Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

GEOFISICA

Ciclo 33

Settore Concorsuale: 04/A4 - GEOFISICA

Settore Scientifico Disciplinare: GEO/12 - OCEANOGRAFIA E FISICA DELL'ATMOSFERA

A NEW GLOBAL OCEAN CLIMATOLOGY

Presentata da: Kanwal Shahzadi

Coordinatore Dottorato

Nadia Pinardi

Supervisore

Nadia Pinardi

Co-supervisore

Marco Zavatarelli
Simona Simoncelli
Vladyslav Lyubarstev

Esame finale anno 2021

Acknowledgements

First of all, thanks to Allah Almighty for giving me strength and an opportunity to study one of his wonderful creation, Ocean.

My special thanks goes to Prof. Nadia Pinardi for all her valuable advices, encouragements and continuous supports throughout the journey of this Ph.D.
My sincere thanks and gratitude goes to Dr. Alexander Barth and Dr. Charles for their guidances and useful suggestions during my stay at University of Liege. Moreover, thanks to Prof. Marco Zavatarelli for helping me to finish the 4th chapter of the thesis.

I would like to say thank you to a my best colleague and friend, Dr. Umesh P. A. for reviewing my writings and listening calmly to all my research hurdles.

In addition, Thank you to Mr. Muhammad Owais for introducing me to Pandas and Dataframes libraries in Python at the very end of Ph.D project which really speed up my calculations.

A special gratitude to my nephews and sisters especially Lubna Asghar for all of their evening calls to cherish me.

The acknowledgement would be incomplete without expressing a special gratitude to my mother, for all her prayers, good words and moral support throughout my life. In addition, thanks to my quiet Dad who always appreciated me for higher studies.

At the end special thanks to Institute of advanced studies for providing me wonderful accommodation during my Ph.D. It gave me a chance to live with other international Ph.D fellows which made me feel like home. In addition, thank you to university of bologna for providing me all the resources and funding to complete this study.

Yay, I'm finally done.. Now on to the papers.. :)

Abstract

Global Ocean climatologies are fundamental for our understanding of climate variability and trends, essential for initialisation and validation of numerical models. The thesis aims to compute a new global ocean monthly climatology of basic physical climate state variables such as temperature, salinity, density and dissolved oxygen from in-situ based historical datasets collected in the World Ocean Database 2018. The novelty of these climatologies stems from the implementation of new quality control procedure, called "Nonlinear Quality Control" (NQC) thereafter. NQC is applied to the database that is used to compute the climatology and the improvements in the analysis discussed.

The climatologies presented in the thesis are processed by a statistical interpolation tool, the Data Interpolating Variational Analysis (DIVA) that is applied on the global domain for the first time. Two different versions of temperature and salinity climatologies are estimated based on the different temporal coverage of the data: a long term average (1900 to 2017) using multiple platforms, and a shorter time estimate (2003 to 2017) using data from ocean drifting platforms such as profiling floats. Sensitivity experiments are carried out to choose the key parameters of DIVA. The computed climatologies show consistency with well-known reference climatologies such as World Ocean Atlas 2018 and World Argo Global Hydrographic Climatology.

Preliminary health of the ocean indicators are presented for the historical data set, using the validated NQC and mapping algorithms. First, the mapping of density has been carried out from NQC processed profiling float dataset and Brunt-Väisäla frequency profiles were computed from the density and averaged in 5° square boxes for Atlantic and Pacific Oceans. Brunt-Väisäla frequency is quantifying the ocean stratification strength which impacts ocean ventilation and vertical mixing in general. The computation of dissolved oxygen climatology and saturation oxygen in the water column was analysed and the Apparent Oxygen Utilisation (AOU) was estimated from several different measuring platforms.

Evidence for ocean warming is present in the two global ocean climatologies as expected as well as the deep increase in salinity. The preliminary comparison of averaged Brunt-Väisäla profiles with Emery et al. (1984) shows significant stratification changes in the water column. Moreover, the comparison of O2 and AOU calculated only for the period of 2003-2017 seems to indicate lower and higher values respectively in comparison with WOA18. These differences

are found to be large at intermediate depth for both O2 and AOU. These results could be important for health of the ocean considerations and further extensive analysis will be required before a solid conclusion is reached.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Introduction

The natural variability of the ocean is due to time-space scales that range from seconds to decades, from tens of meters to several hundred kilometers. Estimating the variability of such a system at time scales of decades to centuries is a formidable task. Observations usually involve sampling the highest space-time frequency possible given the measuring platform and the sensor accuracy. To estimate climatologies, information on the long term variability of the oceans needs to be extracted from these data. However it is difficult to accurately compute information on the observed statistical distributions and their moments, as well as the mean and the standard deviation.

Operational ocean satellites (Le Traon, 2011) have started to map the surface ocean variability for a few ocean variables and they will continue to increase in scope in the few next years. However only in situ observations can explore the variability of the deep ocean. A key issue in estimating climatology from in situ data is that the observations undersample the oceanic variability both in space and time. In situ observations are also affected by errors in the measurement procedures, the calibration of the instrument, or by the limited precision of the instrument. However, data quality issues can be analyzed in the framework of quality control methods applied to the observations: quality control is an essential component of climate research, which is often overlooked.

In general, the Global Ocean Observing System (Moltmann et al., 2019) has three different streams of observations, namely, real time, delayed mode and historical. In the real time stream,

observations are available immediately after the measurement and very little quality control is carried out. This stream of data supports the ocean operational forecasting systems throughout the world (Pinardi et al., 2018). The delayed mode data, is a higher quality control stream, and the delay in delivery is mainly due to data policy agreements[1]. The historical data are stored in the real time and delayed mode data in a common historical archive database. The historical observational collections can thus undergo a higher quality control check based on statistical information deduced from the observations themselves.

As in situ data are primary building blocks in understanding the ocean's climate variability, the challenge is to manage them at a global level with standard, international protocols. The NOAA global ocean database (World Ocean Database-WOD, Garcia et al. (2018)) was the first to start collecting data from different observation producers and to establish a historical, publicly available database[2]. For Europe, SeaDataNet (Schaap and Lowry, 2010) and SeaDataCloud (Simoncelli et al., 2018) have started to tackle the problem of European ocean historical databases and climatologies. In all these projects, quality control procedures have been defined and applied to ocean data sets starting from the 1900s to today, and climatological estimates have been carried out, as described in the following section.

Finally, long-term time series of ocean variables have also be reconstructed using reanalysis techniques (Storto et al., 2011). These techniques are applied to historical databases and use a numerical ocean model to dynamically "interpolate" observations on a regular grid.

In this thesis we focus on climatological estimates from historical databases using statistical methods derived from the observations themselves and not from a numerical model. This is a complementary methodology with respect to reanalysis, which we hope will contribute in understanding the uncertainties in the estimation of ocean climate trends. However these statistical realisation of inappropriate sampled observations are highly dependent upon several factors such as mapping algorithm, corresponding parameters and the dataset being used. Consequently several uncertainties will emerges among these estimates due to these choices. In order to understand and minimise them, it is always desirable to have various realisation of these estimate which should be reproducible. The previous estimates are unable to be reproduced because of various ad-hoc assumptions. The motivation of the thesis is to have a long term climatologies of ocean essential variables with an advanced mapping algorithm called

[1] see for example https://www.nodc.noaa.gov/woce/wdiu/wocedocs/datapol.htm
[2] https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html

Data Interpolating variational Analysis with new a quality controlled (introduced in chapter 2) processed data and its comparison with previous estimates. Multiple approaches in mapping the ocean climate into regularly gridded data sets will support an ensemble-based estimate of the ocean climate, which is necessary given the turbulent, episodic nature of ocean climate variability is proposed and given in this Thesis.

1.2 Methods in Ocean Climatological Studies

Historical ocean observational databases enable climatologies to be constructed with different space and time scales. Several steps are required to develop a long-term climatological estimate of the ocean state, for example:

(i) Quality assurance of the observations;

(ii) Understanding of the spatial and temporal data gaps;

(iii) Mapping to a regular grid.

In this thesis we start by developing a new method for the quality assurance of the historical observations. The latter are affected by several kinds of uncertainties. The main source of errors in the observations are:

(i) Sensor accuracy, problems in the data transmission system, sensor calibration and drifts, and human errors in the assembly procedures;

(ii) Non-uniform sampling or errors in representativeness.

In order to capture these sources of uncertainties, sophisticated quality control procedures are exploited, which depend on the data stream used. Real time quality procedures are automatic and normally only check for instrumental and transmission errors. When real time data streams enter a delayed mode data stream, they are further quality controlled on the basis of statistical checks. Statistical checks are needed to capture the representativeness errors. In this thesis, a new statistical quality control procedure is developed.

A suitable interpolation scheme plays a vital role in constructing a reliable climatology. Objective, statistical schemes used to map the observations in a regular grid started in the second half of the 20th century. In Objective Analysis (OA), the interpolated field is a combination of the first guess field and the weighted sum of the observations. Different weight functions were considered by Cressman (1959) who defined them as the ratio of the distance between the grid location and the observation location. However, Cressman's weight functions are too

simple and suboptimal, are not founded on basic statistical principles and do not consider observational errors. Later, Barnes (1964) improved Cressman's method and defined the weight functions as exponentials. But the empiricism of the technique remained. In 1963, Gandin introduced objective analysis (OA) based on the Gauss-Markov least square theory. This was a real step toward a comprehensive estimation theory of a field on a regular grid based on irregularly spaced data.

The next step was Optimal Interpolation (OI) which facilitated the optimal combination of observations with a first guess field, unlike quantities derived from the observations themselves, assuming first guess and observations with their own errors. OA and OI require prior knowledge of the error variances and covariances of the first guess field and the observations. In addition, the basic assumption in OI is that the field under investigation is stationary and homogeneous, i.e. its statistical characteristics must remain unvaried for the time interval considered. This also implies that its second order statistics, or correlation function, is homogeneous in time.

In 2010, Data Interpolating Variational Analysis (DIVA; Barth et al., 2014) was developed which is based on the Variational Inverse Method (VIM; Brasseur, 1991). VIM estimates a climatological field on a regular grid starting from irregularly spaced observations with the characteristics. The interpolated field is smooth but close to observations within their error statistics. The innovative feature of DIVA is the inclusion of realistic land sea boundaries and coastal constraints such as a no slip boundary condition at the coasts. It works on unstructured grids that are able to resolve the complex coastal shapes. The OI interpolation technique is equivalent to the field constructed by VIM under the conditions explained in (McIntosh, 1990) i.e. the weights used by VIM are the inverse of the error covariances used by OI.

1.3 A Brief Overview of Global Ocean Climatologies

Several global ocean climatologies have been constructed from historical observations stored in large databases.

The first global ocean climatology was produced by (Levitus, 1982) in 1982 and its modern version is the World Ocean Atlas (WOA) produced by the U.S. National Oceanographic Center (NODC) using the World Ocean Database (WOD). The WOA has several versions starting from 1994 with the latest being WOA18 (Locarnini et al., 2018). WOA18 is produced using the interpolation scheme of Barnes briefly described above.

Recently a World Argo global hydrographic climatology (WAGHC) was produced by (Gouretski, 2019) for temperature and salinity on both isobaric and isopycnal levels. Argo is a profiling float program[3] that revolutionized the number of observations in the open ocean (Belbeoch et al., 2010). The Gandin (1963) interpolation method is used with a Gaussian correlation function. WOA and WAGHC are both on a regular $0.25^\circ \times 0.25^\circ$ spatial resolution grid.

Another gridded dataset for temperature and salinity, called EN4, was created by the UK Met office (Good et al., 2013). EN4 uses data from profiling floats and many other data sources and applies objective analysis. The International Pacific Research Center (IPRC; Hacker et al., 2010) has its own tool called the Monthly Objective Analysis for Argo (MOAA; Hosoda et al., 2008). In addition, global monthly isopycnal upper-ocean climatology (MIMOC) was computed by Sunke 2013 (Schmidtko et al., 2013). Climatological estimates such as EN4, IPRC, MOAA and MIMOC are available on coarse 1° or 0.5° spatial resolution grids, for 2000 to 2015 because they only use Argo data. In this thesis we also develop a climatology from Argo data, but we use the DIVA advanced interpolation scheme for a high resolution $0.25^\circ \times 0.25^\circ$ grid.

A gridded climatology can be constructed also from reanalysis described in the previous section. The Copernicus Marine Environment Monitoring Service (CMEMS; Lellouche and Regnier, 2015) provides ocean reanalyses using in situ and satellite observations at $(1/12)^\circ$ resolution. Estimating the Circulation and Climate of the Ocean (ECCO; Fukumori et al., 2018) provides a dynamical mean state of the ocean for 1994 to 2013 on an 1° resolution grid. The Global Ocean Data Assimilation Experiment (GODAE; Bell et al., 2009) provides a global ocean climatology of temperature and salinity along with biogeochemical variables using DIVA on $1^\circ \times 1^\circ$ grid resolution. The Simple Ocean Data Assimilation (SODA; Carton and Giese, 2008) analysis consists of a global ocean reanalysis from 1958 to 2008. The Ocean Reanalysis and system4 (ORA4; Balmaseda et al., 2013) produced by ECMWF is a global Ocean Reanalysis, and finally the CMCC Global Ocean Reanalysis System (C-GLORS; Storto et al., 2011) are two other global reanalyses at 0.25° resolution.

1.4 Objectives of the Thesis

This thesis contributes with two essential steps in the creation of a climatology from in situ observations: (i) an improved data quality control procedure, (ii) an advanced interpolation

[3] https://www.seabird.com/profiling-floats/

algorithm for the depiction of monthly mean temperature and salinity climatologies for the world ocean.

The three primary aims of this thesis are thus:

• To develop an advanced control procedure of in-situ data, thereby eliminating the observations that are not suitable for large-scale ocean climatologies.

• To implement the above quality control procedure on temperature and salinity profiles from WOD18 and construct monthly mean global ocean climatologies using DIVA.

• On the basis of the experience gained, to develop climatologies for derived quantities from temperature and salinity such as ocean density, Brunt-Väisäla frequency profiles, Dissolved oxygen and Apparent Oxygen Utilisation which will be used in the future to estimate the health of the ocean indicators.

The thesis is structured as follows:

1. Chapter 2 describes a new quality control procedure with its detailed schematics. The procedure is tested in four different regional domains and the results are analysed.

2. Chapter 3 presents a new global ocean climatological estimate of temperature and salinity using DIVA from the quality control processed data. The results are analysed and compared with previous climatologies.

3. Chapter 4 gives an account of the computed mapping of derived quantities. The details of the Density, Mixing index Brunt-Väisäla frequency and AOU are discussed.

4. Chapter 5 provides the overall conclusions and recommendations for future work.

Chapter 2

A Nonlinear Quality Control for large-scale ocean temperature and salinity climatologies

2.1 Introduction

In situ ocean observations are fundamental to our understanding of oceanic processes, from short to climate time scales. The climatological analysis of in-situ observations requires high quality and uniform spatial and temporal coverage for repeated periods. This is why historical data sets where observations are collected from different sources and are partially harmonised, are essential components of the global ocean observing system. Such historical data sets provide the longest records of in situ measurements that can be used to reconstruct the ocean climate. The World Ocean Database (WOD, Boyer et al., 2019) is one of the largest collections of in-situ observations from multiple observing platforms . In this chapter we concentrate on this large collection in order to derive an advanced monthly mean temperature and salinity climatology for the world ocean. We examine the required quality control procedures for historical observations before spatial mapping and other types of analysis are carried out. Measurements in WOD derive from multiple platforms with non-uniform spatial and temporal sampling and from different sensors with substantially different accuracies. As a result, WOD measurements contain various kinds of uncertainties. Given that a measurement is an approximation of the truth because of the errors inherent in the specific observation methodology, the sources of

errors need to be quantified in order to use them to estimate the climatological signals.

Uncertainties in oceanographic in-situ data are quite complicated because the oceanic variability is the sum of several processes occurring at a very wide range of spatial and temporal scales. These multiple scales are not adequately sampled by the instruments because of sensor accuracy and the sampling scheme. The uncertainties in the observational historical database are the sum of two major errors: Gross Errors and Representativeness Errors (REs) (Janjić et al., 2018). Gross errors are related to limited instrument precision, sensitivity, calibration, failure, transmission errors or incorrect archiving, while REs are caused by poor sampling by the platform and/or inability to resolve the spatial and temporal scales. As described by Daley (1993); Janjić and Cohn (2006); Schutgens et al. (2016), REs are connected to the inappropriate sampling of the space and time scales. In addition, Barth et al. (2008) explained that REs occur when the observations used in the analysis are not collected at the same time on the spatial area of interest (non-synoptic data).

In terms of estimating a "time mean" or climatology, REs are particularly important because none of the observations collected are repeated in time at the same position and for a sufficiently long time, e.g. CTD casts taken in an eddy field of a certain region and never repeated for that region. The CTD profiles taken in a region and not repeated several times in the same region will not be a good estimate of the climatology. Probably such CTD profiles should be rejected in the climatological analysis because of large REs. Gouretski and Koltermann (2004) defines the observational ocean climatology as a statistical estimate starting from inadequately sampled characteristics of a dynamical fluid. In order to estimate the low frequency variability of the ocean, the observations that do not accurately sample the time mean value, i.e. they are not close to the mean, should be eliminated from the database collection. Disentangling these unresolved scales from the observational data before a spatial and temporal analysis is referred to as the quality control procedure to eliminate non-representative data.

There are several ways to account for REs in a spatial climatological analysis:

(i) By eliminating errors using robust quality control procedures before performing any analysis.

(ii) By accounting for these uncertainties directly in the analysis algorithm. For example, the objective or variational analysis (Barth et al., 2014) uses error covariance matrices of the observations and background to grid the data, accounting directly for such uncertainties. For the background errors, the data assimilation algorithms use the standard deviation of the

oceanic fields, assuming that the standard deviation is an error proxy.

In this chapter we explore the first method, i.e. the quality control procedure for the observations before using them in the analysis algorithms. This a priori quality control is a trade-off between a low RE and a sufficient number of measurements. We use the Gross Error checks already defined by WOD for temperature and salinity observations, and we develop a new set of RE checks that eliminates outliers from the dataset.

WOD and its 2018 version (WOD18) is the most comprehensive collection of historical global ocean observations from 1772 to the present. WOD18 has its own QC that was developed by Levitus (1982) and updated by Boyer et al. (2019). In WOD18, a quality flag value, based on the Gross error check and a statistical quality check, is assigned to each individual measurement on a profile of temperature and salinity. If these flags are used to reject data, several outliers cannot be eliminated, as is shown for the salinity values in Fig. 2.1. Some prominent outliers and anomalous values are evident regardless of the WOD quality flags. An advanced and robust quality control procedure is therefore needed in order to account for these outliers that are most likely due to REs. We call the RE quality control procedure that we defined and implemented in this study the Nonlinear Quality Control (NQC) procedure, similar to the one used by Jia et al. (2016) but now used to compute a global ocean climatology for temperature and salinity. The main questions addressed by this paper are:

(i) How the NQC be formulated so that it rejects the outliers/non representative data?

(ii) Can we apply the NQC to very different ocean regions and how sensitive are the results to the specific choices in the NQC method?

Section 2.2.1 describes the WOD18 quality control procedures and introduces the NQC algorithm. In section 2.3 the NQC is applied to the North Western Pacific, the North Atlantic, and the South Atlantic regions. Section 2.4 discusses the sensitivity experiments, and Section 2.5 provides the conclusions.

2.2 Quality Control Procedures

2.2.1 WOD18 Quality Control Procedures

Profiling floats have dramatically increased the amount of data in the world oceans and increased the basic data set for the estimation of climatology due to the almost uniform, random

sampling scheme of the floats (Riser et al., 2016). For this study, we selected only profiling float (PFL) dataset from WOD18 because we wanted to try the procedure on a uniform quality data set. In addition, PFL are the most numerous measurements in the global ocean nowadays and we wanted to show that the method worked well with this basic monitoring system. Profiling float (PFL) measurements, as with other types of sensors and platforms, suffer from salinity drift, thermal lag offset, pressure hysteresis and data stream transmission errors (Boyer et al., 2019). In WOD18 there are two types of quality control flags associated with the PFL: (a) an individual value flag (WODf) for each variable measured at a certain depth; (b) a profile flag (WODfp) which considers a statistical checks explained below.

The WODf consist of eight different checks:

(i) Format conversion;

(ii) Profile position, date and time checks,

(iii) Assignment of cruise and cast numbers,

(iv) Speed check,

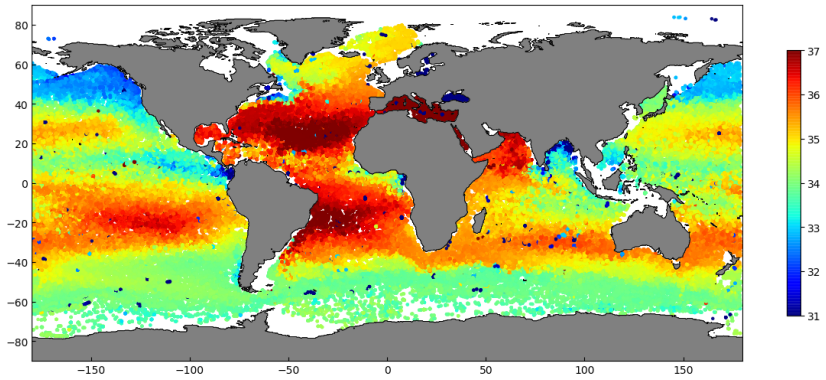
(v) Duplicate cast check,

(vi) Depth Inversion,

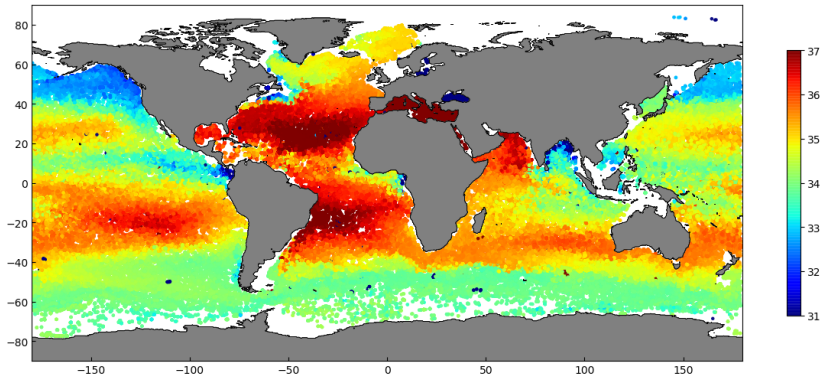
(vii) Excessive gradient check, and

(viii) Range checks on observed data.

The WODf quality check is clearly aimed at controlling the gross error. On the other hand, the WODfp checks for measurements that do not exceed a statistical threshold computed from the data in different subregions. The world ocean is subdivided into the rectangles, $5^\circ \times 5^\circ$, where the mean and standard deviation of the observations is computed as a function of depth. The measurement value is then compared to the standard deviation and if it exceeds a pre-set number of standard deviations, it is flagged as an "outlier". One of the problems in this procedure was highlighted by Smolyar et al. (2018) that is, the random occurrence of variable water masses in regular squares. As a result, the statistics of the regular box will not be an effective threshold to eliminate the outliers as shown in Fig. 2.1.



(a)



(b)

Fig. 2.1: Quality controlled salinity observations selected from WOD18 in a layer of 10m (between 5m to 15m): (a) WODf, (b) WODf together with WODfp.

2.2.2 The Nonlinear Quality Control

The NQC aims to eliminate data that are not representative of a climatology of temperature and salinity in a $0.25^{\circ} \times 0.25^{\circ}$ grid. Taking as the initial point the WODf and WODfp flags explained above and eliminating the rejected data with these checks, NQC adds a novel procedure to estimate the REs based on a new statistical estimate of the standard deviation.

In the definition of statistics for ocean observations, there are two degrees of freedom: the spatial averaging area and the number of standard deviations to be used as a threshold to reject the observations that exceed these values. Starting from the spatial averaging area, Fig. 2.1 shows that the WODf and WODfp checks do not eliminate all the outliers using an arbitrary rectangular grid as the subdivision to compute the standard deviations. We argue that a selection of the averaging domain that is "regime oriented" would be better. For example, in the Kuroshio extension, Jia et al. (2016) defined 5 areas containing different subregions around

the Kuroshio western boundary current, its extension and the Japan Sea. We show here how important this is.

The second issue is how to compute the mean with which to calculate the standard deviation in each subregion. In the NQC this is performed by estimating the mean by first interpolating the fields on the target grid and then estimating the mean. To interpolate the irregular space data, we use the Objective Analysis (OA, see Appendix A) method applied to the observations. This provides a smoothed mean of the field in each subregional averaging area. The standard deviations are then computed from the mean of the objectively interpolated fields in each subregion. The observations are then rejected if they are three standard deviations away from the mean thus calculated, and they are then eliminated from the data set. The detailed schematics of the procedure is shown in Fig. 2.2.

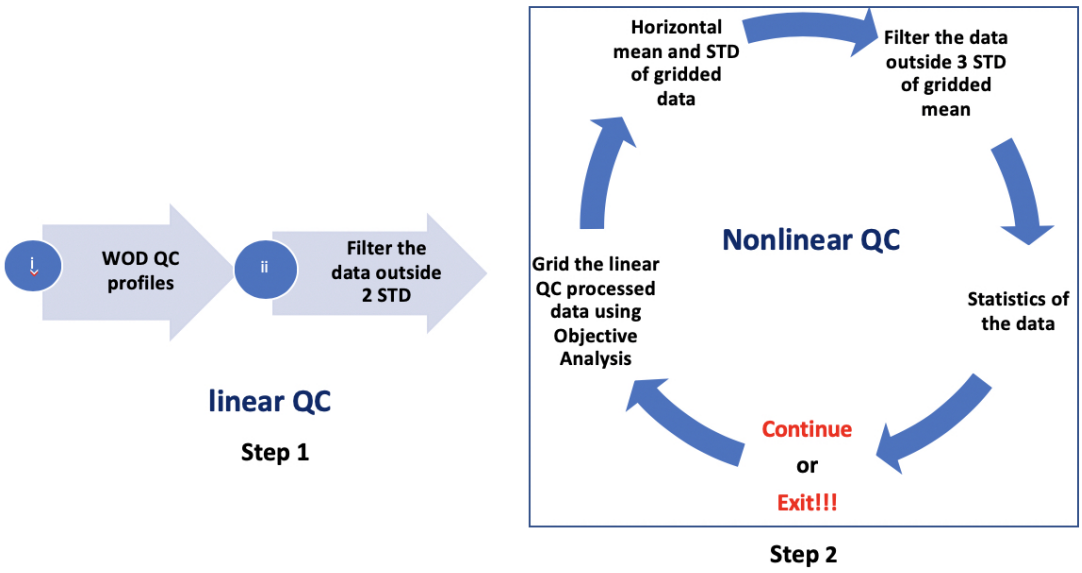


Fig. 2.2: NQC schematic. The procedure is composed of two steps: **linear QC**, where the profiles are checked by the WODf and WPDfp flags; and a second, so-called **NQC**, that is iterative. Observations are gridded in the area of interest using OA, the mean and the standard deviation computed for each subregion and the observations are rejected on the basis of exceeding 3 standard deviation. At the next iteration step, the observations with the eliminated data from the previous step are then used again to compute the average and the std and the data are flagged again. At a certain point of the iterative process no more data will be rejected and we will consider the NQC to have converged.

The procedure of objective analysis and standard deviation computation is repeated using the steps outlined in the previous paragraph. The iterations continue until there are no more observations rejected, thus ending the NQC procedure. This procedure is shown in Fig. 2.2. The proposed quality control procedure is nonlinear because in the iterative procedure each step depends on the results of the previous one. We called this procedure NQC starting from this Thesis. The success of the NQC relies on three main factors such as: the appropriate subdivision of the domain,in order to calculate the standard deviation, the threshold standard deviation to eliminate the non representative data and the choice of OA parameters to grid the data at intermediate steps. The above choices will directly impact on the convergence of the procedure and the number of iterations required. In contrast, classical quality control procedure simply recalculates the standard deviation after the data removal without any gridding that changes the structure of the error field.
Here we focus on producing monthly mean climatologies of temperature and salinity in a $0.25^\circ \times 0.25^\circ$ grid for three regions of the world ocean using the profiling float data from WOD18.

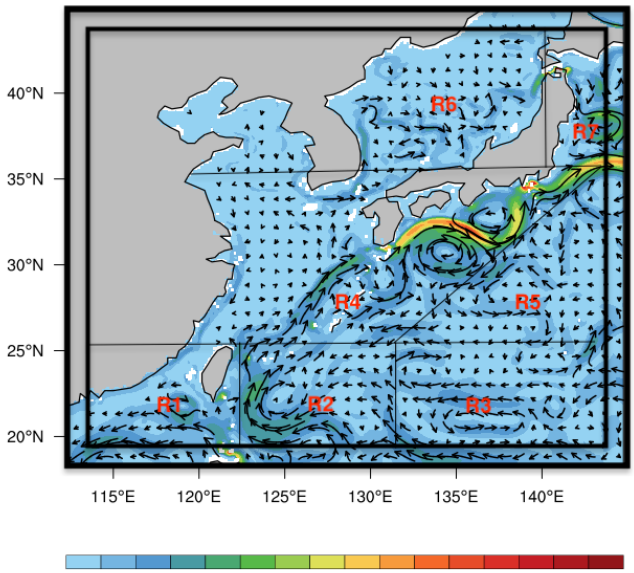
2.3 Application of the NQC Procedure

Three study regions were chosen for the NQC application : North West Pacific (NWP), North Atlantic (NA) and South Atlantic (SA), as shown in Figures. 2.3, 2.6 and 2.9. The WOD18 PFL profiles for temperature and salinity are used for the period 2003-2017. We have selected all those profiles and observed values for which we have BOTH quality flag value (WODf and WODfp) equal to "0". The target is a monthly mean climatology in a $0.25^\circ \times 0.25^\circ$ grid.
The NQC procedure is carried out using regime oriented subdivisions to compute the standard deviations. The regime oriented subdivision of these regions is based mainly on the ocean currents and the topography, considering the difference between open ocean and shelf/slope areas.

2.3.1 NQC applied in the North Western Pacific

The NWP domain is selected from 18°N to 44°N and 115°E to 144°E. The subdivision of NWP is taken from the study by Jia et al. (2016) with some modifications, as shown in Figure 2.3. Regions R2 and R3 include the North Pacific subtropical counter current ,and R1 the South

China Sea. Region R4 is the Kuroshio western boundary current region, and R5 contains North Pacific subtropical mode waters and the Kuroshio recirculation. Regions R6 and R7 are the southern part of the Japan sea and a small part of the Oyashio current, respectively. The histogram of the number of PFL profiles available from WOD18 is shown in Fig.2.4 which highlights that a relatively scarce number of PFLs are available for the NWP region.



(a)

Fig. 2.3: Regime oriented subdivision of North West Pacific.

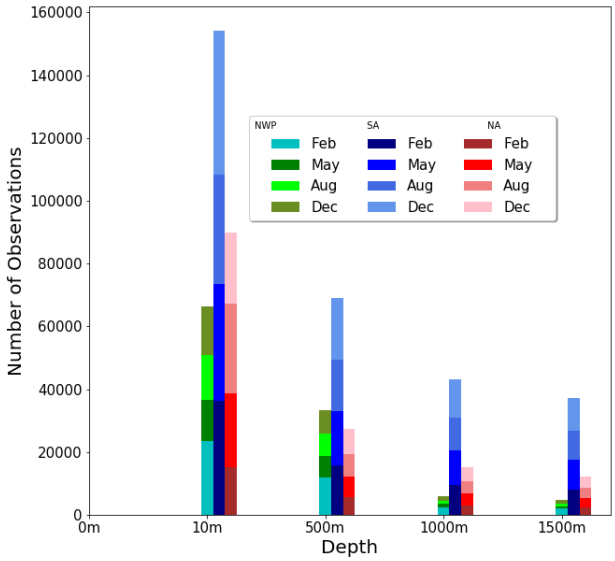


Fig. 2.4: Number of observations with WOD QC for different depths and months used in this study.

The gridded field of temperature and salinity before and after the application of NQC is shown in Fig. 2.5. The WODf and WODfp quality control procedure clearly does not eliminate outliers that are evident on the western side of the Okinawa island chain, especially for salinity. For temperature, the smoother estimate of climatology is evident in the noise in the 25°C isoline contour in the open ocean subtropical recirculation area.

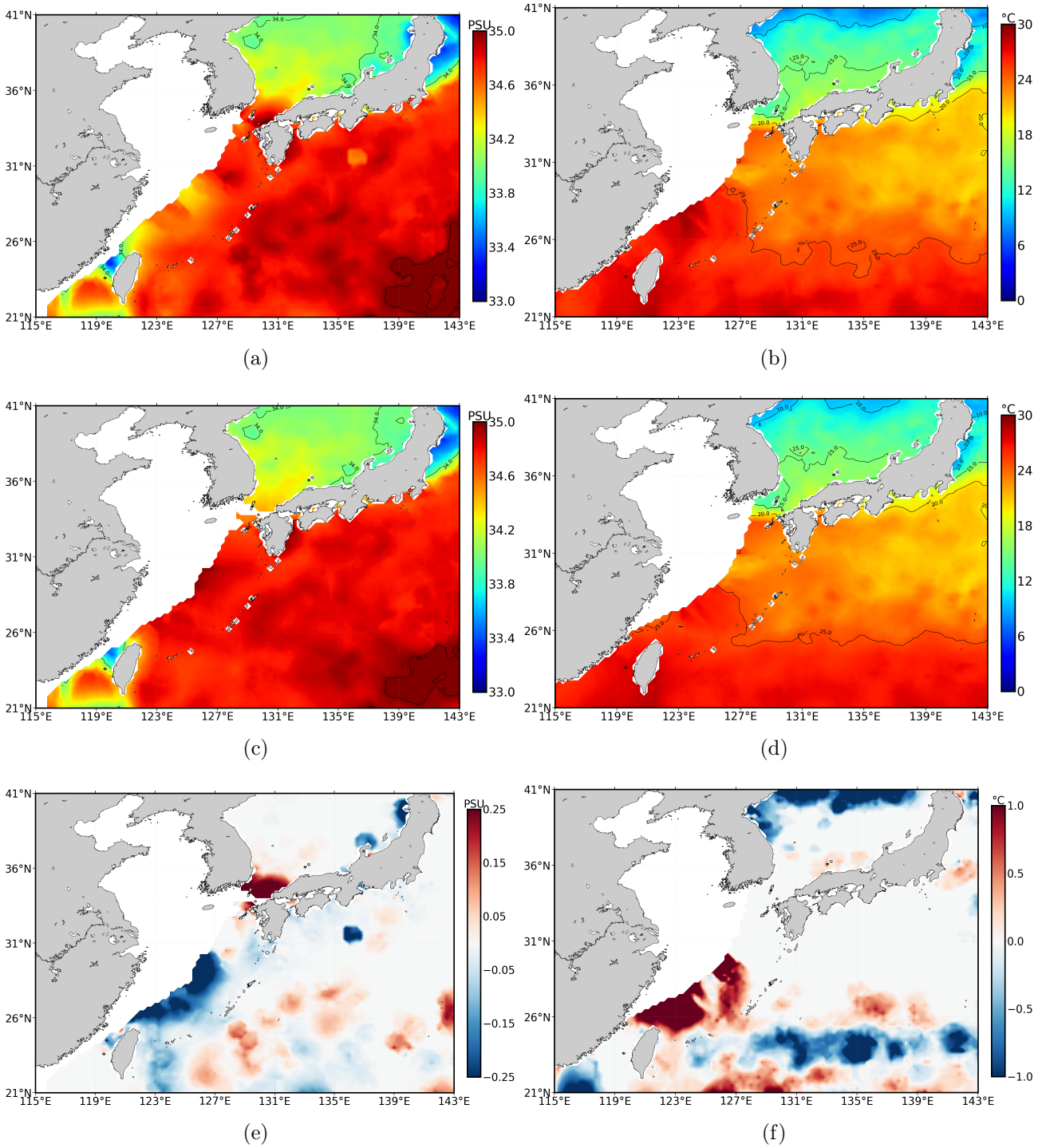


Fig. 2.5: Salinity (May)((a),(c) & (e)) and Temperature (February) ((b),(d) & (f)) mapping at 10m for North West Pacific using regime oriented division scheme. First row represents OA mapping using only WOD QC check (before application of NQC), second row is NQC after third iteration, while third row is difference of WOD QC and NQC check.

2.3.2 NQC applied in the North Atlantic

The NA domain was selected from 29°N to 51°N and 76°W to 6°W, including the North Atlantic subtropical gyre. The subdivision of North Atlantic follows the major ocean currents of the region. The Gulf stream off Cape Hatteras and its free jet part in R1 and R2, comprehensive of the rings and mesoscale eddies in the vicinity of the Stream, in the subtropical Gyre area. R3 and R4 contain part of the North Atlantic drift area, comprehensive of the Azores current system and the Gibraltar area inflow/outflow system. R5 and R6 consider the northern side of the North Atlantic drift, the first flowing northeastward toward the Gulf of Biscay while the other dominated by open ocean mesoscale eddies. Last R7 contains the eastern portion of the subpolar gyre with the consideration of the Labrador sea western boundary current system.

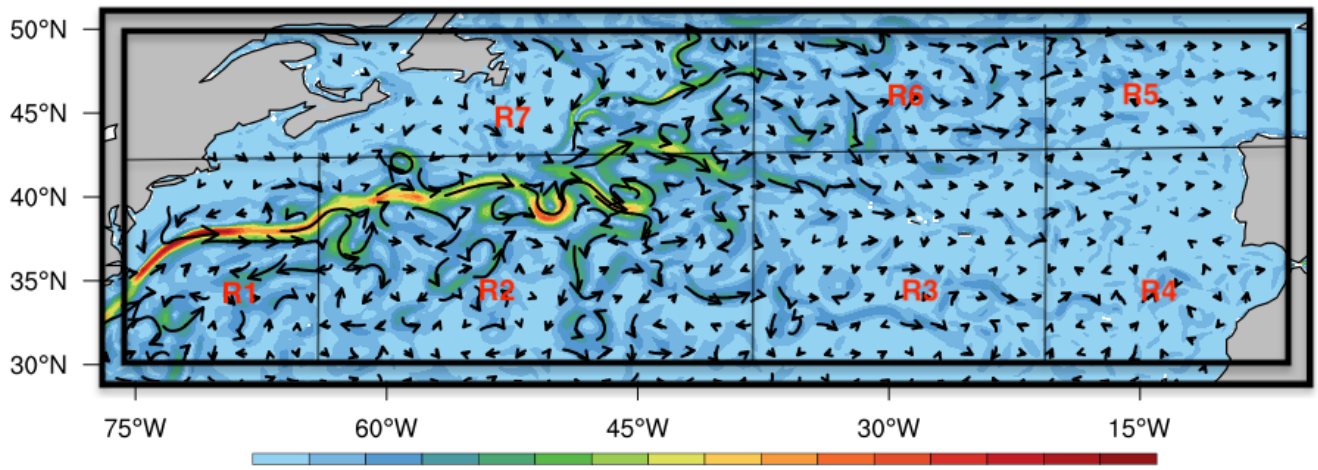


Fig. 2.6: Regime oriented subdivision of North Atlantic.

Similarly to the NWP region, the NQC removes the outliers, as shown in Fig. 2.8. The removal is particularly evident for the anomalous salinity values around 56°W and 43°N, as well as 46°W and 42°N. It is now evident that these signals are not consistent with climatological estimates, but are due to a subsampling of the oceanic eddy variability. Essentially, these observations are not representative of climatological temperature and salinity values. The difference between WOD QC and NQC in the mapped fields is shown in Fig. 2.8 highlighting the locations where data is being removed.

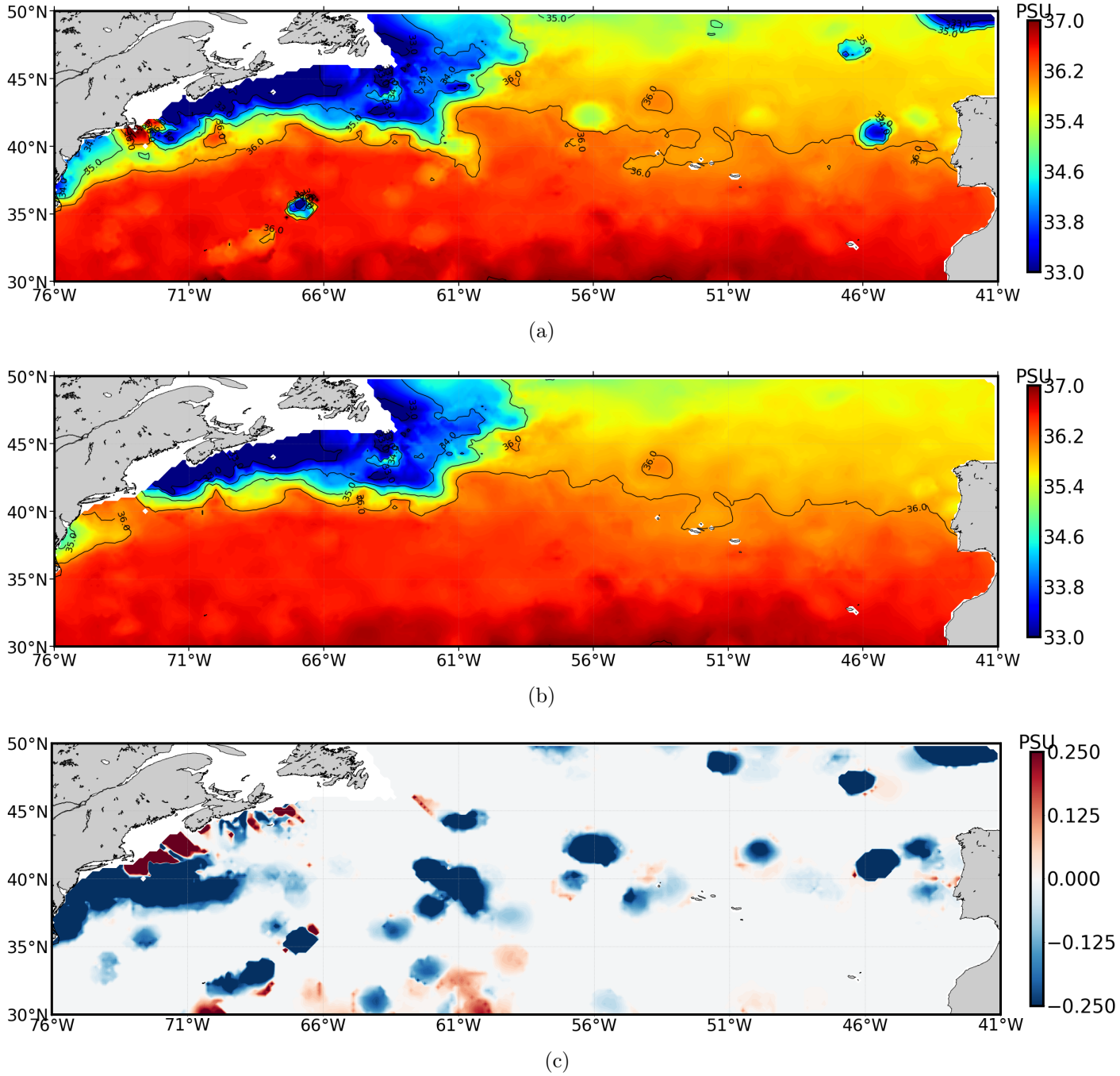
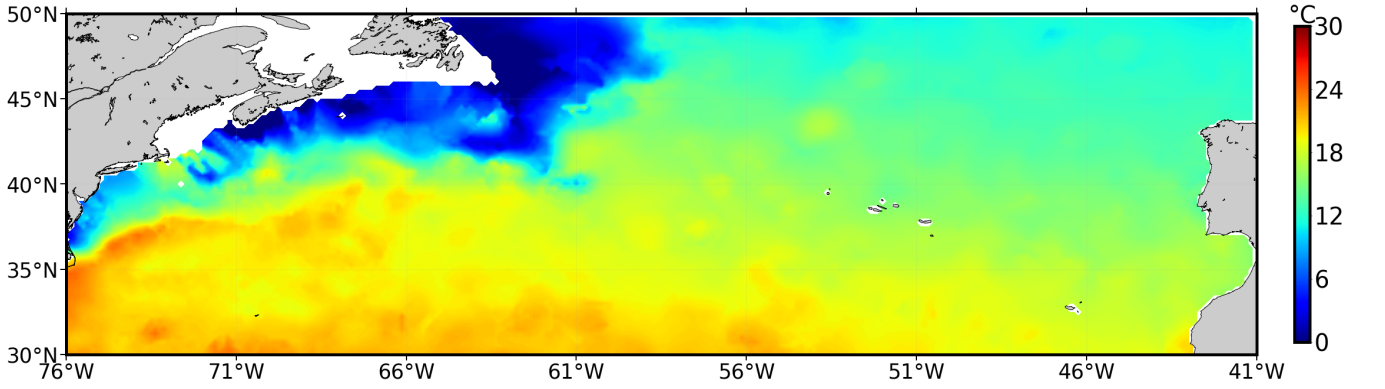
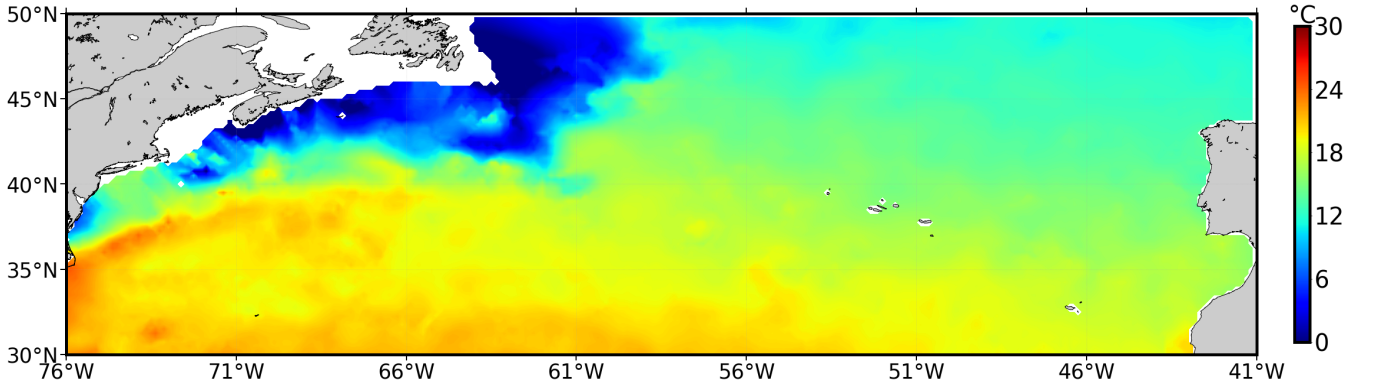


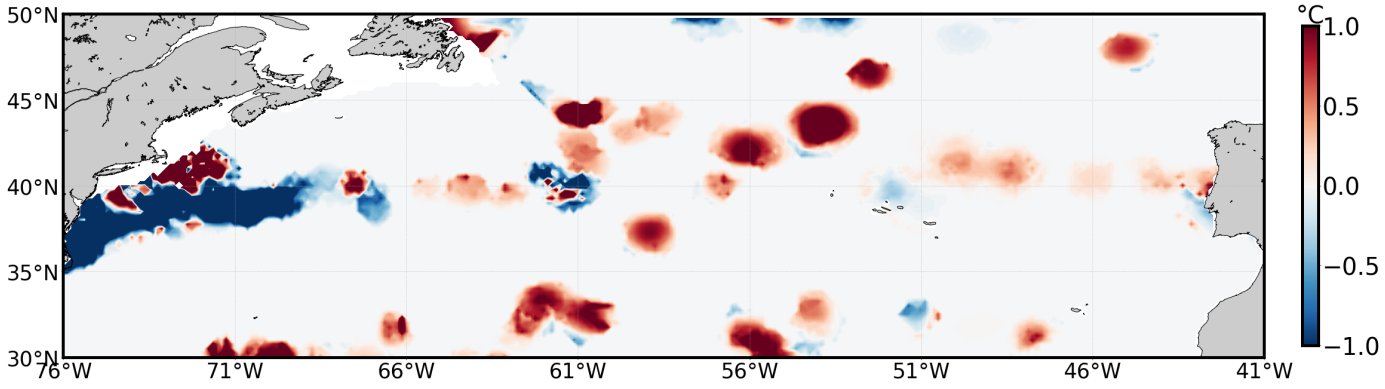
Fig. 2.7: Salinity (February) (a),(b) mapping at 10m for North Atlantic using regime oriented division scheme. (a) represents mapping using only WOD QC (before application of NQC), (b) is NQC processed, while (c) is the difference of WOD QC and NQC check.



(a)



(b)



(c)

Fig. 2.8: Temperature (February) (a),(b) mapping at 10m for North Atlantic using regime oriented division scheme. (a) represents mapping using only WOD QC (before application of NQC), (b) is NQC processed, while (c) is the difference of WOD QC and NQC check.

2.3.3 NQC applied in the South Atlantic

The SA domain extends from 10°S to 70°S and 73°W to 7°E. It is divided into ten dynamical regime oriented subregions (Lothar and Matthew, 1999), as shown in Fig. 2.9. R1, R2 and R3 contain the South Equatorial current region. R7 is the region of confluence between the Brazilian current moving southward in R4 and the Malvinas current moving toward north in R8. R6 and R8 represent the Benguela and Agulhas current and its retroflection, respectively. R5 and R9 represent a central part of the South Atlantic subtropical gyre. R10 contains the Falkland current and the border of the Antarctic circumpolar current.

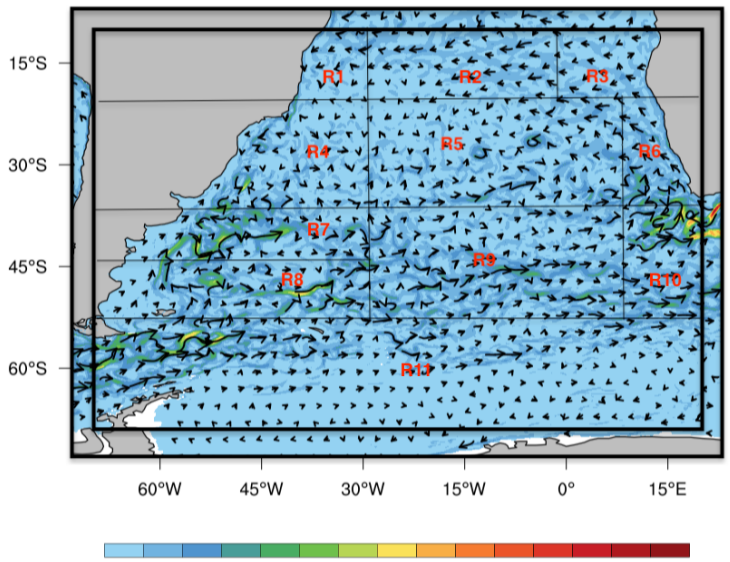


Fig. 2.9: Regime oriented subdivision of South Atlantic.

The gridded field of temperature and salinity for the SA is shown in Fig. 2.10. In this region, different types of outliers remain after the initial WODf and WODfp check. They are removed by the NQC as seen from the difference field in Fig. 2.10. In this region it is particularly important to apply NQC if a smooth climatology needs to be estimated.

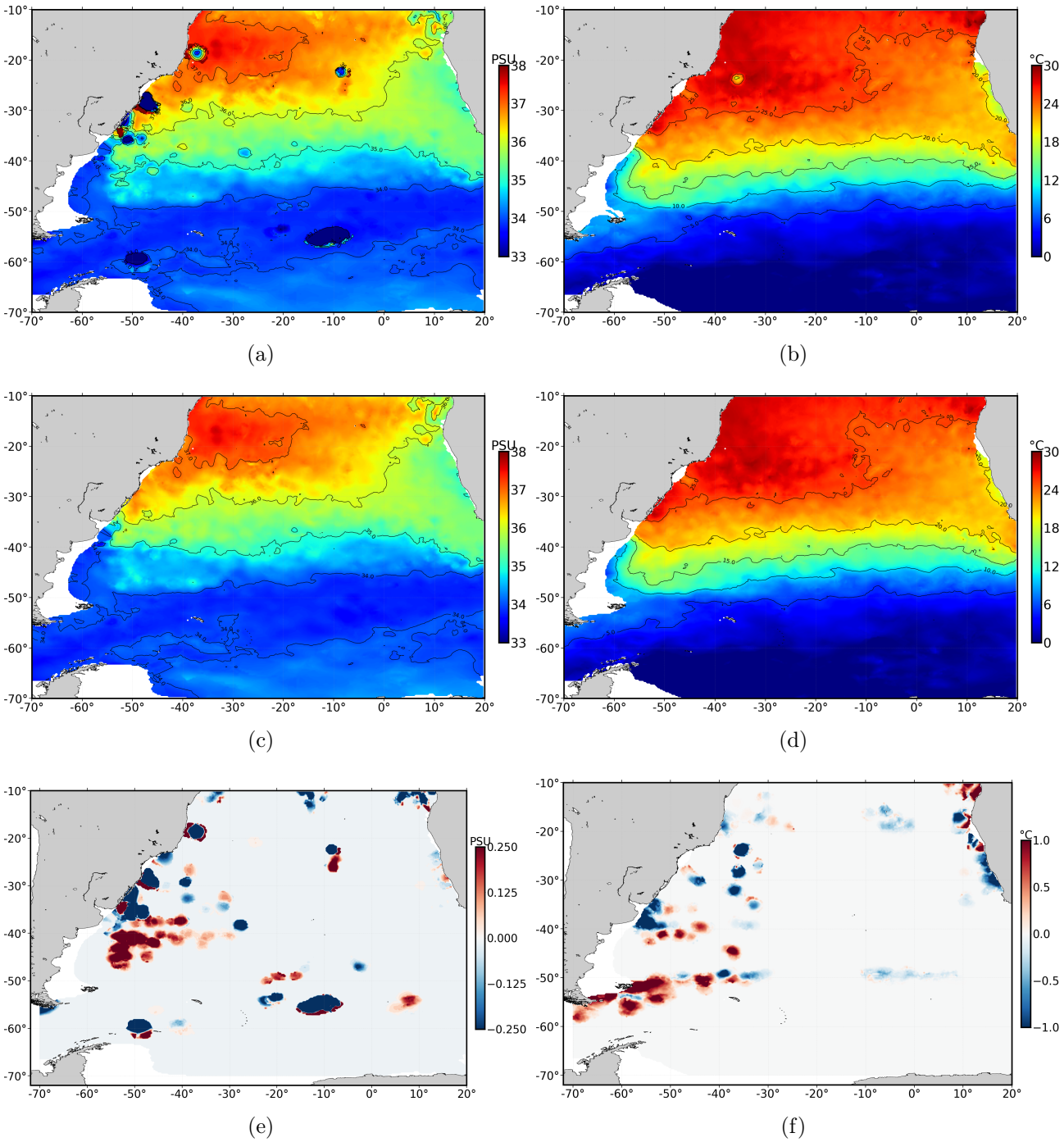


Fig. 2.10: Salinity (May) ((a),(c) & (e)) and Temperature (February) ((b),(d) & (f)) mapping at 10m for South Atlantic using regime oriented division scheme. First row represents mapping using only WOD QC (before application of NQC), second row is after application of NQC, while third row is difference of WOD QC and NQC check.

As the ocean variability is large in the upper ocean layers, the REs or sampling errors are more frequent in the surface layers, and larger amounts of data are rejected for the 10m and 500m layers (not shown) compared to other depths. The resulting standard deviation with NQC should therefore decrease with respect to the simple WOD quality control checks. Figure 2.12 shows for all three regions that in fact the difference in standard deviations between the two data sets is more prominent for the upper layers.

2.4 Sensitivity of the NQC Procedure to Subregions

In this section we show the sensitivity of NQC to the different domain subdivisions and to the different methods used to calculate the mean in each subregion.

Regarding the choice of different subregions, we compared a classical $5^\circ \times 5^\circ$ regular subdomain decomposition to the regime oriented subdivision previously discussed. The number of observations rejected after NQC with these two different subdivisions are shown in Fig. 2.11. Both subdivision schemes for NQC eliminate the outliers, however for the regular subdivision, the method eliminates larger amounts of data with respect to the regime oriented subdivision in all domains. For regime oriented subdivisions, the NQC scheme converges successfully at the fourth iteration for all domains and rejects a maximum of 15% of data for all the three regions. NQC with regular division did not converge at all in most of the cases even after seventh or eighth iteration, while NQC with regime oriented division is converging mostly before 4th iteration. This is clearly due to random occurrence of water masses in the regular division and as a result the statistics of a regular box is not an effective threshold either rejecting so much data or not eliminating any outliers. The NQC with regular division in SA domain has resulted to eliminate 85% of data after 4th iterations and this shows that the method does not converge. In Figures 2.13, and 2.14 we show the distribution of the rejected data using the two methods, the regular and the regime oriented NQC. Moreover, Tables 2.1 and 2.2 also show the distribution of the rejected data using the two methods, the regular and the regime oriented NQC. It is clear that the regular domain subdivision rejects much more uniformly across the different subareas while the regime oriented NQC is single point rejected data as expected for outliers.

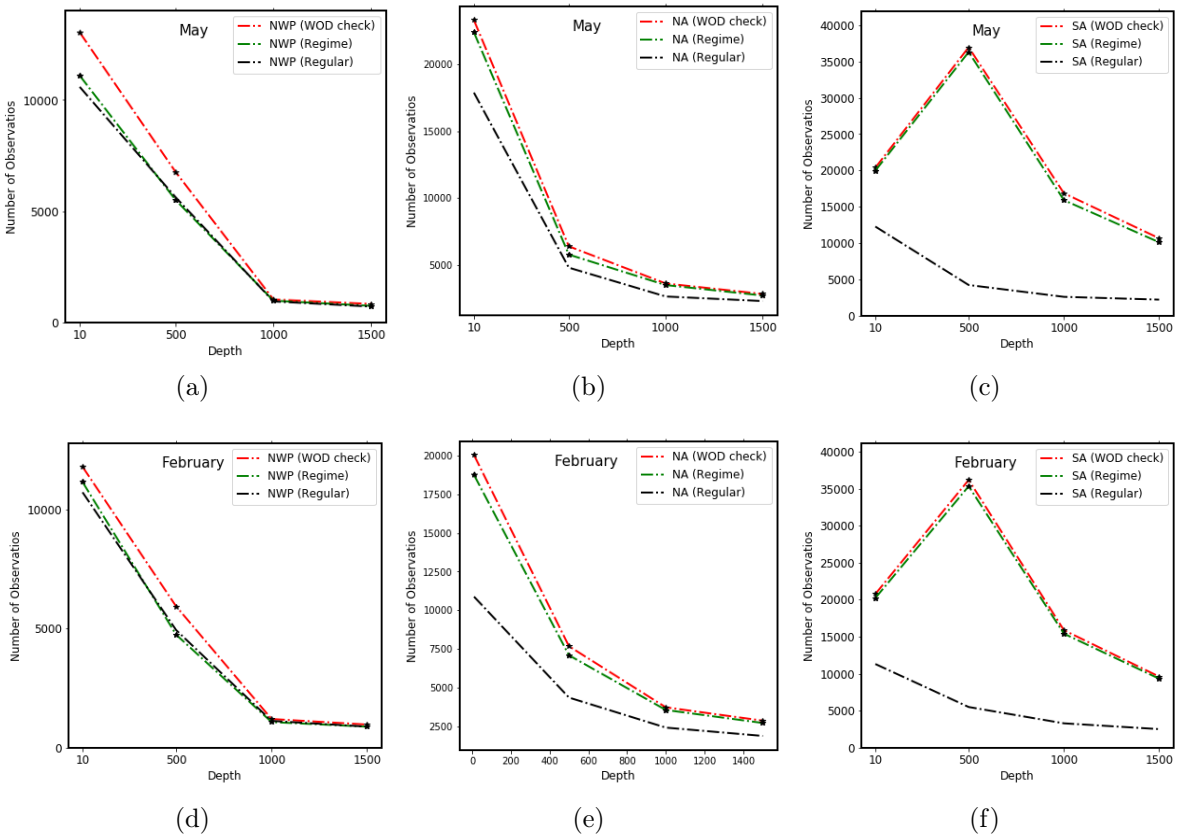


Fig. 2.11: Number of observations for Salinity (May) (first row) and Temperature (February) (second row) with WOD QC and NQC check (fourth iteration) with regime and regular division schemes for NWP ((a) and (d)), NA((b) and (e)), and SA((c) and (f)).

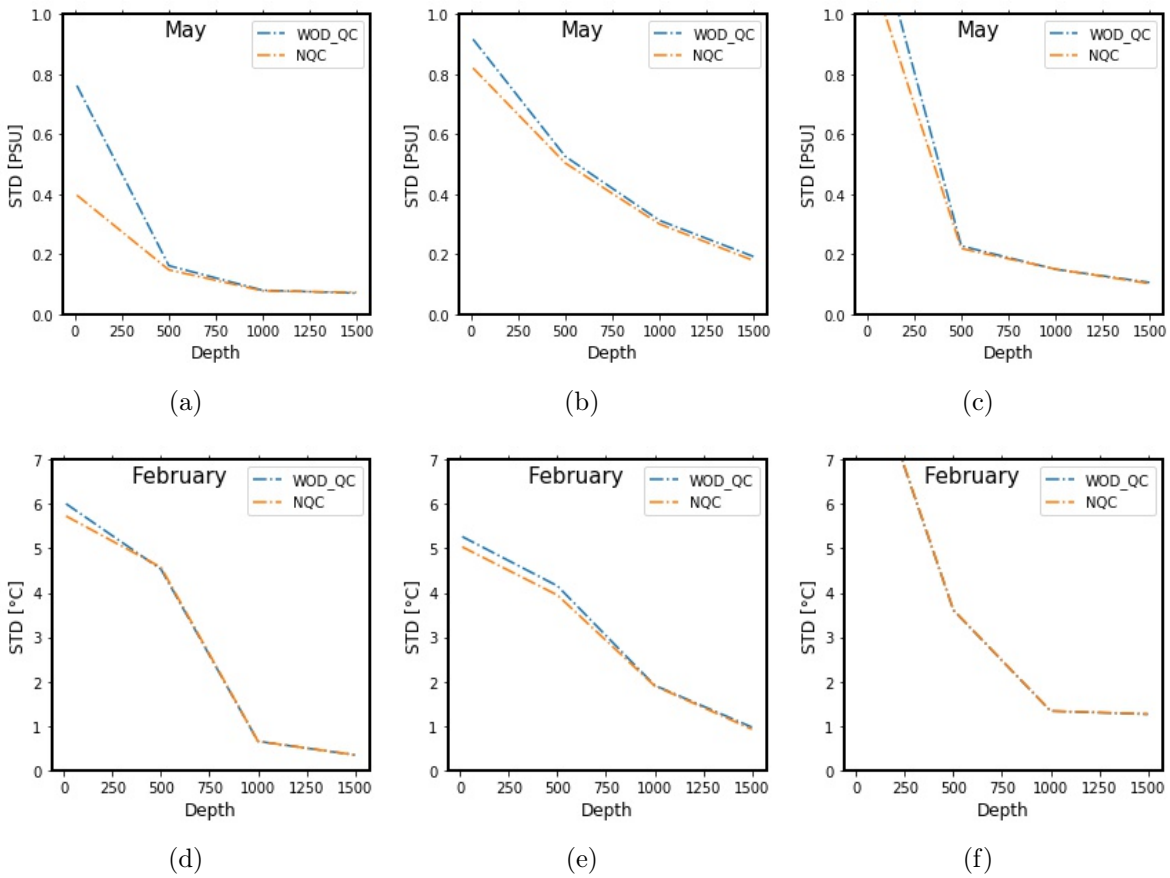


Fig. 2.12: Standard deviation for different depths for Salinity (May, first row) and Temperature (February, second row) for NWP ((a), (d)), NA ((b), (e)), and SA ((c), (f)).

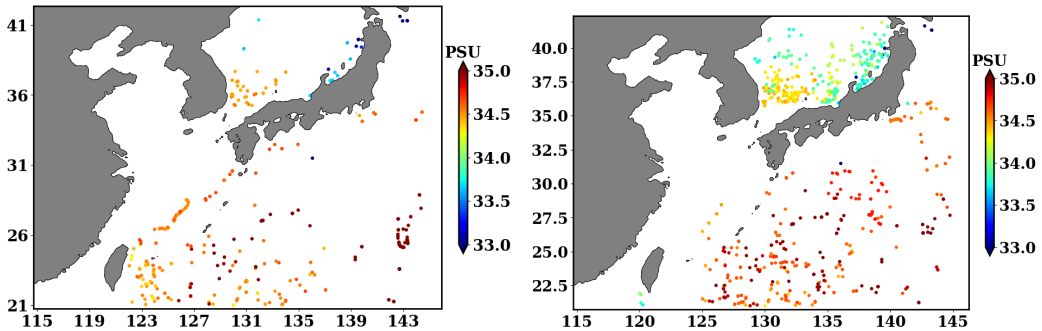
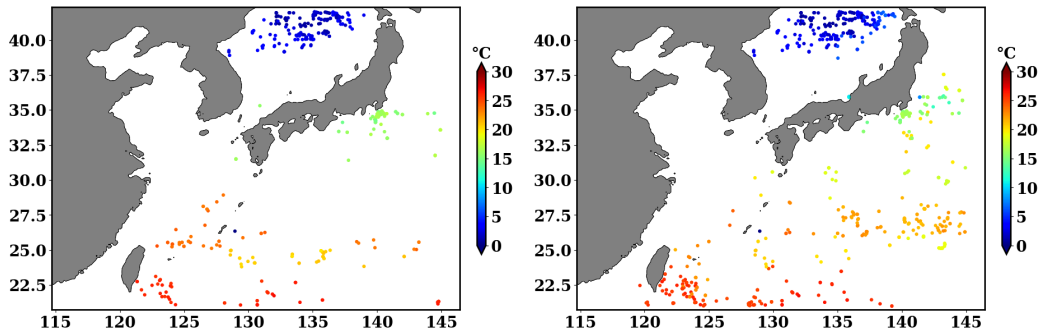


Fig. 2.13: Scatter plot of outliers eliminated for regime oriented division (a) and regular division in (a) at 10m for Salinity for month of February,



(a) (b)

Fig. 2.14: Scatter plot of outliers eliminated for regime oriented division (b) and regular division in (c) at 10m for Temperature for month of February,

Depth	%Salinity (Regime)	%Temperature (Regime)	%Salinity (Regular)	%Temperature (Regular)
10m	10%	5%	9%	12%
500m	11%	20%	20%	26%
1000m	12%	10%	16%	15%
1500m	10%	9%	9%	12%

Table 2.1: Percentage of rejected observations in NWP for Temperature and Salinity for regime and regular divisions at different depths.

Depth	%Salinity (Regime)	%Temperature (Regime)	%Salinity (Regular)	%Temperature (Regular)
10m	6%	5%	28%	24%
500m	7%	6%	22%	22%
1000m	4%	4%	20%	28%
1500m	5%	3%	22%	23%

Table 2.2: Percentage of rejected observations in NA for Temperature and Salinity for regime and regular divisions at diffrent depths.

2.5 Summary and Conclusions

We have described a new method for the quality control of historical observational data in order to produce long term climatological estimates of temperature and salinity in the world's oceans. The new method is characterised by an iterative procedure that computes the mean

and standard deviations of a series of interpolated fields and rejects outliers on the basis of a threshold standard deviation value.

The NQC detects outliers that are due to the non-representativeness of observations with respect to climatological estimates. The procedure entails the subdivision of the domain into dynamical homogenous regions, "regime oriented" regions, and the computation of gridded fields using OA to estimate the subregional mean. The procedure iterates and continues to eliminate the data until the convergence is reached, i.e. no more data are rejected.

This new procedure was applied in the North West Pacific, North Atlantic and South Atlantic regions. A previous application was carried out only for the NWP and the sensitivity to the choice of subregion was not assessed. The sensitivity of the NQC to the regional subdivisions, i.e., a regime oriented with respect to a regular, $5 \times 5^{\circ}$ subdivision, was also evaluated. The NQC with regime oriented subdivisions was found to converge after four iterations and to reject a relatively low number of profiles compared to regular ones which did not even converge. We argue that a regime oriented quality control is necessary to estimate climatology for temperature and salinity in the world oceans. We also demonstrated that this NQC procedure is applicable to very different regions of the world oceans, and thus it will enable us to define a low frequency climatology with more representative data. The choice of the division of the domain is one of the key steps in the elimination of non- representative data. It would therefore be interesting to apply a machine learning algorithm such as self-organised maps (Maze et al., 2017) in order to have the division of the dynamically homogenous regions based on water mass properties and compare our results with the previous studies such as Bhaskar et al. (2017). The procedure could also be applied to biogeochemical data. In the future we plan to consider both new data sets and machine learning algorithms to help in the automatisation of the NQC algorithm for the global ocean.

Chapter 3

A New Temperature and Salinity Climatology for the Global Ocean

3.1 Introduction

Defining the climatological state of the ocean is a formidable task. Climatology can be defined as the study of the statistics of environmental variables characterizing the ocean's physical and biochemical state. Specifically, this work concentrates on the estimation of the monthly mean values of temperature and salinity in the global ocean derived from historical observational records. Climatology is an essential input to numerical ocean models for initialization and validation purposes and it is intrinsically useful to understand climate anomalies.

The first challenging task for climatological studies is collecting observational records and harmonizing them, in terms of metadata and quality control. This was described in the second chapter of this thesis and the same concepts will be used in this third chapter where we look at the second step in building climatology, that is, spatial analysis. As defined by Daley (1993): "Spatial analysis is the estimation by numerical algorithm of atmospheric state variables on a three-dimensional regular grid from observations available at irregularly distributed locations." These numerical algorithms are based on statistical assumptions and theory which have greatly evolved over the past twenty years. These techniques are being referred to as interpolation schemes.

The first global ocean gridded climatology was reported by Levitus (1982) and has been the basis for all other subsequent estimates. The World Ocean Atlas (WOA) is an improved version

of Locarnini et al. (2018) and, since 1994, this has been regularly updated every four years. WOA uses the basic interpolation schemes defined by Barnes (1964). Another global ocean climatological estimate is the World Argo Hydrographic Global Ocean Climatology (WAGHC, Gouretski (2019)), a first global ocean climatology that is produced on isopycnal and isobaric levels. The WAGHC interpolation scheme is based on Objective Analysis (OA) following Gandin (1960).

Some of the challenges in previous climatologies are related to the specific algorithm that interpolates observations across land-sea boundaries. As most of the past interpolating algorithms do not naturally inherit an objective method that forbids using observations across the land-sea boundaries, we analysed the differences among climatologies in the vicinity of peninsulas. For instance, the Isthmus of Panama, a narrow land area between the Caribbean Sea and the Pacific Ocean, is a critical area where observations could be improperly used across disconnected oceans. Figure 3.1 shows several available climatologies in this area. It is clear that the climatologies give a very different estimate: for WAGHC and OA schemes it is evident that information spreads from the Pacific to the Atlantic along the Columbian coasts. On the other hands, WOA, despite the usage of separate first guess field across the Isthmus (Tim Boyer, personal communication) shows low salinities anomalies in the Gulf of Mexico. By contrast, Figure 3.1 demonstrates that the DIVA interpolation algorithm has completely suppressed the contamination of Pacific Ocean data in the Caribbean Sea and vice-versa. Another difference between the climatologies is evident along the Louisiana coasts of the Gulf of Mexico, where the Mississippi river outflow dominates. The climatologies are very different and this could be due to the algorithm used, the first guess and the number of data used in the analysis. Thus it is clear that climatology uncertainty has several sources that are difficult to point out and if the statistical interpolation method uses ad hoc assumptions, specific hypothesis reproducibility of the results is at stake. DIVA solves in an objective way the problem of interpolation of oceanographic observations across land boundaries but it is, as the other statistical models, making assumptions about the statistical distribution of ocean variables. Therefore, it is likely that a multi-model ensemble of all available climatologies would provide a more accurate solution, a shown later in this Chapter.

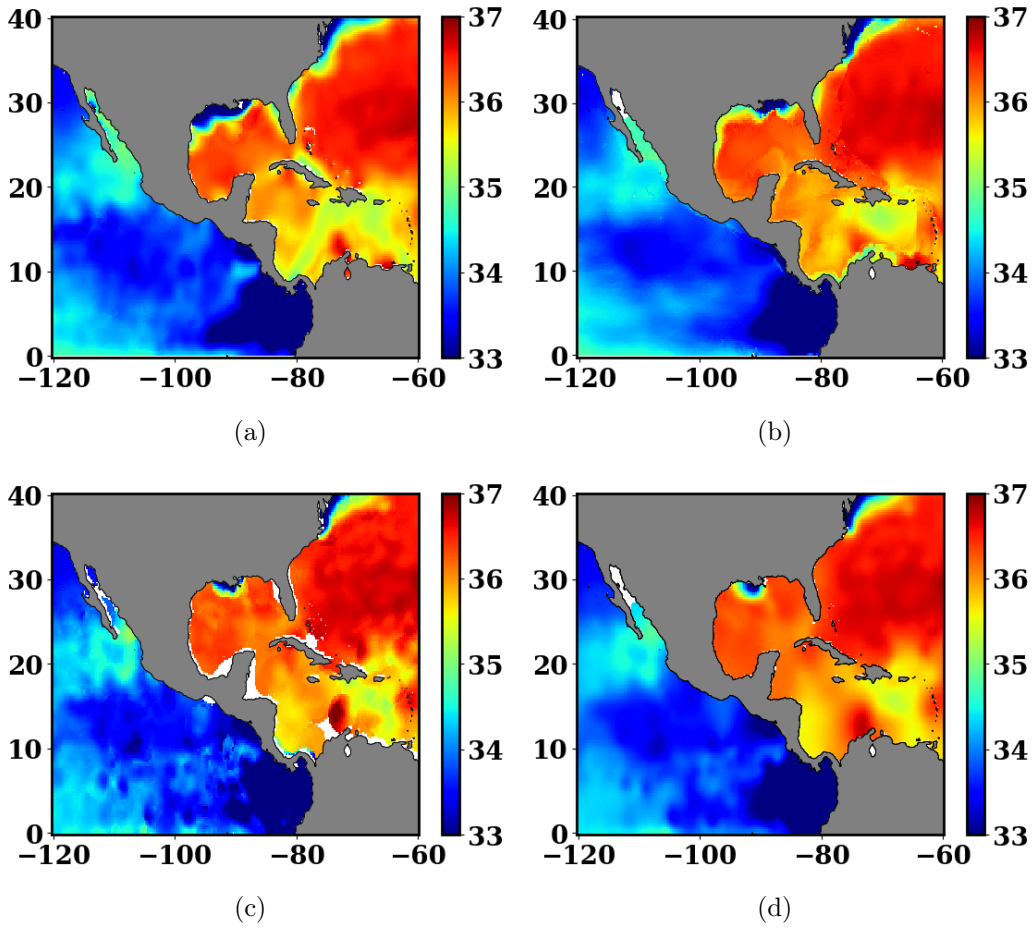


Fig. 3.1: Salinity(January) mapping of Panama Isthmus (a) WOA(correlation length = 214 km, uses all data from WOD18), (b) WAGHC (correlation length = 333 km, signal to noise ratio = 0.5 , uses data from WOD13, in particular OSD, CTD, PFL and APB and additional data from the Alfred Wegener Institute, Bremerhaven, and from different institutions in Canada), (b) OA estimate (correlation length = 300 km and error in the observations = 0.3, uses Dataset1), and (d) DIVA(correlation length = 300 km and N/S = 0.5, uses Dataset1).

The chapter's main objective is to estimate global ocean climatology using an advanced interpolating tool called Data-Interpolating Variational Analysis (DIVA), following proper quality control undertaken on the historical data set. In fact, the Nonlinear Quality Control algorithm developed in Chapter 2 of this thesis is applied. Furthermore, sensitivity experiments to interpolation parameters, such as the signal to noise ratio and the field correlation length, are also carried out. Finally, results are compared with the WOA18 and WAGHC datasets.

In section 3.2, the historical data set used for climatology is reviewed together with the quality control procedure. The interpolation scheme and the implementation domain, together with the choices of the interpolation parameters, are discussed in section 3.3. Monthly mean

temperature and salinity fields are compared with existing climatologies in section 3.4, while sections 3.6 and 3.7 conclude the Chapter.

3.2 Historical Databases

Two different versions of climatologies are being estimated based on two datasets extracted from the World Ocean Database 18 (WOD18, Garcia et al. (2018)). The first, called Dataset1 (see Table 5.1) uses multiple platforms, such as bottle data from OSD (Ocean Station Data) and CTD (Conductivity Temperature and Depth) from ship surveys, MRB (Mooring Buoy) and PFL (Profiling Floats). MRB profiles are only distributed across the equatorial and tropical region, while CTD, OSD and PFL profiles cover instead the global domain. The data from other available platforms were not used because we wanted to consider the concomitant measurement of temperature and salinity and an approximately equal number of profiles for the surface and the upper pycnocline. Therefore, XBT and MBT were discarded because only temperature measurement are available from these platforms. DRB, UOR and SUR were also not selected because they were only at the surface. Moreover, APB and GLD (gliders) were not used because they consisted in high temporal resolution measurements that were not considered to be appropriate for the climatological estimate. The observations in Dataset1 are taken from 1900 to 2017: the climatology estimated from this data set will be named SDC_V1. The second dataset, called Dataset2 (see Table 5.1), only contains profiles from floats that are autonomous neutrally buoyant vehicles equipped with several oceanographic sensors. It contains data from different floats such as PLACE, MARVOR, SOLO and APEX, etc. A real revolution in ocean observations started with the Argo program in 2000 where these floats started to be numerous in all the world ocean basins. In Dataset2, only profiling floats from 2003 to 2017 were considered, the majority of PFL being APEX floats. The PFL measurements earlier than 2003 were not considered because strongly affected by different problems such as pressure drift (Barker et al., 2011), an offset in the salinity due to biofouling (Wong et al., 2003), (Wong, 2008) and transmission errors. Therefore, we selected the consolidated profiles only from 2003 to 2017 to avoid erroneous observations. The amount of PFL data in the last fifteen years exceeds the amount of data available from all the other platforms, as shown in Figure 3.2.

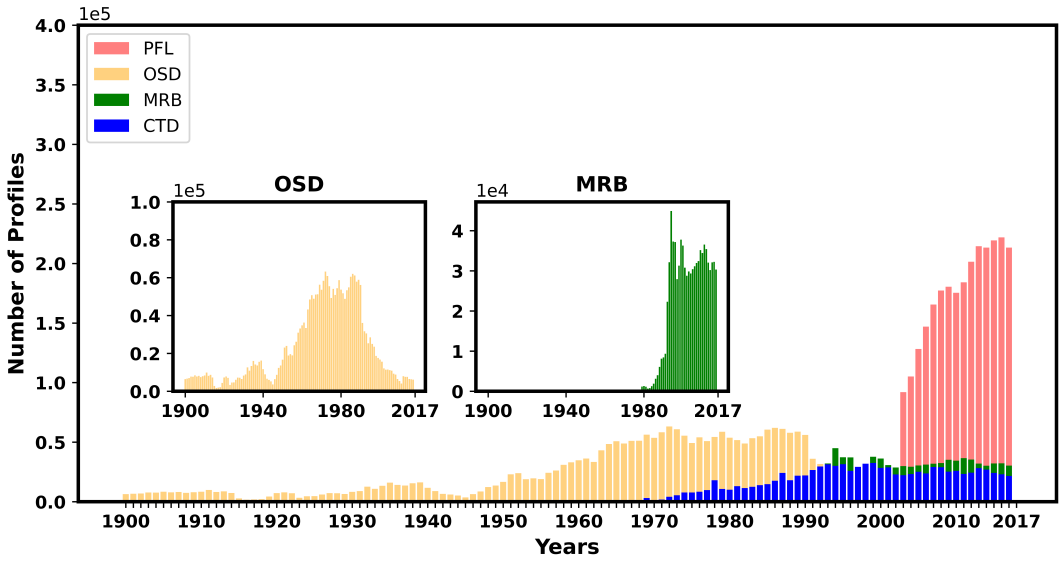


Fig. 3.2: Number of profiles from four different measuring platforms used in this Thesis extracted from WOD18.

3.2.1 Application of Nonlinear Quality Control Procedure

In this climatology, we only accepted profiles flagged as positive by the WOD18 quality flags. Two types of quality flags are available: a) an individual value flag (WODf) available for each variable measured at a certain depth; b) a profile flag (WODfp) which considers a statistical quality check that was described in Chapter 2.

An additional quality control step, the Nonlinear Quality Control procedure (NQC) described in Chapter 2, was implemented with the following assumptions:

(i) The domain is divided into 5×5°boxes, where mean and standard deviations (std) are computed.

(ii) Data is eliminated outside 2 std in each box and the procedure is repeated until convergence is achieved.

The number of observations before and after the application of NQC are show in Figure 3.3. The application of NQC has eliminated outliers and non-representative data, as demonstrated in Chapter 2. Data eliminated by means of NQC approximately constitute less than 15% of the total for all months.

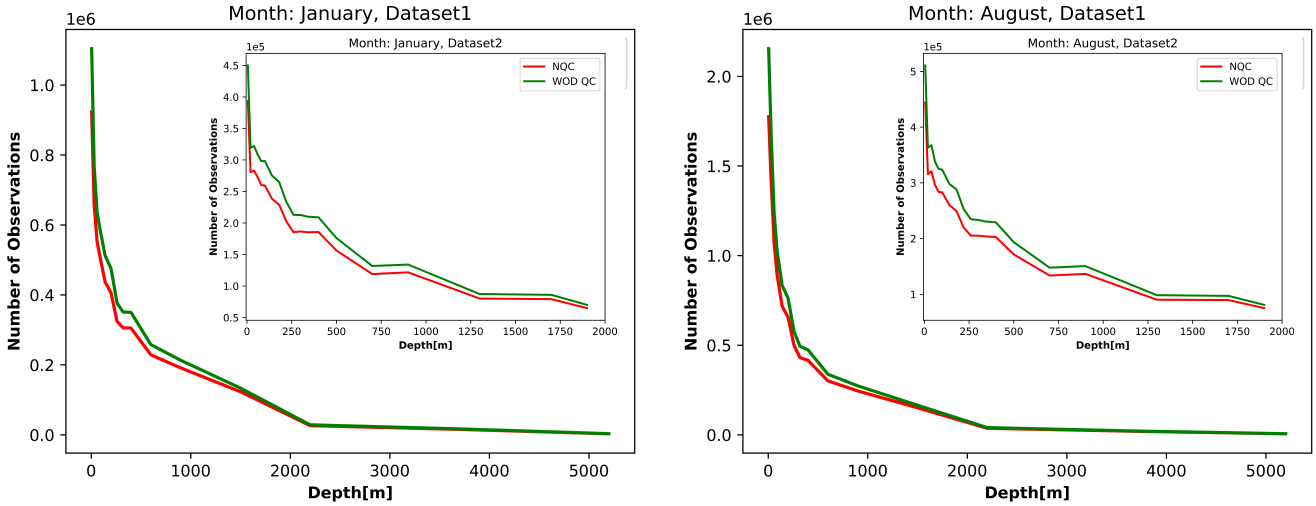


Fig. 3.3: Number of observations usind WOD QC and NQC for both Dataset1 and Dataset2 for Temperature: January (left) and August(right).

Dataset name	Temperature profiles	Temperature measurements	Salinity profiles	Salinity measurements
Dataset1	6,012,750	803,362,255	5,265,504	757,320,791
Dataset2	1,658,955	384,430,391	1,557,989	362,928,173

Table 3.1: Number of profiles and measurements in Dataset1 and Dataset2.

3.3 Climatology Construction

3.3.1 An Overview of the Interpolation Algorithm

DIVA is based on the Variational Inverse Method (VIM) applied on a curvilinear orthogonal grid using a Finite Element Method (Barth et al., 2014). The method is equivalent to Optimal Interpolation (OI). The major difference between DIVA and OI is in the proper consideration of land boundaries, as explained in the introduction.

In DIVA, a cost function is minimized and contains three terms: the misfit between the observations and reconstructed field, the regularity or smoothness constraint and the advection constraint. This cost function can be written as:

$$J[\phi] = \sum_{i=1}^{N} \mu_{\mathrm{i}}[d_{\mathrm{i}} - \varphi(x_{\mathrm{i}}, y_{\mathrm{i}})]^2 + ||\varphi - \varphi_{\mathrm{b}}||^2 + J_{\mathrm{c}}(\varphi), \tag{3.1}$$

where d_i are the observations at the location $(x_\mathrm{i}, y_\mathrm{i})$, φ is the target field in the regular grid, also called analysis, φ_b is the first guess field or "background" and μ_i are weights derived from specific error estimates Troupin et al. (2012) and the correlation length L, which will be described later. J_c is the advection constraint which considers that variable gradients should be along the coasts. Furthermore, the smoothness constraint is defined as:

$$||\varphi - \varphi_\mathrm{b}||^2 = \alpha_2 \int_\Omega (\Delta\Delta\varphi : \Delta\Delta\varphi + \alpha_1 \Delta\varphi.\Delta\varphi + \alpha_0 \varphi^2) d\Omega, \tag{3.2}$$

The non-dimensional form of the cost function is:

$$\tilde{J}[\phi] = \sum_{i=1}^{N} \mu_\mathrm{i} L^2 [d_\mathrm{i} - \varphi(x_\mathrm{i}, y_\mathrm{i})]^2 + \int_{\tilde{\Omega}} (\tilde{\Delta}\tilde{\Delta}\varphi : \tilde{\Delta}\tilde{\Delta}\varphi + \alpha_1 L^2 \tilde{\Delta}\varphi.\tilde{\Delta}\varphi + \alpha_0 L^4 \varphi^2) d\tilde{\Omega} + J_\mathrm{c}(\varphi). \tag{3.3}$$

In DIVA the following values are assumed:

$$\alpha_0 L^4 = 1 \tag{3.4}$$

$$\alpha_1 L^2 = 2 \tag{3.5}$$

$$\alpha_2 = 1 \tag{3.6}$$

$$\mu_i L^2 = 4\pi \frac{\sigma^2}{\epsilon_i^2} \tag{3.7}$$

From (3.7), we see that μ are defined as the signal variance σ^2 of the background with respect to the error variance of the observations, ϵ_i^2. The kernel, equivalent to the correlation function in optimal interpolation is defined as:

$$K(r) = \frac{r}{L} K_1(\frac{r}{L}), \tag{3.8}$$

where r is the distance between the data point and the analysis grid point and K_1 is the Bessel function of the second kind. For more details on the solution method see Barth et al. (2014).

The best estimate or analysis is dependent on the values of two key parameters that are decided a priori, the **Correlation Length** (CL) and the **N/S**, that is, $\frac{1}{\mu_i}$. The CL determines the distance over which observations will influence a particular estimate of the ϕ field at the

given grid point. Large CL values indicate a larger-scale weighted average of the observations resulting in a smoother field, while smaller values will allow for the resolution of smaller-scale features, resulting in a noisier field.

Large values of N/S imply larger deviations in the analysis field from observations or, in other words, the analysis field closer to the background or first guess field. On the other hand, small values of N/S mean an analysis field closer to observations relative to the first guess field. The relative error field μ values in DIVA are computed using the method called the clever poor man approach Troupin et al. (2012). Moreover, the correct specification of the first guess field is of paramount importance in the estimation of the final analysis field and will be examined in the next sections.

3.3.2 Horizontal and Vertical Analysis Domain

The global domain for the analysis extends from 0°E to 360°W and from 80°N to −80°S. The grid spacing is $\frac{1}{4}°$ in latitude and longitude. Bathymetry is specified from the GEBCO 30 sec data set (IOC, 2003). The number of non-uniform depth layers considered in this analysis are 45 (surface to 6000m) and 36 (surface to 2000m) for SDC_V1 and SDC_V2, respectively, as listed in Table 3.2. We considered a vertical discretization of layers, thus all measurements belonging to a 10m thick layer around each nominal depth are considered for interpolation at that specific depth. This prevents vertical interpolation of data which is a problematic issue when there are large vertical data gaps in a profile near the thermocline surface.

To better resolve the upper thermocline structure, a larger number of layers are defined from the surface to 500m, while the remaining levels are at a 100m (500m to 1900m) and 500m (1900 to 6000) distance apart. Data are grouped in monthly files in order to estimate the monthly climatology.

3.3.3 Background Fields

The choice of the first guess field or background field could be important when data are irregularly-spaced horizontally and vertically. Two types of backgrounds were tested in this study. The first type, called Background1, is a vertical profile corresponding to a spatial mean of the observations over the entire global ocean for each layer given in (see Table 5.1). The fields are shown in Fig. 3.4 (a) and (b) for Dataset1 and (c) and (d) for Dataset2, respectively.

The differences in these two backgrounds at the surface are due to the absence of observations on the continental shelves in Dataset2 since the PFL parking depth is 1000m or deeper. Thus, the surface salinity-minimum is absent in Background1 for Dataset2 due to river runoff in the continental shelves. The second choice of background, called Background2, is estimated by using DIVA starting from Background1, and using a correlation length of 1000 km and a small N/S ratio of 0.5. Figure 3.6 shows the difference of the SDC_V1 analysis with respect to Background1 and Background2 at the surface. In the case of the Background1, the difference is at large scales, adding and removing temperature and salinity across the latitudinal extension of the ocean basins. In the case of the analysis done with Background2, observations are used twice in the DIVA interpolation algorithm, adding only small scale increments to the initial guess. We can define this as a progressive refinement procedure that should better converge locally to the observations. We conclude that SDC_V1 is a progressive refinement of Background1, using twice the same observations with different correlation lengths to arrive to a statistically optimal solution. One way to decide which background is best is by computing analysis residuals, that is, the difference between the observations and the analysis interpolated back at observation locations. Figure 3.5 illustrates an important result in this chapter: when NQC is used, the root mean square residual analysis is insensitive to the choice of background. This is because the NQC eliminates outliers or non-representative data and this reduces the analysis' sensitivity to the specification of background. The difference of residuals between the analysis with Background1 and Background 2 using the NQC data set illustrate the insensitivity of the resulting analysis to this choice, except for some regions of intense mesoscale activity as the western boundary currents and the Antarctic Circumpolar Current northern border. As illustrated before, the Background2 is a refinement of Background1 using a large correlation function and using the data set twice, the second time with smaller correlation to produce the analysis. The convergence of this process depends on the quality of the input data set: if only the WOD QC input data set is used, i.e. outliers/non-representative data are left in the analysis, the choice of background becomes more important and the difference between residuals becomes more non-uniform and larger, especially for salinity. This is shown very clearly in Fig. 3.6 where the difference of residuals in the case of WOD QC input data set shows sensitivity to the Background chosen, especially for the salinity residuals. Thus, our analysis is carried out for both Dataset1 and Dataset2 with the spatial mean profile of Fig. 3.4.

No.	**Nominal Depth(m)**	**Layer**	**No.**	**Nominal Depth(m)**	**Layer**
1	5	0-10	24	370	365-375
2	10	5-15	25	400	395-405
3	20	15-25	26	450	445-455
4	30	25-35	27	500	495-505
5	40	35-45	28	600	595-605
6	50	45-55	29	700	695-705
7	60	55-65	30	800	795-805
8	70	65-75	31	900	895-905
9	80	75-85	32	1100	1095-1105
10	90	85-95	33	1300	1295-1305
11	100	95-105	34	1500	1495-1505
12	120	115-125	35	1700	1695-1705
13	140	135-145	36	1900	1895-1905
14	160	155-165	37	2200	2195-2205
15	180	175-185	38	2700	2695-2705
16	200	195-205	39	3200	3195-3205
17	220	215-225	40	3700	3695-3705
18	240	235-245	41	4200	4195-4205
19	260	255-265	42	4700	4695-4705
20	280	275-285	43	5200	5195-5205
21	300	295-305	44	5700	5695-5705
22	320	315-325	45	6000	5995-6005
23	340	335-345			

Table 3.2: Depth layers used for SDC climatology, the nominal depth is selected at the middle of each layers. The levels of SDC_V1 extend from 5m to 6000m while for SDC_V2 from 5m to 2000m

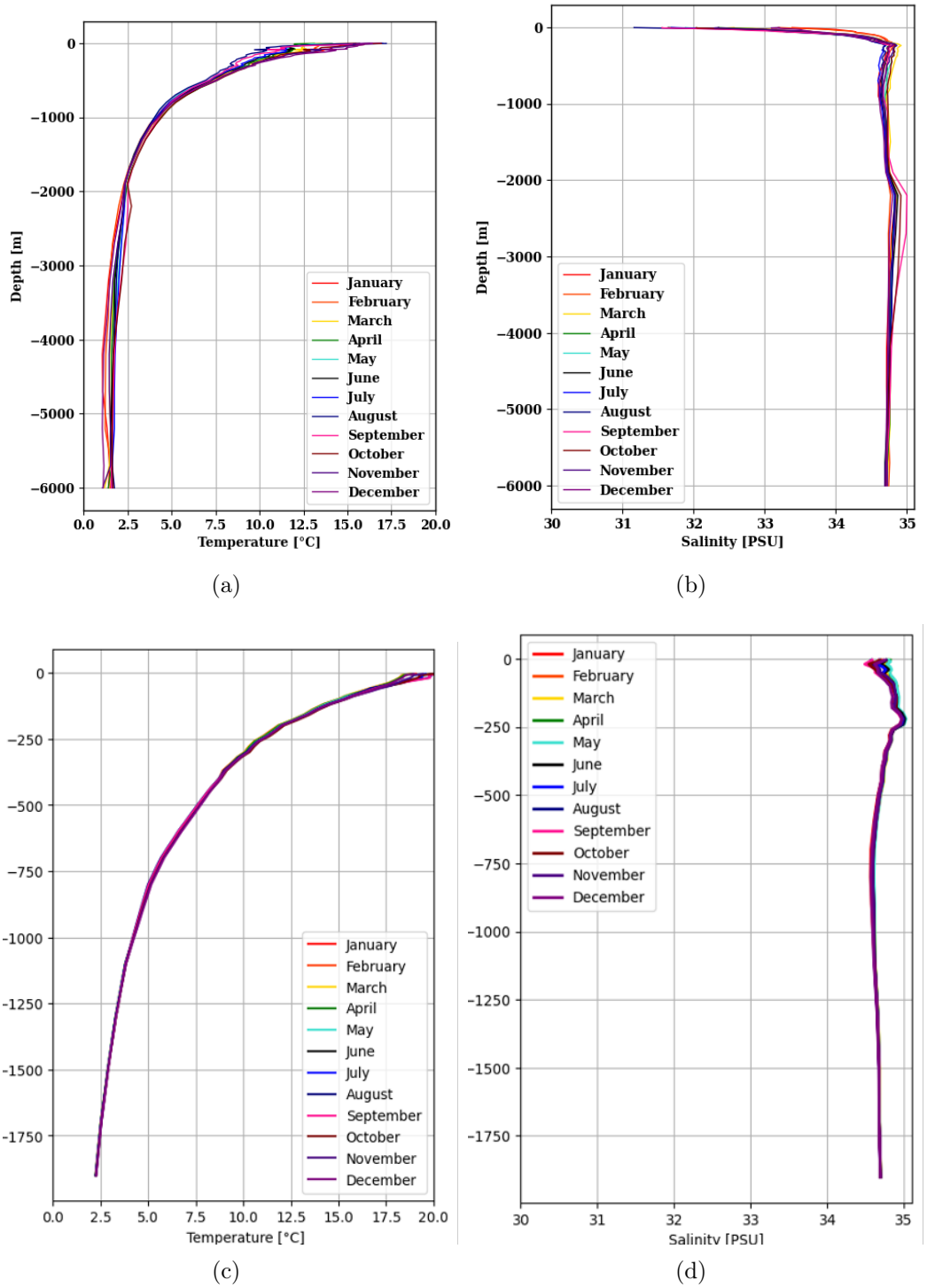


Fig. 3.4: Background1(spatial mean of data at each layer) for SDC_V1 computed from Dataset1: Temperature (a) and Salinity (b). Background1 for SDC_V2 computed from Dataset2: Temperature(c) and Salinity(d).

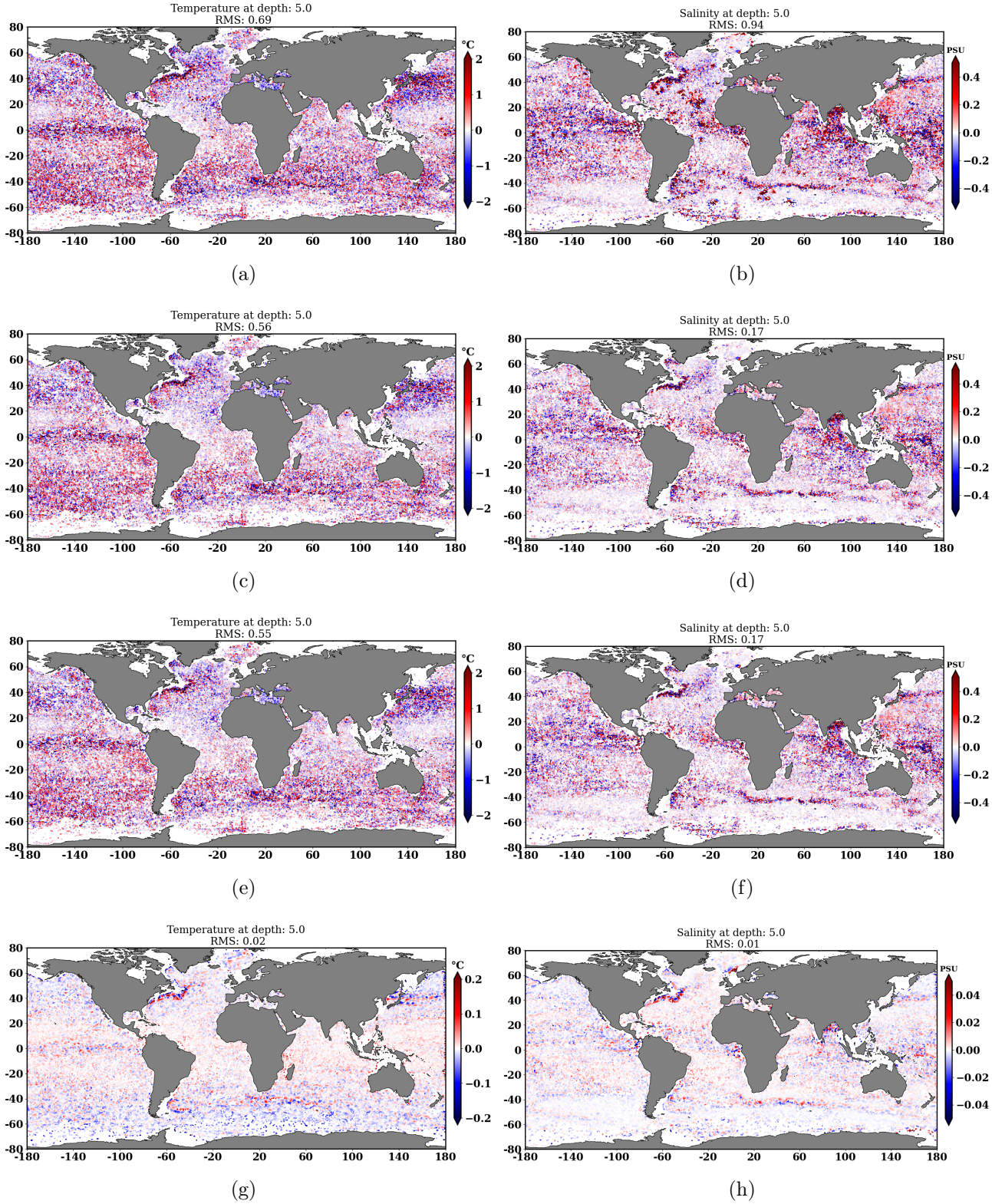


Fig. 3.5: RMS of Residuals for the month of January: Temperature(left column) and Salinity(right column) at 5m mapping with WOD QC dataset with choice of Background1 (a),(b), with NQC dataset with choice of Background1 (c),(d) and with NQC dataset with choice of Background2 (e). And (g) and (h) show differences of (c) and (e) for temperature, (d) and (f) for salinity respectively.

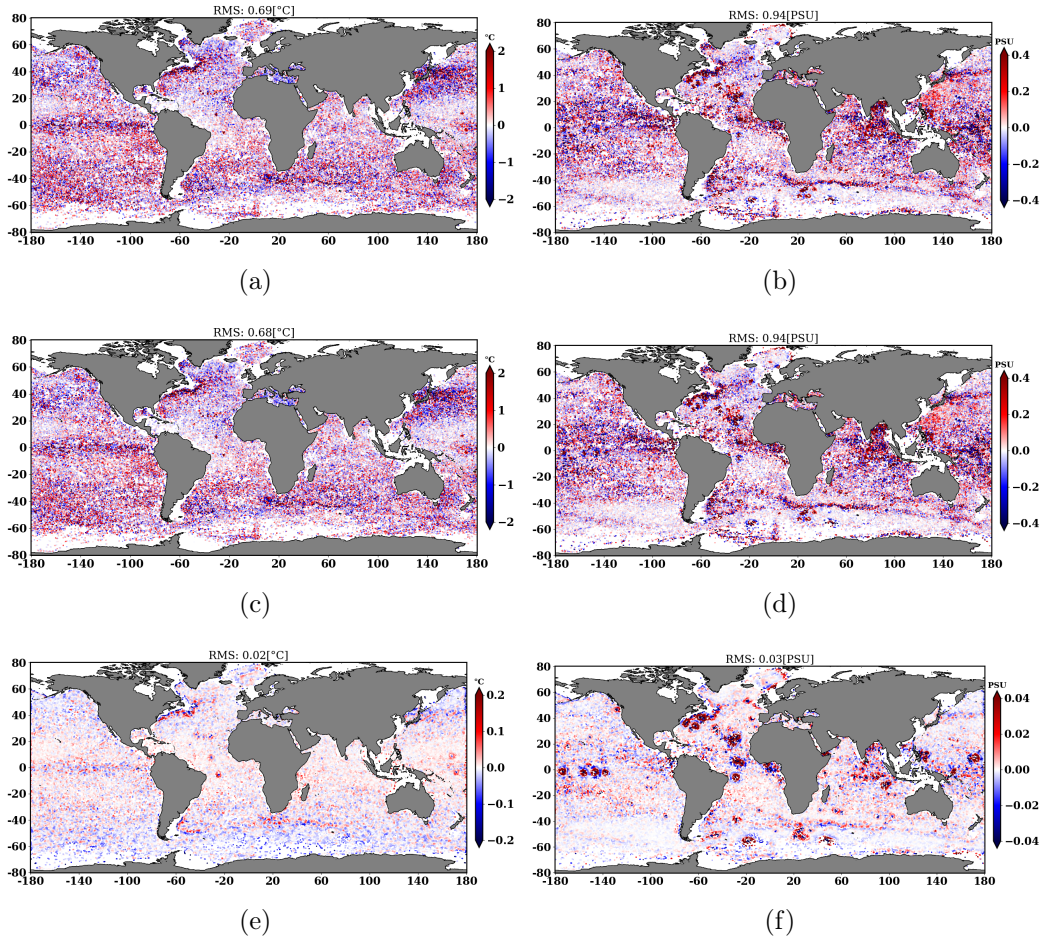


Fig. 3.6: RMS of residuals for the month of January using the WOD QC input data set: Temperature (left column) and Salinity(right column) at 5m. (a) and (b): with choice of Background1. (c) and (d): with choice of Background2. (e) and (f): difference of residuals (Background1-Background2).

3.3.4 Sensitivity Experiments for DIVA Parameter Choices

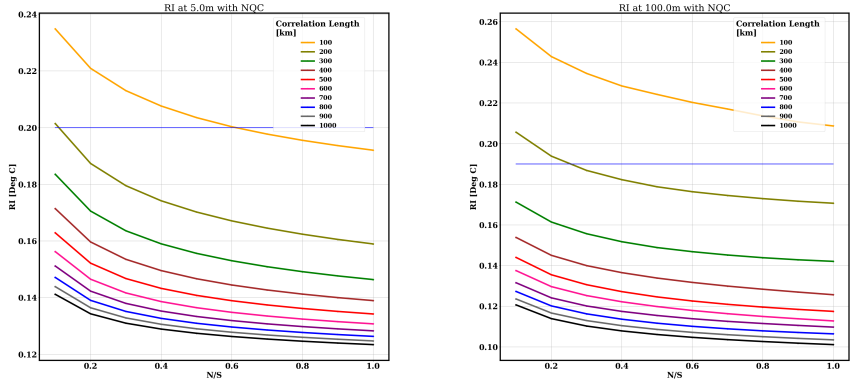
The choice of CL and N/S for a global ocean domain is quite challenging. The global ocean contains a multiplicity of scales. Therefore, a single CL value could either overly smear the general circulation fronts (such as the western boundary currents) or contaminate the climatology with mesoscale eddies or other higher frequency processes. In the past CL was also estimated by the data itself, binning the data and fitting analytical curves (reference). However in the case of the global ocean the data is so non-uniformly spaced that the CL estimation quality will be very different among ocean areas. Thus we preferred to take a new but traditional approach that uses equal CL everywhere as in WOA18 Locarnini et al. (2018). In order to choose reasonable values, several sensitivity experiments were performed with different CL values ranging from

100 to 1000 km and N/S values varying from 0.1 to 50. A roughness index is defined as the mean of the derivative of field in the two directions and is defined as:

$$RI = \frac{1}{N} \sum_{i,j=1}^{n,m} \sqrt{(\Delta_{\mathrm{x_i}} f)^2 + (\Delta_{\mathrm{y_j}} f)^2} \tag{3.9}$$

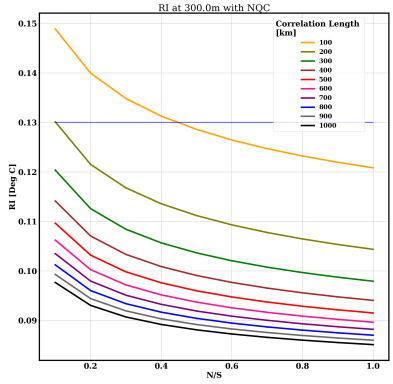
where Δ is the finite difference derivate in the latitudinal and longitudinal directions, x_{i} is the grid location in longitude and y_{j} in the latitudinal direction and $N = n * m$ is total number of the interpolating grid points. RI gives a measure of the spatial scale of the field, for instance a field with mesoscale features will have high RI while a smoother field with large scale features have low values of RI. We do not believe that the usage of Rossby radius of deformation and/or its corresponding wavelength is a correct way to define the correlation length for a climatology. The latter is the result of many propagating waves in the ocean, which sum up in a mean field that is necessarily smooth. Thus a roughness index, or its inverse, the smoothness index, is a better choice to decide the correlation length of the interpolating algorithm with respect to the wavelength of the primary process that has created the climatology. Many climate indices are in fact "smoothed" to extract the basic long term signals.

As expected, for large CL values the analysis has a small RI value, as shown in Fig. 3.7 and 3.8. Furthermore, we decided that RI should not exceed the standard deviation (std) of the data itself. The results for the month of January are shown in Fig. 3.7 and 3.8: it is evident that the criteria of accepting a value of RI less than the field STD eliminates only CL at 100 km, varying a little bit with depth. The "elbow" of all the curves sits between 0.4 and 0.6 for N/S ratio, thus we choose 0.5. Choosing this N/S value, and taking RI equal approximately half of the field STD, the value of CL of 300 km emerges. However, in the case of N/S, we took an intermediate value of 0.5 which always give a relatively low RI.

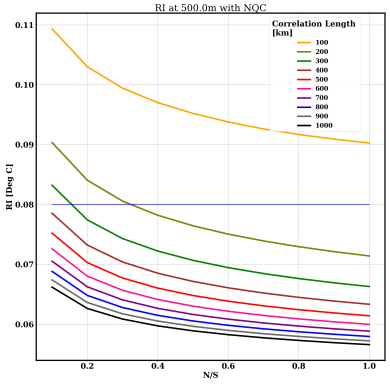


(a)

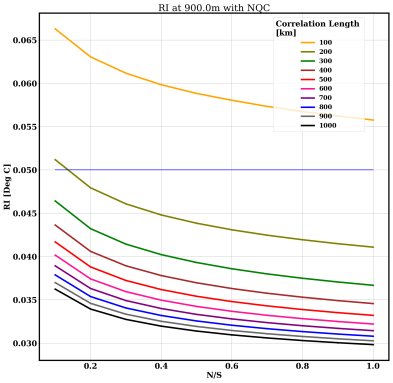
(b)



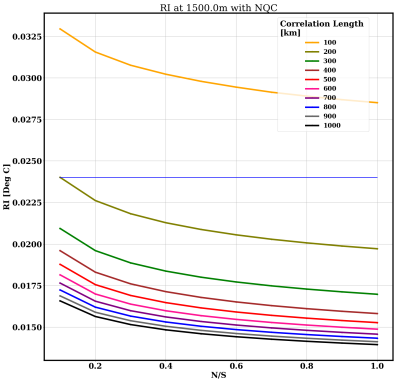
(c)



(d)



(e)



(f)

Fig. 3.7: Roughness Index of SDC_V1 for Temperature and the month of January at 5m (a), 100m (b), 500m (c), 1100m (d), 1500 (e) and 1900m (f) for different CL and N/S.

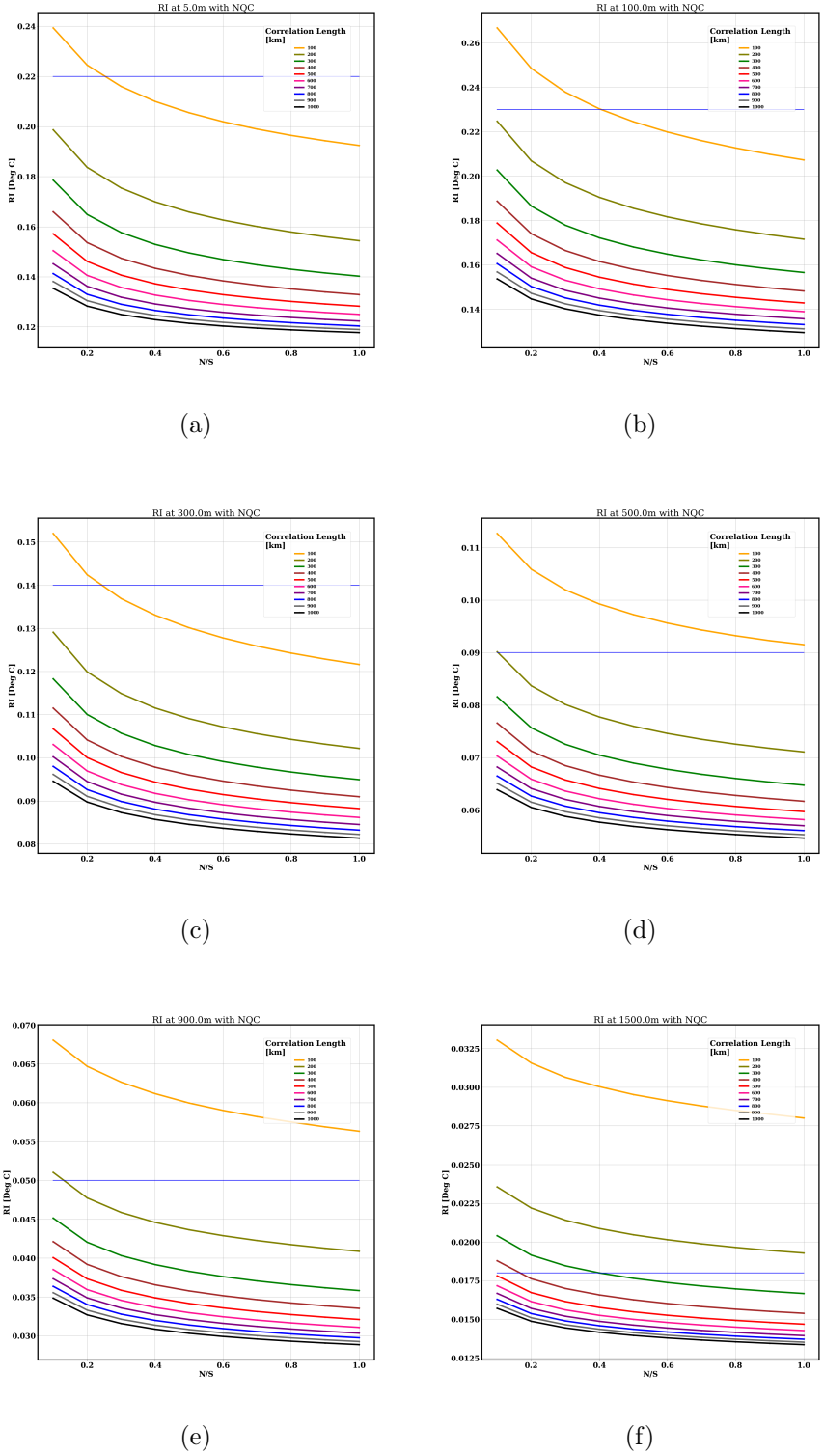


Fig. 3.8: Roughness Index of SDC_V2 (Temperature) for the month of January at 5m (a), 100m (b), 500m (c), 1100m (d), 1500 (e) and 1900m (f) for different for different CL and N/S.

3.4 Temperature and Salinity Climatology

Temperature and salinity mapping with a CL of 300km and an N/S of 0.5 was carried out for Dataset1 and Dataset2 for all depths and months. Figures 3.9 and 3.10 show the mapped fields of temperature and salinity for Dataset1 and Dataset2 and for January and August at different depth levels. Fields are masked where analysis errors are greater than 30% (relative to the field standard deviation). It is immediately evident that SDC_V1 provides more information in polar and coastal regions with respect to SDC_V2, since more observations are available in these areas.

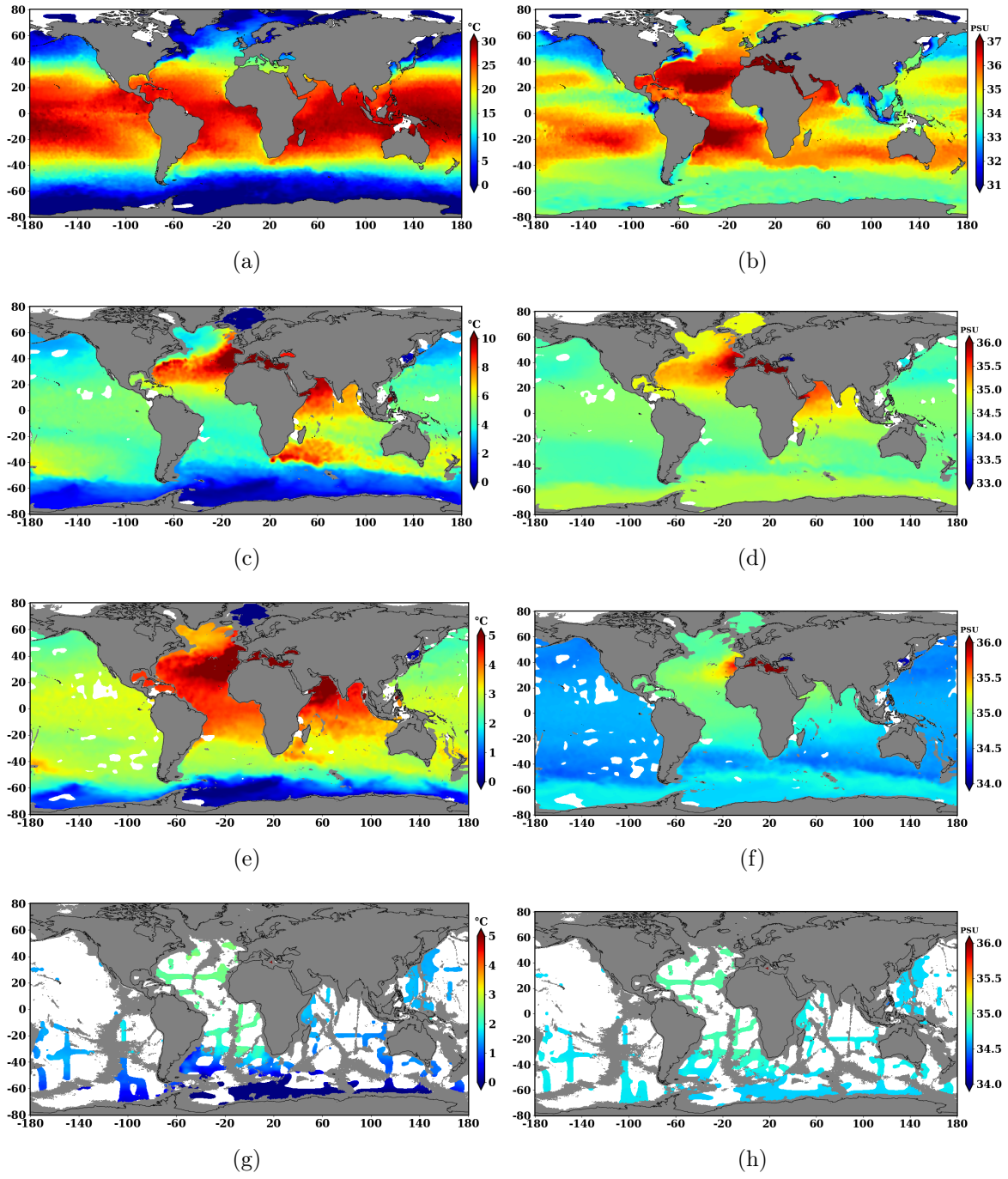


Fig. 3.9: SDC_V1 January mapping for Temperature (left column), (a), (c), (e) and (g) and Salinity (right column) (b), (d), (f) and (h) at 5, 900, 1050 and 3700m respectively.

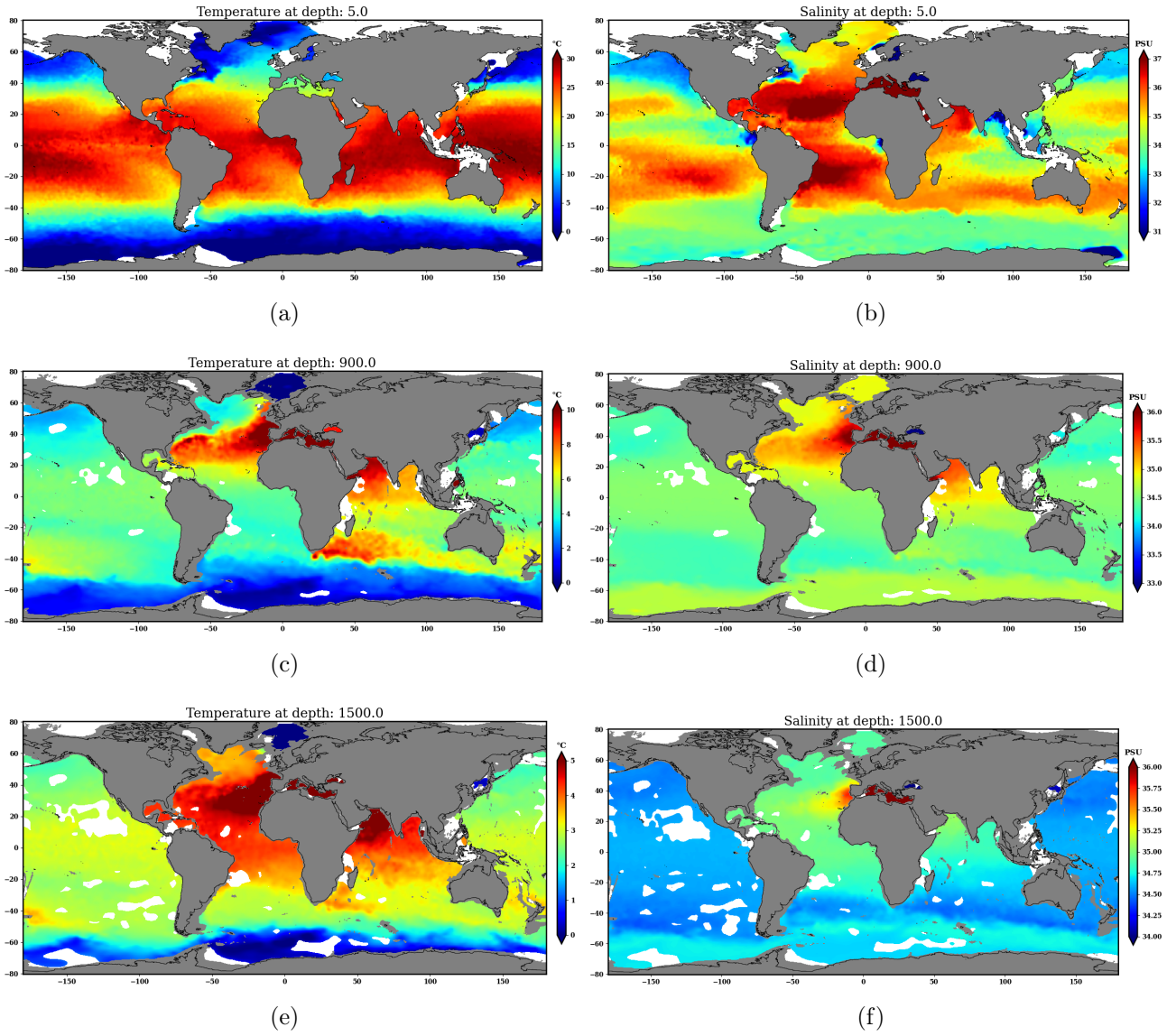


Fig. 3.10: SDC_V2 January mapping for Temperature(left), ((a), (c) and (e)) and Salinity ((d), (f) and (b)) at 5, 900 and 1050m respectively.

SDC_V1 and SDC_V2 are both plausible climatological estimates. SDC_V1 is a longer term average while SDC_V2 is an estimate of the last 15 years. The difference between these estimates is shown in Figure. 3.11 and 3.12 : SDC_V2 is warmer and more saline than SDC_V1 and the root mean square (RMS) difference varies from 0.4 ° to 0.5°C and 0.7 to 0.6 PSU for temperature and salinity, respectively. Dataset1 uses bottle data therefore the difference field along ship tracks is evident. The differences are quite localised because of the difference in the input datasets but there is also a large scale signal which probably refers to the last fifteen years warming. In the case of salinity the signal is a bit more complicated, with no clear sign, with a clear increase/decrease only in localised areas near shelf breaks and the coastal areas.

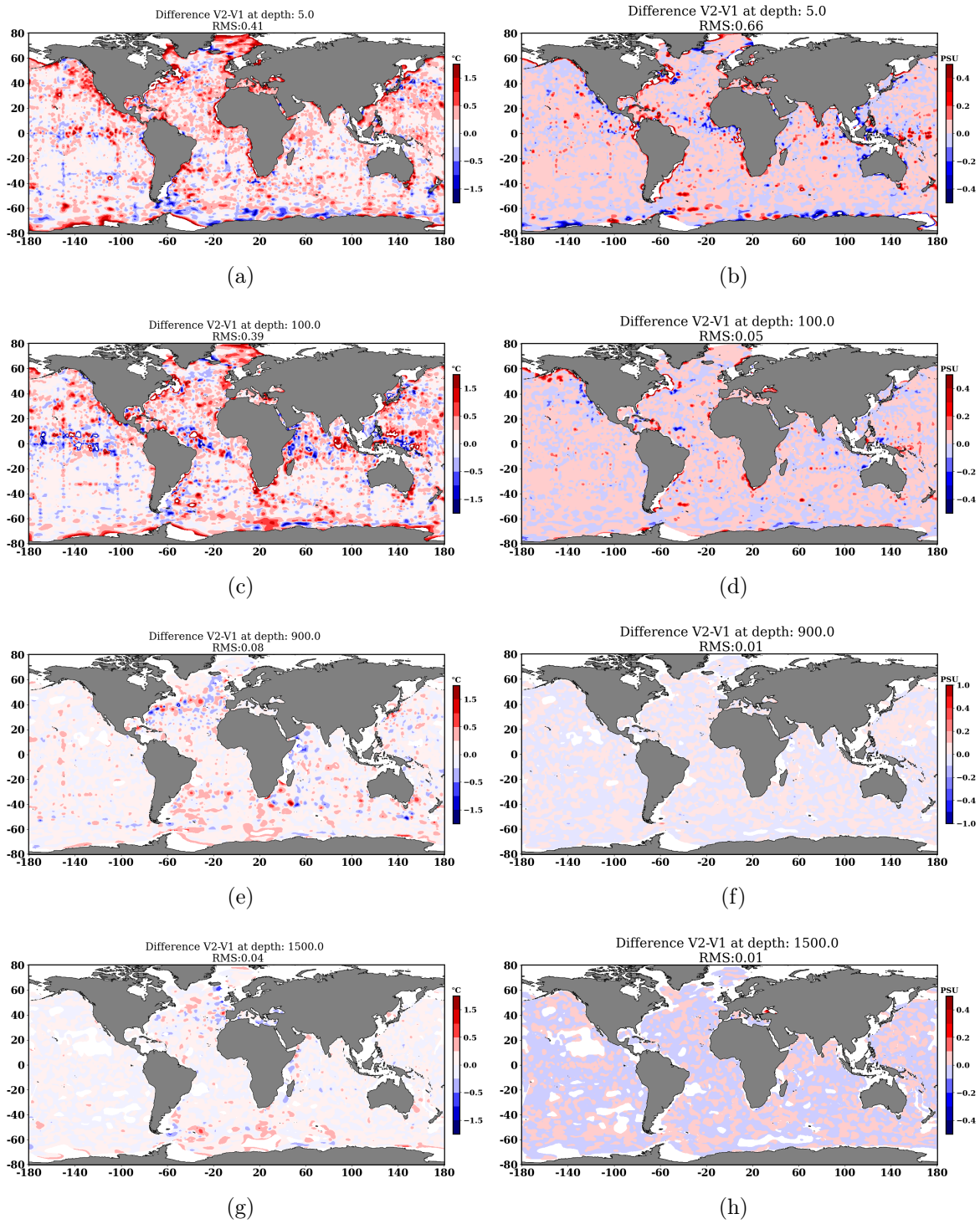


Fig. 3.11: Difference between SDC_V2 and SDC_V1 for January. Temperature (left column) (a), (b), (c), and (d) and Salinity (right column) (e), (f), (g), and (h) for 5, 100, 900 and 1500m respectively.

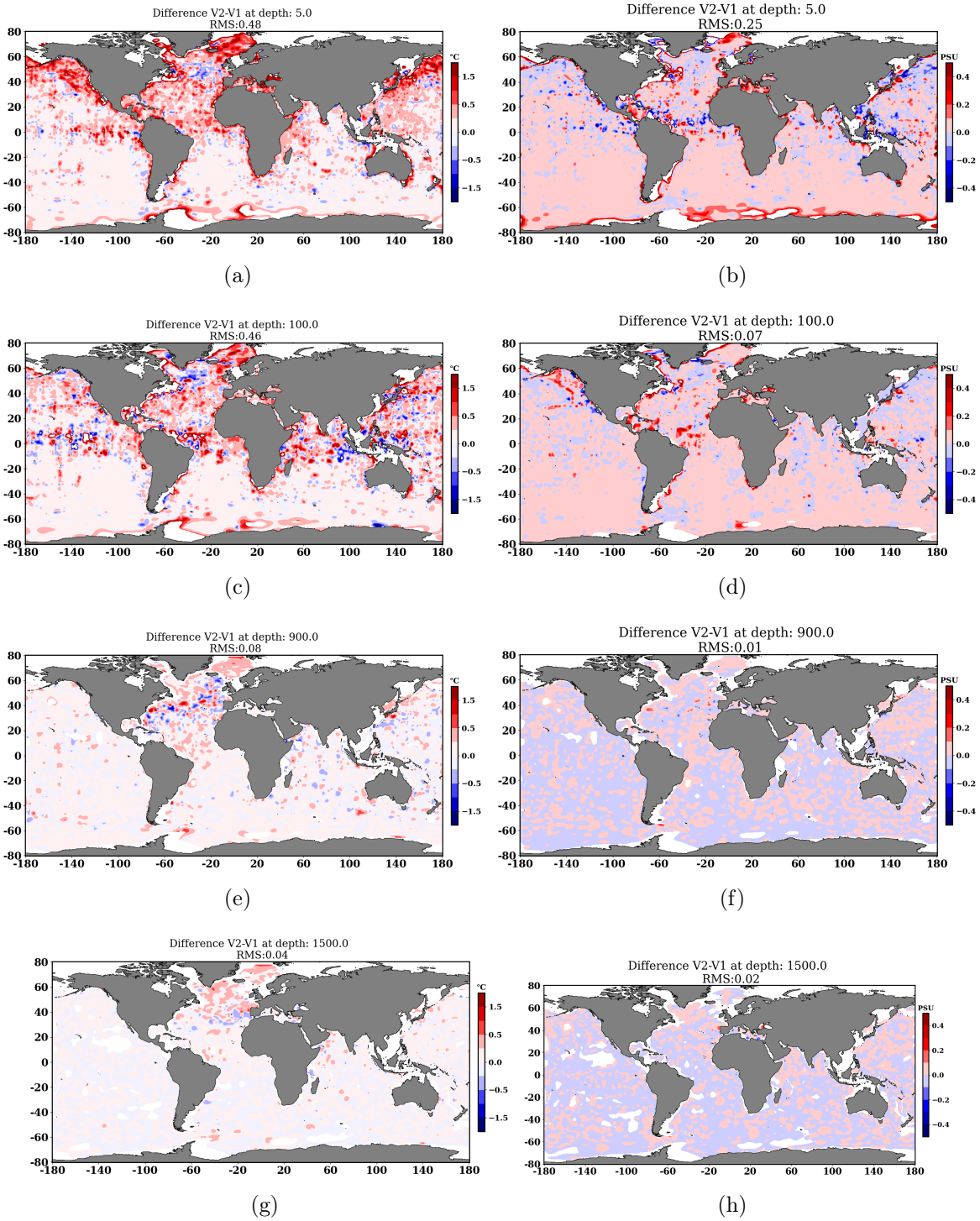
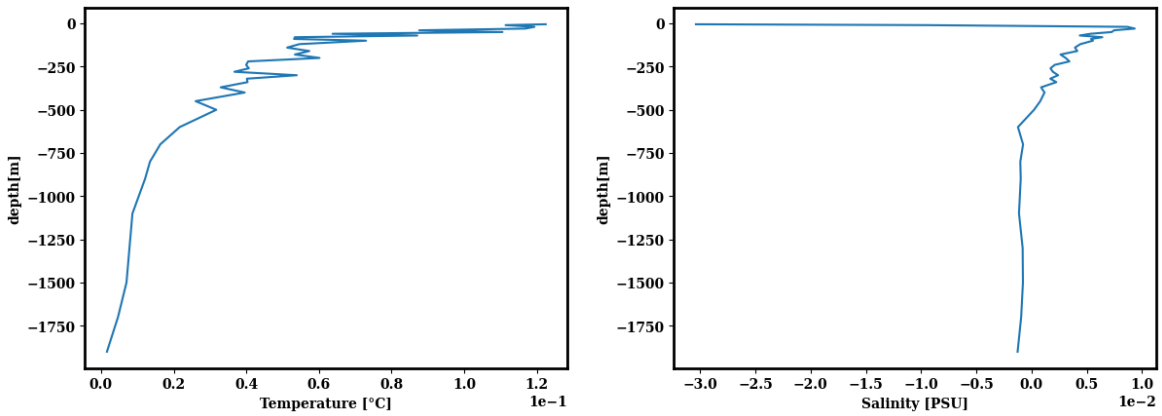


Fig. 3.12: Difference of SDC_V2 and SDC_V1 for August, Temperature(left column) at (a), (b), (c), and (d) at 5, 100, 900 and 1500m and Salinity(right column) at (e), (f), (g), and (h).

Furthermore we would like to point out that SDC_V1 is also affected by a large nonuniform distribution of observations in the two hemispheres. Such non-uniform distribution is not present in the SDC_V2 estimates so this is also a difference between the two climatologies.

However, the statistical multi-model estimate that we propose here would be a way to control such an uncertainty and a potential solution to get a lower error estimate due to the skewed input data distribution. To quantify the temperature and salinity differences between the two analyses, we evaluated the horizontally averaged bias between SDC_V2 and SDC_V1 as shown in Fig. 3.13. It is evident now that while temperature is warming at all depth, salinity is lower except for the intermediate depths between 50 and 500 m that, if confirmed is a new result. This might be due to compensating effects between warming and salinity in the upper water column as described by Chen et al. (2019).



(a) (b)

Fig. 3.13: Bias between SDC_V2 and SDC_V1 for January, Temperature in (a), and Salinity in (e)

3.5 Validation with other Climatologies

The analysis' validation is an essential step for an indication of the reliability of results.

The main source of data in WAGHC is extracted from WOD13, in particular OSD, CTD, PFL and APB. However, additional data were added from the Alfred Wegener Institute, Bremerhaven, and from different institutions in Canada for the time period between 1900 to 2016 Gouretski (2018). Similarly, the analysis was performed monthly at a 0.25 degree spatial resolution. In WAGHC, the objective analysis scheme is used with a CL of 333km and an N/S value of 0.5. The WAGHC background is an analysis carried out with a correlation length of 555km. The difference field between SDC_V1 and WAGHC is shown in Fig. 3.14. The comparison shows overall consistency, even if differences are quite large in the tropical and equatorial areas around 100m. This is probably due to the vertical smoothing in WAGHC which considered the interpolation of observations at levels instead of thin layers as in our case.

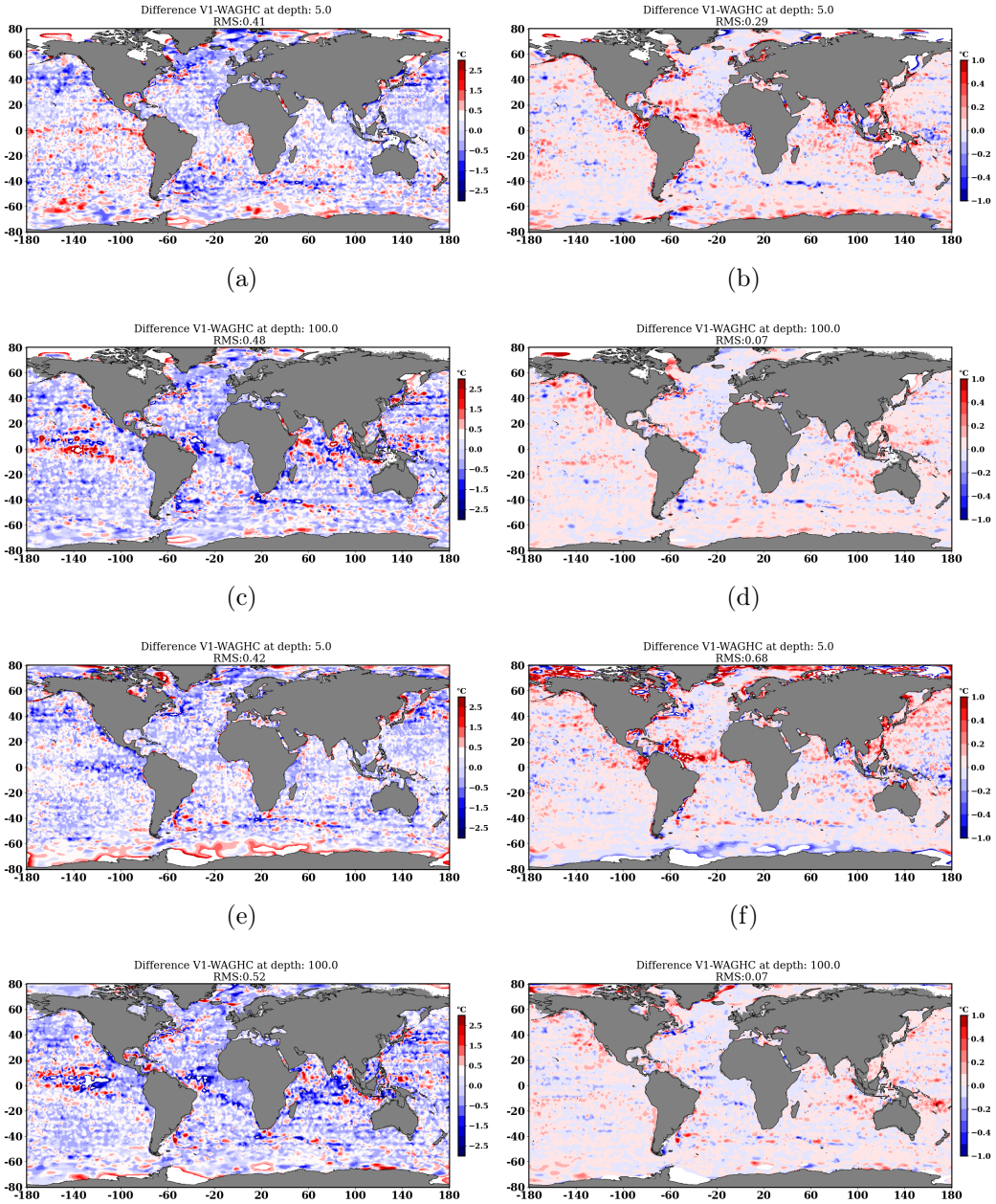


Fig. 3.14: Difference of SDC_V1 and WAGHC for January and August at 5 and 100m: Temperature ((a), (c), (e) and (g)) and Salinity ((b), (d) , (f) and (h)).

The second comparison of our climatology is with WOA18. The data considered in WOA18 are profiles from OSD, CTD, PFL, MRB, Mechanical Bathythermographs, Digital Bathythermographs, Expendable Bathythermographs, moored and drifting buoys, gliders, undulating oceanographic recorders (UOR), pinniped mounted CTD sensors and surface only data Locarnini et al. (2018) and Zweng et al. (2019). WOA18 is a monthly climatology defined at

57 depth levels at the spatial resolution of 0.25° for six decades i.e. 1955-1964, 1965-1974, 1975-1984, 1985-1994, 1995-2004 and 2005-2012. To compute the difference between the climatologies, WOA18 time average fields over the six decades were interpolated on the DIVA analysis grid using linear interpolation. In WOA, the Barnes interpolation scheme was used with a CL of 333km. The comparison of SDC_V1 with WOA18 is shown in Fig. 3.15. Again, the comparison shows consistency, even if a larger temperature positive difference is evident with respect to WAGHC differences.

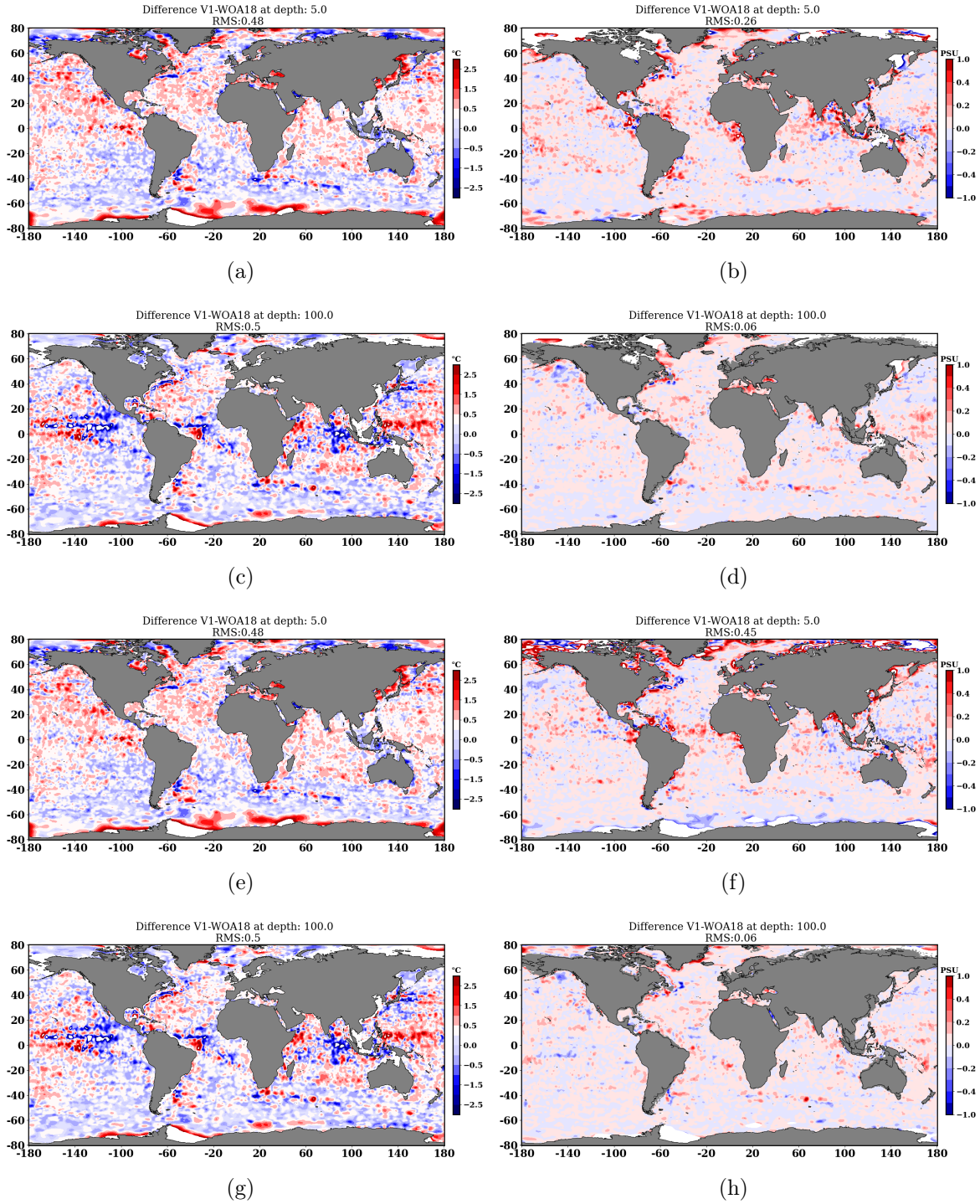
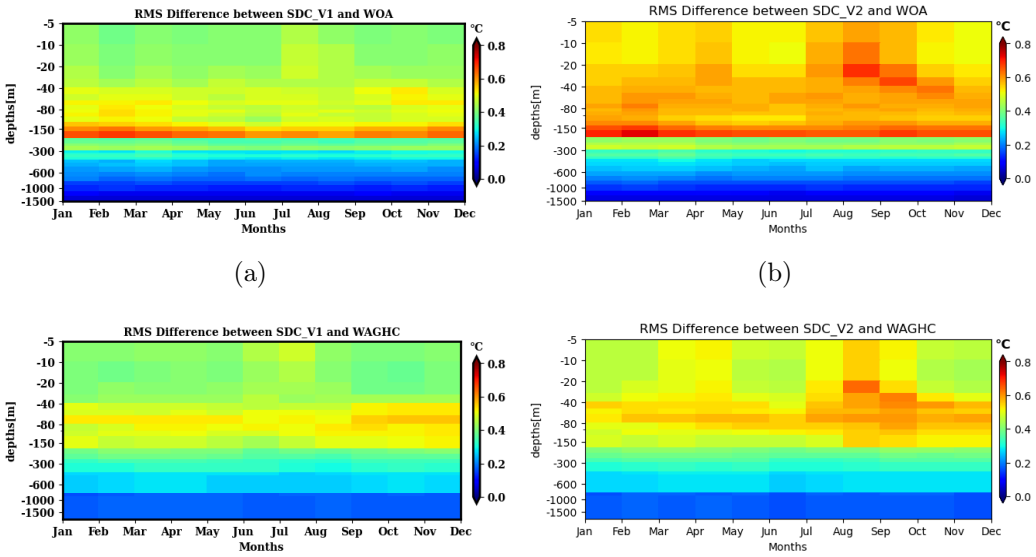


Fig. 3.15: Difference of SDC_V1 and WOA18 for January and August at 5 and 100m: Temperature ((a), (c), (e) and (g)) and Salinity ((b), (d), (f), and (h)).

Finally, we constructed an Hovmoller diagram of RMS differences between WOA18 and WAGHC to show the major differences in the vertical profile (Fig. 3.16 and 3.17). The largest RMS differences of temperature are found with SDC_V1 at the thermocline depth for both

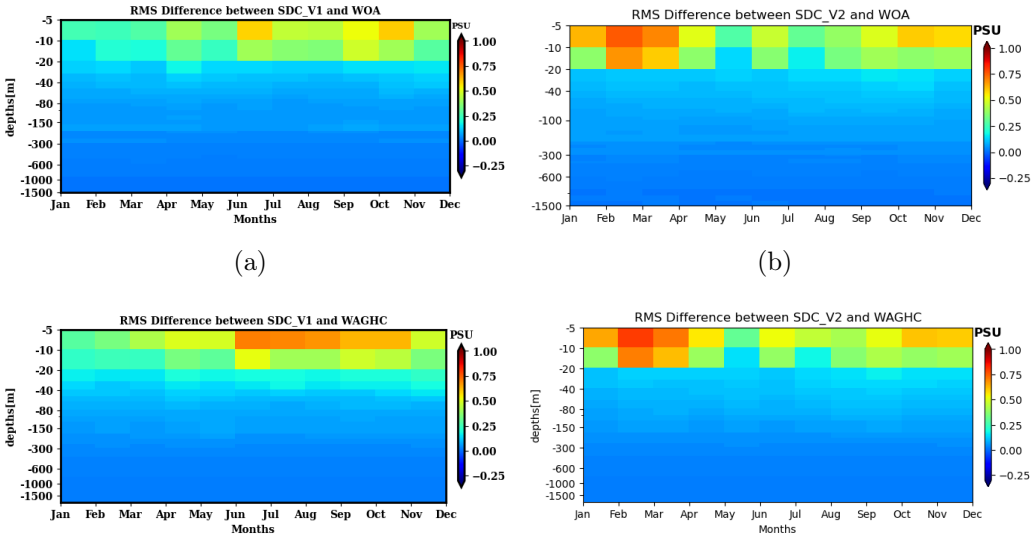
WOA18 and WAGHC, yet the differences are more prominent with WOA18. We argue that the reason for this difference at the thermocline is due to the different interpolation method used in the vertical profile which might create different thermocline gradients.



(a) (b)

(c) (d)

Fig. 3.16: Hovmoller diagram of RMS difference for temperature, SDC - WOA18 (a) & (b) and SDC - WAGHC (c) & (d).



(a) (b)

(c) (d)

Fig. 3.17: Hovmoller diagram of RMS difference for salinity, SDC - WOA18 (a) & (b), and SDC - WAGHC (c) & (d).

In this chapter, we developed a multi-model ensemble methodology to improve the global climatology estimate. Uncertainties in global climatologies are due to the details of the input data set chosen, to the specific choices of background and statistical interpolation algorithms, to the specific quality control used. Like for numerical models, a multi-model statistical estimate can diminish the errors with respect to specific quality assessments indices. Thus having a diversity of climatological estimates could provide in the future the best estimate of the climatological state of the ocean. This was demonstrated here using a simple multi-model mean of the 4 climatology fields available showing that the multi-model ensemble spread is lower than the root mean square of residuals. In the future it could be desirable to use more estimates for the multi-model climatology. The discrepancies among the available estimates are the combination of different hypothesis in the mapping algorithm, the number of observations used, the first guess or the interpolation in space and time of the observations, if considered. In order to reduce the single climatology estimate uncertainties, ensemble multi-model methodologies could be devised. The multi-model ensemble approach is very simple: considering each member of the ensemble to be a different climatology from a different statistical interpolating model, the ensemble mean of these models will be superior to the single models within a certain evaluation matrix Krishnamurti et al. (1999). The evaluation matrix used consists of the comparison between the ensemble mean error and residual root mean square (RMS) errors. The multi-model ensemble mean error is represented by the standard deviation of the members around the ensemble mean, also called the spread. If the spread is lower than the RMS of residuals, then the ensemble estimate will be more accurate. Residuals, r^i are defined as the difference between the climatology, so-called ${\theta^i}_{\rm c}$, interpolated at the observation location and the reference observations y_0 .

$$r^i = H({\theta^i}_{\rm c}) - y_{\rm o}, \tag{3.10}$$

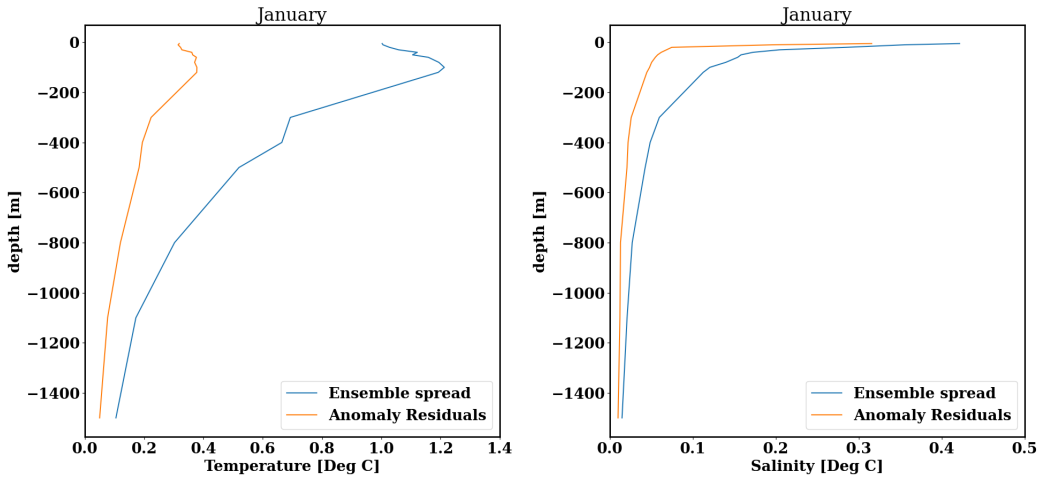
In the formula above, H is the bilinear interpolation or observation operator. It is clear that r^i contain a mixture of errors coming form the estimate ${\theta^i}_{\rm c}$, from the observations and from H. The climatology multi-model ensemble mean, ${\theta^E}_{\rm c}(x, y, z)$ is defined:

$${\theta^E}_{\rm c}(x, y, z) = \sum_{i=1}^{N} \frac{{\theta_{\rm c}}^i(x, y, z)}{N}, \tag{3.11}$$

where $N = 4$, corresponding to the number of climatological estimates under consideration. Furthermore, ensemble spread is computed as the standard deviation from the ensemble mean of each individual estimate, average over the number of grid points m, and is given by:

$$\tilde{s}(z) = \frac{1}{4}\sum_{i=1}^{4}\left(\sqrt{\sum_{k=1}^{m}\frac{(\theta^{i}{}_{\mathrm{k}} - \theta^{E}{}_{\mathrm{ck}})^2}{m}}\right). \tag{3.12}$$

Similarly, the residual RMS, averaged between the four realizations is computed and its comparison with the ensemble spread is shown in Figure 3.18(a) and 3.18(b). The figure shows clearly that the ensemble spread for temperature and salinity is smaller than the anomaly residual standard deviation, proving that the multi-model ensemble mean is a better estimate of the climatology.



(a) (b)

Fig. 3.18: Ensemble spread and RMS of residuals for Temperature (a) and Salinity (b) as a function of depth for the different layers of Table 2.

3.6 Discussions

Two different versions of global ocean climatologies were estimated using DIVA, with the application of the new NQC developed in Chapter 2. Two different backgrounds were analyzed: a spatial mean of observations in the vertical profile and an analysis carried out with a correlation length of 1000km and an N/S of 0.5. Results show that, if pre-processing is carried out by the NQC algorithm, the interpolated field becomes less dependent on the choice of background field

(see Figure 3.5).

Furthermore, for the first time, the choice of DIVA parameters is deduced from a new roughness index (RI) which quantifies the analysis' degree of smoothness as a function of CL and N/S values.

Inter-comparison of the SDC_V1 with the SDC_V2 climatology shows that the 2003-2017 climatology is warmer and saltier. It is to be fully assessed if this is a trend caused by climate change and global warming. The comparison of the SDC_V1 climatology with WOA and WAGHC shows reasonable agreement but also relevant differences in the tropical and polar areas. Moreover, the SDC_V1 climatology is closer to WAGHC than WOA18 for both temperature and salinity. One reason could be connected to the fact that the objective analysis parameters used and the technique itself are similar to DIVA.

3.7 Conclusions and Future Work

The estimation of temperature and salinity climatology is a key element for improving our understanding of the ocean state. Historical data sets available today enable an almost complete reconstruction of global ocean fields. For the first time, this chapter has outlined the combined application of a nonlinear quality control method and a DIVA interpolation scheme which can consider coastal constraints. We believe the computed SDC_V1 and SDC_V2 climatologies have the potential to become a reference data set for the oceans temperature and salinity mean state. Worth considering for future work is the application of an improved NQC version using a regime-oriented division instead of regular 5° square rectangles on a global domain. An optimised choice of DIVA parameters that are different for each level might improve results. Moreover, further validation with independent dataset such as satellite observations or randomly subsampled input dataset will enable to understand whether the analysis under fit or over fit the observations.

3.8 Data Availability Statement

The estimated temperature and salinity climatologies has resulted as a data products and can be accessed using the following link:
https://doi.org/10.12770/98d22ac0-5398-4889-8f8e-8f28273b548b *(Global Ocean Climatology -*

Temperature and Salinity Climatology V2)

Chapter 4

Global Ocean Health Indicators

4.1 Introduction

Ocean health indicators were advocated in the Sustainable Development Goals framework (Rosa, 2017). These indicators are designed to measure the level of pollution, acidification, etc. Two of these indicators, stratification and oxygen level indicators, are easily connected to the work in this thesis and this Chapter shows their preliminary computation.

According to the latest IPCC report (Poloczanska et al., 2018), the upper oceans seem to have been gaining stratification and losing oxygen ever since the 1970s due to global warming. The water column's stratification is related to its vertical mixing and is one of the main processes affecting upper ocean biochemistry, the lower levels of the marine ecosystem. Water column stratification is related to physical properties, such as temperature and salinity, through the density and the derived Brunt-Väisäla frequency. In this context, this chapter focuses on the computation of water column density and evaluates the averaged Brunt-Väisäla frequency profiles in the Atlantic and Pacific oceans.

In addition, the WOD18 (Garcia et al., 2019) historical database also contains the largest collection of in situ dissolved oxygen measurements. Again, as for temperature and salinity, the new DIVA mapping technique has never been applied to this data set and this will be attempted in this thesis.

This chapter will first define a methodology to compute high quality density profiles from the WOD18 PFL data set (section 4.2). Climatological seasonal estimates of the derived Brunt-Väisäla frequency will then be constructed for the Pacific and Atlantic oceans (section 4.2.3),

followed by the mapping of dissolved oxygen and Apparent Oxygen Utilization (AOU) (Chapter 4.3). A summary and section on future steps is concluded in this chapter.

4.2 Global Ocean Density Mapping

Density is a key ocean dynamical variable which affects circulation through induced changes in the pressure field, forcing momentum equations. It is difficult to measure density directly and, since the beginning, oceanographers developed temperature and salinity measurements that are converted to density by means of an empirical equation of state, that is, the Equation of Seawater developed by UNESCO in the 1970s and 1980s.

The Equation of State of Seawater used here is referred to as EOS-80 (Fofonoff and Millard Jr, 1983) or the in situ equation for seawater since in situ temperature and salinity are used. It is a nonlinear function of temperature and salinity thus, it is improper to use interpolated data to find density. Instead, a density profile should be computed at the depth where temperature and salinity are measured, as pointed out by Gouretski (2019).

4.2.1 Preprocessing of the Dataset

The basic data set used in this chapter is the Profiling FLoats (PFL) subset from the WOD18 historical database, called Dataset2 in Chapter 3. In addition to Nonlinear Quality Control (NQC) applied to temperature and salinity as described in Chapter 3, we considered here only the profiles where both temperature and salinity values were present at different depths. The number of profiles and observations are given in Table 5.1. The number of profiles that include temperature and salinity together are reduced to $1/2$ of the total number: we argue that this is due to the fact that the NQC procedure is performed separately on temperature and salinity data. This problem must be addressed in future by considering the application of NQC with the constraint of only retaining profiles containing both temperature and salinity measurements.

Density profiles are computed using the Python Seawater Library which uses the EOS-80 equation for seawater (Fofonoff and Millard Jr, 1983), that is:

$$\begin{aligned}
\rho(S,T,P) &= \frac{\rho(S,T,0)}{1-\frac{P}{K(S,T,P)}} \\
K(S,T,P) &= K_0(S,T,0) + AP + BP^2 \\
\rho(S,T,0) &= \rho_0 + [b_0 + b_1T + b_2T^2 + b_3T^3 + b_4T^4]S + \\
&\quad [c_0 + (c_1T + c_2T^2)]S\sqrt{S} + d_0S^2 \\
\rho_0 &= a_0 + a_1T + a_2T^2 + a_3T^3 + a_4T^4 + a_5T^5
\end{aligned} \tag{4.1}$$

where T,S,P are the temperature, salinity and pressure in $°C$, PSU and db, respectively and the density is in $kg\,m^{-3}$. K_0, A, B are given by:

$$K_0 = KW + f_0 + f_1T + f_2T^2 + f_3T^3 + g_0\sqrt{S} + g_1T\sqrt{S} + T^2\sqrt{S})S, \tag{4.2}$$

$$KW = e_0 + e_1T + e_2T^2 + e_3T^3 + e_4T^4, \tag{4.3}$$

$$A = AW + (i_0 + i_1T + i_2T^2 + j_0\sqrt{S})S, \tag{4.4}$$

$$AW = h_0 + h_1T + h_2T^2 + h_3T^3. \tag{4.5}$$

$$B = BW + m_0S + m_1TS + m_2T^2S, \tag{4.6}$$

$$BW = k_0 + k_1T + k_2T^2, \tag{4.7}$$

where $f_0, f_1, f_2, f_3, g_0, g_1, e_0, e_1, e_2, e_3, e_4, m_0, m_1, m_2, i_0, i_1, i_2, j_0, k_0, k_1, k_2, h_0, h_1, h_2, h_3$, $a_0, a_1, a_2, a_3, a_4, a_5, b_0, b_1, b_2, b_3, b_4, c_0, c_1, c_2$ and d_0 are constants provided in McDougall et al. (2009).

Variable	**Total Profiles**	**Total Observations**
Temperature	1,652,136	380,970,230
Salinity	1,564,877	343,619,337
Temperature & Salinity together	688,186	196,536,424

Table 4.1: Number of profiles and observations for temperature and salinity after NQC from Dataset2 defined in Chapter 3 (2003-2017 period and only PFL platforms).

4.2.2 Interpolated Density Fields

After the computation of density profiles, DIVA is applied at each depth layer given in Chapter 3 with a correlation length of 300km and noise-to-signal (N/S) value of 0.5. Temporal and spatial resolution is monthly and $0.25^\circ \times 0.25^\circ$ in the global domain (similar to Temperature and Salinity Climatologies). The background is given by the spatial mean of all observations as a function of depth and it is presented in Figure 4.1 for all months.

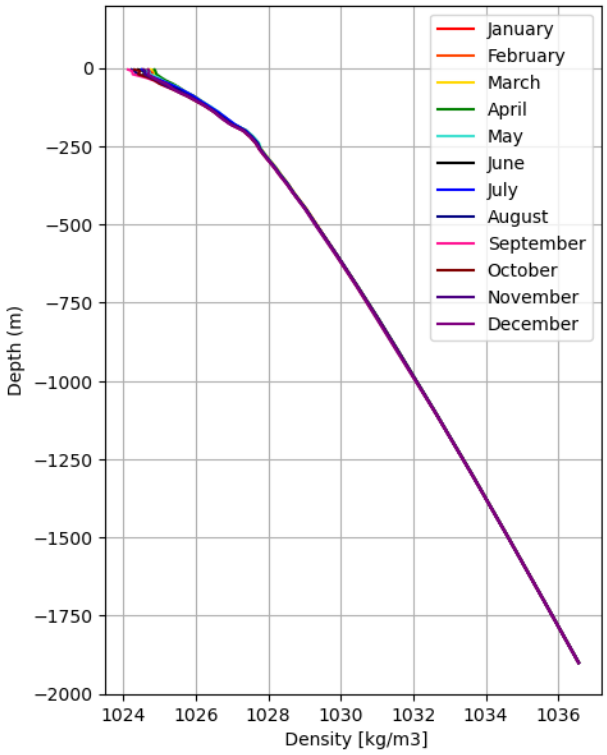
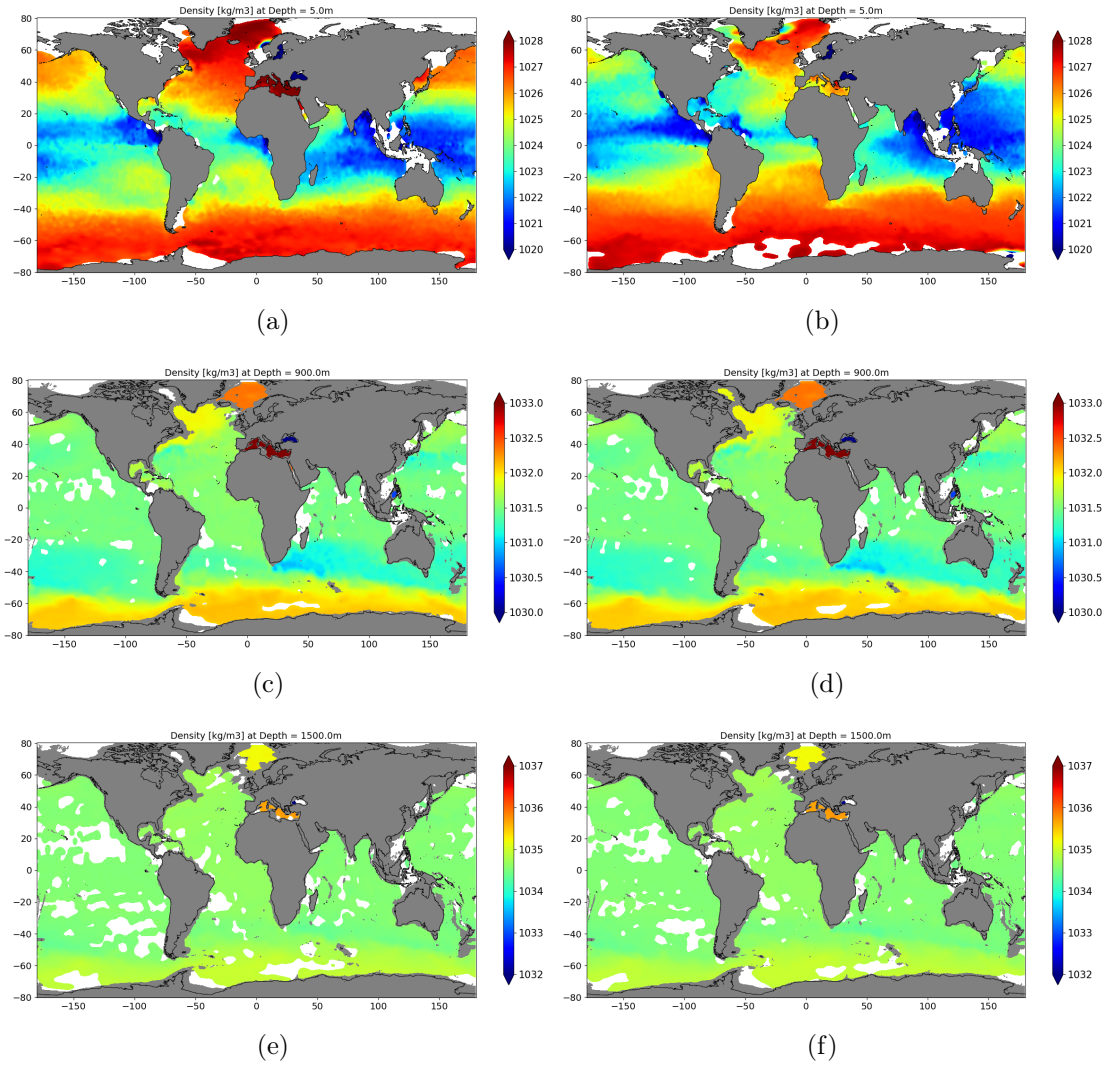


Fig. 4.1: Background density profile used for the interpolation.

Climatological estimates of density for January and August are shown in Fig. 4.2 at 5m, 900m and 1,500m. Several large area gaps appear at 900 and 1500m. The field shows the typical meridional gradient between lighter waters at the equator and heavier waters at the Poles and in the Mediterranean Sea. More analysis will be conducted in the future to study the structures of this density field.



(a) (b) (c) (d) (e) (f)

Fig. 4.2: Interpolated density field: January(left column)(a), (c) and (e), August(right column), (b), (d), and (f) at 5m, 900m, and 1500m. The areas with normalised error greater than 30% are masked.

4.2.3 Brunt-Väisäla Frequency Profiles

Brunt-Väisäla Frequency (BVF) profiles are computed for each density profile described previously for density mapping. Preprocessing considers the following steps:

(i) Only density profiles with no vertical gaps larger than 40m in the first 500m of the water column are considered.

(ii) The resulting profiles are interpolated vertically at every metre from the surface up to 2000m using a linear interpolation between two adjacent values.

(iii) A finite difference scheme is applied to the vertically interpolated profile to estimate the

Brunt-Väisäla frequency which is then spatially and temporally averaged.

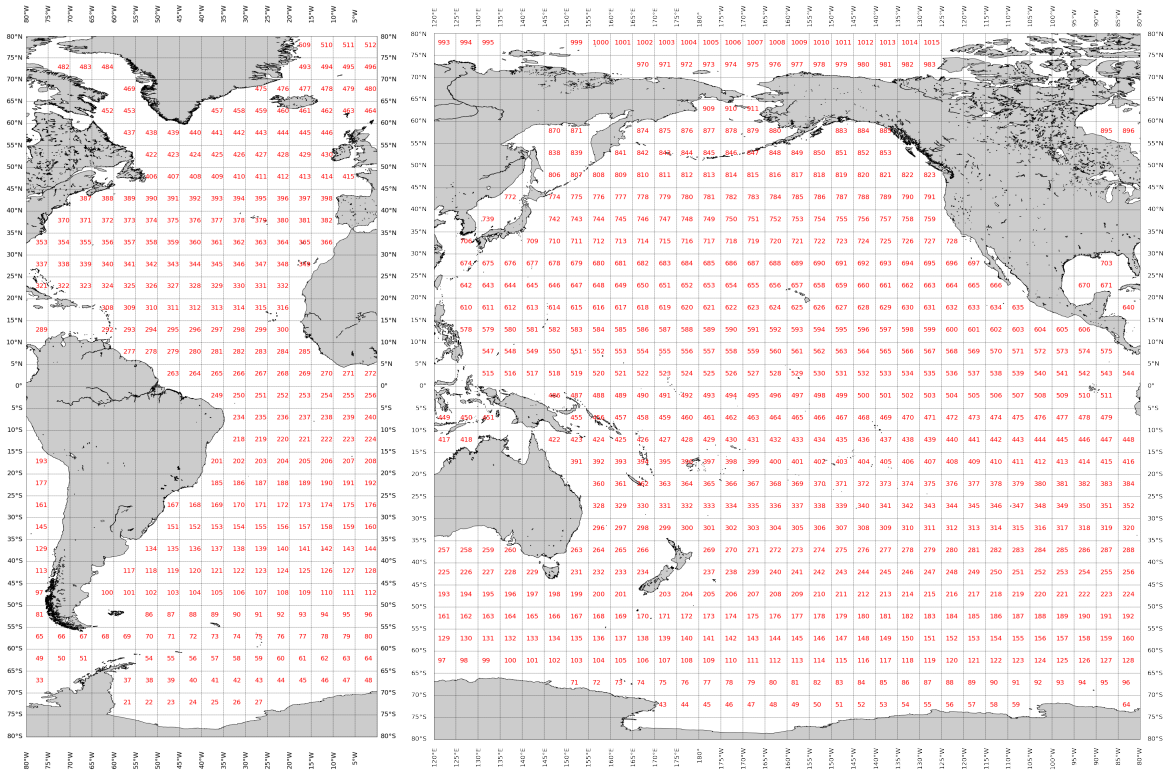
BVF squared is defined by:

$$N^2 = -\frac{g}{\rho}\frac{d\rho}{dz} \tag{4.8}$$

where g is the gravitational constant, $9.8m/s^2$, ρ is the density, z is the depth in metres and N^2 is in s^{-2}. The finite difference form of (4.8) is:

$$N^2 = -\frac{g}{\rho_l}\frac{(\rho_{l+1} - \rho_l)}{\Delta z_l} \tag{4.9}$$

where the index $l = 1, 2, ..., N$ indicates the levels from the surface to 2000m and Δz is 1m. We decided to calculate the box averages of BVF profiles for the Pacific and Atlantic ocean areas and seasonal averages only. Each domain is subdivided into 5 by 5 degree square boxes as shown in Fig. 4.3 and BVF is averaged in the boxes. Seasons are defined as: winter for Jan-Feb-Mar, spring for Apr-May-June, summer for Jul-Aug-Sept and autumn for Oct-Nov-Dec.



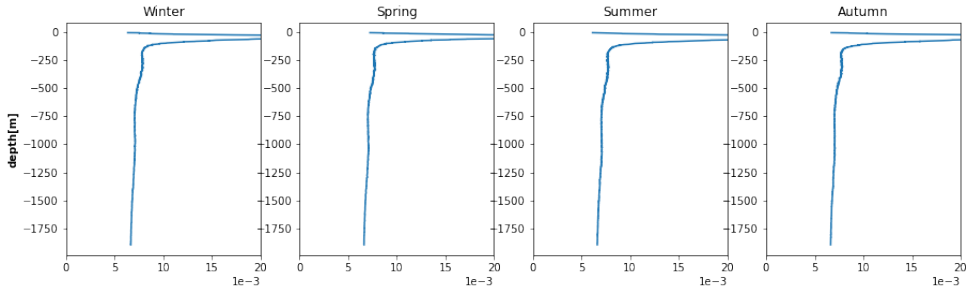
(a) (b)

Fig. 4.3: Brunt-Väisäla Frequency study 5×5 degree boxes for the Atlantic (a) and Pacific (b) ocean areas.

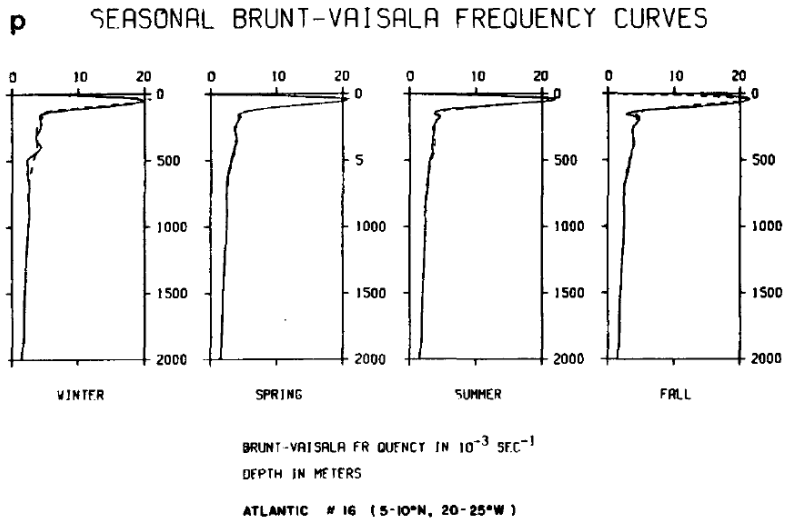
The resulting BVF profiles are shown in Figures (4.4 & 4.5) for the Equatorial Atlantic, Fig. (4.6 & 4.7) for the Equatorial Pacific, Fig. (4.8 & 4.9) for the North Atlantic, Fig. (4.10) for the North Pacific and Fig. (4.11 & 4.12) for the South Atlantic and Pacific, respectively. They are compared with the results reported in Emery et al. (1984).

Seasonally averaged BVF profiles show maxima in the upper 300m everywhere except in winter and spring in the North Atlantic and Pacific. Maximum BVF peaks occur in the summer followed by autumn, while minimum peaks are found in winter followed by spring, respectively, in the North Atlantic and Pacific. Minimal seasonal variations are noted in equatorial regions as shown in Fig. (4.4 & 4.5) for the Atlantic and in Fig. (4.6 & 4.7) for Pacific areas. For the southern hemisphere, the minimum values of BVF are reached during summer (of the northern hemisphere in our definition) and the maximum occur in winter. The comparison with Emery et al. (1984) profiles reveals overall good congruence. An increase in the magnitude of the peaks during summer (northern hemisphere) and winter (southern hemisphere) are worth noting. This can be due to the use of in situ density instead of potential density in (4.8) and climate warming.

This important aspect will be analyzed in the near future.

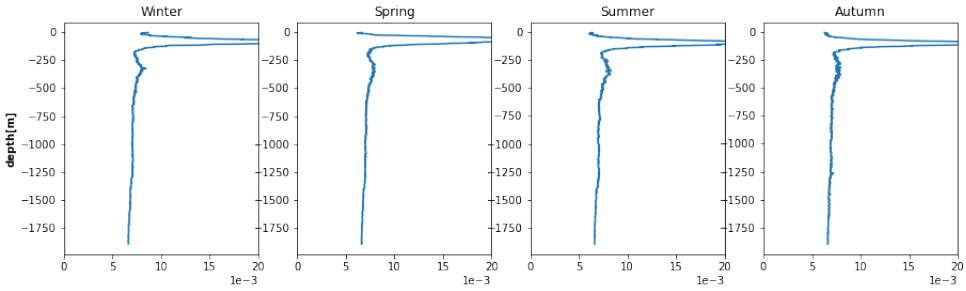


(a)

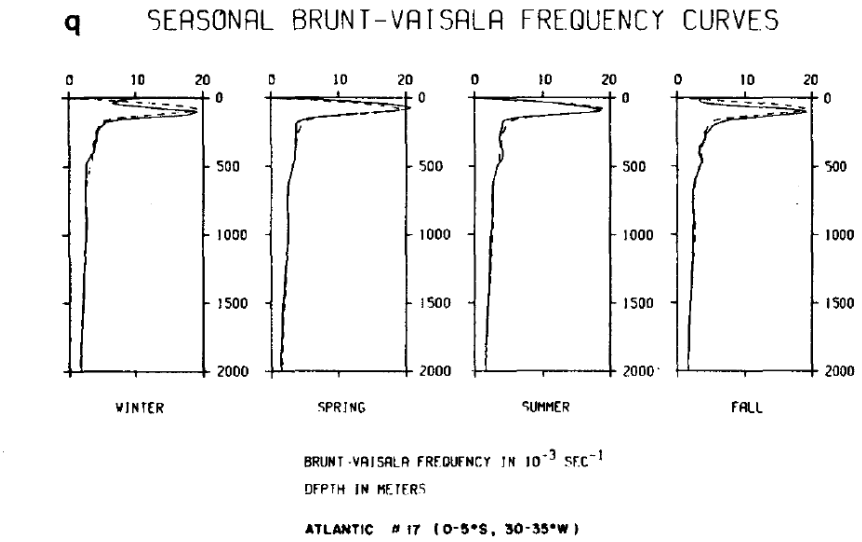


(b)

Fig. 4.4: Averaged BVF (s^{-1}) profiles for the Equatorial Atlantic: box no. 284 (f) from Fig. 4.3 and corresponding profiles from Emery et al. (1984) in (b).

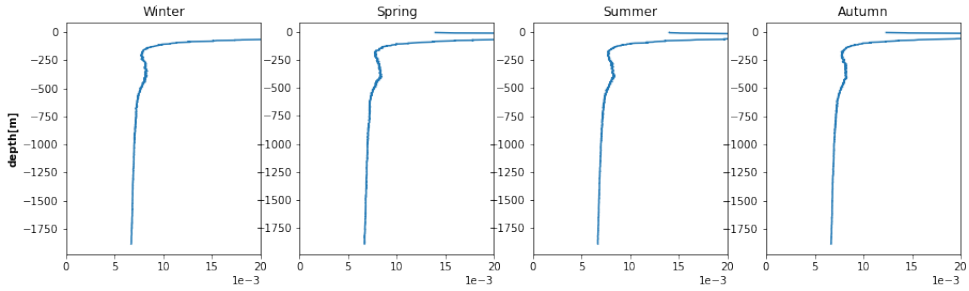


(a)

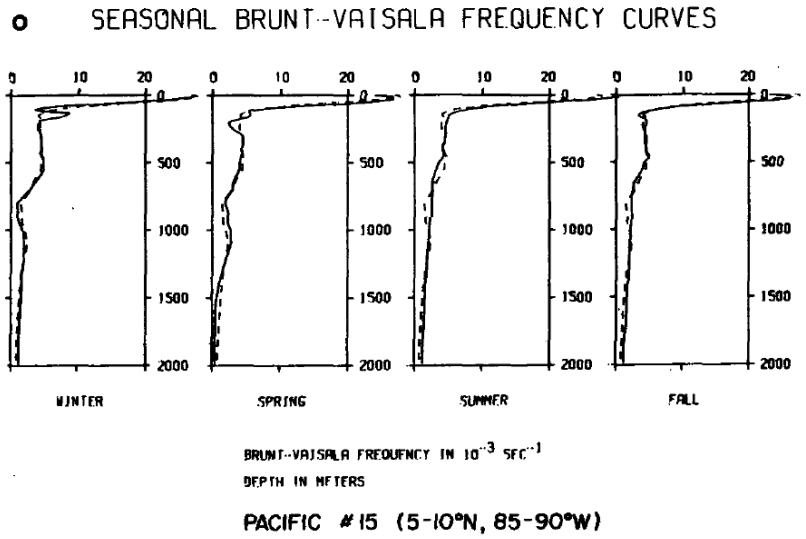


(b)

Fig. 4.5: Averaged BVF (s^{-1}) profiles for the Equatorial Atlantic: box no. 250 (a) from Fig. 4.3 and corresponding profiles from Emery et al. (1984) in (b).

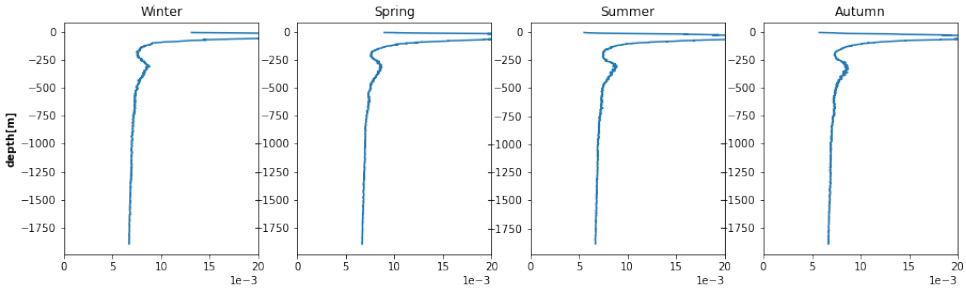


(a)

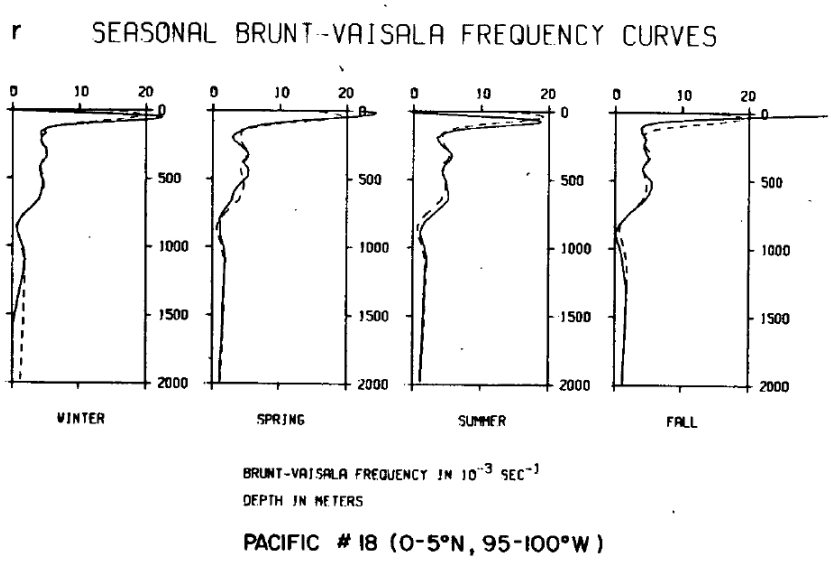


(b)

Fig. 4.6: Averaged BVF (s^{-1}) profiles in the Equatorial Pacific: Box no. 575 (a) from Fig. 4.3 and corresponding profiles from Emery et al. (1984) in (b).

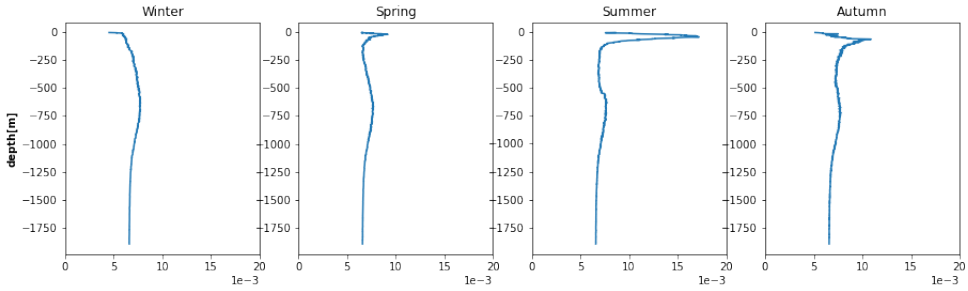


(a)

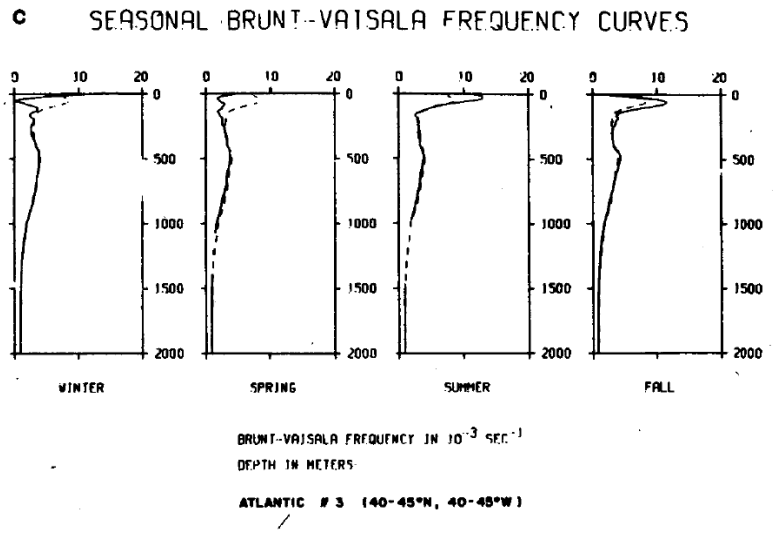


(b)

Fig. 4.7: Averaged BVF (s^{-1}) profiles in the Equatorial Pacific: Box no. 541 (a) from Fig. 4.3 and corresponding profiles from Emery et al. (1984) in (b).

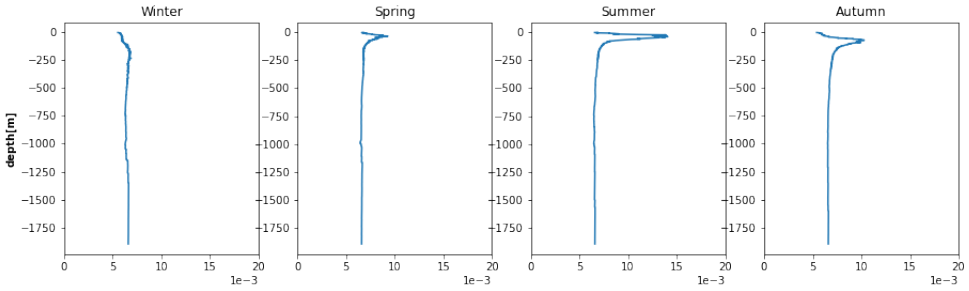


(a)

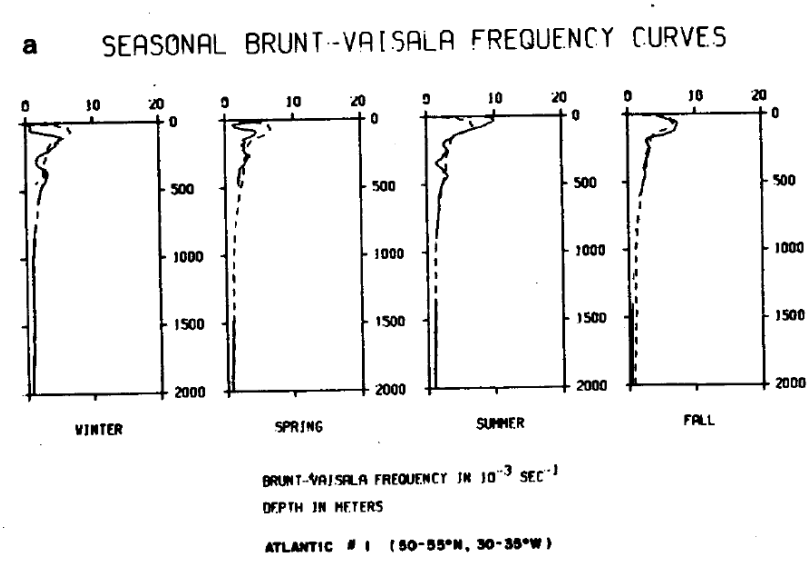


(b)

Fig. 4.8: Averaged BVF $[s^{-1}]$ profiles in North Atlantic: box no. 392 (a) from Fig. 4.3 and corresponding profiles from Emery et al. (1984) in (b).

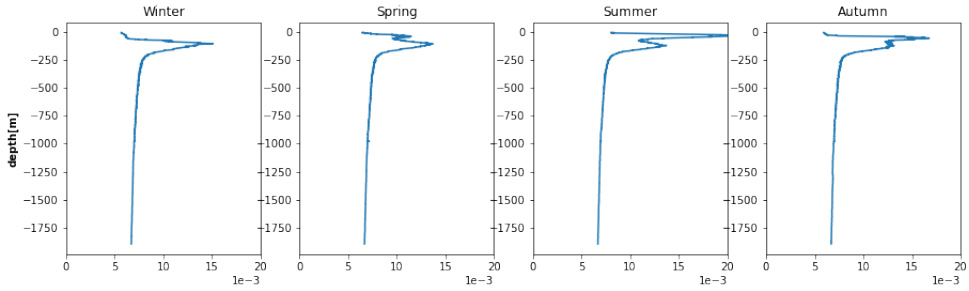


(a)

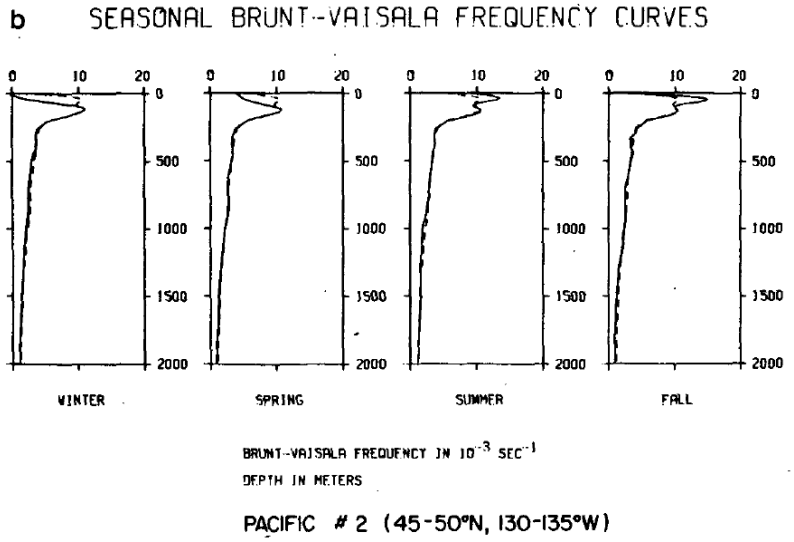


(b)

Fig. 4.9: Averaged BVF (s^{-1}) profiles in North Atlantic: box no. 426 (a) from Fig. 4.3 and corresponding profiles from Emery et al. (1984) in (b).

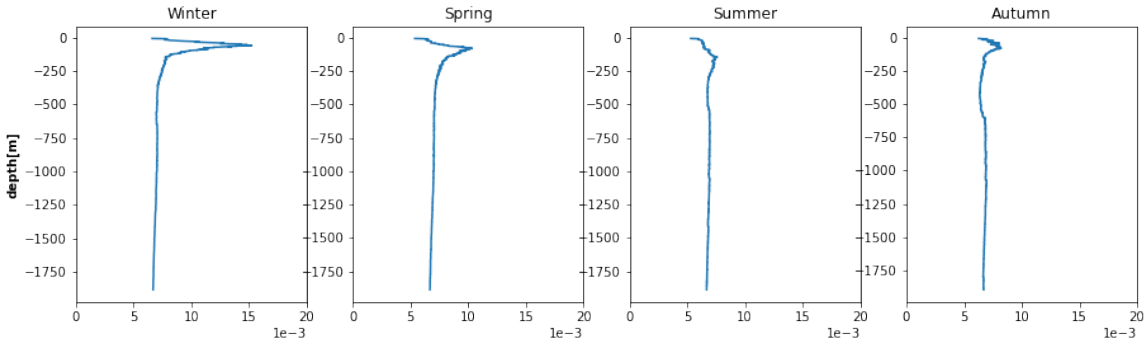


(a)

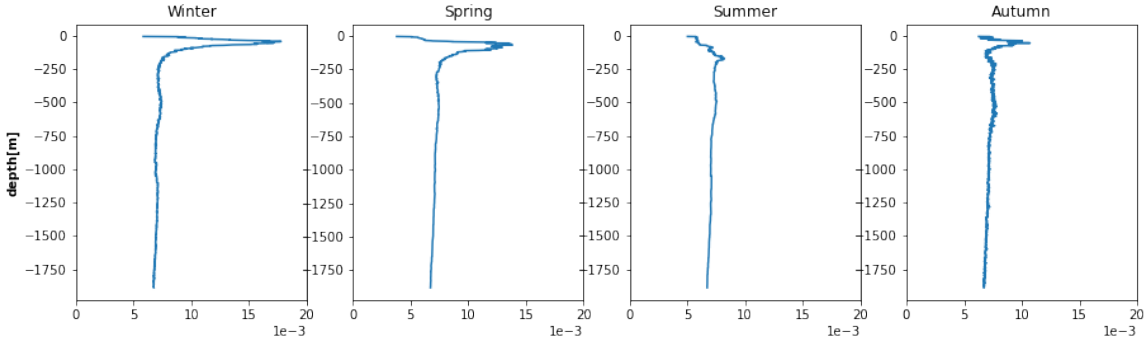


(b)

Fig. 4.10: Averaged BVF $[s^{-1}]$ profiles in North Pacific: box no. 822 (a) from Fig. 4.3 and corresponding profiles from Emery et al. (1984) in (b).

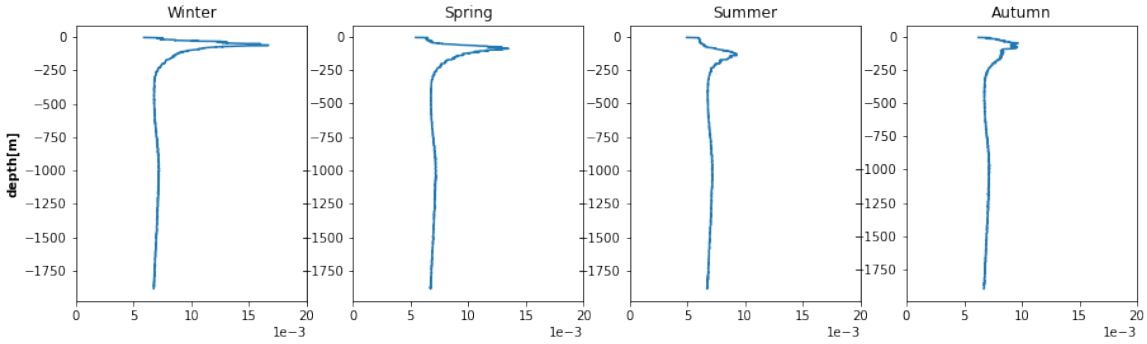


(a)

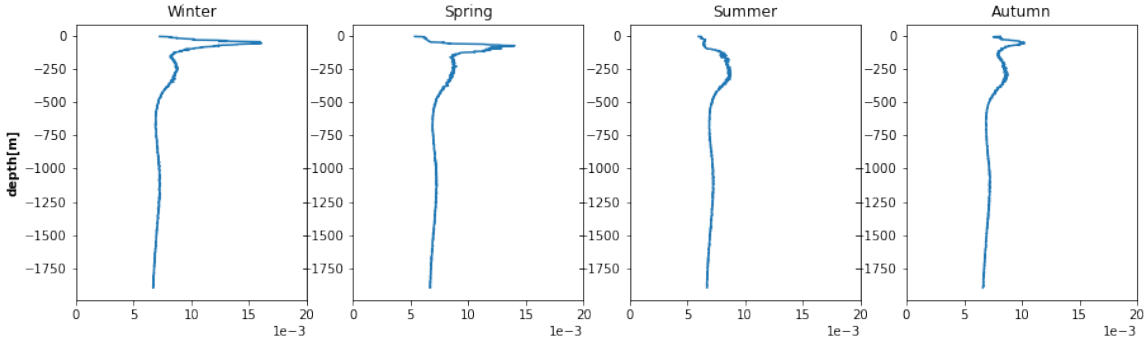


(b)

Fig. 4.11: Averaged BVF (s^{-1}) profiles in South Atlantic: box no. 122 (a) and 156(b) from Figure. 4.3.



(a)



(b)

Fig. 4.12: Averaged BVF (s^{-1}) profiles in South Pacific: box no. 246 (a) and 306 (b) from Figure. 4.3.

4.3 Global mapping of Dissolved Oxygen and AOU

As stated in the introduction, dissolved oxygen ($\mu mol\,kg^{-1}$) is a very important indicator of the state of ocean health. It is a non-conservative quantity since it is absorbed from the atmosphere by air-sea exchanges, produced by photosynthesis inside the water column and consumed by living organisms trough respiration.

Another important quantity is Apparent Oxygen Utilisation (AOU) which is the difference between saturation and in situ concentration. AOU ($\mu mol\,kg^{-1}$) at depth is used to infer oxygen consumption trough (mostly bacterial) respiration. It is defined as "apparent" because it is assumed that a water particle at depth left the surface in condition of oxygen saturation. In the next section, in situ dissolved oxygen observations and AOU are mapped using DIVA and the results are discussed.

4.3.1 Preprocessing of the Data

The input database used for this climatology is the WOD18 (Garcia et al., 2019), profiles from multiple platforms such as CTD and PFL, and bottle data from OSD for the 2003 to 2017 time periods with WOD QC i.e. WODf and WODfp (Table 4.2). No additional QC is applied in this analysis due to the scarcity of data. The only profiles selected are those where O2, Temperature and Salinity are present together in the datasets, as this is a requirement for the computation of the equilibrium saturation concentration as described below. This restriction has further decreased the amount of O2 profiles as listed in Table 4.2, where the total number of profiles and measurements are shown. Figure 4.13 shows O2 spatial distribution: measurements are frequent in the north Atlantic, North-East Pacific and Southern Ocean while there are mainly single transects in the rest of the ocean.

Total Profiles O_2	**Total Observations** O_2
225,434	60,050,946

Table 4.2: Number of profiles and observations of O2 (PFL, CTD and OSD) available from WOD18 when also temperature and salinity measurements are considered .

The spatial resolution of mapping is global and only a seasonal analysis of O2 and AOU is carried out due to lack of data at higher frequencies. The Apparent Oxygen Utilisation (AOU) is the difference between the in situ dissolved oxygen and its equilibrium saturation

concentration with the same physical and chemical properties. AOU represents the sum of the biological activity of sampled seawater experienced with respect to its last equilibrium with the atmosphere and is computed by the following equation.

$$AOU = O_2^s - O_2, \tag{4.10}$$

where O_2^s is the equilibrium saturation concentration with the same physical and chemical properties. The saturation concentration, O_2^s, is calculated by using the equation of Garcia and Gordon (1992) which uses solubility coefficients derived from the data of Benson and Krause Jr (1984).

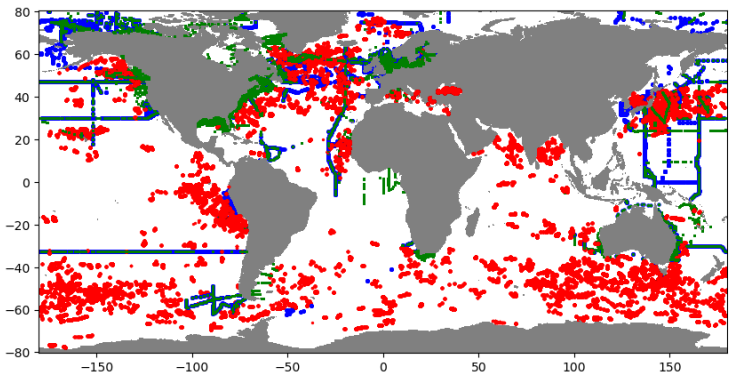
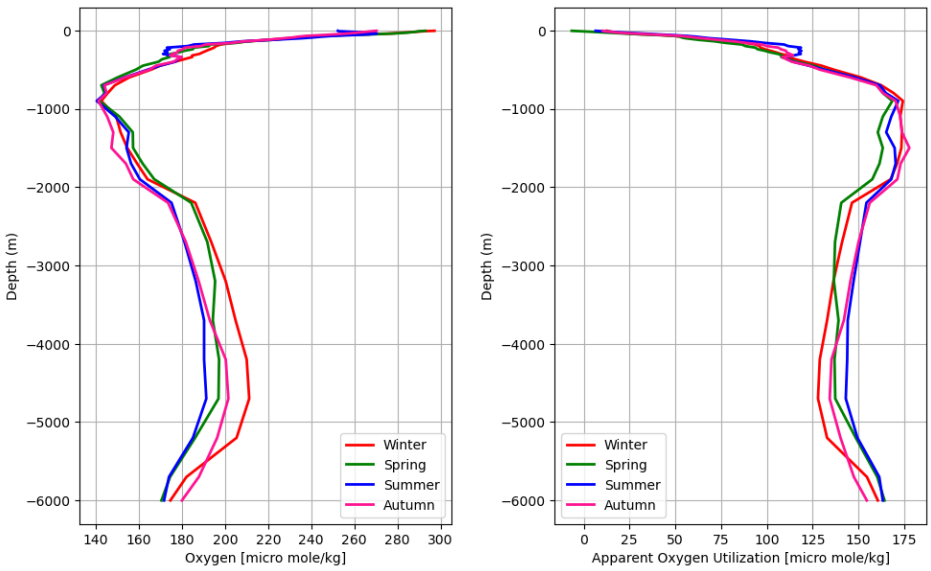


Fig. 4.13: Spatial distribution of observations for dissolved Oxygen for Summer.

4.3.2 Interpolation Algorithm Set-up

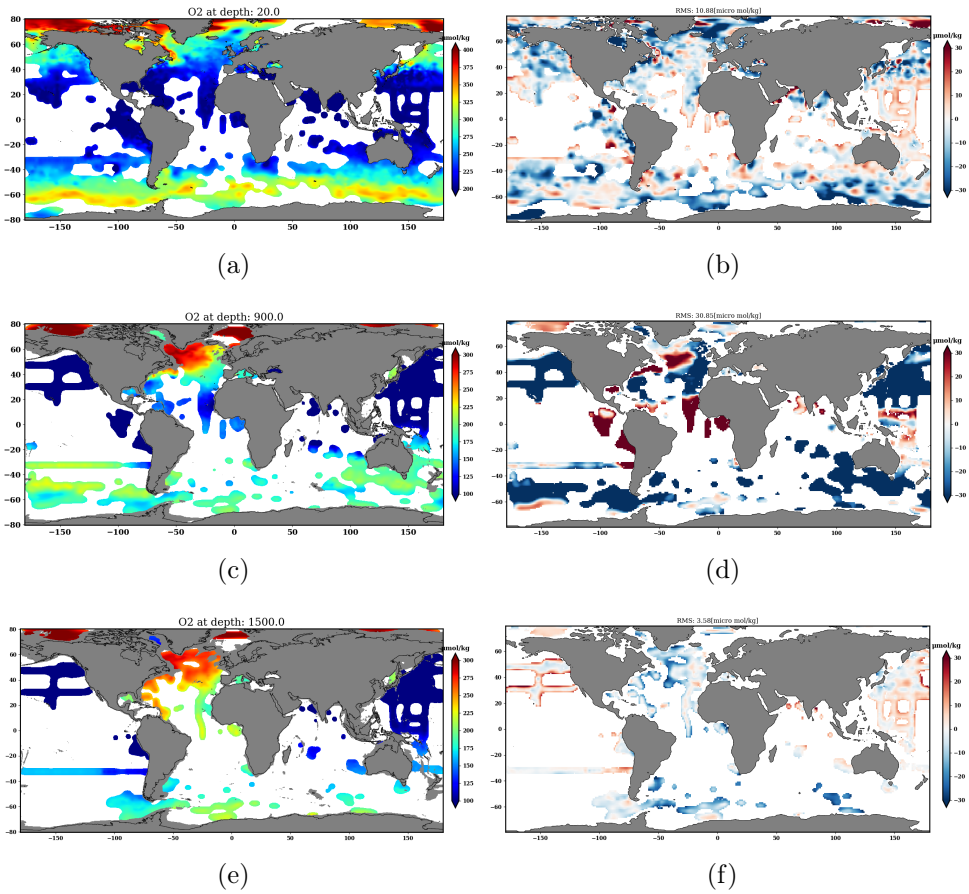
The DIVA mapping of dissolved oxygen and AOU is carried out seasonally, with seasons defined as described in the previous section. For the global domain, it is carried out with a spatial resolution of 0.25 degrees. The correlation length and noise-to-signal ratio are 300km and 0.5, respectively. These are the same values chosen in Chapter 3. The mapped fields of O2 and AOU for summer and winter are shown at 20m, 900m, and 1500m in Figures 4.15, 4.16, 4.17 and 4.18 along with the root mean square of the difference between the DIVA interpolated fields and the WOA18 available at a 1° resolution. The DIVA interpolated fields are masked with a 50% percentage error and are interpolated in the WOA18 grid for comparison purposes.



(a) (b)

Fig. 4.14: Vertical Profile used as background field for DIVA mapping: Dissolved Oxygen (a) and Apparent Oxygen Utilisation (b). Units are µmol kg^{-1}.

Overall, the distribution of O2 from the two data sets shows consistent patterns such as a high amount of O2 ranges between 300 to 400 µmol kg^{-1} at high latitudes, particularly in the Arctic region. Mid- and subtropical latitudes show minimum amounts of oxygen of around 100 to 200 µmol kg^{-1}. AOU depends on biological respiration activity and AOU is generally low, or even negative, at the surface because of the air-sea oxygen exchange processes. Negative values of AOU implies that dissolved oxygen (DO) has not been utilised by biological activity and it is larger than the saturation DO. This could be due to several physical and biological processes such as a strong intake of oxygen from air-sea sea interaction processes and low respiration of primary production. On other hand, deeper layers have a high AOU because the water parcel has not been in contact with the atmosphere for a long period of time and respiration is active.



(a) (b)

(c) (d)

(e) (f)

Fig. 4.15: O2 (µmol kg^{-1}) mapping for Summer (left column) and difference between this thesis mapped field and the WOA climatology (right column) at different depths: for 20m for ((a) and (b)); 900m ((c) and (d)); and for 1500m ((e) and (f)).

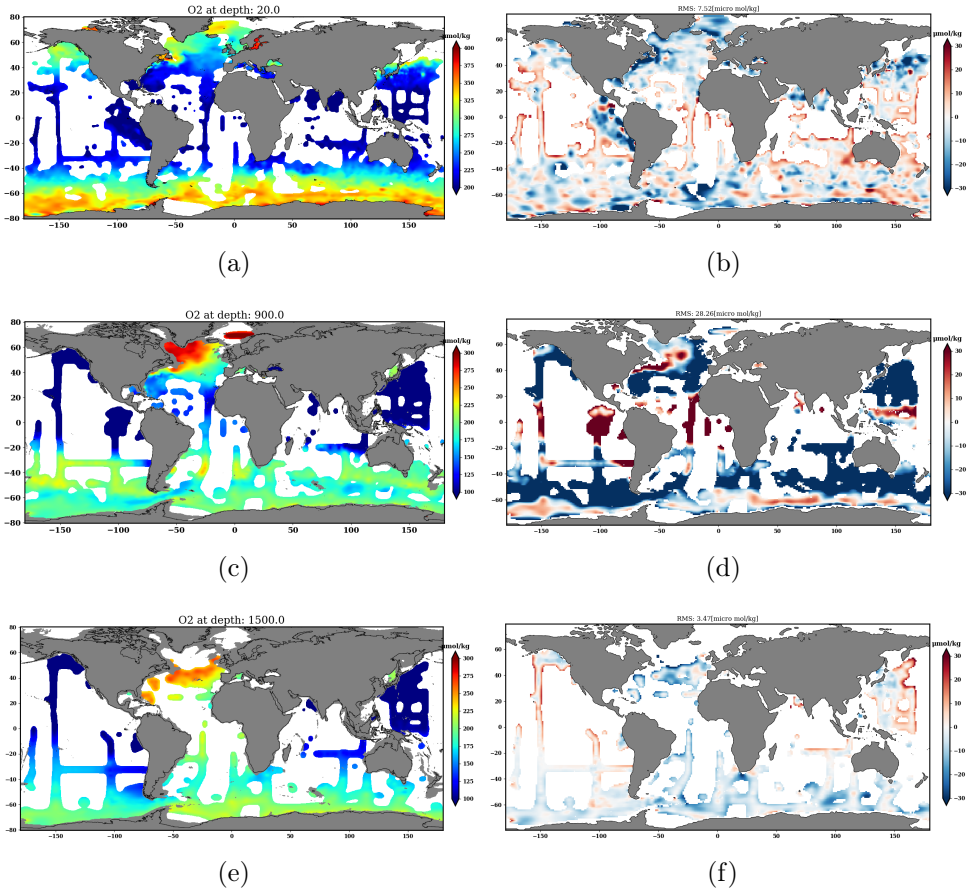
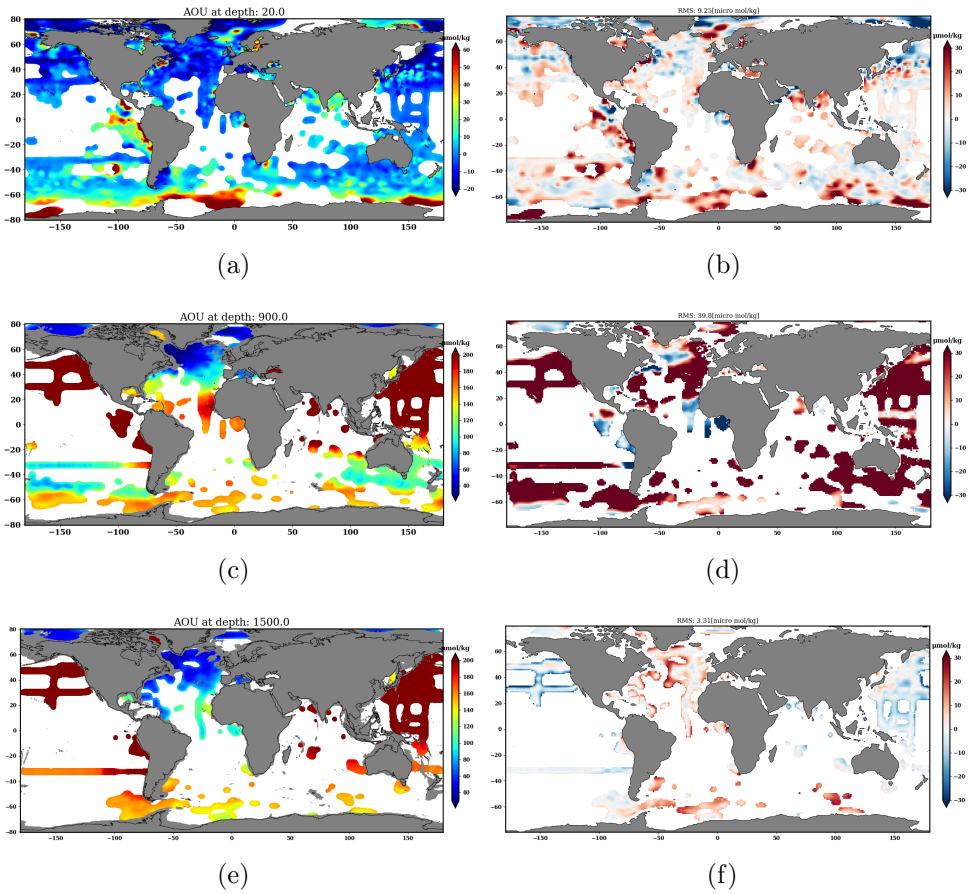


Fig. 4.16: O2 (μmol kg^{-1}) mapping for Winter (left column) and difference between this thesis mapped field and the WOA climatology (right column) at different depths: for 20m ((a) and (b)); 900m ((c) and (d)); and for 1500m ((e) and (f)).

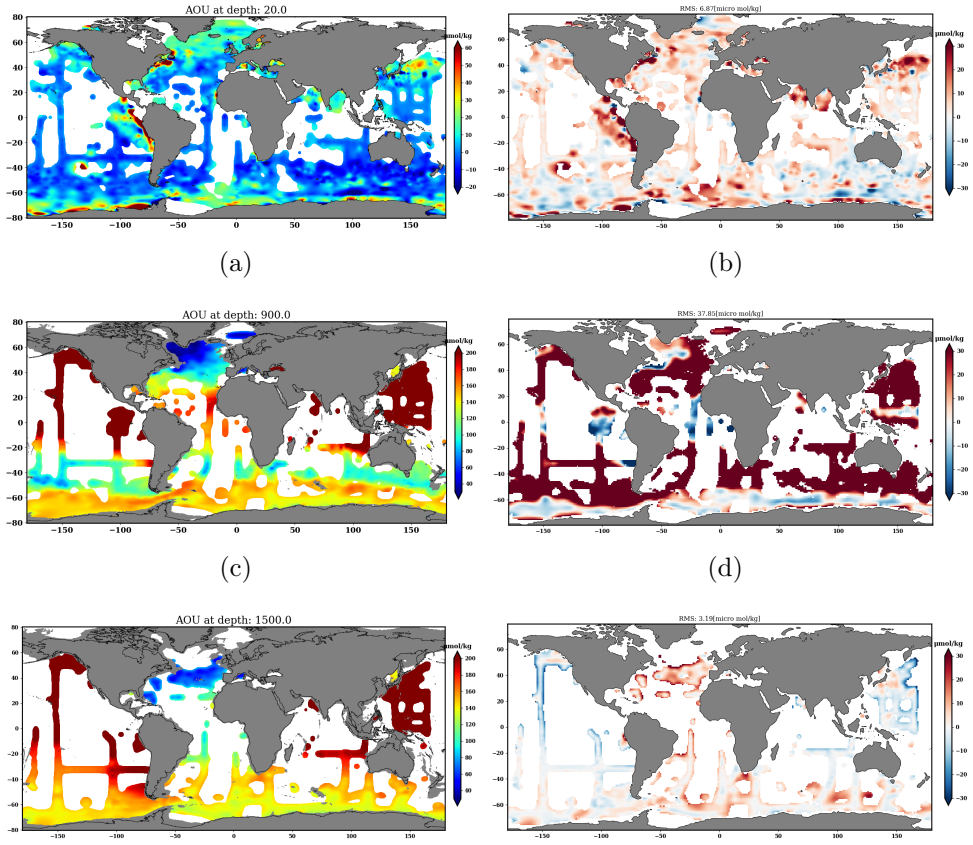


(a) (b)

(c) (d)

(e) (f)

Fig. 4.17: AOU (μmol kg^{-1}) interpolated field for Summer (left column) and difference between the interpolated and the WOA mapped field (right column) at different depths: for 20m ((a) and (b)); 900m ((c) and (d)); and for 1500m ((e) and (f)).



(a) (b) (c) (d) (e) (f)

Fig. 4.18: AOU ($\mu mol\,kg^{-1}$) interpolated field for Winter (left column) and difference between the interpolated and the WOA mapped field (right column) at different depths: for 20m((a) and (b)); 900m ((c) and (d)); and for 1500m ((e) and (f)).

The difference between the DIVA interpolated oxygen or AOU and WOA18 shows a negative difference of 10 to 30 $\mu mol\,kg^{-1}$ at all the depths and seasons, except for a positive difference in the eastern equatorial regions that are characterised by oxygen minimum zones. Moreover, the comparison of the DIVA interpolated fields with WOA18 shows maximum positive differences at around 900m which implies higher oxygen utilisation when compared to the WOA18 mapping. This is also a relevant theme to be further addressed in the near future.

4.4 Summary and Future Steps

In this chapter, density profiles are computed from NQC-processed in situ temperature and salinity profiles. Averaged Brunt-Väisäla frequency profiles are computed in 5° boxes for the Pacific and Atlantic Oceans. Stratification was found to be higher than previously reported.

O2 and AOU DIVA mapping was carried out using WOD18 data for the past 15 years. The comparison of the O2 and AOU DIVA mapping with WOA18 shows that the water column is possibly less oxygenated and biological respiration activity might also be higher. These differences could be attributed to two main reasons:
(i) The averaging time period of WOA18 is from 1964 to 2017 and only bottle data have been considered in computing the estimate.
(ii) The temperature in the recent data set shows warming temperature trends when compared to the long-term average from 1900 to 2017 (see Chapter 3). This potentially supports the fact that warmer water has less ability to dissolve oxygen, thereby confirming the fact that less oxygen is entering the ocean.

These issues demand an extensive analysis that was not carried out in this thesis. Now that we have shown that the DIVA interpolation algorithm works for all parameters and that this algorithm has given consistent results with respect to published climatologies, we will undertake a specific study on the changing ocean conditions in the different decades.

In this chapter, density profiles are computed from NQC processed in-situ temperature and salinity profiles. Averaged Brunt-Väisäla frequency profiles are computed in 5° boxes for Pacific and Atlantic Ocean and it is found that stratification is higher than previously reported.

4.5 Data Availability Statement

The analysis carried out in this chapter has resulted as three data products and can be accessed using the following link:
https://doi.org/10.12770/725d02ed-2c93-44e8-a1ac-48a9cd5ac883 *(Global Ocean Climatology - Density Climatology)*
https://doi.org/10.12770/ee42b875-d615-42ed-9023-e1e7d7d27250 *(Averaged Brunt-Vaisala Frequency profiles for the Atlantic and Pacific Oceans)*
https://doi.org/10.12770/b7af83fb-83ba-46bf-9c4f-26167b5b3fda *(Global Ocean Climatology - Dissolved Oxygen and Apparent Oxygen Climatology)*

Chapter 5

Summary and Conclusions

The knowledge of the climatological state of the global ocean is essential for climate studies giving an estimate of the "reference" variability of the essential ocean variables. Furthermore in this thesis we have designed and analyzed the methods to produce such climatological estimates only from observational data using statistical algorithms that can map the irregularly spaced data into regular grids. Such methods are complementary to reanalysis methodologies that produce the optimal melding of observations with numerical ocean models. These climatological studies are possible because historical data sets are now available with common formats and they are freely available. The historical data set used in this thesis is WOD18 (Garcia et al., 2018) for temperature, salinity and dissolved oxygen. These essential ocean variables are the basic ones to start a proper investigation of health of the ocean issues as required for the Sustainable Development Goals targets (Rosa, 2017). As the mean/climatological state of the ocean is a virtual representation of the reality, i.e. a statistical mean of measurements taken at different space-time scales, several approaches are needed in order to reduce uncertainties in the estimates. Recently a variational methodology has been developed (DIVA, Barth et al. (2014)) for observational databases that estimates the climatological state of the oceans only on the basis of statistical assumptions about errors in the observations, in the background and the correlation length. The method is optimal from the point of view of the least square theory because it weights observations and background with their errors and produces an estimate of the gridded-climatological errors. The following sections we will summarise the research carried out in this thesis, followed by conclusion and directions of future work.

5.1 Summary

In this thesis, a new global ocean climatological estimate for basic physical parameters such as temperature, salinity, density, and dissolved oxygen has been computed. The reliability of estimate is closely tied to the quality assurance of the in-situ observations and statistical interpolation schemes of mapping. Therefore in the present work, a special focus has been given to following:

(i)- A Nonlinear Quality Control (NQC) procedure is developed (Chapter 2) in order to eliminate non- representative or high frequency signals from the historical datasets . The procedure eliminates the non-representative data and outliers based on the first order statistics of homogenous dynamical region. The subdivision of the target domain into dynamical subregions, and the computation of the standard deviations (std) used as a threshold to eliminate non-representative data are the two key steps of the procedure. In the present study, the domain subdivision is considered from our knowledge of ocean currents and topography. The observations are first gridded in the subdomain, then the subregion std is used to discard the non-representative data. The procedure is iterative and statistics of data changes at each step and eliminate the data until convergence is reached, i.e. no more data is rejected by comparison with the standard deviation in the subregion. The method is tested in regional domains such as North Atlantic, South Atlantic and North West Pacific and results are compared with and without NQC mapping procedures.

(ii)- The temperature and salinity mapping using the variational algorithm called DIVA (Chapter 3) is carried out after an application of NQC on the WOD18 datasets. In this study, two different versions of the climatological estimates are produced: a long term (1900 to 2017) climatology using multiple platforms in-situ data, a shorter time estimate (2003-2017) using data from ocean drifting platforms such as profiling floats. It is found that the choice of background is not so important for the quality of the final estimate if NQC is applied prior the mapping algorithm is applied. Sensitivity experiments are carried out to choose the key parameters of DIVA, that are the horizontal correlation lengths and the Noise to Signal ratio (N/S). Furthermore a "thin layer" discretization is used in vertical, without any vertical interpolation thus avoiding the smearing of the vertical temperature and salinity profiles. Furthermore two new indices are designed to show the impact of the correlation and N/S ration choices. The computed climatologies show consistency with well-known reference climatology such as WOA18

(Locarnini et al., 2018) and WAGHC (Gouretski, 2019).

(iii)- An initial development of health of the ocean indicators from the historical data set, with the validated NQC and mapping algorithms, has also been presented (Chapter 4). In this context, mapping of density has been carried out from NQC processed profiling float dataset, using DIVA with same parameters as selected for the temperature and salinity mapping. Brunt-Väisäla frequency profiles were then computed from the density profiles and averaged in 5° square boxes for Atlantic and Pacific Oceans to have an estimate of stratification. The climatology of Brunt-Väisäla frequency has been compared with previous literature results of Emery et al. (1984). The computation of dissolved oxygen climatology and saturation oxygen in the Brunt-Väisälaater column was analyzed and the Apparent Oxygen Utilization (AOU) values were analyzed. AOU contains information about biological respiration in the ocean and it is directly connected to ocean health indicators.

5.2 Conclusions

The major findings of the Thesis are summarised as follows:

(i)- The NQC algorithm developed in this thesis has eliminated the prominent outliers and non-representative temperature and salinity values from the data set. Only four iterations are required for the algorithm to converge and only about 15% of the data are eliminated. NQC has shown thus the potential of being a robust quality control procedure for the global ocean and for computation of large scale ocean climatologies.

(ii)- The application of NQC to the historical data set shows that the choice of background field in the interpolation algorithm becomes less important for the quality of the resulting climatology. The comparison of SDC with available estimates has highlighted various uncertainties that are combination of various hypothesis in the mapping algorithms or the datasets being used. In order to address these discrepancies, a multimode ensemble mean is computed and found a more reliable for global ocean climatologies.

(iii) This thesis has shown that the application of variational interpolation algorithms to the biogeochemical data such oxygen data and apparent oxygen utilisation. The comparison of Brunt-Väisäla Frequency with available estimates in the literature has shown high stratification for all the seasons. In addition, preliminary results of comparisons of O2 and AOU with available literature has shown several differences that needs further investigations.

5.3 Future Prospects

The success of the NQC procedure developed in this thesis relies on the correct determination of homogenous dynamical subregions of the target interpolating domain. In this context, there is a need to develop an advanced machine learning algorithm that will be able to define the subdomains for the optimal control of the data. The extension of NQC to biogeochemical data should be also tried. Optimisation of DIVA parameters is essential for a quality estimate of the global ocean climatology. In the future it will be preferable to use different correlation functions for the different regions of the world ocean, given the different dynamics that prevail in the tropics with respect to the polar regions, etc.

Evidence for ocean warming is present in the two global ocean climatologies calculated for the two periods, 1900-2017 and 2003-2017. Furthermore the preliminary results of comparison of averaged Brunt-Väisäla profiles with Emery et al. (1984) shows significant stratification changes in the water column. Moreover, the comparison of O2 and AOU calculated only for the period 2003-2017 seems to indicate lower and higher values respectively in comparison with WOA18. The differences are found to be large at intermediate depth for both O2 and AOU. These results could be important for health of the ocean considerations and further extensive analysis is required before a solid conclusion is reached.

Appendix A

The ocean was first measured in the 17th century by Count Luigi Ferdinando Marsili (Pinardi et al., 2018). Considerable progress was then made with the four-year Challenger expedition (1872-1876) (Deacon et al., 2013), during which physical, biochemical and geological in-situ observations were collected for scientific ocean studies.

The modern but early ocean measurements started with reversing mercury thermometers in 1874 and advanced to Nansen bottles in 1894, which were replaced by Niskin bottles in 1966(Warren, 2008). The advent of the rosette sampler in 1930 and later the Conductivity, Temperature, Depth (CTD) instrument in the 1970s (Wallace, 1974) substantially increased the capacity to measure the ocean variables at several time and space scales. The CTD uses a platinum thermometer to measure temperature, a conductivity sensor to measure the salinity of water, and a pressure sensor to measure depth. CTD led to the possibility of continuous measurements of temperature and salinity along the water column and at the surface using a thermosalinograph (Schloesser et al., 2016). For temperature only, the Mechanical Bathythermograph (MBT), an instrument developed during the late 1930s, introduced the concept of semi-autonomous sampling of the water column. It was replaced by the Expendable Bathythermograph (XBT) in 1966 (Spilhaus, 1938).

In order to carry out observations from these sensors, a platform is required for deployment at the sea surface, the sea floor or in the ocean interior. The platform varies depending on the type of measurement required in the space and time. For instance, a research vessel is an Eulerian platform that is used to deploy rosette bottles, and also for MBTs and XBTs. An Anchor is a mooring cable attached to a floating element that keeps it close to vertical positions; and moored buoys attached with various sensors such as CTDs record seawater measurements at a fixed position. Lagrangian measurements are taken by buoys and floats that drift with the ocean currents. Autonomous, vehicles such as gliders (Davis et al., 2002) and profiling floats (Gould et al., 2004) equipped with different sensors are also used for oceanographic measurements. In order to fill the temporal and spatial gaps in the harsh environmental areas such as the southern ocean, marine animals such as northern and southern Weddell elephant seals are used as platforms (Boehlert et al., 2001). In addition, satellite technologies can be used to measure the dynamic properties from space such as sea surface temperature, color of the ocean, coral reefs, and sea ice.

Appendix B: Objective Analysis

Objective Analysis was introduced in oceanography by Bretherton et al. (1976) and Carter and Robinson (1987). It is a mapping technique to interpolate irregularly space data in a regular grid. OA is based on the Gauss-Markov theorem that consider the minimization of the variance of the observational errors in the interpolated field, also called an analysis. OA requires the knowledge of the correlation function of the field to be interpolated. In order to simplify the procedure, an isotropic and homogenous Gaussian correlation function is normally used which assumes that the targeted field is stationary and homogenous. The spatial correlation function used in the OA is:

$$F(r) = (1 - \frac{r^2}{a^2})exp(-\frac{r^2}{2b^2}) \tag{5.1}$$

$$r^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 \tag{5.2}$$

where a is de-correlation distance, while b is a folding scale and r is the distance between two coordinate points (x_1, y_1) and (x_2, y_2). In order to fulfil the condition of having a positive definite correlation matrix, a and b must satisfy condition $a > b\sqrt{2}$. In addition, OA provides an estimation of the analysis error field. The specific OA parameters chosen in our interpolation are shown in Table 5.1.

Objective Analysis Parameters	
Grid Spatial Resolution	0.25° latitude, longitude
Temporal Resolution	monthly
Zero Crossing Distance (a)	700 km
Decay Length (b)	400 km
Radius of Influence	300 km

Table 5.1: OA Parameters for the three region of application of NQC algorithm

This study also used the R radius of influence to interpolate the data, i.e. the maximum radius where observations are considered in order to estimate the field at a given grid point. Equation 5.1 coverts the longitude and latitude into spherical coordinates that involves radius of earth hence it is taking the curvature of the earth into account. The interpolated data are masked by an analysis percentage error variance greater than 30%.

Bibliography

Balmaseda, M. A., Mogensen, K., and Weaver, A. T. (2013). Evaluation of the ecmwf ocean reanalysis system oras4. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1132–1161.

Barker, P. M., Dunn, J. R., Domingues, C. M., and Wijffels, S. E. (2011). Pressure sensor drifts in argo and their impacts. *Journal of Atmospheric and Oceanic Technology*, 28(8):1036–1049.

Barnes, S. L. (1964). A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, 3(4):396–409.

Barth, A., Azcárate, A. A., Joassin, P., Beckers, J.-M., and Troupin, C. (2008). Introduction to optimal interpolation and variational analysis. *SESAME Summer School, Varna, Bulgaria.*

Barth, A., Beckers, J.-M., Troupin, C., Alvera Azcarate, A., and Vandenbulcke, L. (2014). divand-1.0: n-dimensional variational data analysis for ocean observations. *Geoscientific Model Development*, 7:225–241.

Belbeoch, M., Viola, H., Freeland, H., Clark, C., Meldrum, D., Charpentier, E., Alverson, K., Dexter, P., and Fellous, J.-L. (2010). The jcomm in situ observing programme support centre: A decade of progress and remaining challenges.

Bell, M. J., Lefebvre, M., Le Traon, P.-Y., Smith, N., and Wilmer-Becker, K. (2009). Godae: The global ocean data assimilation experiment. *Oceanography*, 22(3):14–21.

Benson, B. B. and Krause Jr, D. (1984). The concentration and isotopic fractionation of oxygen dissolved in freshwater and seawater in equilibrium with the atmosphere 1. *Limnology and oceanography*, 29(3):620–632.

Bhaskar, T. U., Shesu, R. V., Boyer, T. P., and Rao, E. P. R. (2017). Quality control of oceanographic in situ data from argo floats using climatological convex hulls. *MethodsX*, 4:469–479.

Boehlert, G. W., Costa, D. P., Crocker, D. E., Green, P., O'Brien, T., Levitus, S., and Le Boeuf, B. J. (2001). Autonomous pinniped environmental samplers: using instrumented animals as oceanographic data collectors. *Journal of atmospheric and oceanic technology*, 18(11):1882–1893.

Boyer, T., Baranova, O., Coleman, C., Garcia, H., Grodsky, A., and Locarnini, R. (2019). World ocean database 2018. av mishonov, technical editor. *NOAA Atlas NESDIS*, 87.

Brasseur, P. P. (1991). A variational inverse method for the reconstruction of general circulation fields in the northern bering sea. *Journal of Geophysical Research: Oceans*, 96(C3):4891–4907.

Bretherton, F. P., Davis, R. E., and Fandry, C. (1976). A technique for objective analysis and design of oceanographic experiments applied to mode-73. In *Deep Sea Research and Oceanographic Abstracts*, volume 23, pages 559–582. Elsevier.

Carter, E. F. and Robinson, A. R. (1987). Analysis models for the estimation of oceanic fields. *Journal of Atmospheric and Oceanic Technology*, 4(1):49–74.

Carton, J. A. and Giese, B. (2008). A reanalysis of ocean climate using simple ocean data assimilation (soda). *Mon Weather Rev*, 136(8):2999–3017.

Chen, L., Huang, J.-G., Ma, Q., Hänninen, H., Tremblay, F., and Bergeron, Y. (2019). Long-term changes in the impacts of global warming on leaf phenology of four temperate tree species. *Global change biology*, 25(3):997–1004.

Cressman, G. P. (1959). An operational objective analysis system. *Mon. Wea. Rev*, 87(10):367–374.

Daley, R. (1993). *Atmospheric data analysis*. Number 2. Cambridge university press.

Davis, R. E., Eriksen, C. C., Jones, C. P., et al. (2002). Autonomous buoyancy-driven underwater gliders. *The technology and applications of autonomous underwater vehicles*, pages 37–58.

Deacon, M., Rice, T., and Summerhayes, C. (2013). *Understanding the oceans: a century of ocean exploration*. Routledge.

Emery, W., Lee, W., and Magaard, L. (1984). Geographic and seasonal distributions of brunt–väisälä frequency and rossby radii in the north pacific and north atlantic. *Journal of Physical Oceanography*, 14(2):294–317.

Fofonoff, N. P. and Millard Jr, R. (1983). Algorithms for the computation of fundamental properties of seawater.

Fukumori, I., Heimbach, P., Ponte, R. M., and Wunsch, C. (2018). A dynamically consistent, multivariable ocean climatology. *Bulletin of the American Meteorological Society*, 99(10):2107–2128.

Gandin, L. (1960). On optimal interpolation and extrapolation of meteorological fields. *Trudy GGO*, 114:75–89.

Garcia, H., Boyer, T., Locarnini, R., Baranova, O., and Zweng, M. (2018). World ocean database 2018: Users manual (prerelease). *AV Mishonov, Technical Ed., NOAA, Silver Spring, MD (Available at https://www. NCEI. noaa. gov/OC5/WOD/pr_ wod. html)*.

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, M., Zweng, M., Mishonov, A., Baranova, O., Seidov, D., et al. (2019). World ocean atlas 2018, volume 3: Dissolved oxygen, apparent oxygen utilization, and dissolved oxygen saturation.

Garcia, H. E. and Gordon, L. I. (1992). Oxygen solubility in seawater: Better fitting equations. *Limnology and oceanography*, 37(6):1307–1312.

Good, S. A., Martin, M. J., and 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.

Gould, J., Roemmich, D., Wijffels, S., Freeland, H., Ignaszewsky, M., Jianping, X., Pouliquen, S., Desaubies, Y., Send, U., Radhakrishnan, K., et al. (2004). Argo profiling floats bring new era of in situ ocean observations. *Eos, Transactions American Geophysical Union*, 85(19):185–191.

Gouretski, V. (2018). World ocean circulation experiment-argo global hydrographic climatology. *Ocean Science*, 14(5).

Gouretski, V. (2019). A new global ocean hydrographic climatology. *Atmospheric and Oceanic Science Letters*, 12(3):226–229.

Gouretski, V. and Koltermann, K. (2004). Woce global hydrographic climatology: A technical report, vol. 35. *Bundesamtes fur Seeschiffahrt und Hydrographie, Hamburg, Germany*.

Hacker, P., Maximenko, N., Potemra, J., Lebedev, K., DeCarlo, S., and Shen, Y. (2010). Argo and synthesis products developed and served at the asia-pacific data-research center. *Proceedings of OceanObs' 09: Sustained Ocean Observations and Information for Society (Annex)*.

Hosoda, S., Ohira, T., and Nakamura, T. (2008). A monthly mean dataset of global oceanic temperature and salinity derived from argo float observations. *JAMSTEC Report of Research and Development*, 8:47–59.

IOC, I. (2003). Bodc: Centenary edition of the gebco digital atlas, published on cd-rom on behalf of the intergovernmental oceanographic commission and the international hydrographic organization as part of the general bathymetric chart of the oceans. *British Oceanographic Data Centre, Liverpool, UK*.

Janjić, T., Bormann, N., Bocquet, M., Carton, J., Cohn, S., Dance, S., Losa, S., Nichols, N., Potthast, R., Waller, J., et al. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278.

Janjić, T. and Cohn, S. E. (2006). Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, 134(10):2900–2915.

Jia, W., Wang, D., Pinardi, N., Simoncelli, S., Storto, A., and Masina, S. (2016). A quality control procedure for climatological studies using argo data in the north pacific western boundary current region. *Journal of Atmospheric and Oceanic Technology*, 33(12):2717–2733.

Krishnamurti, T., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S. (1999). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, 285(5433):1548–1550.

Le Traon, P.-Y. (2011). Satellites and operational oceanography. pages 29–54.

Lellouche, J. and Regnier, C. (2015). For the global ocean sea physical analysis and forecasting products global_analysis_forecast_phys_001_002. product user manual.

Levitus, S. (1982). *Climatological atlas of the world ocean*, volume 13. US Department of Commerce, National Oceanic and Atmospheric Administration.

Locarnini, M., Mishonov, A., Baranova, O., Boyer, T., Zweng, M., Garcia, H., Seidov, D., Weathers, K., Paver, C., Smolyar, I., et al. (2018). World ocean atlas 2018, volume 1: Temperature.

Lothar, S. and Matthew, E. (1999). On the water masses and mean circulation of the south atlantic ocean. *Journal of Geophysical Research: Oceans*, 104(C9):20863–20883.

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

McDougall, T., Feistel, R., Millero, F., Jackett, D., Wright, D., King, B., Marion, G., Chen, C., Spitzer, P., and Seitz, S. (2009). The international thermodynamic equation of seawater 2010 (teos-10): Calculation and use of thermodynamic properties. *Global Ship-based Repeat Hydrography Manual, IOCCP Report No*, 14.

McIntosh, P. C. (1990). Oceanographic data interpolation: Objective analysis and splines. *Journal of Geophysical Research: Oceans*, 95(C8):13529–13541.

Moltmann, T., Turton, J., Zhang, H.-M., Nolan, G., Gouldman, C., Griesbauer, L., Willis, Z., Piniella, A. M. n., Barrell, S., Andersson, E., Gallage, C., Charpentier, E., Belbeoch, M., Poli, P., Rea, A., Burger, E. F., Legler, D. M., Lumpkin, R., Meinig, C., O'Brien, K., Saha, K., Sutton, A., Zhang, D., and Zhang, Y. (2019). A global ocean observing system (goos), delivered through enhanced collaboration across regions, communities, and new technologies. 6.

Pinardi, N., Özsoy, E., Latif, M. A., Moroni, F., Grandi, A., Manzella, G., De Strobel, F., and Lyubartsev, V. (2018). Measuring the sea: Marsili's oceanographic cruise (1679–80) and the roots of oceanography. *Journal of Physical Oceanography*, 48(4):845–860.

Poloczanska, E., Mintenbeck, K., Portner, H. O., Roberts, D., and Levin, L. A. (2018). The ipcc special report on the ocean and cryosphere in a changing climate. In *2018 Ocean Sciences Meeting*. AGU.

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M., Gilbert, D., Xu, J., Pouliquen, S., Thresher, A., et al. (2016). Fifteen years of ocean observations with the global argo array. *Nature Climate Change*, 6(2):145–153.

Rosa, W. (2017). Goal 14. conserve and sustainably use the oceans, seas, and marine resources for sustainable development. *A New Era in Global Health: Nursing and the United Nations 2030 Agenda for Sustainable Development*, page 359.

Schaap, D. M. and Lowry, R. K. (2010). Seadatanet–pan-european infrastructure for marine and ocean data management: unified access to distributed data sets. *International Journal of Digital Earth*, 3(S1):50–69.

Schloesser, F., Cornillon, P., Donohue, K., Boussidi, B., and Iskin, E. (2016). Evaluation of thermosalinograph and viirs data for the characterization of near-surface temperature fields. *Journal of Atmospheric and Oceanic Technology*, 33(9):1843–1858.

Schmidtko, S., Johnson, G. C., and Lyman, J. M. (2013). Mimoc: A global monthly isopycnal upper-ocean climatology with mixed layers. *Journal of Geophysical Research: Oceans*, 118(4):1658–1672.

Schutgens, N. A., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., and Stier, P. (2016). Will a perfect model agree with perfect observations? the impact of spatial sampling. *Atmospheric Chemistry and Physics*, 16(10).

Simoncelli, S., Coatanoan, C., Myroshnychenko, V., Pinardi, N., Bäck, Ö., Sagen, H., Scory, S., Barth, A., Schaap, D., Schlitzer, R., et al. (2018). From seadatanet to seadatacloud: historical data collections and new data products. *EGUGA*, page 17818.

Smolyar, I., Boyer, T., and Reagan, J. R. (2018). Structural criteria for quality control of oceanographic data and its potential impact on the construction of climatic fields. *AGUFM*, 2018:OS21C–1586.

Spilhaus, A. F. (1938). A bathythermograph. *The International Hydrographic Review*.

Storto, A., Dobricic, S., Masina, S., and Di Pietro, P. (2011). Assimilating along-track altimetric observations through local hydrostatic adjustment in a global ocean variational assimilation system. *Monthly Weather Review*, 139(3):738–754.

Troupin, C., Barth, A., Sirjacobs, D., Ouberdous, M., Brankart, J.-M., Brasseur, P., Rixen, M., Alvera-Azcárate, A., Belounis, M., Capet, A., et al. (2012). Generation of analysis and consistent error fields using the data interpolating variational analysis (diva). *Ocean Modelling*, 52:90–101.

Wallace, W. J. (1974). *The development of the chlorinity/salinity concept in oceanography*. Elsevier.

Warren, B. A. (2008). Nansen-bottle stations at the woods hole oceanographic institution. *Deep Sea Research Part I: Oceanographic Research Papers*, 55(4):379–395.

Wong, A. P. (2008). An improved calibration method for the drift of the conductivity sensor on autonomous ctd profiling floats by θ-s climatology w. brechner owens woods hole oceanographic institution, woods hole, massachusetts. *Deep-Sea Research*.

Wong, A. P., Johnson, G. C., and Owens, W. B. (2003). Delayed-mode calibration of autonomous ctd profiling float salinity data by θ–s climatology. *Journal of Atmospheric and Oceanic Technology*, 20(2):308–318.

Zweng, M., Seidov, D., Boyer, T., Locarnini, M., Garcia, H., Mishonov, A., Baranova, O., Weathers, K., Paver, C., Smolyar, I., et al. (2019). World ocean atlas 2018, volume 2: Salinity.