YouTube clips of drummers in pink animal costumes—a perfect mix, if you ask me. I am truly grateful that you agreed to take me on as a student once again, and I believe the thesis turned out great. Thank you; this would not have been possible without you!

I also want to extend my gratitude to all the PhD students (and now post-docs), Birte, Estelle, Johan, Romain, Theo, and Vincent, for assisting me with all my “weird” Python questions. I think you are all just as excited as I am that this project is now officially over.

Further, thank you to everyone at Natrium. Sometimes, relying solely on my computer during the thesis—through analysis and writing—was isolating. You all helped me stay sane and avoid getting lost in a sea of text. Sometimes, all it took to keep going was a lunch with Axel, a playful fistfight with Elias, a Disney moment with Simon, or a chat with the bacteria expert herself, Hedda. And Sofia, I’m incredibly grateful that you took on the role of my manager, luring me to the university with cakes and baked goods whenever my motivation waned.

Lastly, I wish to thank Fionn, my wonderful boyfriend, who still might not fully understand what I’m doing. Thank you for tolerating my weighted blanket during stressful times, even though I know you’re not a fan and all you wanted was a peaceful night’s sleep. I appreciate your unwavering belief in me when I doubted myself. Your patience is commendable, especially when I’ve turned into a human burrito demanding snacks after a long day that has been just as long for you. I promise to return the favour by letting you pick the next movie for date night, and yes, you don’t have to choose a Disney movie!

Plain Language Summary

The upper layer of the ocean is nearly homogenous, with the same physical properties, and is often referred to as the Mixed Layer (ML). This layer is important because it is the only link between the deep ocean and the atmosphere. Thus, it has an impact on the overall climate by controlling heat and gas exchange. The depth of this layer, the Mixed Layer Depth (MLD), varies based on factors like season and location, and while it is often assumed that global warming should make the MLD shallower by increasing ocean stratification (layering the ocean more distinctly by density), some recent studies have suggested that it might be deepening, especially during summer months.

Historically, estimating global MLD trends has been challenging due to sparse and uneven sampling across the ocean, with most data collected in the northern hemisphere focusing on coastal areas. However, with the advent of autonomous profiling floats ocean monitoring expanded dramatically after 2000. These floats measure temperature and salinity continuously, providing data that is more frequent and covers more remote areas than previous methods, such as CTD (Conductivity-Temperature-Depth) casts. CTDs, deployed primarily along coastlines, have provided reliable data since the 1960s but lack the wide spatial and temporal range of floats. Although profiling floats provide valuable insights, they also come with their own issues, such as sensor errors in salinity and pressure, that can skew MLD measurements. This study examines the statistical robustness of observed global MLD trends, as well as the impact of these differences in sampling methods.

The results reveal that MLD varies seasonally, reaching its greatest depths in winter and shallower levels in summer, which also agrees with previous literature. When the data from profiling floats is excluded, MLD is generally shallower, especially in winter. This suggests that data from profiling floats estimates deeper MLDs than measurements from traditional CTD casts, potentially due to sensor-related issues or because the floats sample differently. Over time, MLD measurements indicate a general deepening when all data is included, but when float data is removed, this trend is non-prominent, and the MLD slightly decreases or could even be considered constant. These differences show that profiling floats have a strong effect on MLD trends, making it essential to account for their biases in any long-term climate analysis.

The findings also highlight how sampling imbalances—more data from the northern hemisphere than the southern hemisphere and a concentration of observations in well-studied areas like the North Atlantic—affect our understanding of MLD trends. After 2000, increased data from the Southern Hemisphere helps improve the balance, but gaps in coverage remain, particularly in the Southern Ocean, which plays a key role in global climate dynamics. Therefore, any conclusions of MLD changes need to carefully consider these regional and instrument-related biases, as the increased deepening trends found in other studies may not represent a global change but rather reflect these sampling challenges.

The study ultimately concludes that, despite of increased data availability and improved instruments, MLD trends still cannot be definitively characterised on a global scale, especially not before 2000. The seasonal and regional variability in MLD, combined with the potential biases from newer sampling methods, means that caution is necessary when interpreting MLD changes as indicators of climate-driven oceanic transformations. Thus, while profiling floats have revolutionised ocean observations, they bring their own complexities that must be thoroughly accounted for in climate studies.

Supervisor: Fabien Roquet
Degree project, 60 hp in Physical Oceanography, 2024
Inst. for Marine Science, Gothenburg University

Abstract

The upper layer of the ocean is nearly vertically homogeneous and is often referred to as the Mixed Layer (ML). Changes in the ML can significantly influence climate and its variability, as it serves as the only link between the ocean interior and the atmosphere. The thickness of this layer, known as the Mixed Layer Depth (MLD), determines rates of exchange of heat and carbon, and strongly affects the primary production. It has recently been claimed in the literature that the MLD has increased globally in summertime. This is a surprising result, not only because global warming should act to decrease the MLD. But even more so because the MLD is known for its complex, highly regional, variability which remains poorly sampled. The introduction of autonomous vehicles, such as Argo floats, fundamentally changed the spatial and temporal structure of sampling between 2000 and 2005. This study investigates the interannual variability of the MLD, assessing potential trends since 1970 and their statistical robustness. Using the EN4.2.2 dataset, a co-location approach revealed a skewness in profiling floats (PF) compared to CTD data, with PF estimates systematically deeper than those from CTD. While it is challenging to establish whether this discrepancy is due to sensor biases (e.g. salinity or pressure) or to spatial heterogeneity in sampling methods, the potential bias is of similar order of magnitude as the observed deepening trend. Contrary to claims in the literature, no trend in MLD can be considered robust prior to 2000, even in relatively well-sampled regions such as the North Atlantic.

Table of contents

1. Introduction

Studies on global warming and its impacts on the planet are more important than ever. Greenhouse gases drive Earth's radiative imbalance, with the ocean playing a crucial role by storing over 90% of the excess heat in the climate system. However, this storing of heat is not uniform, neither across different regions nor through the ocean's depths. Certain areas take up more heat than others, which has significant implications for the upper layer of the ocean (Somavilla et al., 2017). This upper layer is nearly vertically homogenous, sharing the same physical properties, and it is, therefore, often referred to as the Mixed Layer (ML) (Dong et al., 2008; Kara et al., 2000). It is the only link between the ocean interior and the atmosphere, and by controlling heat and carbon exchange rates, changes in the surface layers have a strong influence on the climate and its variability. The Mixed Layer Depth (MLD) is an indication of how much water directly interacts with the atmosphere and has temporal and spatial variability (de Boyer Montégut et al., 2004; Dong et al., 2008; Kara et al., 2003). The deepest MLs are found in certain areas of the subpolar ocean during the annual winter months, while the shallowest are observed during summer, and the difference between the two hemispheres can reach a magnitude of hundreds of meters (Caneill & Roquet, 2023; de Boyer Montégut et al., 2004; Kara et al., 2000). Processes such as surface forcing and lateral advection are some of the elements affecting the temporal fluctuations of the MLD (Alexander et al., 2000; de Boyer Montégut et al., 2004; Kara et al., 2000). On the contrary to the atmospheric boundary layer, the variability of the MLD is not as well understood (Kara et al., 2003). Investigating and describing changes is therefore of high importance in Earth climate sciences to understand the oceans' contribution in climate, both when it comes to climate change and impacts on ecosystems (Buongiorno Nardelli et al., 2017; Diehl, 2002; Dong et al., 2008).

Hydrographic sampling techniques have rapidly progressed alongside technological advancements. Over the past two decades, innovations like autonomous vehicles have substantially increased the amount of collected data by reducing costs and improving access to previously under-sampled regions (Boyer et al., 2018). Autonomous profiling floats (Figure 1A) have become a central component of in situ ocean observation, largely due to the Argo program's implementation (Barker et al., 2011; Liu et al., 2024). The vehicles use oceanographic sensors to measure vertical profiles of oceanographic variables like pressure, temperature and conductivity (for calculating salinity), regardless of sea or weather conditions, by passively floating in the water column. In comparison, CTD (Figure 1B) casts, which have been a well-used measuring method since the 1960s, mainly sample in coastal regions of the northern hemisphere, which makes profiling floats a valuable add-on in sparse or none, sampled regions (Boyer et al., 2018). However, challenges due to sensor problems, such as salinity drift (Boyer et al., 2018; Liu et al., 2024) or pressure error drift (Barker et al., 2011), introduce data problems that need to be corrected (Boyer et al., 2018) and taken into consideration when doing global change research (Barker et al., 2011).

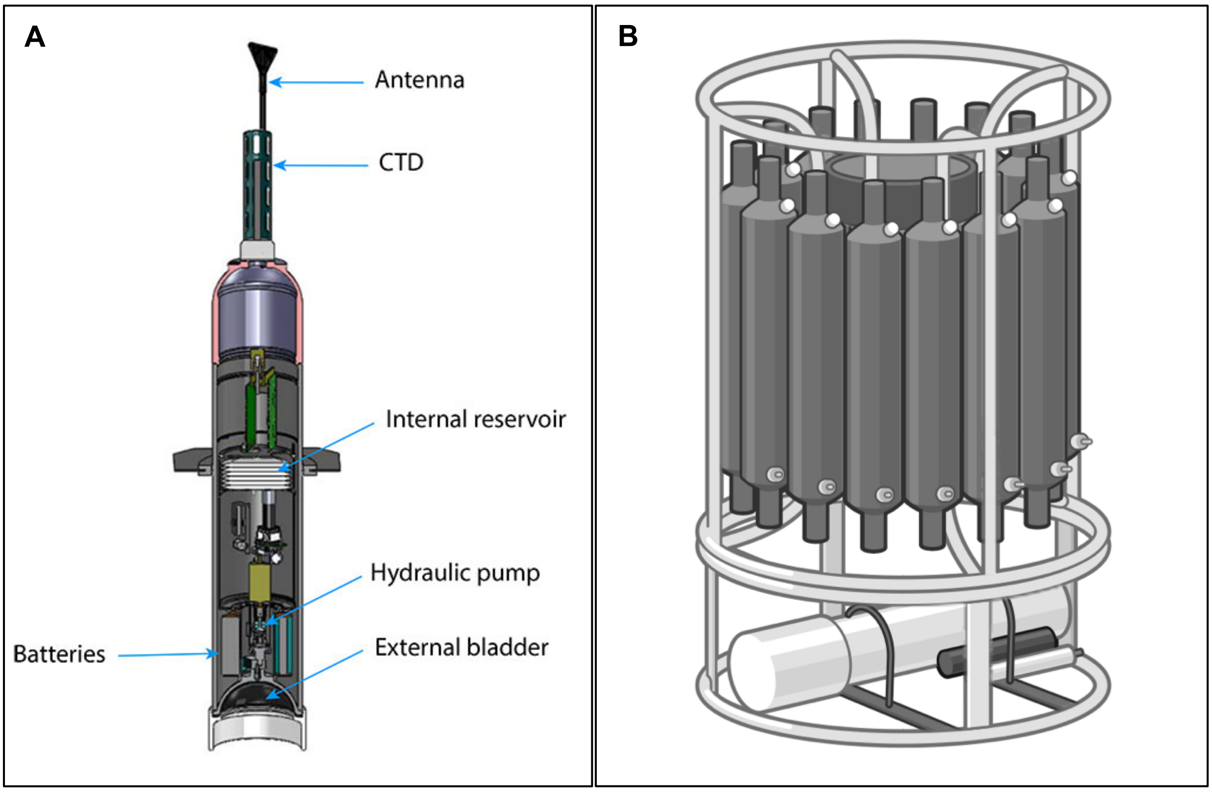

Figure 1. (A) Illustration of an Argo float labelled with hardware components. (B) Illustration of CTD/Rosette.

As mentioned earlier, the heterogeneity of heat storage affects the MLD. In the context of global warming, intensified surface warming increases stratification in the pycnocline, weakening vertical exchanges between surface and deep waters, which leads to a shoaling of the MLD. This process is accelerated by more rapid surface warming compared to deeper layers, along with freshwater inputs from ice melt and precipitation, further enhancing surface stratification. As a result, shoaling of the mixed layer reduces air-sea gas exchange, limits deep-ocean ventilation, and diminishes biological productivity, all critical factors in regulating the global climate system (Sallée et al., 2021). Another key trend over the last 30 years has been the poleward intensification of the westerly winds in the Southern Ocean and the poleward shift of the jet stream in the Northern Hemisphere, both of which may cause changes in wind-driven divergence and ocean circulation. Such wind-driven divergence can offset or even reverse the effects of surface warming by competing with it (Somavilla et al., 2017). As a result, the interplay between surface warming and wind-driven divergence complicates the overall picture, making it uncertain whether there is a consistent deepening or shoaling trend of the MLD. Therefore, the long-term trends in MLD remain inconclusive and may vary regionally depending on the relative dominance of these competing forces.

Research on changes in upper ocean stratification is often focused on a regional scale, or over fixed depth ranges, leading to knowledge gaps on the global effects of climate change (Sallée et al., 2021). A previously made study by Sallée et al. (2021) claimed a global deepening of the mixed layer depth during summertime, at a mean 3% per decade rate (typically 5-10 m per decade). This comes as a surprise as global warming tends to increase the upper ocean stratification, due to increased freshwater forcing in high latitudes and surface warming in the lower latitudes (Roch et al., 2023; Somavilla et al., 2017). However, uncertainties are expected due to the paucity and heterogeneity of the sampling. Different regional dynamics are expected due to variable forcings and stratification types (Sérazin et al., 2023), thus it is puzzling that this deepening trend would be observed so consistently over so many different regions. Therefore, this study aims to investigate the interannual variability of the mixed layer depth, assessing potential trends over the last 50 years and assess their statistical robustness by using the same EN4 dataset as Sallée et al. (2021). Based on the EN4.2.2 dataset, global trend analyses of the mean mixed layer depth (MLD) were conducted. A co-location approach highlighted a skewness in profiling floats compared to CTD, potentially due to salinity drift in the conductivity sensor in combination with temporally and spatially inhomogeneous sampling coverage. The report is structured as follows: First, the two datasets used in this study are presented. Second, the method for trend and potential bias analysis is described, followed by the results, which underscore the potential bias due to sensor drift. Third, the findings are discussed in context, and finally, the report concludes with a summary of key insights.

2. Material and Methods

2.1 Data

2.1.1 EN4

The EN4.2.2 (Good et al., 2013) dataset was obtained from *https://www.metoffice.gov.uk/hadobs/en4/*. The dataset contains measurements from all instruments used in the World Ocean Database, excluding instruments only measuring at the ocean surface. It includes profiles from Argo and other profiling floats, bathythermographs, bottles, CTDs, drifting buoys, gliders, moored buoys, STDs, instruments attached to animals, undulating oceanographic recorders and XCTDs (Sun, 2010). The dataset uses the Gouretski and Cheng (2020) MBT, and Gouretski and Reseghetti (2010) XBT corrections. In this study, data from 1970 to 2019 was used.

2.1.2. ETOPO1

The bathymetry dataset used in this study was obtained from NOAA (Amante & Eakins, 2009). Prior to analysis, the data was regridded to a 1/10° resolution.

2.2 Methods

All pre-processing was done using Snakemake. The script used for the initial data processing was built on a previous work by Romain Caneill (2023), with minor adjustments for this project. His methods were as follows. Firstly, a longitude/latitude grid with a 1° resolution was created. All timestamps, longitude and latitude measurements were extracted from the EN4 dataset only to select data in the given spatial and temporal range (90°N to 85°S and 1970 to 2019). Only the profiles with a good quality control flag (=1) were used. Secondly, all profiles consisting of only NaNs were removed. Using the temperature and conductivity measurements the conductive temperature and absolute salinity were computed using the Gibbs SeaWater toolbox. The same toolbox was used to calculate the potential density. The Mixed Layer Depth (MLD) was then computed using a density threshold criterion of 0.03 kgm^{-3} referenced to 10 m depth (de Boyer Montégut, 2004; Sallée et al., 2021). The equation was as follows:

$$MLD = \rho_{z_{10}} - \rho_{z_i} \geq 0.03\ kgm^{-3}.$$

This was done for all profiles. If the MLD was equal to the depth of the profile it was replaced with a NaN. Lastly, the pre-processing was completed by saving the MLD output as individual monthly NetCDT files for further analysis.

Additionally, all performed analyses were done using Python (version 3.11.3). Using the ETOPO1 bathymetry dataset, each profile was assigned a max depth based on the closest longitude and latitude coordinates. All coastal data, measurements taken in areas with a depth shallower than 200 m, were removed to focus on the deep ocean. All profiles were thereafter grouped based on hemisphere (northern or southern), year and month. A 1° by 1° grid was also created to get a spatial count per latitude. The percentage of total observations in the northern and southern hemispheres was calculated for each year and plotted to show changes in data distribution over time.

As shown in Figure 2, a high temporal variability in data quantity can be observed. Up until the early 2000s the great majority of all observations are made in the northern hemisphere, as also visible in Figure 3, where observations from the northern hemisphere contribute to approximately 85% of all collected data. Additionally, the measurements are exclusively TESAC (Temperature, Salinity and Current) data collected from CTD, STD, bottles and others. After the 2000s there is a remarkable increase in the number of observations, in both hemispheres, mainly due to the introduction of profiling floats. The inhomogeneity of sampling methods is also shown in Figure 4. The measurements in the Southern Ocean (SO) are mainly from XBT, MBT, APB (animal mounted data), Micro-BT and profiling floats, areas around the equator by buoys, profiling floats and TESAC, and the north TESAC and profiling floats. Measurements are also highly concentrated around 30°N to 40°N, with quantities doubled compared to south of the equator.

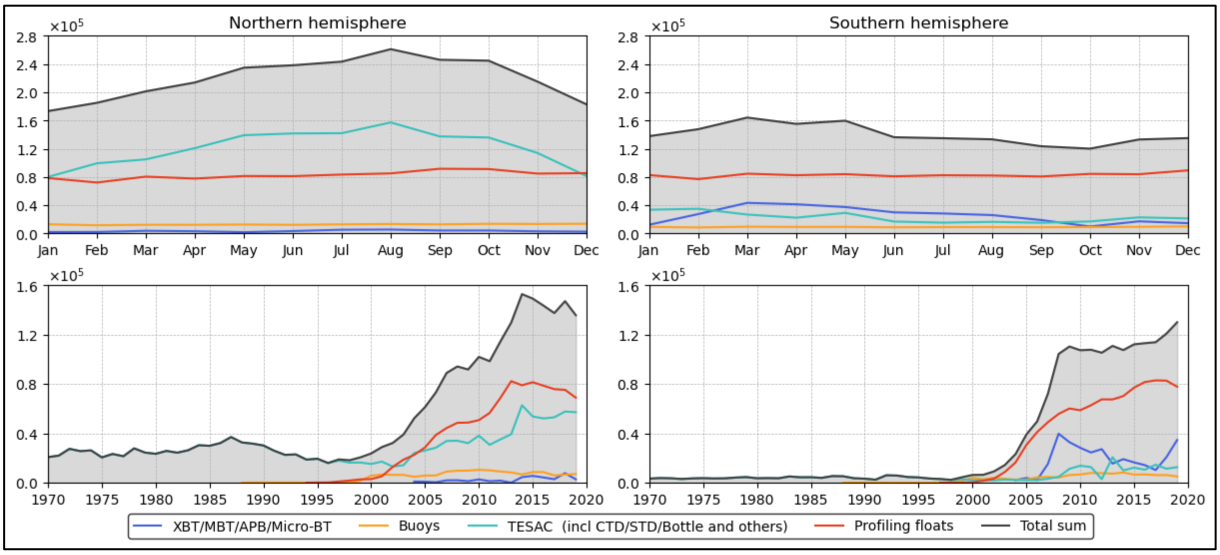

Figure 2. The number of profiles acquired in the northern (left panel) versus southern hemisphere (right panel) each month (top), and per year (bottom). Data obtained from XBT/MBT/APB/Micro-BT in blue, from buoys in orange, TESAC in turquoise, profiling floats in red and the total sum in black.

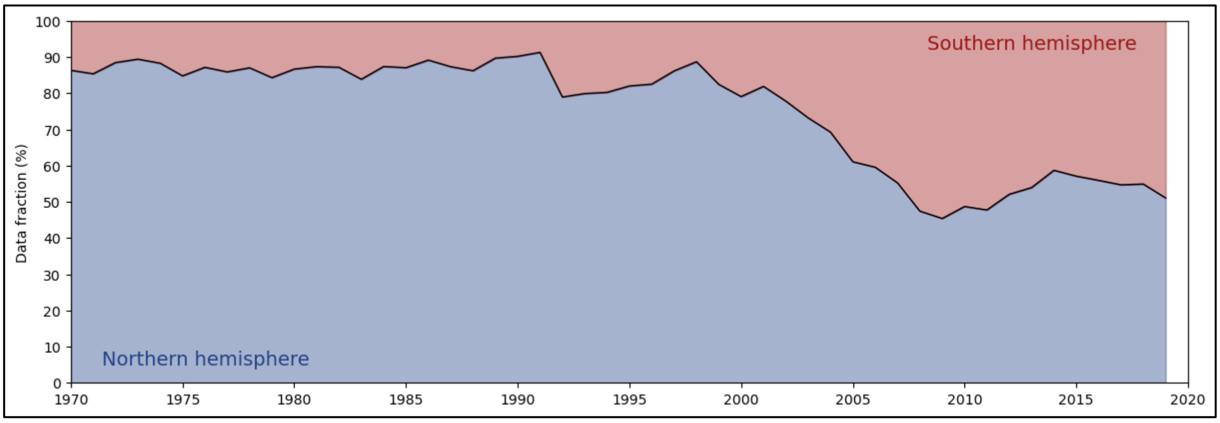

Figure 3. Fraction of the number of profiles acquired in the northern (blue) versus southern (red) hemisphere over time.

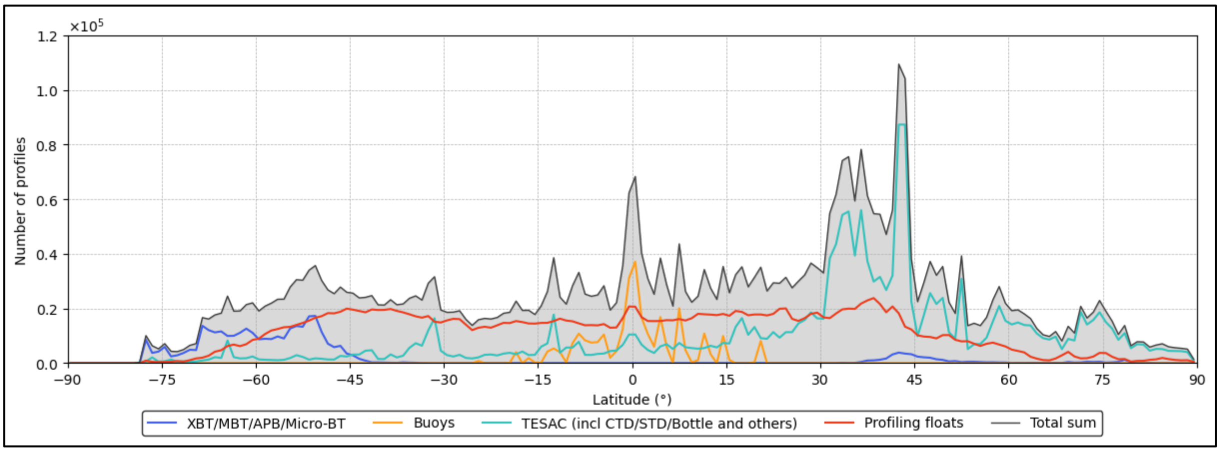

Figure 4. The number of profiles acquired per degree of latitude. Data obtained from XBT/MBT/APB/ Micro-BT in blue, from buoys in orange, TESAC in turquoise, profiling floats in red and the total sum in black.

To analyse interannual changes a frequency conversion was applied before computing both the yearly and monthly mean MLD. Two additional datasets–one excluding profiling floats (EPF) and one with only profiling floats (OPF)–was created for comparative reasons. Statistical analysis such as anomaly, linear least-squares regression and standard deviation was performed using SciPy. The datasets were further divided into two periods: period 1 (1970–2004) and period 2 (2005–2019), to highlight differences during the most significant increase in data availability. Separate linear regressions were computed for each period. All plots were created using Matplotlib.

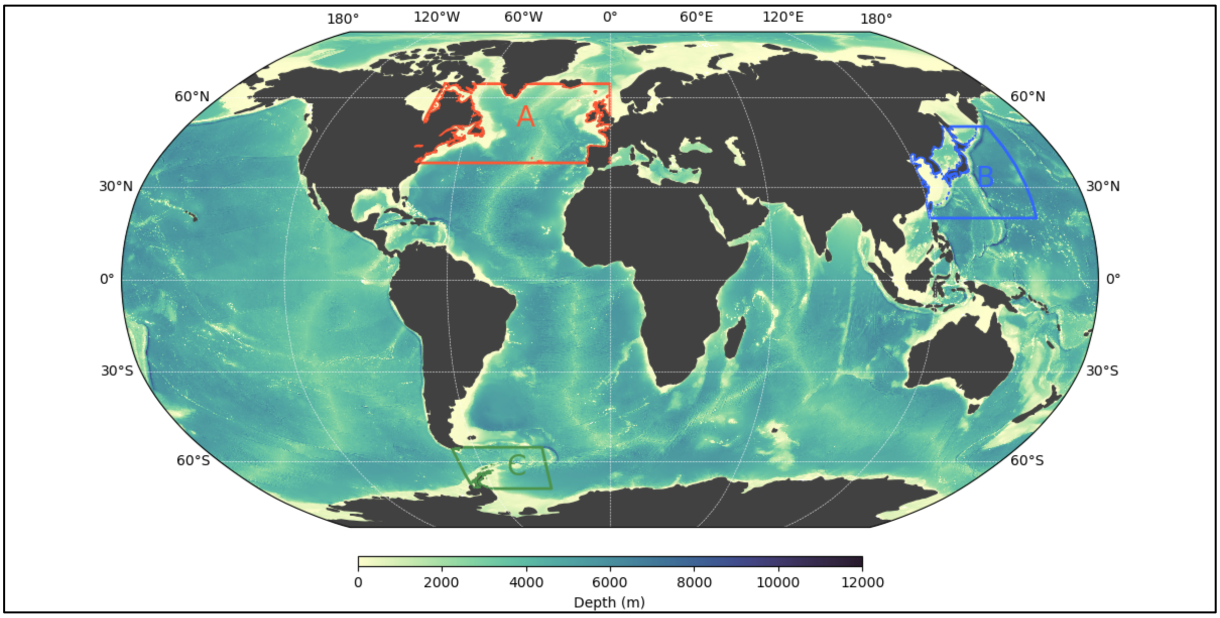

Figure 5. Bathymetry plotted at a 0.1°x0.1° resolution with the three chosen subregions highlighted. Region A (North Atlantic) is in red, Region B (Northeast Pacific) is in blue and C (Drake Passage) is in green.

Three subregions (A, B and C) were selected for investigations of trends on regional scales, see Figure 5. The criteria for the chosen regions were to have as high spatial homogeneity as possible and measurements from all 5 decades. The final choices were; Region A, in the North Atlantic (80°W to 0° and 38°N to 65°N), Region B, in the Northwest Pacific (120°E to 160°E and 20°N to 50°N) and Region C, at the Drake Passage in the Southern Ocean (30°W to 70°W and 55°S to 70°S). The same methodology as described above was thereafter used, with computations of monthly and yearly mean MLD anomaly as well as linear regressions for the whole time series, period 1 and period 2. However, it is worth mentioning that the data quantity differs between regions and especially region C required interpolation for missing months, which might have an impact on the results.

For comparison analysis, first, a co-location was made between the PF measurements and the closest measurement from any other instrument in both time and space. The co-location was done by applying a spatiotemporal distance using the query function from Pandas. A threshold of 1 for the spatiotemporal distance was applied, along with a maximum difference of 2° in both longitude and latitude. Second, for more relevant comparisons, only the co-location pairs with CTD as the comparing sampling method were chosen, resulting in a dataset with co-location pairs that have spatiotemporal values lower than or equal to 1, a maximum spatial difference of 2°, and CTD as the comparing sampling method to the PF. Thereafter, the difference for each PF-CTD pair was computed. The overall mean difference and standard deviation were analysed with an additional two-tailed paired t-test for testing significance levels ($\alpha = 0.05$).

Additionally, the co-location output was used to analyse the spatial distribution of the potential MLD bias. The filtered dataset mentioned above was binned to a resolution of 2° x 2°. The mean MLD difference was computed for each bin and plotted on a map. The zonal mean, as well as the global mean, was also calculated and plotted.

3. Results

3.1 Intra-annual variability

The Mixed Layer Depth (MLD) varies seasonally, with deeper depths during the annual winter months, January, February and March in the northern hemisphere and July, August and September in the southern hemisphere, as seen in Figure 6. A similar pattern is observed in the global monthly mean but with a smaller depth range. For both hemispheres, the two datasets follow each other well but with shallower depths during the annual winter months when data from profiling floats (PF) is excluded. This difference is not observed during the summer months. However, looking at the global mean, the biggest difference is observed during the annual winter months (summer months) of the southern hemisphere (northern hemisphere), from June to October.

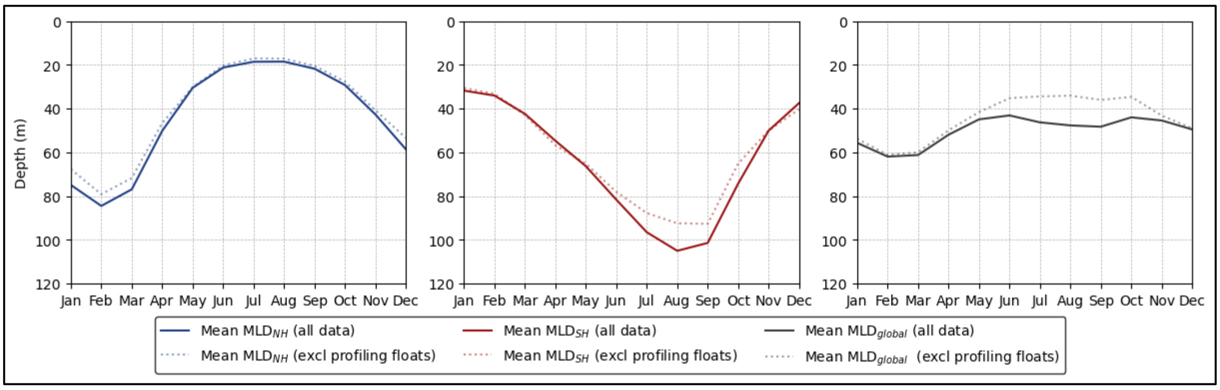

Figure 6. Monthly mean Mixed layer Depth over the whole period (1970 to 2019) for the northern hemisphere (left), southern hemisphere (middle) and globally (right). Mean depths calculated from all observations as a solid line, and mean depths calculated excluding profiling floats as a dotted line.

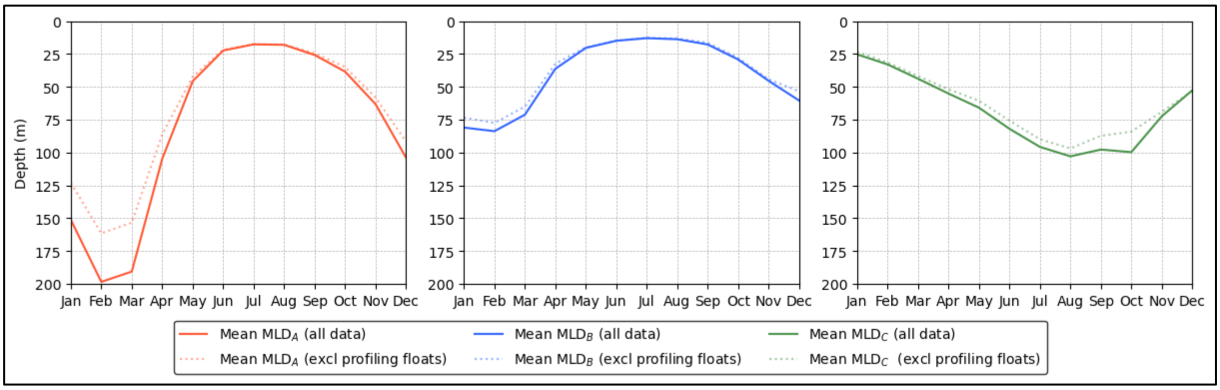

Figure 7. Monthly mean Mixed layer Depth over the whole period (1970 to 2019) for Region A (left), Region B (middle) and Region C (right). Mean depths calculated from all observations as a solid line, and mean depths calculated excluding profiling floats as a dotted line.

Similarly, the MLD varies seasonally in the three subregions—Region A, Region B, and Region C—as shown in Figure 7, with shallower depths during the summer months and deeper depths in winter. The two datasets, again, follow similar trends, with shallower depths during winter months, when PF data is excluded across all three regions. As observed in the global dataset, and when divided into the northern and southern hemispheres, this difference is not apparent during the summer months. However, the largest difference is found in Region A.

3.2 Interannual variability and trend detection

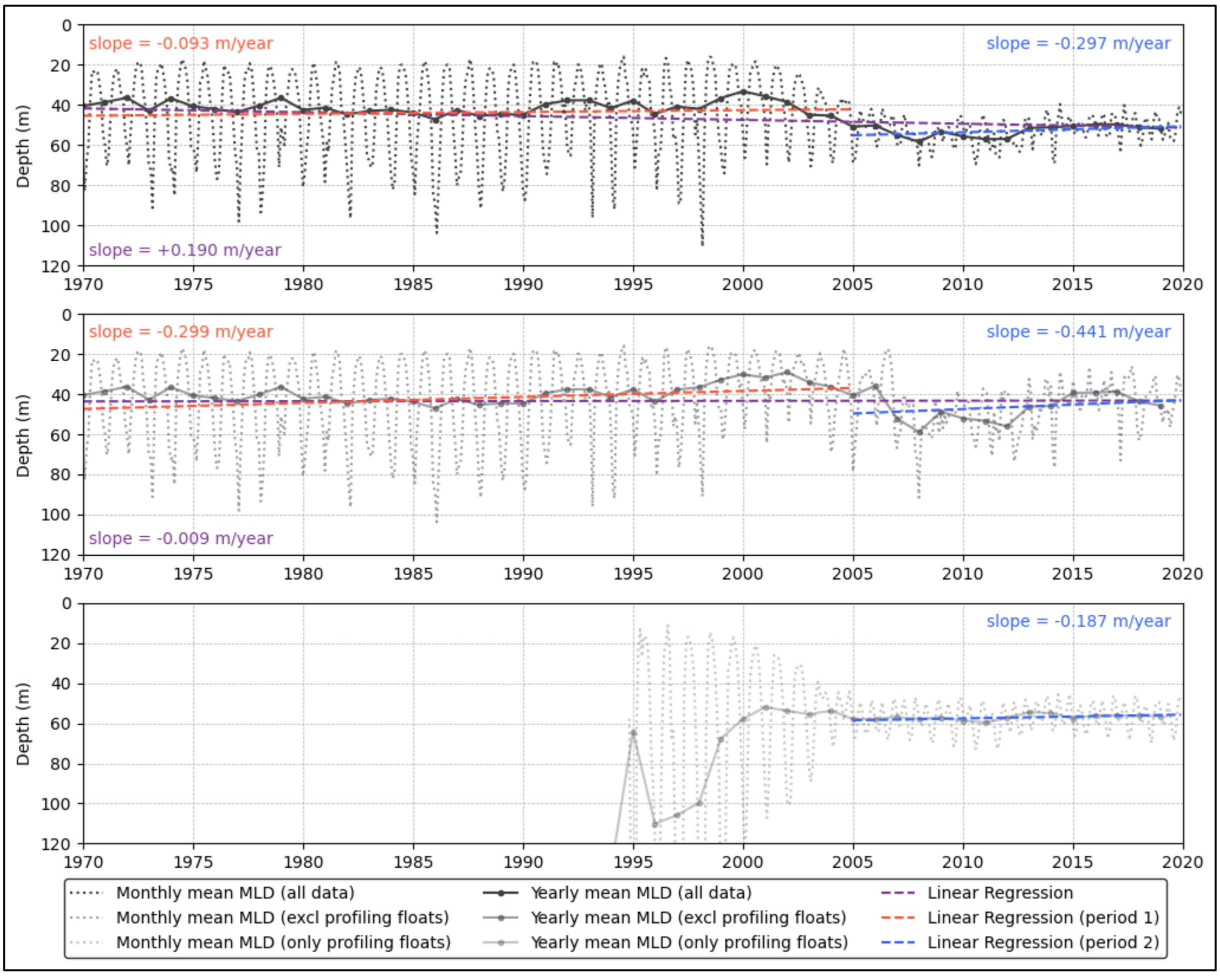

Figure 8. Mean Mixed Layer Depth (MLD) over time with all data included (top), mean MLD with data excluding profiling floats (middle) and mean MLD with only profiling floats data (bottom). The monthly mean is represented as a dotted line and the yearly mean as a solid line. The linear regression for the whole time series is shown as a dashed purple line with the slope printed in the bottom left corner. The linear regression for period 1 is shown as a dashed red line, and the linear regression for period 2 as a dashed blue line, with respective slopes printed in matching colours.

A seasonal variability is equally observed for both the all observations (AO) and the exclusion of profiling floats (EPF) dataset, as seen in Figure 8. However, the range of the variability decreases from the early 2000s until 2005, and after 2005 the variability range is nearly constant. Remarkably, the post-2005 variability range seems to have a larger magnitude in the EPF dataset than in the AO. The largest variability, however, is observed in the dataset only consisting of data obtained from profiling floats (OPF), especially up until 2005. The deepest monthly mean maximum of the MLD occurred in March 1998 for AO (111 m), in February 1986 for EPF (104 m), and in March 1997 for OPF (498 m). The shallowest monthly mean was recorded in August 1994 for both AO and EPF (16 m and 11 m, respectively) and in August 1996 for OPF. The highest yearly mean MLD was observed in 2008 for the AO and EPF datasets (58 m and 59 m, with a 1 m difference) and in 1994 for the OPF dataset (133 m). Conversely, the year with the shallowest mean depth was 2000 for AO (33 m), 2002 for EPF (29 m), and 2001 for OPF (52 m).

The calculated linear regression had a slope of +0.190 m per year in the AO dataset and -0.009 m per year in the EPF dataset, corresponding to a global deepening of the mean MLD over time when all data were included and an essentially constant mean depth, or a slight shoaling, when PF measurements were excluded. When divided into period 1 and period 2, the linear regression showed a slope of -0.093 m per year in period 1 in AO, corresponding to a relatively stagnant MLD with a shoaling of less than a meter per decade. In EPF, the linear regression had a slope of -0.299 m per year, implying a shoaling of ~3 m per decade. In period 2, the slope was -0.297 m per year in AO, -0.441 m per year in EPF, and -0.187 m per year in OPF, all indicating a shoaling, with the greatest magnitude in EPF, ~4.5 m per decade.

For the MLD anomaly (Figure 9), where the seasonal variability has been removed, the highest peaks were observed in March 1998 for AO (+42 m), June 2012 for EPF (+39 m), and March 1997 for OPF (+393 m). The lowest peaks occurred in January 1993 for AO (-24 m), March 2002 for EPF (-31 m), and February 2012 for OPF (-56 m). In terms of yearly means, the maximum anomaly was recorded in 2008 for both AO (12 m) and EPF (16 m), and in 1997 for OPF (68 m). The minimum anomaly occurred in 2000 for AO (-11 m), 2002 for EPF (-14 m), and 2013 for OPF (-12 m).

The calculated linear regression had a slope of +0.201 m per year for AO and +0.002 m per year for EPF. Similar to the mean MLD, a global deepening of the MLD was observed when all data were included and a relatively constant MLD when PFs were excluded. When divided into period 1 and period 2, the linear regression showed a slope of -0.071 m per year for AO and -0.277 m per year for EPF in period 1. Both correspond to a shoaling, with a magnitude of ~0.7 m per decade when all data were used and ~2.8 m per decade when PFs were excluded. In period 2, the slope was -0.178 m per year for AO, -0.322 m per year for EPF, and +0.058 m per year for OPF. Both AO and EPF showed shoaling, with a greater magnitude when PFs were excluded. However, an opposite trend was observed when only PF data were used.

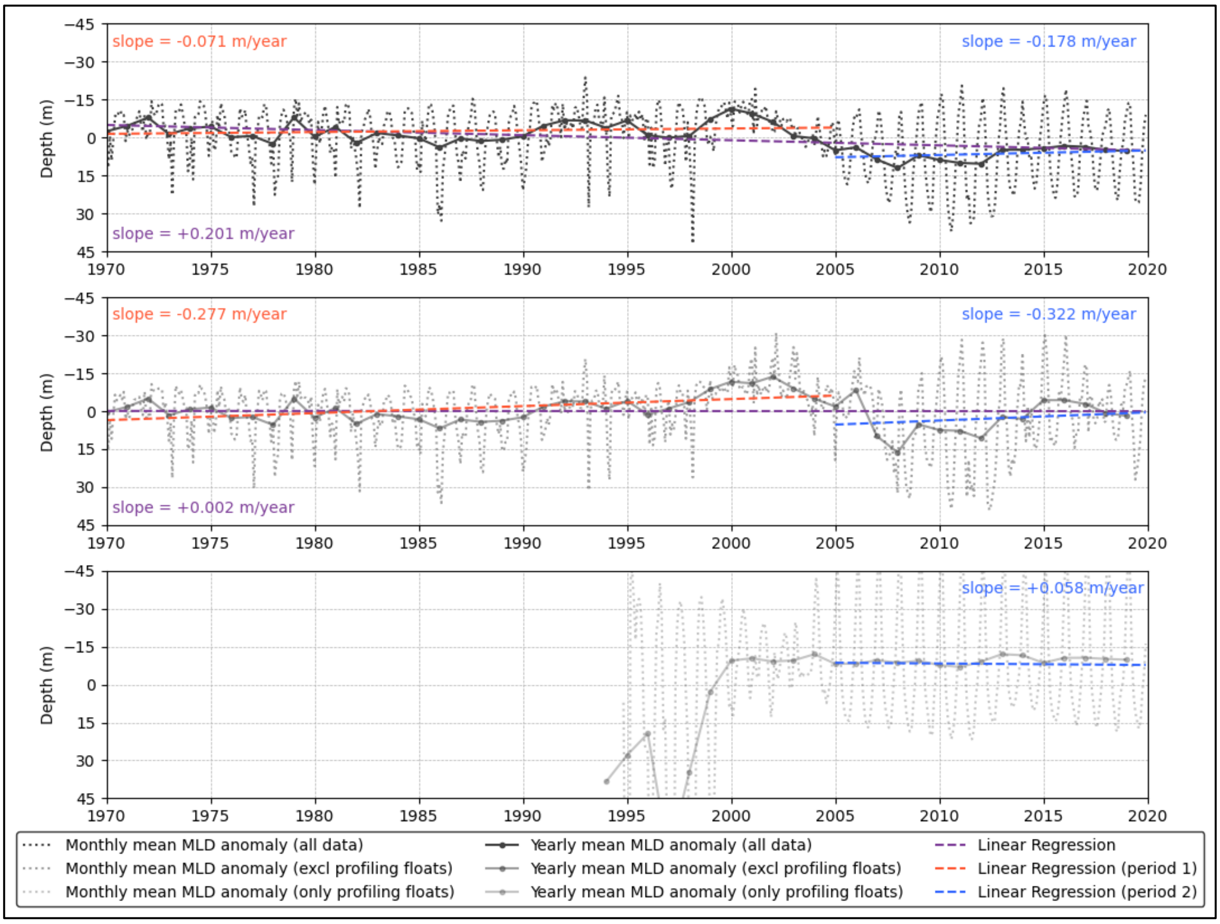

Figure 9. Mean Mixed Layer Depth (MLD) anomaly over time with all data included (top), mean MLD anomaly with data excluding profiling floats (middle) and mean MLD anomaly with only profiling floats data (bottom The monthly mean is represented as a dotted line and the yearly mean as a solid line. The linear regression for the whole time series is shown as a dashed purple line with the slope printed in the bottom left corner. The linear regression for period 1 is shown as a dashed red line, and the linear regression for period 2 as a dashed blue line, with respective slopes printed in matching colours.

As for the global monthly mean of the MLD anomaly (Figure 10), when studying the linear regression, the mean MLD tends to get deeper in all three regions (+0.407 m/year in Region A, +0.117 m/year in Region B and +0.099 m/year in Region C) when all data is included. Similarly to the global mean values, opposite trends are observed when PFs are excluded (-0.194 m/year in Region A, -0.068 m/year in Region B and -0.231 m/year in Region C).

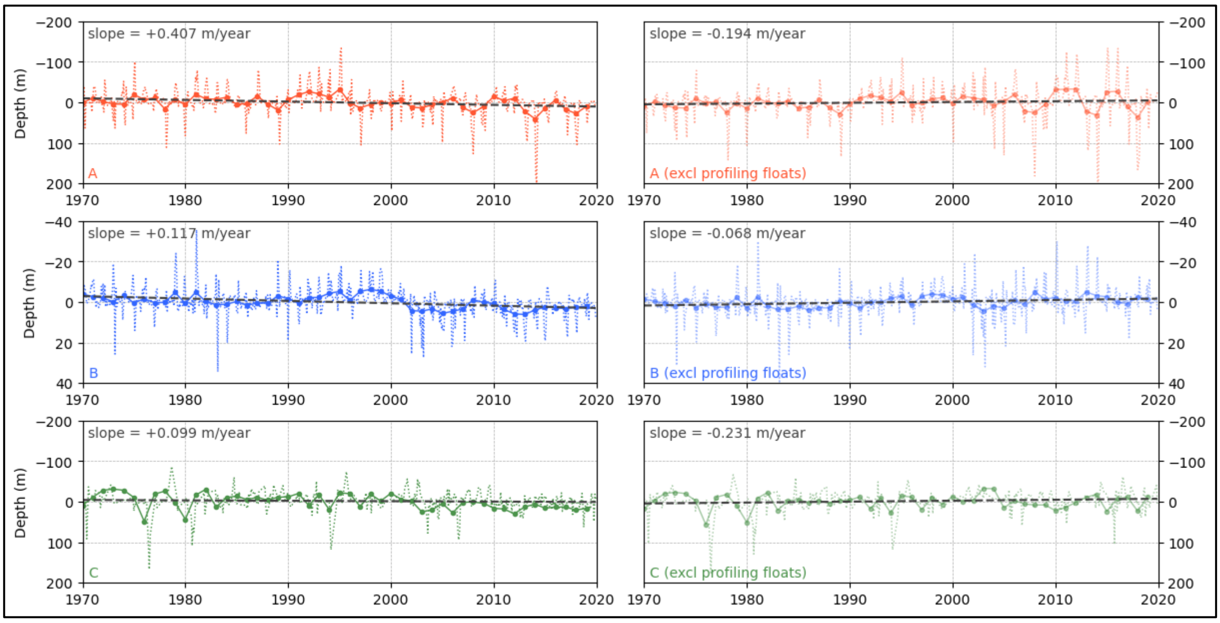

Figure 10. Mean Mixed Layer Depth anomaly over time with all data included (left panel) and with data excluding profiling floats (right panel). Region A (North Atlantic) is in red (top row), Region B (Northeast Pacific) is in blue (middle row), and C (Drake Passage) is in green (bottom row). The monthly mean anomaly is represented as a dotted line and the yearly mean anomaly as a solid line. The linear regression is shown as a dashed grey line with the slope printed in the top left corner.

3.4 Co-location and spatial distribution of potential bias

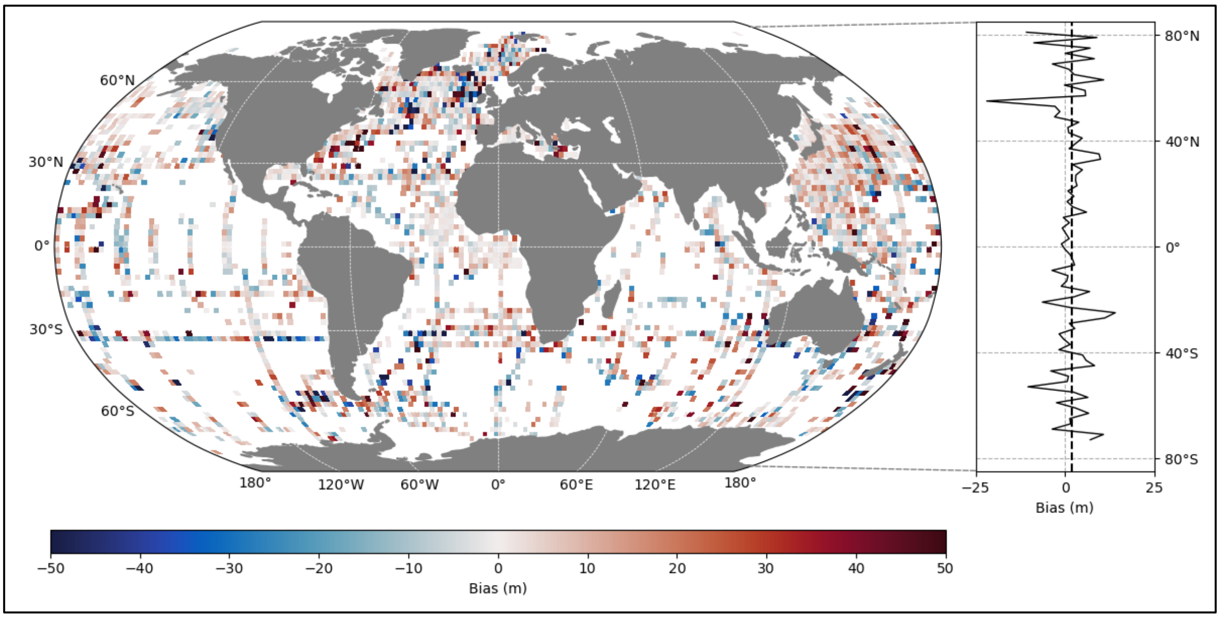

Figure 11. Mixed Layer Depth bias between CTD and profiling floats averaged on 2° by 2° grid (left), and the zonal and global mean (as a solid line and dashed line respectively) (right). Blue colour corresponds to negative values, i.e., that profiling floats output shallower MLD than CTD, and red to positive values, meaning that profiling floats output deeper MLD than CTD.

A clear patchiness could be observed when looking at the spatial distribution of the potential bias (Figure 11). Looking at, e.g. 45°N, a negative bias was observed in the North Atlantic while the Northwestern Pacific shows positive biases. On the contrary, nearly no biases are seen around the equator in the Atlantic. This is also seen when studying the zonal mean when larger peaks are seen in the higher latitudes and smaller ones around the equator. The maximum is seen at 77°S, with a value of 14.75 m, and the minimum at 55°N, with a value of -21.87 m. The sampling heterogeneity is also observed, where sampling is mainly concentrated in the North Atlantic and Western Pacific, with big holes in the Southern Ocean and the centre of the Pacific and Indian Ocean.

The co-location output indicates deeper MLD estimates from the PF measurements than CTD in all studied regions (globally, when divided into the northern-, and southern hemisphere and the three subregions A, B and C) (Table 1). Globally a significant difference ($p < 0.01$) of 2.64 (±47.94) m is observed. Comparisons between the two hemispheres highlight a regional difference, where the northern hemisphere has a significant difference ($p < 0.01$) of 2.90 (±50.23) m and the southern ($p < 0.01$) 1.04 (±30.14) m, indicating that the potential profiling float bias shows greater effects in the areas north of the equator. However, the standard deviation is greater in the northern than in the southern. These regional differences are also well visualised in the three subregions. The smallest mean difference is in Region A (0.62 ±88.36 m) and the highest in B (3.82 ±38.91 m). In Region C, the mean difference is 1.42 (±37.23) m. Nevertheless, it is only in Region B, where this difference is significant ($p < 0.01$), since the p-values for Region A and Region B are greater than 0.05.

Table 1. The MLD co-location pairs' mean difference and standard deviation (PF minus CTD) with paired t-test output. The output is divided into global, northern, and southern, hemispheres and the smaller subregions A, B and C.

Region		Mean diff. (m)	Std (m)	t-statistic	p-value	df
Global		+2.64	47.94	12.16	< 0.01	48891
Hemisphere	Northern	+2.90	50.23	11.83	< 0.01	42088
	Southern	+1.04	30.14	2.83	< 0.01	6674
Smaller regions	A	+0.62	88.36	0.51	0.61	5284
	B	+3.82	38.91	14.38	< 0.01	21461
	C	+1.42	37.23	0.57	0.57	222

The difference distribution between the co-located data points shows a skewness of -0.71, as seen in the histogram in Figure 12. Negative values indicate shallower depths and positive values represent deeper depths compared to the referenced PF measurements. The negative skewness, therefore, implies a higher median than the mean difference, indicating a potential positive bias where PF estimates a deeper MLD than the CTD.

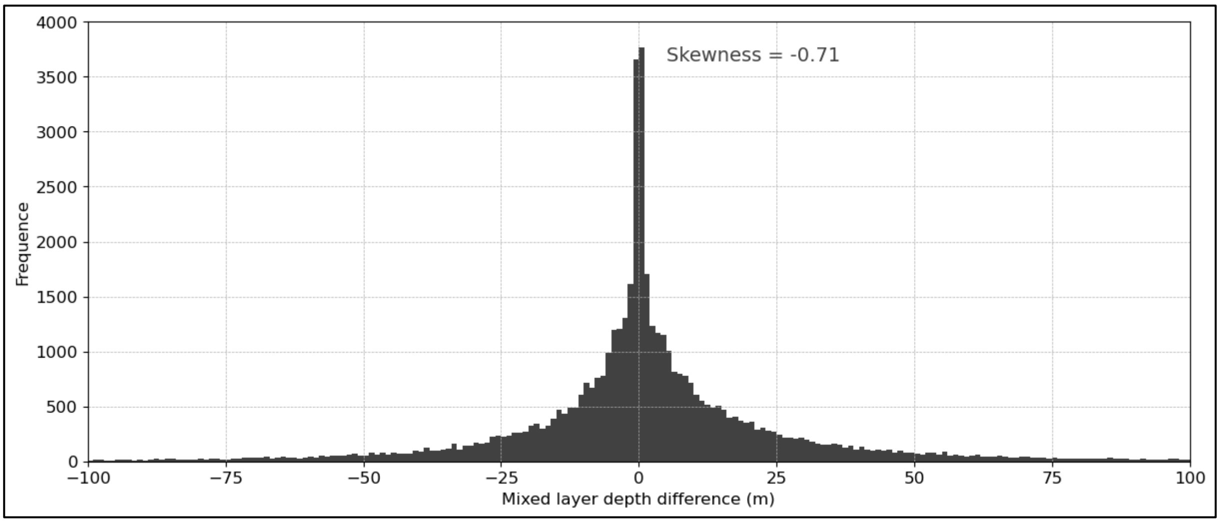

Figure 12. Histogram displaying the distribution of the MLD co-location pairs' differences in metres. The skewness is -0.71.

4. Discussion

Seasonal patterns of Mixed Layer Depth (MLD) variability, as shown in Figure 6 and Figure 7, align with expected climatological behaviour, with the deepest MLDs observed during winter months in both hemispheres and across all three subregions. A similar seasonal cycle is visible in the global mean, which closely mirrors the seasonality of the northern hemisphere, possibly reflecting an imbalance in data quantity between the hemispheres. This seasonality is consistent with previous studies, which demonstrate that winter mixing deepens the MLD due to stronger winds and surface cooling, both of which reduce stratification (de Boyer Montégut et al., 2004; Somavilla et al., 2017). Interestingly, the exclusion of Profiling Float (PF) measurements results in shallower MLDs during winter in both hemispheres and in each subregion, with this difference being especially noticeable in the global mean and in the North Atlantic, which is the most uniformly sampled region. This disparity may reflect variations in the spatial and temporal coverage of the datasets or potential biases in the PF measurements, as will be discussed further later.

Similarly, these seasonal patterns are evident when examining the global mean MLD over time (Figure 8). However, the variability, as mentioned above, is heavily influenced by the northern hemisphere, and it is not until after 2005 that the two hemispheres begin to cancel each other out, leading to a significantly smaller range between the yearly maximum and minimum values. This trend is also visible in the MLD anomaly data, where variability increases substantially after 2005, indicating larger extremes in the following years (Figure 9). This could be linked to the broader spatial coverage after 2005, due to the introduction of Argo floats, which provided more consistent and homogeneous data between the hemispheres (see Figure 3) (Liu et al., 2024). The inclusion of deeper MLDs, particularly in regions like the Southern Ocean (SO), which were previously under-sampled, might therefore contribute to this change in variability. This could also explain why the post-2005 variability range is larger when PF data is excluded.

The linear regression computed for the mean MLD indicates a deepening of the MLD when all data are used, and a relatively stagnant depth when PF measurements are excluded (Figure 8). Notably, when the seasonality has been removed (the MLD anomaly, Figure 9), this difference is even greater, with an even flatter trend (+0.002 versus -0.009 per year) when PFs are excluded. When divided into period 1 and period 2, the linear regression results reveal large differences dependent on which dataset is used (Figure 8 and Figure 9). In Period 1, for the global mean MLD, a shoaling of ~1 m per decade is observed (Figure 8), corresponding to a relatively stagnant MLD behaviour. When PF are excluded, an even greater shoaling is observed, with a magnitude of ~3 m per decade. Indicating large impacts from sampling and data origin. In Period 2, again, a shoaling is seen in all three datasets, but with large differences in magnitude. The non-PF dataset (EPF) shows a greater shoaling than both the all observations (AO) and the only-PF (OPF) ones, suggesting that the increased spatial and temporal data coverage after the introduction of floats plays a significant role in the observed trends. Similar results are found when examining the MLD anomaly (Figure 9). Remarkably, the dataset consisting of only profiling floats measurements, which is arguably the most homogenous dataset, we see a mild shoaling trend in the last 20 years. However, it is not clear whether this trend is significant. Hence, the fact that sampling and data origin have a strong impact on observed trends is certain, but the co-location results, additionally, indicate that even at the same time and space PFs estimate significantly deeper MLDs than measurements from CTD. This suggests there may be biases in the technology itself.

Measurement biases are a known issue, with new sensor drifts frequently highlighted. Barker et al. (2011) identified a pressure error drift in the Autonomous Profiling Explorer (APEX), which accounted for the majority of Argo floats used in 2009. Their study demonstrated that pressure corrections effectively reduced significant regional errors in ocean temperature and salinity, which are key density-dependent variables essential for MLD estimates, particularly when using the density threshold method applied in this study. They also noted that not all pressure profiles allow for corrections, and that uncorrectable profiles should be excluded from global change research. Similarly, Liu et al. (2024) recently highlighted a 'salty drift,' characterized by an increase in global mean salinity and growing inconsistencies between different gridded salinity products from Argo floats. They emphasized the distinction between real-time and delayed-mode Argo data, where real-time data are subject to less extensive quality control. This benefits faster updates but carries the risk of uncorrected issues introducing inaccuracies intro gridded salinity products. Despite certain batches of Argo floats being flagged for rejection in the latest EN4 dataset, the dataset used in this study and in Sallée et al. (2021), documentation indicates that not all floats have been identified, and some may still exhibit salinity drift (Killick, 2021).

When analyzing PF data exclusively, the bias may be less relevant since comparisons are made with the same slight bias across the dataset. However, the bias becomes apparent when comparing data from different sampling methods, especially since data quantity is not spatially or temporally homogeneous. The sampling method with the highest data quantity becomes a driving factor, and if that method has a bias, it can skew analyses, resulting in trends that may not exist. Because the salinity drift is not fully accounted for in the EN4 dataset, this dataset may be unsuitable for global change analyses or additional methods need to be implemented to account for such biases.

Correcting or calibrating the biases in Argo profiling float MLD estimates would require comprehensive adjustments at several levels. As mentioned by Barker et al. (2011), pressure errors and salinity drifts can be corrected for some floats, but not all profiles allow for reliable corrections. Further advancements in sensor calibration techniques and quality control measures, particularly for real-time data, are crucial. In the future, more extensive quality control procedures, such as those proposed by Liu et al. (2024), could help mitigate salinity drift, making the data more reliable for global change analyses. Enhanced calibration efforts would also need to consider regional differences in stratification and how biases in salinity or pressure might manifest differently in polar versus tropical regions. However, even with improved calibration, significant challenges remain. The coarse vertical resolution of Argo float profiles (Dong et al., 2008) may not adequately capture shallow MLDs typical of tropical regions, while they are more suited for deeper MLDs in the SO. Addressing these limitations would require a combination of better sensor technology, more spatially comprehensive sampling strategies, and improved methods for identifying and correcting sensor drifts.

Despite previous research highlighting sensor drift in Argo floats (Barker et al., 2011; Liu et al., 2024), it remains uncertain whether this drift fully explains the observed differences in the MLD trends. Although PF data subject to salinity drift is included in the EN4 dataset, the question remains whether this drift is significant enough to skew MLD trends on a global scale, particularly since the analysis shows different regional patterns (Figure 10). While the framework of a deepening MLD when all data is included, and shoaling when PF data is excluded, is consistent, the magnitude of the differences is not uniform across subregions A, B, and C. This suggests that if a bias exists, it may not be prominent everywhere. Nevertheless, a significant difference between the MLDs computed from CTD versus PF is evident in the co-location analysis (Table 1 and Figure 12). The spatial distribution of the difference, and potential bias (Figure 11), illustrates regional disparities, with the largest biases in higher latitudes and relatively smaller ones near the equator. This patchiness could be a consequence of the varying sensitivity of ocean stratification to salinity versus temperature. Polar regions are more strongly stratified by salinity, while subtropical and subpolar regions are more stratified by temperature (Caneill & Roquet, 2023). Therefore, a salinity drift would likely have a greater impact on the MLD estimates in high-latitude regions, particularly where PF data dominate, and be a candidate for the explanation of the regional differences between the three subregions. While the differences are statistically significant in some areas, particularly in the northern hemisphere, the standard deviations are high, indicating considerable variability within the datasets. This underscores the need for caution in attributing trends solely to salinity drift, and further analysis on other plausible parametrical drifts, e.g. pressure, is required.

Another plausible explanation for these differences is of course the spatial inhomogeneity of data coverage. Until the early 2000s, most observations were concentrated in the northern hemisphere, which accounted for approximately 85% of the data. Following the introduction of PFs, data quantity increased substantially in both hemispheres, especially in the SO. However, a notable disparity in observation density persists, with data collection primarily focused between 30°N and 40°N, resulting in roughly double the data compared to southern latitudes. Additionally, data collection methods have varied over time and space, with earlier data primarily from TESAC, CTD, and bottle measurements, and later data increasingly from XBT, MBT, APB (animal-mounted data), and PF. This variability in sampling methods, illustrated in Figure 4, adds further complexity. While northern hemisphere measurements rely heavily on TESAC and profiling floats, in the SO and near the equator, there is greater reliance on other platforms, including buoys and APB. The inhomogeneity in data quantity, location, and methodology could contribute to biases in coverage, impacting the interpretation of hemispheric changes. As described above, and as seen in both Figure 8 and Figure 9, data origin has a very strong impact on the observed trends and drawing robust conclusions would therefore be premature. However, this inhomogeneity does not fully explain the differences seen in the co-location output.

To build on this analysis, I would suggest a deeper investigation into several factors driving these potential biases. First, it would be valuable to examine the role of temperature and salinity profiles to identify which variable is the primary driver of the observed potential biases. Further analysis of stratification (N^2) and buoyancy fluxes could help clarify the dynamics contributing to the MLD variability. Additionally, a more focused comparison between summer and winter conditions would reveal whether seasonal stratification patterns significantly influence these trends.

While models provide valuable insights into long-term MLD changes, the observed variability in MLD trends complicates direct comparisons. For example, models predict a shoaling of the MLD due to increased upper ocean stratification under global warming scenarios (Gao et al., 2023), yet some regions, as observed in the results of this study and those of Sallée et al. (2021), exhibit a deepening trend. This paradox could be due to competing forces: while surface warming tends to increase stratification, wind-driven turbulence or other mechanical mixing processes might counteract this (Somavilla et al., 2017). Thus, caution must be exercised when interpreting model projections, particularly because models often rely on parameterizations that may not fully capture the complexity of ocean dynamics. Improvements in mixed-layer parameterization and more accurate estimates of air-sea fluxes are necessary to reconcile observed trends with model predictions. Furthermore, better characterization of ocean stratification, wind forcing, and surface buoyancy fluxes, coupled with high-quality observational data, will be key to refining model predictions and understanding future MLD changes.

Conclusion

To conclude, the full extent of salinity drift and its impact on global Mixed Layer Depth (MLD) estimates remains an open question. Although sensor drifts may influence some of the observed trends, attributing these results solely to this factor would be premature. Caution is essential when interpreting these results, as sensor drift or other biases in PF data could influence variability trends. The combination of sampling inhomogeneity, data coverage variability, and evolving measurement techniques complicates drawing definitive conclusions. However, sampling and data origin certainly have a very large impact on the observed trends. Depending on what dataset analysis is based on, different results emerge, and not even in well-sampled regions like the North Atlantic, any global trend can be considered robust. Further investigation into calibration techniques and a careful evaluation of data quality is necessary to interpret these trends with greater confidence. Future work would benefit from addressing these complexities through region-specific calibration techniques or by adjusting for data gaps and biases resulting from uneven spatial and temporal sampling.

References

Alexander, M. A., Scott, J. D., & Deser, C. (2000). Processes that influence sea surface temperature and ocean mixed layer depth variability in a coupled model. *Journal of Geophysical Research: Oceans*, *105*(C7), 16823–16842. https://doi.org/10.1029/2000JC900074

Amante, C., & Eakins, B. W. (2009). *ETOPO1 arc-minute global relief model: Procedures, data sources and analysis*. https://repository.library.noaa.gov/view/noaa/1163

Barker, P. M., Dunn, J. R., Domingues, C. M., & Wijffels, S. E. (2011). Pressure Sensor Drifts in Argo and Their Impacts. *Journal of Atmospheric and Oceanic Technology*, *28*(8), 1036–1049. https://doi.org/10.1175/2011JTECHO831.1

Boyer, T. P., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Locarnini, R. A., Mishonov, A. V., Paver, C. R., Reagan, J. R., Seidov, D., Smolyar, I. V., Weathers, K. W., & Zweng, M. M. (2018). *WORLD OCEAN DATABASE 2018*. NOAA Atlas NESDIS 87.

Buongiorno Nardelli, B., Guinehut, S., Verbrugge, N., Cotroneo, Y., Zambianchi, E., & Iudicone, D. (2017). Southern Ocean Mixed-Layer Seasonal and Interannual Variations From Combined Satellite and In Situ Data. *Journal of Geophysical Research: Oceans*, *122*(12), 10042–10060. https://doi.org/10.1002/2017JC013314

Caneill, R., & Roquet, F. (2023). *Temperature versus salinity: Distribution of stratification control on the global ocean* [[Unpublished manuscript]]. [Unpublished manuscript], Department of Marine Sciences, University of Gothenburg.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans*, *109*(C12). https://doi.org/10.1029/2004JC002378

Diehl, S. (2002). PHYTOPLANKTON, LIGHT, AND NUTRIENTS IN A GRADIENT OF MIXING DEPTHS: THEORY. *Ecology*, *83*(2), 386–398. https://doi.org/10.1890/0012-9658(2002)083[0386:PLANIA]2.0.CO;2

Dong, S., Sprintall, J., Gille, S. T., & Talley, L. (2008). Southern Ocean mixed-layer depth from Argo float profiles. *Journal of Geophysical Research: Oceans*, *113*(C6). https://doi.org/10.1029/2006JC004051

Gao, Z., Long, S.-M., Shi, J.-R., Cheng, L., Li, G., & Ying, J. (2023). Indian Ocean mixed layer depth changes under global warming. *Frontiers in Climate*, *5*. https://doi.org/10.3389/fclim.2023.1112713

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), 6704–6716. https://doi.org/10.1002/2013JC009067

Gouretski, V., & Cheng, L. (2020). Correction for Systematic Errors in the Global Dataset of Temperature Profiles from Mechanical Bathythermographs. *Journal of Atmospheric and Oceanic Technology*, *37*(5), 841–855. https://doi.org/10.1175/JTECH-D-19-0205.1

Gouretski, V., & Reseghetti, F. (2010). On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep Sea Research Part I: Oceanographic Research Papers*, *57*(6), 812–833. https://doi.org/10.1016/j.dsr.2010.03.011

Kara, A. B., Rochford, P. A., & Hurlburt, H. E. (2000). An optimal definition for ocean mixed layer depth. *Journal of Geophysical Research: Oceans*, *105*(C7), 16803–16821. https://doi.org/10.1029/2000JC900072

Kara, A. B., Rochford, P. A., & Hurlburt, H. E. (2003). Mixed layer depth variability over the global ocean. *Journal of Geophysical Research: Oceans*, *108*(C3). https://doi.org/10.1029/2000JC000736

Killick, R. (2021). *EN.4.2.2 Product User Guide*. https://www.metoffice.gov.uk/hadobs/en4/EN.4.2.2_Product_User_Guide_v1.0.pdf

Liu, C., Liang, X., Ponte, R. M., & Chambers, D. P. (2024). "Salty Drift" of Argo Floats Affects the Gridded Ocean Salinity Products. *Journal of Geophysical Research: Oceans*, *129*(9), e2023JC020871. https://doi.org/10.1029/2023JC020871

Roch, M., Brandt, P., & Schmidtko, S. (2023). Recent large-scale mixed layer and vertical stratification maxima changes. *Frontiers in Marine Science*, *10*. https://www.frontiersin.org/articles/10.3389/fmars.2023.1277316

Sallée, J.-B., Pellichero, V., Akhoudas, C., Pauthenet, E., Vignes, L., Schmidtko, S., Garabato, A. N., Sutherland, P., & Kuusela, M. (2021). Summertime increases in upper-ocean stratification and mixed-layer depth. *Nature*, *591*(7851), Article 7851. https://doi.org/10.1038/s41586-021-03303-x

Sérazin, G., Tréguier, A. M., & de Boyer Montégut, C. (2023). A seasonal climatology of the upper ocean pycnocline. *Frontiers in Marine Science*, *10*. https://www.frontiersin.org/articles/10.3389/fmars.2023.1120112

Somavilla, R., González-Pola, C., & Fernández-Diaz, J. (2017). The warmer the ocean surface, the shallower the mixed layer. How much of this is true? *Journal of Geophysical Research: Oceans*, *122*(9), 7698–7716. https://doi.org/10.1002/2017JC013125

Sun, C. (2010). The Data Management System for the Global Temperature and Salinity Profile Programme. *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2)*. OceanObs'09: Sustained Ocean Observations and Information for Society, Venice, Italy. https://doi.org/10.5270/OceanObs09.cwp.86