University of Southampton

Faculty of Natural and Environmental Sciences

Deep ocean heat content observing strategy: evaluating the past and preparing for the future

Thesis for the degree of Doctor of Philosophy

Author:

Freya Garry MSci Oceanography

Supervisory Panel (alphabetised):

Dr Adam Blaker (National Oceanography Centre Southampton)

Associate Professor Eleanor Frajka-Williams (University of Southampton)

Dr Elaine McDonagh (National Oceanography Centre Southampton)

Dr Christopher Roberts (Met Office Hadley Centre)

Chair: Dr Brian King (National Oceanography Centre Southampton)

Submitted 11 January 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Ocean and Earth Science

Doctor of Philosophy

DEEP OCEAN HEAT CONTENT OBSERVING STRATEGY: EVALUATING THE PAST AND PREPARING FOR THE FUTURE

by Freya Kelly Garry

Changes in full depth ocean heat content (OHC) serve as a key metric of climate change with accurate measurements vital to reduce uncertainties in other climate metrics such as the planetary energy imbalance and to improve global and regional sea level rise prediction. The below 2000 m ocean experiences low frequency sampling in time and space; over the last 30 years hydrographic ship transects were typically repeated every 5 - 10 years, covering a small fraction of the oceans.

Uncertainties that exist in repeat hydrographic sampling are quantified here by deploying a pseudo hydrographic sampling array in high resolution ocean model output. Through comparison to the complete model output, this work presents an estimation of temporal and spatial biases that may exist in calculations of decadal OHC change below 2000 m from hydrographic sampling. In addition, longer timescale changes in OHC and its spatial variations are revealed using centennial length simulations from four coupled climate models. The spatial structure and depth structure of OHC change are evaluated in anthropogenically unforced and forced scenarios to identify key regions and depths for future deep OHC observing.

This work suggests that there may be significant biases in hydrographic sampling in the upper deep ocean such that during the period 1990 - 2010 only around a third to a half of the magnitude of the OHC warming trend was captured by hydrography between 2000 - 2700 m. Biases are much smaller below 3500 m, but still differ in magnitude regionally, with spatial and temporal biases dominating in different basins. On average 82% of the trend between 2000 m and the seafloor is captured by hydrographic style sampling. Past and future coupled climate model scenarios suggest that decadal ocean heat content change experiences the largest variations across the Southern and Atlantic Oceans. This work clearly motivates the deployment of a deep ocean observing strategy with higher temporal and spatial resolution than current hydrography, especially between 2000 - 4000 m, and with particularly high resolution sampling in the Southern and Atlantic Oceans.

Contents

Abstract . . . . . **i**

Contents . . . . . **ii**

List of Tables . . . . . **vi**

List of Figures . . . . . **vi**

Declaration of authorship . . . . . **ix**

Acknowledgements . . . . . **i**

1 Introduction **1**

1.1 Earth's energy imbalance 1

1.2 The role of the oceans in the 'hiatus' 3

1.3 Deep ocean variability 5

1.4 Measuring the oceans 7

1.4.1 The rise of autonomous float technologies 7

1.4.2 OHC change from hydrographic sections 10

1.5 The scope of this thesis 12

2 Model derived uncertainty in observed deep ocean heat content trends in the North-East Atlantic **15**

2.1 Abstract 15

2.2 Introduction 16

2.3 The NEMO model and simulations used 20

2.4 Method 24

2.4.1 Hydrographic sampling in a model 24

2.4.2 Temperature trends on sections 26

2.4.3 Estimating the basin trend 27

2.4.4 Generating uncertainty estimates due to the timings of pseudo sections . . . 28
2.4.5 Calculating thermosteric sea level rise from temperature trends 29
2.5 Results & Discussion . . . 30
2.5.1 Temperature on sections . . . 30
2.5.2 Temperature trends along A16N . . . 32
2.5.3 Temperature trends along A05E . . . 40
2.5.4 Temperature trends along OVIDE . . . 43
2.5.5 Estimating basin trends from hydrographic sections . . . 46
2.5.6 Further analysis of spatial variability . . . 51
2.5.7 Sea level rise . . . 55
2.6 Conclusions . . . 60

3 Did hydrographic sampling capture global and regional deep ocean heat content trends accurately between 1990-2010? . . 63
3.1 Abstract . . . 63
3.2 Introduction . . . 64
3.3 Methodology . . . 66
3.3.1 Choice of model . . . 66
3.3.2 Hydrographic section choice . . . 70
3.3.3 Hydrographic section coordinates . . . 70
3.3.4 Masking sections by basin . . . 71
3.3.5 Extracting sections from the model . . . 71
3.3.6 Section weightings . . . 71
3.3.7 Calculating trends on sections . . . 72
3.3.8 Calculating the error of the average trend on a section . . . 74
3.3.9 Basin trends as estimated from sections and associated errors 75
3.3.10 Average trend in basin - model truth . . . 76
3.3.11 Ocean trends from section estimates and basin truth and associated errors . . . 77
3.3.12 Heat fluxes . . . 78
3.3.13 Uncertainties in the estimation of basin heat content change from hydrographic sections . . . 79
3.3.14 Monte Carlo analysis of sensitivity to pseudo section timing 82
3.3.15 Assessing whether the same results are gained from an updated analysis of observations to 2015 . . . 85

3.4 Results 85
3.4.1 Trends on sections 85
3.4.2 Basin trends and section based estimates 91
3.4.3 Heat fluxes and estimating biases 93
3.4.4 Global and major ocean trends 102
3.4.5 Using additional sections and timings up to 2015 112
3.5 Discussion 113
3.5.1 Capturing trends on individual sections or in individual basins 113
3.5.2 Capturing trends globally or across oceans 116
3.5.3 Using additional sections and timings up to 2015 119
3.6 Conclusions 120

4 The evolution of spatial patterns of ocean heat content in CMIP5 models and implications for deep ocean observing . . 123
4.1 Abstract 123
4.2 Introduction 125
4.3 Models 129
4.4 Method 134
4.4.1 Aim and clarification of associated issues in OHC sampling that are not covered 134
4.4.2 Calculating OHC 134
4.4.3 Variability in models compared to observations during 1990 - 2010 135
4.4.4 Using EOF analysis to identify modes of variability 136
4.4.5 Emergent signals of ocean climate change 137
4.4.6 How deep do we need to measure to resolve full depth OHC? 137
4.4.7 Estimating planetary energy imbalance from OHC 138
4.5 Results 139
4.5.1 21 year trends in OHC in models and observations 139
4.5.2 EOF analysis in the North Atlantic and Southern Oceans . . 146
4.5.3 Emergent signals of ocean climate change 150
4.5.4 How deep do we need to measure to resolve full depth OHC? 153
4.5.5 Estimating planetary energy imbalance from OHC 155
4.6 Discussion 158
4.6.1 Further questions surrounding ocean sampling 163
4.7 Conclusions 164

5 Conclusions 165
5.1 Summary 165
5.2 Implications for deep ocean observing 167
5.3 Further work 168

Bibliography 173

List of Tables

2.1 NEMO model simulations used 23
2.2 Overestimation of average trend below 2000 m by pseudo sections . 51

3.1 Isolating specific uncertainties by subtracting data calculated by different methods 82
3.2 Repeating pseudo section analysis with differing dates 83
3.3 Uncertainties in measuring global and oceanic trends in the deep ocean 104
3.4 Uncertainties in measuring global and oceanic trends in the abyssal ocean 105

4.1 CMIP5 models used 130

List of Figures

1.1 Schematic representation of energy flows in Earth's climate system . 2
1.2 Schematic of below 1000 m ocean heat content observing technologies 8
1.3 Spatial frequency of temperature profiles below 1000 m and 2000 m 9
1.4 Deep ocean temperature trends during the period 1980 - 2015 . . . 12

2.1 Location and timings of hydrographic sections in NEA 25
2.2 Temperatures at 2000 m and 3000 m in NEMO083 (DFS) 33
2.3 Temperatures at 4000 m and 5000 m in NEMO083 (DFS) 34
2.4 Temperatures at 2000 m and 3000 m in NEMO025 (ERA Interim) . 35
2.5 Temperatures at 4000 m and 5000 m in NEMO025 (ERA Interim) . 36
2.6 Temperatures at 2000 m and 3000 m in NEMO025 (CORE2) 37

2.7 Temperatures at 4000 m and 5000 m in NEMO025 (CORE2) 38
2.8 Temperature trends along A16N 41
2.9 Horizontally averaged temperature trends along A16N 42
2.10 Temperature trends along A05E 44
2.11 Horizontally averaged temperature trends along A05E 45
2.12 Temperature trends along OVIDE 47
2.13 Horizontally averaged temperature trends along OVIDE 48
2.14 Horizontally averaged temperature trends in NEA 49
2.15 Map of temperature trends across NEA in NEMO083 (DFS) 53
2.16 Map of temperature trends across NEA in NEMO025 (ERA Interim) 54
2.17 Map of temperature trends across NEA in NEMO025 (CORE2) . . 56
2.18 Thermosteric sea level rise contribution from deep and abyssal oceans 59

3.1 Hydrographic sections and basins in the global oceans 68
3.2 Occupation timings for each hydrographic section 69
3.3 Hydrographic sections occupied at least twice during the period 1980 - 2015 84
3.4 Temperature trends along A16N 87
3.5 Horizontally averaged temperature trends along A16N and S04 . . . 88
3.6 Horizontally averaged temperature trends along P21 and I08 89
3.7 Horizontally averaged temperature trends in largest basins 92
3.8 Heat fluxes through 4000 m, comparison to Purkey and Johnson (2010) 95
3.9 Heat fluxes into deep and abyssal ocean layers 96
3.10 Biases in the heat fluxes into the deep ocean 98
3.11 Biases in the heat fluxes into the abyssal ocean 101
3.12 Horizontally averaged temperatures trends in the global ocean . . . 103
3.13 Horizontally averaged temperatures trends in the major oceans . . . 106
3.14 Horizontally averaged temperatures trends in the major oceans with uncertainty due to shifting timings 108
3.15 Horizontally averaged temperatures trends in the Southern Ocean with uncertainty based on different timings 109
3.16 Horizontally averaged temperatures trends in the global ocean using sections in Desbruyères *et al.* (2016b) 111
3.17 Horizontally averaged temperatures trends in the major oceans using sections in Desbruyères *et al.* (2016b) 112

4.1 Global OHC in 22 CMIP5 models 131
4.2 Trends in OHC 1990 - 2010, 0 - 2000 m 140
4.3 Std. of 21 yr trends in OHC (control), 0 - 2000 m 141
4.4 Trends in OHC 1990 - 2010, 2000 m - seafloor 143
4.5 Std. of 21 yr trends in OHC (control), 2000 m - seafloor 144
4.6 EOFs - North Atlantic Ocean 148
4.7 EOFs - Southern Ocean 149
4.8 Trends in OHC during 2010 - 2100 151
4.9 Fraction of variability, control simulations 152
4.10 Schematic of possible vertical profiles of OHC trends 155
4.11 Fraction of variability, RCP 8.5 156
4.12 Bias in OHC captured by different OHC observing strategies 157

Academic Thesis: Declaration Of Authorship

I, Freya Garry, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Either none of this work has been published before submission, or parts of this work have been published as:
   Desbruyères DG, McDonagh EL, King BA, **Garry FK,** Blaker AT, Moat BI & Mercier H. 2014. Full-depth temperature trends in the northeastern Atlantic through the early 21st century. *Geophysical Research Letters*, 41(22), 7971-79.

Signed:_____

Date:_____

Acknowledgements

This thesis was funded by the University of Southampton through a Vice Chancellor's Award, by the Marine Physics and Ocean Climate group at the National Oceanography Centre, and by the Met Office Hadley Centre. Additional funding for attending conferences came from a Peter Killworth Memorial Scholarship. I am very grateful for all the funding received, and to those who helped me acquire it as an Isle of Man student. I am also grateful for the prior funding of the Isle of Man Government who enabled me to study for my Master of Oceanography at the University of Southampton.

Many thanks to Professors Keith Haines and Bob Marsh for examining this thesis, and for their helpful discussions and comments.

To all I have met along the way, thank you for your various inspirations. A particular thank you to my team of supervisors, for all your help, advice and time. I was delighted to get the chance to continue working with Chris Roberts and Adam Blaker during my PhD, and I have learnt so much working with Elaine McDonagh and Eleanor Frajka-Williams. For your advice and support, and for taking me to the Southern Ocean, thank you Brian King. There are so many knowledgeable brains and friendly faces at both the National Oceanography Centre Southampton and the Met Office Hadley Centre; I'm grateful for the interactions with all of you. I have been privileged to be working on this PhD on the same timescale as the DEEP-C NERC consortium led by Richard Allan have been studying the very related subject of energy flows in the global climate system during the 'hiatus'; I've really enjoyed project meetings and thank the consortium for their support and helpful comments on my work throughout the project.

Thank you to the various modellers and observationalists who have generated the data used in this thesis. Particular thanks to Sarah Purkey and Damien Desbruyères, for the use of data, and especially to Damien for the advice and company on this journey of deep ocean temperature exploration. I acknowledge the Nucleus of European Modelling of the Oceans framework, and in particular the developers and those who work so hard to produce model runs in the Marine Systems Modelling group at the National Oceanography Centre Southampton. I acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for the Coupled Modelling Intercomparison Project (CMIP), and thank the climate modelling groups involved for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

A huge thank you to the PhD community at the National Oceanography Centre Southampton; I have been immensely privileged to spend so much time with such an interesting and fun group of people. A special thank you to my often travel companion, Zoe, and to Helen for knowing so much about how to make computers behave and for giving advice so freely. I am very lucky to have enjoyed living with Anieke, Helen, Jesse, Marina, Jon, Tianya and Matt in number 48. Sam, thank you for the companionship in 344/26. Thank you to all of you who have been there for me when times were tough, whether you realised it or not.

Finally, thank you to my parents for everything (there just aren't enough words). Robin, thank you for being you; I know you'll always be there for me. And to Tucker, who holds a very special place in my heart.

CHAPTER 1

Introduction

1.1 Earth's energy imbalance

The Earth's climate system has been undergoing rapid global change for over a century, largely due to anthropogenic emissions of greenhouse gases (carbon dioxide, methane and nitrous oxide; IPCC, 2013). Increased levels of greenhouse gases in the atmosphere have led to a global energy imbalance (more heat is entering the earth system than leaving it, Figure 1.1) with an estimated magnitude of between 0.6 - 0.9 Wm^{-2} over the last decade (Allan *et al.*, 2014; Johnson *et al.*, 2016; Trenberth *et al.*, 2016). The state of global climate change can be summarised by the Earth's energy imbalance, making the imbalance a key property of the Earth's system and a priority for observational systems to measure (Von Schuckmann *et al.*, 2016). Satellite measurements alone are not sufficient to resolve the absolute magnitude of Earth's energy imbalance (Hansen *et al.*, 2011), with current estimations relying primarily on ocean heat content (OHC) measurements together with radiation measurements from satellites (Von Schuckmann *et al.*, 2016).

The oceans are an excellent metric for Earth's energy imbalance because they have absorbed up to 93% of anthropogenically forced global warming since the 1960s (Abraham *et al.*, 2013; Rhein *et al.*, 2013). Figure 1.1 illustrates the oceans' dominant role in storing accumulated heat due to the global energy

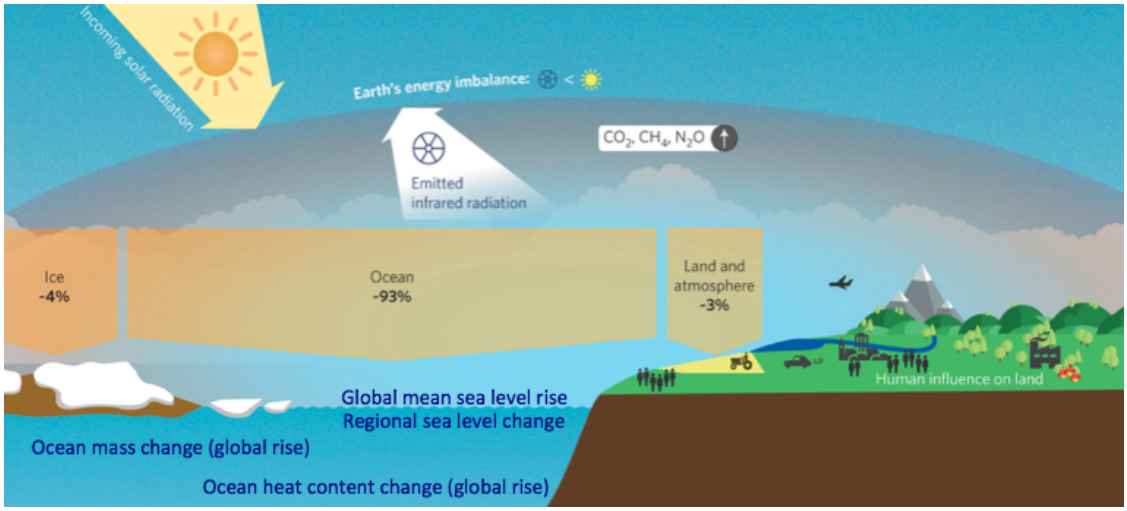


Figure 1.1: Schematic representation of energy flows and stores in the Earth's climate system and some of the physical consequences for the global oceans (blue text), adapted from Von Schuckmann *et al.* (2016).

imbalance as well as illustrating the other stores of heat in the climate system (ice, land and atmosphere). In addition to taking up heat accumulating in the climate system due to anthropogenic forcing, modelling studies have demonstrated that the oceans dominate the planetary energy budget for internal variability on time scales of a year or more (Palmer *et al.*, 2011; Palmer and McNeall, 2014). Internal variability is defined as variations in the climate system that are not due to external forcing. External forcing includes both anthropogenic forcing and that from 'natural' forcing factors such as large volcanic eruptions or variability in solar intensity.

Accumulation of heat by the oceans moderates the rate of atmospheric temperature change, but has potentially negative societal consequences primarily due to global and regional sea level rise (Figure 1.1). Levitus *et al.* (2012) estimate that the thermosteric component of sea level rise was 0.54 ± 0.05 mm yr^{-1} for the 0 m to 2000 m layer during the period 1955 - 2010. Surface ocean temperature anomalies are associated with extreme weather events (e.g., Grist *et al.*, 2015). Density changes due to temperature change in the oceans may lead to ocean circulation changes which in turn result in changes to regional climate (e.g., Rahmstorf, 2002). Ocean temperature adjustments will feed back on the surface climate through several mechanisms including changes to carbon uptake and storage (e.g., Lashof, 1989; Sarmiento *et al.*, 1998) and sea-ice melt

and associated albedo changes (e.g., Curry *et al.*, 1995).

Warming has been observed in the upper ocean (0-700m: e.g., Domingues *et al.*, 2008; Ishii and Kimoto, 2009; Levitus *et al.*, 2009), intermediate oceans (700-2000m: e.g., Von Schuckmann *et al.*, 2009; Levitus *et al.*, 2012) and deep and abyssal oceans (2000m-seafloor: Purkey and Johnson, 2010; Kouketsu *et al.*, 2011; Desbruyères *et al.*, 2016b) emphasising that understanding ocean behaviour at all depths is vital to understanding and predicting climate change. A quarter to a third of the OHC increase since 1971 is attributable to warming below 700 m (Levitus *et al.*, 2012; Rhein *et al.*, 2013; Gleckler *et al.*, 2016). During the twenty-first century, approximately 30% of ocean warming has occurred below 700m (Balmaseda *et al.*, 2013). However, a study of OHC change in an ensemble of dynamical and statistical ocean reanalyses highlights their limitations below 700 m due to past observational constraints, emphasising the need for improved deep ocean observations (Palmer *et al.*, 2015). Using climate models, Palmer *et al.* (2011) reveal that to gain information relating to top of the atmosphere changes from OHC it is necessary to measure down to at least 4000 m. Improvements to the accuracy of OHC variability estimates can be made by measuring over a longer period of time or by measuring deeper into the oceans, so if information about OHC variability is required quickly, measuring into the deep and abyssal oceans is imperative.

1.2 The role of the oceans in the 'hiatus'

Despite the net downward radiative flux imbalance at the top of Earth's atmosphere there has been a 'hiatus' in global surface temperature rise over the first decade of the twenty-first century (Simmons *et al.*, 2010) relative to 1998 (notably a warm year due to strong El Niño conditions). Although 'hiatus' is a misleading term for the climatic variability observed over the period (other metrics of climate change show that global warming did not slow) and so the term may not be useful for the climate science community to continually refer to (Lewandowsky *et al.*, 2015), this thesis does refer to the 'hiatus' as the term is now ubiquitous. Some of the proposed mechanisms behind the 'hiatus' follow to highlight the role of the oceans in surface climate variability on decadal

timescales and further emphasise the necessity for widespread ocean observation to understand climatic variations.

A multi climate model study by Palmer and McNeall (2014) highlights the importance of internal variability in ocean heat rearrangement, showing that it can be responsible for OHC changes of 0.1 W m^{-2} on decadal timescales. Palmer and McNeall (2014) suggest that it had a vital role in shaping the recent observed trends in global surface temperature and ocean heat uptake. Rather than the ocean heat uptake changing, the 'hiatus' appears to have been largely due to internal redistribution of heat that changed the rate of surface temperature warming without significantly changing the global energy imbalance.

Changes in the Pacific Ocean have been identified as the key to the 'hiatus' (Kosaka and Xie, 2013). Watanabe *et al.* (2013) and England *et al.* (2014) find that wind driven changes in the Pacific are likely to be the main driver. Over the last two decades, strengthened trade winds in the Pacific helped facilitate subsurface ocean heat uptake as well as driving shallow overturning cells in the Pacific (further cooling the surface ocean). This behaviour represents a change to the state of the Pacific Decadal Oscillation and was responsible for regional climate anomalies (Trenberth *et al.*, 2014). Mann *et al.* (2016) find that North Pacific internal variability (which they consider to be critical to the 'hiatus') was not predictable.

An observational and model study by Lee *et al.* (2015) concluded that the enhanced heat uptake in the Pacific did not result in an accumulation of heat in the Pacific but instead heat was transferred into the Indian Ocean. In contrast to identifying the Pacific Ocean as the driver for the 'hiatus', Chen and Tung (2014) find that transfer of heat into the deep Atlantic and Southern Ocean due to a recurrent salinity anomaly in the subpolar North Atlantic is the key mechanism, with insignificant heat storage in the Pacific. Meehl *et al.* (2011) show that in 'hiatus' periods excess heat is transferred into the oceans below 300 m, with heating also below 750 m in the Atlantic and Southern Ocean. The El Niño Southern Oscillation potentially links heat transfer processes in different basins together (increased subtropical thermocline ventilation in Pacific and Atlantic, weakened convective mixing in the North Atlantic and Southern Ocean) (Meehl *et al.*, 2011).

More recently, Liu *et al.* (2016) conclude that heat uptake in the Atlantic and Southern Oceans is due to consistent anthropogenic forcing and that the recent 'hiatus' was driven by changes in the Indian and the Pacific Oceans. The Southern Ocean is playing a major role in taking up heat accumulating in the climate system over recent years; Wijffels *et al.* (2016) use Argo data from 2006 to show that the Southern Hemisphere ocean is heating four times faster than the northern hemisphere, estimating 75 - 99 % of heat accumulating between 0 - 2000 m is south of the equator (mostly between 30°S and 50 °S). There have certainly been changes to heat transfer into and within the oceans in the last two decades although the precise mechanisms are still under debate and require further research, aided by a comprehensive OHC observing system.

Although this thesis focuses on OHC change, decadal surface climate variability will be forced to some extent by non-ocean mechanisms such as solar forcing variability and atmospheric changes. This recent 'hiatus' has in part been attributed to reduced solar forcing as well as variations in the El Niño Southern Oscillation (Huber and Knutti, 2014; Johansson *et al.*, 2015) and it has been shown that aerosol redistribution may have driven changes in the Pacific Ocean (Smith *et al.*, 2015). Models and observations have suggested that 'hiatus' events are not unusual due to internal variability (e.g., Roberts *et al.*, 2015) with natural forcing (e.g. volcanoes) increasing their likelihood (Schurer *et al.*, 2015). However, Sévellec *et al.* (2016) suggest that this early twenty-first century 'hiatus' was unlikely with 'hiatus' events becoming extremely unlikely by the end of the century assuming business-as-usual future anthropogenic forcing.

1.3 Deep ocean variability

In addition to decadal variability such as that identified in 'hiatus' research, variability in the ocean has been identified on a range of timescales; Balmaseda *et al.* (2013) identify past changes in OHC relating to volcanic eruptions and emphasise the importance of surface winds, for instance intensification of trade winds in subtropical gyres. However for the deep ocean in particular, Balmaseda *et al.* (2013) suggest some deep changes might be related to changes in the Atlantic Meridional Overturning Circulation. Roberts *et al.*

(2016a) use surface flux products and OHC change estimates for the period 1985 - 2012 to deduce that on interannual timescales ocean dynamics drive variability. On seasonal timescales full depth changes in OHC also cannot be explained solely by surface fluxes, with ocean heat transport partly responsible (Roberts *et al.*, 2016a).

Changes to water properties in deep ocean convective regions affect deep water formation mechanisms and export processes (de Lavergne *et al.*, 2014) and change the characteristics of deep water; for example, recent observations show that Antarctic Bottom Water is warming and freshening (Purkey and Johnson, 2013). Variability may be particularly large in the north and west Atlantic and Southern Ocean due to the influence of surface forcing (Wunsch and Heimbach, 2014). Variability at deep water formation regions impacts on ocean circulation, e.g., Roberts *et al.* (2013) identify a link between mid-depth ocean densities in the Labrador Sea and the strength of the Atlantic Meridional Overturning Circulation.

An OHC analysis by Desbruyères *et al.* (2016a) decomposes temperature change into that due to heave (vertical displacements of isopycnals) and spiciness (changes to water mass properties on isopycnals). Both above and below 2000 m, warming in the North Atlantic and Southern Ocean is attributable mainly to heave. A study of deep ocean temperature change over the period 1993 - 2016 from hydrography in the eastern Drake Passage also found that 95% of the year-to-year variance was due to heave (Firing *et al.*, 2016), and that much of the variability can be explained by the position of the polar front.

Geothermal heating from the seafloor increases deep ocean temperatures locally, weakening abyssal stratification and influencing deep ocean circulation (e.g., Adcroft *et al.*, 2001; Mashayek *et al.*, 2013; Downes *et al.*, 2016), although there may not be much variability in this mechanism over decadal timescales. Surface winds may also be important to deep ocean variability in some regions; for instance Sheen *et al.* (2014) link surface eddy activity and abyssal turbulence in the Antarctic Circumpolar Current, suggesting that abyssal mixing is related to surface wind forcing. Meredith *et al.* (2014) emphasise the pressing need to better understand what controls the properties and flows of deep water to understand deep OHC variability. An additional challenge when studying the deep ocean is

the long 'memory' of the ocean; climatic forcing many years ago could influence modern trends (Wunsch and Heimbach, 2014).

1.4 Measuring the oceans

Historical ocean temperature observations have progressed from reversing thermometers and bucket measurements of surface temperatures to highly accurate modern thermometers which are lowered through the ocean from research ships, or deployed on moorings. In addition, less accurate but plentiful measurements of upper ocean temperatures have been made by expendable bathythermographs which are probes deployed from ships of opportunity. Acoustic techniques for estimating large scale integrated temperature change also exist. Temperature data from all historical methods can be integrated with corrections to estimate OHC and its rate of change during the historical period, but during the last decade OHC observing has been revolutionised by the introduction of Argo floats (www-argo.ucsd.edu).

1.4.1 The rise of autonomous float technologies

The upper ocean Argo array of 2 metre long autonomous floats has improved the accuracy of research investigating the role of the oceans in surface climate variability. Upper ocean Argo floats measure temperature down to 2000 m (as described in Figure 1.2) providing unprecedented coverage of the oceans to 2000 m since the array reached its target size of 3000 floats in the year 2007. At the end of 2016, the array consisted of over 3700 floats dispersed throughout the global oceans. OHC change estimates prior to the deployment of upper ocean Argo are largely reliant on data from expendable bathythermographs, which are subject to significant biases (e.g., Ishii and Kimoto, 2009), and from measurements along hydrographic sections by survey ships and moorings, which have comparatively coarse spatial and temporal resolution.

Compared to previous temperature sampling methods, upper ocean Argo provides much greater confidence when assessing OHC change, providing vastly

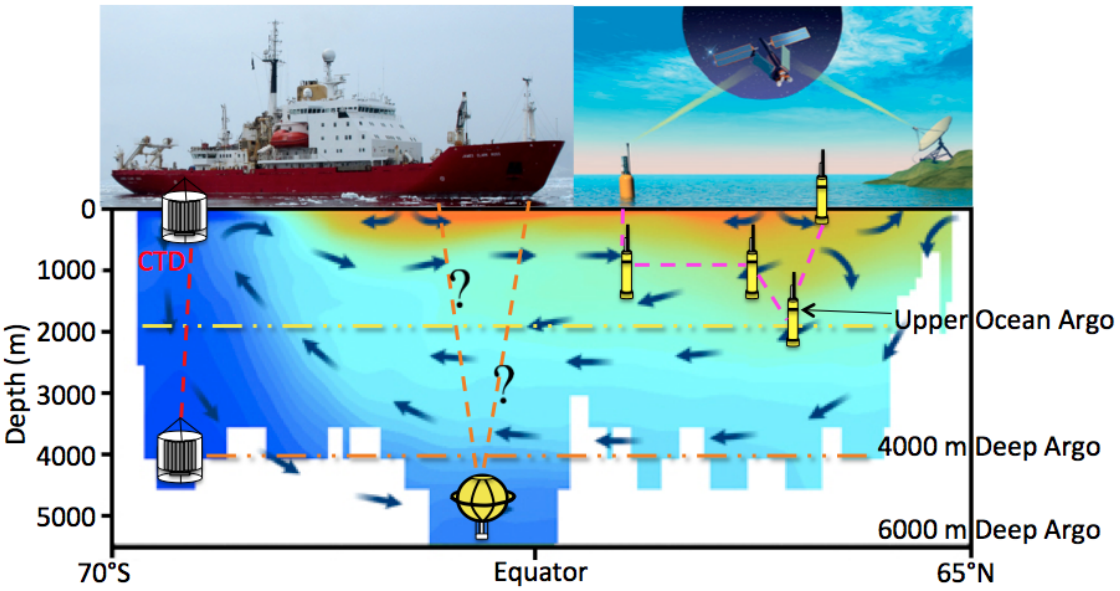


Figure 1.2: Schematic of below 1000 m ocean heat content observing technologies over a cross section of Atlantic Ocean temperatures (cold temperatures - blue, warm temperatures - red) with arrows indicating the time-mean water flow in the Atlantic Meridional Overturning Circulation. Schematic created using figures from Riser *et al.* (2016) and NOAA / The COMET Program (2016). Upper ocean Argo floats descend to a parking depth of 1000 m approximately every 10 days where they drift in ocean currents for over 9 days before descending to around 2000 m. At this point the sampling sensors turn on so that the float measures ocean properties as it rises to the surface where it transmits the data to satellite systems. The CTD is deployed from a research ship to a few metres above the seafloor and provides a profile of highly accurate temperature through the water column. Recent technological developments have led to several types of deep Argo floats which cycle in the same way as upper ocean Argo floats, but that descend to 4000 m or 6000 m depth depending on the float type. Small numbers of deep Argo floats have been deployed, but details of optimal deep Argo float array design (e.g., number of floats, spatial distribution, depth range, parking depth, time underwater) are still under debate with only Johnson *et al.* (2015) providing a potential design for a deep Argo array.

improved temporal and spatial resolution above 2000 m (Abraham *et al.*, 2013; Riser *et al.*, 2016). Figure 1.3 (a) demonstrates the excellent spatial coverage that upper ocean Argo provides compared to all other sampling methods below 1000 m (Figure 1.3 b). However, Argo float measurements are relatively limited in number in regions of high latitude (mainly due to sea ice cover), on continental shelves or where swift western boundary currents exist (very shallow regions and regions of complicated topography are not measured well by the floats). Lyman

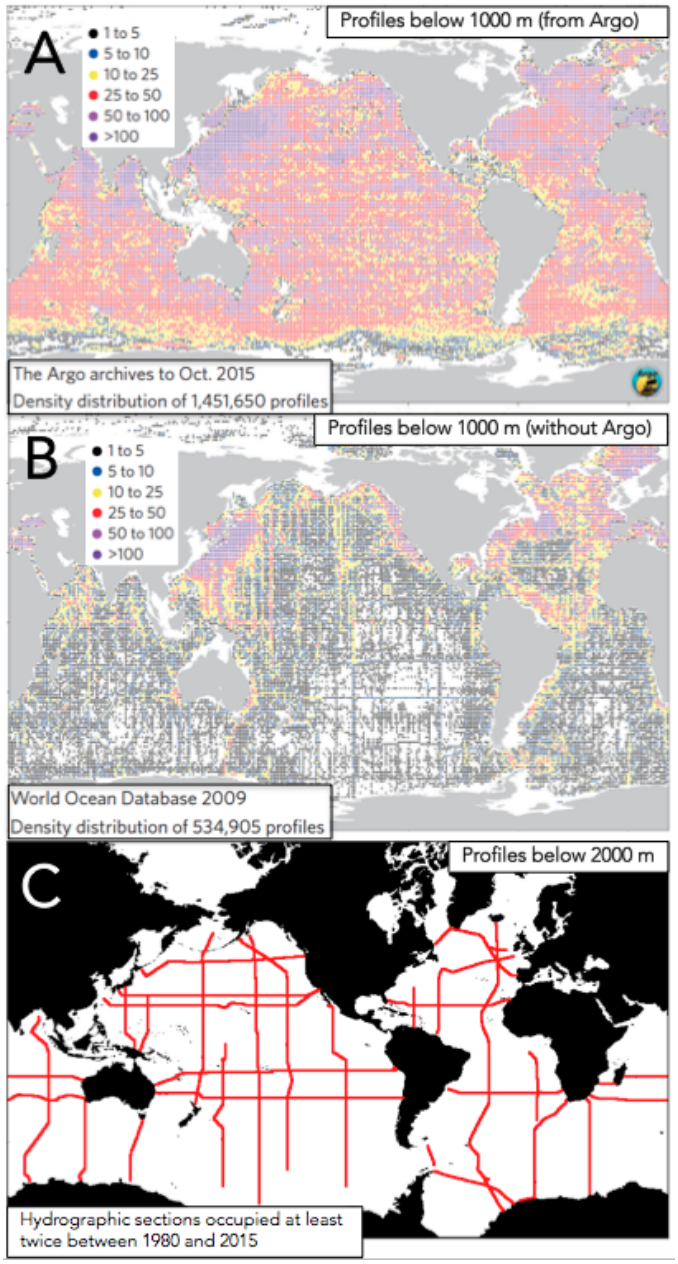


Figure 1.3: Spatial frequency of temperature and salinity profiles generated in recent decades (a) below 1000 m from the upper ocean Argo array during the period January 1999 - October 2015 (b) below 1000 m from all other sources except upper ocean Argo during the historical period (c) below 2000 m from hydrographic sections during the last 35 years. The sampling density shown in (a) and (b) is computed as the total number of profiles in each 1° latitude by 1° longitude square. Adapted from Riser *et al.* (2016).

and Johnson (2014) reveal that assumptions regarding areas poorly sampled by upper ocean Argo can significantly affect global OHC estimates and their trends,

and Durack *et al.* (2014) suggest that 0 - 700 m estimates of OHC change are likely to be biased low because of the poor spatial and temporal sampling in regions such as the Southern Ocean. This is a problem with ocean observing in general as, historically, sampling has always been biased towards the Northern Hemisphere and in particular easily accessible coastal regions (Riser *et al.*, 2016). However, there is a push towards deploying larger numbers of floats in the Gulf Stream and Kuroshio regions (as vigorous ocean turbulence leads to high sampling error), in near-Equatorial regions (where strong atmospheric coupling produces large anomalies) and in marginal seas (Riser *et al.*, 2016). In addition, technological developments have enabled under-ice Argo floats to successfully return data and maintain a lifespan comparable to other floats (e.g., Wong and Riser, 2013), so there is potential to address upper ocean sampling deficiencies given sufficient resources.

Uncertainties in upper ocean observational systems currently prevent the indirect detection of temperature and salinity changes far beneath 2000 m (von Schuckmann *et al.*, 2014). However, deep ocean Argo float ('Deep Argo') technology now exists (e.g., Kobayashi *et al.*, 2012; Petzrick *et al.*, 2013) which can descend to 4000 m or 6000 m depending on the float, and researchers are beginning to start deploying small numbers of these floats. Deep ocean measurement arrays such as Deep Argo are now considered to be a necessary oceanographic tool (e.g., von Schuckmann *et al.*, 2014; Wunsch and Heimbach, 2014) but well thought out deployment strategies will be required to ensure efficient and effective use of the technology to detect subtle changes in the deep ocean. One such deployment strategy has been proposed by Johnson *et al.* (2015), who estimate that a 5° by 5° resolution array with floats cycling to depth every 15 days would provide decadal OHC trends with unprecedented accuracy of 1 - 26 m°C decade^{-1}.

1.4.2 OHC change from hydrographic sections

Though there may be future possibilities for widespread OHC measurement below 2000 m using autonomous floats, currently full-depth estimations of OHC change rely only on repeated hydrographic sections occupied as part of oceanographic

research programs such as the World Ocean Circulation Experiment (WOCE) and its follow up program, the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP; www.go-ship.org). Full-depth hydrographic section data is vital because the very high accuracy measurements (accurate to 0.001°C) are required to calibrate other upper ocean observation systems such as upper ocean Argo. However OHC change estimates below 2000 m (that depend only on hydrographic sections) are likely to have relatively large uncertainties because the measurement frequency in time and space is very low; an ocean basin will be measured by a few transects which are typically repeated at least once a decade. Figure 1.3 (c) indicates just how spatially sparse the profiles used to calculate decadal OHC change in the deep ocean are compared to the spatial distribution of upper ocean Argo (Figure 1.3 a). The Southern Ocean has historically been particularly poorly sampled by ships (Riser *et al.*, 2016).

Purkey and Johnson (2010) derived initial estimates of recent OHC trends in the deep ocean by using data from 28 hydrographic sections spread across all the major ocean basins. Purkey and Johnson (2010) find warming over the time period 1990-2010 in most basins, comprehensively illustrating heat sequestration in the deep (2000 - 4000 m) and abyssal oceans (below 4000 m). In particular, warming across the Southern Ocean was found to be relatively strong.

Desbruyères *et al.* (2016b) provide an updated overview of OHC change in recent decades using additional hydrographic sections and find a similar change in OHC to Purkey and Johnson (2010) of $0.065 \pm 0.040\ Wm^{-2}$ between 1991 - 2010 (averaged over the Earth's surface). The spatial patterns of the deep and abyssal warming identified by Desbruyères *et al.* (2016b) are shown in Figure 1.4. As in the Purkey and Johnson (2010) study, the uncertainties in the magnitude of warming in the deep ocean are nearly as large as the calculated warming. In some basins, Desbruyères *et al.* (2016a) find that independent OHC estimates at 2000 m from upper ocean Argo and the much coarser resolution (in both space and time) hydrographic sampling do not agree.

A residual approach by Llovel *et al.* (2014) used the difference between upper ocean heat change from Argo and total ocean heat content as estimated from satellites to calculate a slight cooling below 2000 m during the period 2005 - 2013. However uncertainties in this method are very large (Johnson and Lyman, 2014)

and though the accuracy of estimates from hydrographic section data is also uncertain, in-situ observations may be more likely to capture a realistic trend as well as providing spatial variation. Improved deep ocean observations are required to constrain the sign and magnitude of deep ocean change and provide detail regarding its spatial variability. More accurate deep OHC estimates will enable cross checking of ocean mass, ocean warming and sea level measurements to give increased confidence to sea level and planetary energy budgets (Johnson and Lyman, 2014).

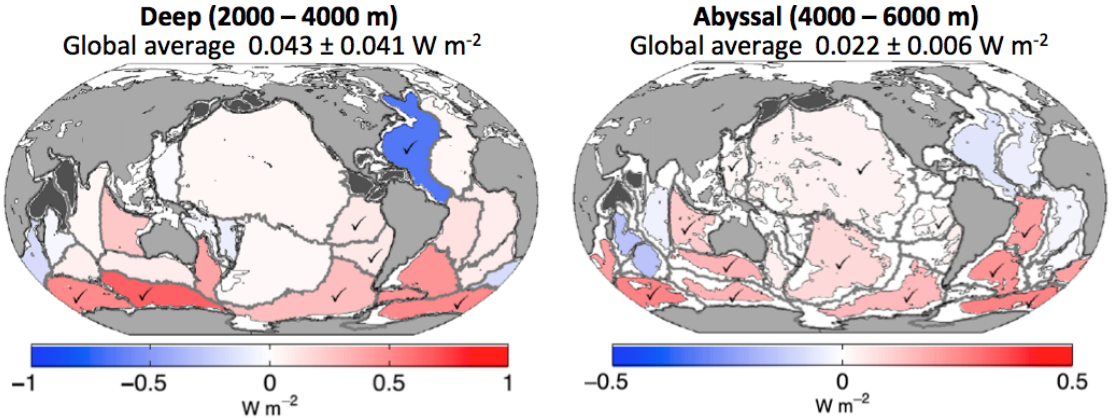


Figure 1.4: Deep and abyssal ocean temperature trends between 1991 and 2010 presented as a heat flux through the upper boundary. Ticks indicate where the trends are statistically different from zero at the 95% confidence interval. Adapted from Desbruyères *et al.* (2016b), magnitude of global average warming in Wm^{-2} from personal communication with Damien Desbruyères (2016).

1.5 The scope of this thesis

The scientific problem addressed in this thesis centres on how deep OHC change estimates need to be improved in the future. The novel work presented in this thesis is entirely model based, presenting an evaluation of the uncertainties likely to have been inherent in deep ocean sampling over the last 30 years, and analysing simulations of unforced and forced past climate scenarios and future scenarios to understand deep OHC change and its variability on multi-decadal timescales.

The first science chapter contributes to a collaboration with an observation based study of deep OHC in the North-East Atlantic by Desbruyères *et al.* (2014).

Using three different simulations of a high resolution ocean model, the chapter presents an evaluation of the temporal and spatial uncertainty surrounding deep OHC change in the North-East Atlantic between 1990 and 2010. In addition to providing an uncertainty estimate for observational sampling (published in Desbruyères *et al.*, 2014), the chapter developed the methodologies for applying a similar approach of estimating uncertainties in deep ocean sampling globally.

The second science chapter extends the first chapter by estimating temporal and spatial uncertainty in deep ocean observations globally. The chapter primarily evaluates uncertainties for the observational analysis of Purkey and Johnson (2010), and compares the uncertainty estimates to those assumed by Purkey and Johnson (2010). In addition to presenting an estimation of the different components of spatial and temporal uncertainty globally, the chapter addresses uncertainty that might exist due to seasonal bias in sampling. Although all full-depth OHC estimates (and thus those of planetary energy imbalance) include deep ocean estimates from hydrography, there has not before been a model based study to estimate the uncertainties inherent in hydrographic sampling below 2000 m and if they might exceed the estimates of uncertainty made by observational studies.

The third chapter considers OHC change over longer timescales, evaluating the spatial patterns of OHC variability in unforced and forced (past and future) climate model scenarios. High resolution coupled climate models produced as part of the most recent release of models by the Coupled Model Intercomparison Project are used to evaluate the depths and regions where measuring deep OHC change is most important to understand full-depth OHC change. This chapter provides a model based understanding of the areas in which high resolution deep ocean sampling (such as that which might be achieved using automonous floats) would be of most benefit. This might be particularly useful in a scenario where not enough funding exists to deploy an evenly spaced network of floats; the current observing systems of GO-SHIP and upper-ocean are projected to decline in size over the next three years due to funding constraints (Durack *et al.*, 2016). In addition, there is also pressure for Argo to deploy other new types of float, including bio-Argo floats to provide biogeochemical data and under-ice Argo floats, and this may further restrict funding for Deep Argo floats.

Together these chapters provide a novel scientific contribution through the provision of uncertainties in hydrographic sampling and analysis of deep OHC change and its spatial variations over multi-decadal to centennial timescales in models. This work therefore provides contributing science to at least three of the World Climate Research Programme Grand Challenges: Regional Sea-Level Change and Coastal Impacts (through uncertainties in deep OHC change), Near-term Climate Prediction (OHC is important for initialising climate models) and Clouds, Circulation and Climate Sensitivity (as changes to ocean heat uptake and storage determine the transient sensitivity of the climate system).

CHAPTER 2

Model derived uncertainty in observed deep ocean heat content trends in the North-East Atlantic

2.1 Abstract

To estimate uncertainty associated with deep ocean temperature trends derived from repeat hydrographic sampling, three high resolution historical simulations from the ocean model NEMO are subsampled during the time period 1988-2010 in the North-East Atlantic. Two types of uncertainty are considered to address representativeness of large-scale heat content change estimates generated from observations:

- Temporal uncertainty exists because there are typically 5-year intervals between repeat hydrographic sections so calculated trends may not be representative of the true trend on a section.

- Spatial uncertainty exists because two or three repeated sections in a basin may not be representative of the heat content variability of the much larger basin area.

At depths below 2000 m, the main features of North-East Atlantic temperature

trends on hydrographic sections are captured well by typically pentadal sampling, except along one hydrographic section (OVIDE) which is only sampled for less than half of the time period of study. This gives some confidence to observation-based trend analysis spanning recent decades, but the magnitude of temperature change in the deep ocean can be overestimated by hydrographic style sampling in model simulations by 25% to 100%. This has implications for regional sea level prediction, as the uncertainty in the deep ocean thermosteric component may lead to error of up to 0.3 mm per year.

Both temporal and spatial uncertainty contribute to the total uncertainty, with the magnitude of temporal uncertainty sensitive to the precise timings of hydrographic section occupation. This implies that the present deep ocean observing network requires additional deep ocean sampling at higher temporal and spatial frequency to better understand the spatial variation of deep ocean temperature change and its variability through time.

2.2 Introduction

Full depth ocean heat content (OHC) is an important metric of the climate system; on interannual to longer timescales, the ocean stores the vast majority of the heat accumulating in the climate system (Rhein *et al.*, 2013). Significant temperature changes have been measured at all ocean depths (Rhein *et al.*, 2013) and ocean temperature changes have potentially severe societal impacts through the thermosteric component of sea level rise (e.g., Church *et al.*, 2013a).

Warming trends have been identified in large parts of the deep (2000 - 4000 m) and abyssal (below 4000 m) oceans over recent decades (Purkey and Johnson, 2010). However, the deep North-East Atlantic (NEA) showed a cooling (calculated as statistically significant) between 1990 - 2010 (Purkey and Johnson, 2010) despite upper ocean warming during the majority of the period, related to changes to the gyre circulation (Häkkinen and Rhines, 2004; Leadbetter *et al.*, 2007). Although influenced by changes in the atmosphere (e.g., due to solar forcing (Ineson *et al.*, 2011), cloud feedbacks (Brown *et al.*, 2016) or aerosols (Booth *et al.*, 2012)) and surface fluxes (e.g., Duchez *et al.*, 2016), ocean

temperatures in the North Atlantic are strongly influenced by the complex ocean circulations and their competing influences. This results in a changeable regime which is the focus of many oceanographic studies. There is a wind driven gyre circulation and the Atlantic Meridional Overturning Circulation (AMOC), which exhibits high frequency variability (McCarthy *et al.*, 2015) and is influenced by both winds and density changes (Kuhlbrodt *et al.*, 2007). Deep convection events in the Labrador and Greenland-Iceland-Norweigan Seas have the potential to change temperatures at intermediate to deep depths in the North Atlantic resulting in interannual to decadal heat content variations.

The North Atlantic has also been proposed as a region where excess heat accumulating in the climate system might have been stored during the 'hiatus' period (e.g., Meehl *et al.*, 2011; Katsman and van Oldenborgh, 2011; Chen *et al.*, 2014; Drijfhout *et al.*, 2014), during which global surface temperatures did not appear to rise significantly for approximately 15 years despite no reduction in the Earth's energy imbalance (Allan *et al.*, 2014). Desbruyères *et al.* (2014) investigate the temperature changes in the NEA between 1988 and 2014 using data from upper ocean Argo and full-depth hydrographic sections. They resolve the vertical structure of the temperature change and identify the oceanic mechanisms at work particularly during the 'hiatus', finding that in the early twenty-first century, whilst the surface (0-450 m) and abyssal (below 4000 m) layers of the NEA cooled, the full-depth temperature change was dominated by a significant warming between 1000 m and 3000 m. This contrasts with the cooling trend observed during the earlier period 1988 - 2003 by Johnson *et al.* (2005), who found Labrador Sea water to get deeper, colder, fresher, denser, and thicker over that period of study.

Desbruyères *et al.* (2014) attribute a cooling in the upper 450 m of the NEA during the 2000s to cooling of the subpolar gyre and warming of the subtropical gyre through the top 1000 m of the ocean. A cooling in the eastern subpolar gyre is consistent with a recent weakening of the oceanic heat transport between Greenland and Portugal identified by Mercier *et al.* (2015). Desbruyères *et al.* (2014) find that intermediate and the upper deep NEA warming was largely between 1000 - 3000 m in the subpolar gyre. This is consistent with warming of Labrador Sea Water in the late twentieth/early twenty-first century which is advected by the North Atlantic Current to the polar NEA within a few years

(Yashayaev *et al.*, 2007). Desbruyères *et al.* (2014) identify a slight cooling in the abyssal layers, mainly in the subtropical region where the influence of the northward flowing Antarctic Bottom Water is felt.

More recent studies in the same region further highlight the complexity of ocean processes affecting this basin on decadal timescales. Robson *et al.* (2016) find cooling in the upper 700 m of the North Atlantic (above 30°N) since 2005, which they show is consistent with a reduction in the strength of the AMOC and hence northward heat transport (linked to low densities in the deep Labrador Sea). Somavilla *et al.* (2016) found that strong winter mixing in 2005 made the eastern North Atlantic modal waters much saltier, warmer and denser, so that upper ocean heat and salt gradually transferred deeper in the basin. By adjusting the zonal dynamic height gradient which reversed southward regional flow and allowed more salty southern waters to penetrate north, the changes further promoted additional heat injection to the deep ocean through deep convection events in the following years (Somavilla *et al.*, 2016).

The upper ocean Argo array provides heat content change studies with very good temporal and spatial resolution temperature data down to around 2000 m in the open oceans. In the NEA region studied by Desbruyères *et al.* (2014), on average each 2° gridbox annually received data from 25 (in the year 2004) to 50 (in the year 2011) Argo floats (highest density in subpolar gyre, lowest density near Africa and the southern reach of the Mid Atlantic Ridge). However, the temperature trends calculated below 2000 m by Desbruyères *et al.* (2014) relied on data from hydrography alone, as have previous studies of deep and abyssal decadal temperature trends (e.g., Purkey and Johnson, 2010). In the NEA basin, hydrographic ships generally only occupy the same section once every 2-8 years, and spatially, the hydrographic sections cover a small area of the basin, although they are designed to measure along representative and oceanographically interesting transects. The uncertainties associated with trends from hydrography are thought to be relatively large compared to those from upper ocean observing methods (Purkey and Johnson, 2010).

Estimates of uncertainty have been made from the observational data (Purkey and Johnson, 2010; Desbruyères *et al.*, 2014) by considering the variation of temperature trends along a section as well as the effective degrees of freedom

(based on the horizontal decorrelation length scale, which reflects how far apart water parcels have to be before their characteristics are no longer correlated). However, these estimates of uncertainty are not able to fully quantify the spatial uncertainty from subsampling a small area of the basin, nor the temporal uncertainty due to sampling infrequently in time. Kouketsu *et al.* (2011) examined virtual observations in a data assimilation system to evaluate uncertainty in heat content trends below 3000 m estimated from hydrography, finding that they could still detect the global heat content increase below 3000 m during recent decades, although the uncertainty is large.

Here, model simulations are used to evaluate how representative the section based estimates of the temperature trends below 2000 m are of the average temperature trend in the NEA. Output from model simulations are sub-sampled along the locations and at the timings that hydrographic sections were occupied, and the temperature trends calculated from the sub-sampled data are compared to temperature trends generated from the basin average temperature data (using the complete model field). This work thus provides an estimate of the uncertainties surrounding trends calculated from deep ocean hydrography observations in the NEA (e.g., those calculated by Desbruyères *et al.*, 2014). In addition, the trends in the model simulations are directly compared to the observational trends of Desbruyères *et al.* (2014) to comment on whether these model simulations agree on the main observed characteristics of decadal OHC change in the NEA.

2.3 The NEMO model and simulations used

Table 2.1 outlines the three existing model simulations that are used in this work. Each simulation was performed with NEMO (Nucleus for European Modelling of the Ocean, Madec *et al.*, 2015), a global ocean-sea ice model with the ocean (ORCA) configuration at either 0.25° (ORCA025) or 0.083° (ORCA083) horizontal resolution. This is equivalent to around 27.8 km resolution at the Equator for ORCA025 and about 8.3 km for ORCA083, becoming finer towards the poles; at 60°N/S the resolution is equivalent to 13.8 km for ORCA025 and 4.6 km for ORCA083. The vertical resolution is the same for each simulation; 75 levels which are less than a metre thick at the surface and up to 250 m thick by 5500 m.

The model grid is tri-polar with one pole in Canada, one in Russia and one at the South Pole (Madec and Imbard, 1996), preventing numerical instability occurring where meridians converge in the northern hemisphere. The grid is isotropic Mercator south of 20°N and quasi-isotropic bipolar further north. For all simulations, bottom topography is represented as partial steps and has been derived from ETOPO2 (U.S. Department of Commerce, 2006). The sea ice representation is the Louvain-la-Neuve Ice Model version 2 (LIM2) sea-ice model (Timmermann *et al.*, 2005) in each simulation.

Climatological initial conditions for temperature and salinity in January are combined from PHC2.1 (Steele *et al.*, 2001) in the high latitudes, MEDATLAS (Jourdan *et al.*, 1998) in the Mediterranean, and Levitus *et al.* (1998) elsewhere. The climatological initial conditions in NEMO have been surface forced with different products depending on the simulation as described and referenced in Table 2.1. Surface forcing comprises 6-hourly mean forcing for 10 m winds and 2 m air temperature and humidity, daily mean radiation forcing and monthly mean precipitation forcing, with the model applying linear interpolation when integrating these fields. The CORE2 forcing set comprises interannually varying air-sea fluxes of momentum, heat, freshwater and their components, calculated from a NCEP reanalysis (near surface vector wind, temperature, specific humidity and density) and various satellite based radiation, sea surface temperature, sea-ice concentration and precipitation products over the period 1984 to 2006.

Biases are reduced through comparison to shorter timescale and regional but higher accuracy products. Prior to this period, climatological annual mean values are used. ERA Interim is a global atmospheric reanalysis product from ECMWF (European Centre for Medium-Range Weather Forecasts). DFS 5.2 is constructed from the ERA Interim reanalysis with additional corrections, as well as using the ERA40 (also from ECMWF) reanalysis to extend it further back in time. As such, the DFS 5.2 version is the most up-to-date set of surface forcings. To guard against excessive drifting in global salinity (due to deficiencies in fresh water forcing) sea surface salinity is relaxed towards climatology (piston velocity of -33.33 mm/day/psu).

The NEMO ORCA025 CORE2 forced (NEMO025-N206-CORE, hereon referred to as NEMO025 (CORE2)) simulation was integrated from 1958-2007, with two additional years (2008 and 2009) subsequently created. The NEMO ORCA025 ERA Interim forced (NEMO025-VN206-ERA, hereon referred to as NEMO025 (ERA Interim)) simulation was integrated from 1989 to March 2011, and was initialised from the 2007 end state of the CORE2 forced simulation. The NEMO ORCA083 DFS forced simulation (NEMO083-N006, hereon referred to as NEMO083) was integrated from rest in 1958. Each simulation was therefore spun up for over 30 years before the data used for this work were generated (temperature data used from the year 1988).

The higher resolution NEMO083 simulation is employed in addition to the lower resolution simulations on the basis that it may provide advantages over the NEMO025 simulations. 1/4° simulations are considered eddy permitting but 1/12° simulations are able to resolve eddies over much of the global oceans. Increased ocean resolution can improve the position of key ocean currents e.g., the Gulf Stream, and reduce associated temperature biases (Marzocchi *et al.*, 2015), blocking frequency (Scaife *et al.*, 2011), equatorial biases (Shaffrey *et al.*, 2009) and El Niño Southern Oscillation teleconnections (Dawson *et al.*, 2013), and surface fluxes (Roberts *et al.*, 2016b), potentially leading to a more realistic climate system. An earlier NEMO083 simulation (N001) to the one used in this work (N006) was compared by Marzocchi *et al.* (2015) to observations in the North Atlantic, finding that compared to NEMO025, NEMO083 provides a more realistic subpolar gyre shape, North Atlantic Current position, and Gulf Stream separation point. The representation of winter mixing in the Labrador Sea and

dense overflow transports across the Greenland-Scotland Ridge are also improved (Marzocchi *et al.*, 2015), potentially improving aspects of the deep ocean.

Model output is available as 5 day averages and is utilised at this frequency for this analysis. References describing the model simulations in further detail are found in Table 2.1.

Ocean temperature trends generated in a model simulation may be due to the trends in the sea surface forcing product associated with global warming, and if the initial trend in the forcing products is not an accurate representation of the real world then there is the potential for error to be introduced into the ocean temperature changes. Each of the simulations may also contain trends due to internal model drift from the initialised deep ocean state. It is very difficult in a forced ocean model hindcast to separate the trend due to internal drift from trends due to global warming. However, initial model adjustment from the initial state over the early decades may be obvious, especially if it results in a trend that opposes that expected from global warming (i.e. cooling). Here I do not attempt to remove any trend due to drift, but note that the 30+ years of model run prior to the period of interest (1990-2010) does allow the model some time to initially adjust to the forcings (there is still likely to be some inherent drift).

Table 2.1: NEMO model simulations used

Simulation Name	NEMO Code Version	Horizontal Resolution	Forcing	Simulation Length	Reference
NEMO083-N006	3.6	1/12°	DFS 5.2 (Dussin *et al.*, 2016)	1958 - 2015	Moat *et al.* (2016)
NEMO025-VN206-ERA	3.2	1/4°	ERA Interim (Dee *et al.*, 2011)	1989 - 2010	Blaker *et al.* (2015)
NEMO025-N206-CORE	3.2	1/4°	CORE2 (Large and Yeager, 2004)	1958 - 2009	Blaker *et al.* (2015)

2.4 Method

2.4.1 Hydrographic sampling in a model

Three repeat hydrographic sections in the NEA basin are considered (Figure 2.1):

- A16N (from Iceland to just north of the Equator)
- A05E (west to east from the Mid Atlantic Ridge to west Africa)
- OVIDE (from the Mid Atlantic Ridge to Spain).

A16N and A05E were sampled every 5 - 8 years since 1988 (for A16N) or 1992 (A05E), so their repeats span the majority of the period 1990 - 2010. OVIDE is sampled more frequently (every two years) but only over the shorter time range of 2002 - 2010 (Figure 2.1).

The observational methodology of Desbruyères *et al.* (2014) is followed, which itself follows the methodology of Purkey and Johnson (2010). The same coordinates for the hydrographic sections and the same times for when the sections were occupied are used. Figure 2.1 shows when each section was occupied during the time interval 1980 - 2011. One exception where the timing in Figure 2.1 differs from observations is the final occupation of A16N which in reality took place in 2011, but not all of the model simulations were integrated through 2011. Instead, for simulations that include 2010, A16N was assumed to be sampled at the end of 2010 (and for the simulation that does not include 2010, any occupations in 2010 were assumed to be sampled at the end of 2009). This preserves the approximate sampling frequency of each hydrographic section during 1988 - 2011. It is assumed that each gridpoint along the hydrographic section is occupied simultaneously at one point in time (the midpoint of the cruise), when in reality occupations of each section take weeks.

Potential temperature (hereafter 'temperature') data along these sections were subsampled in the different ocean model simulations by extracting the data at the model coordinates closest to the section coordinates used in an observational

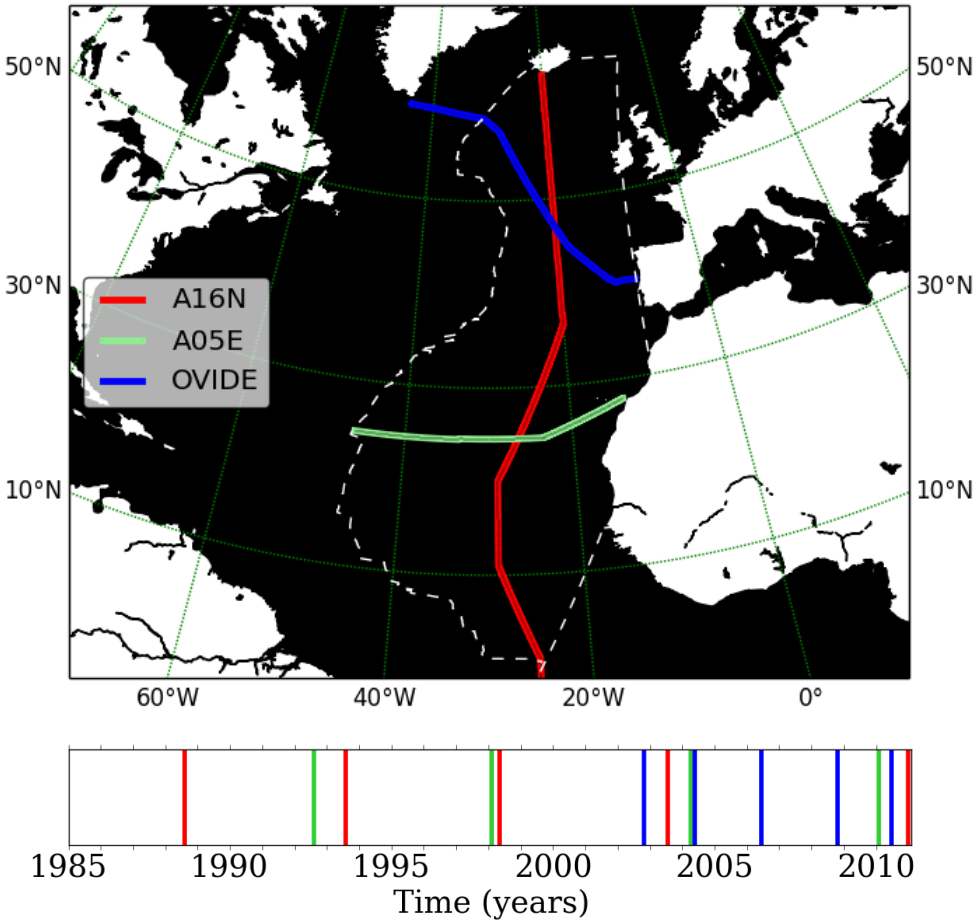


Figure 2.1: Top panel: Map of hydrographic sections sampled using high resolution ocean models. Dashed white lines illustrate the NEA basin area used in this study; this work only considers the portion of a section lying within these bounds. Bottom panel: Timings for occupations of hydrographic sections used in this work.

analysis (Desbruyères *et al.*, 2014). The coordinates used in the observational analysis are derived from the locations of the observations, which are interpolated onto an even standard 2 minute longitudinal/latitudinal grid along the section using a space-preserving piecewise cubic Hermite interpolant (Purkey and Johnson, 2010).

2.4.2 Temperature trends on sections

Linear temperature trends are calculated from section data after subsampling the data in time as follows:

- **Pseudo Sections** The pseudo section trends are calculated from temperature data only for the times of the actual research cruises (the approximate midpoint of the cruise interval is used). The times at which research cruises occupied each section are shown in Figure 2.1 by vertical coloured lines. The pseudo sections span the following time periods: 1988 - 2010 for A16N, 1992 - 2010 for A05E and 2002 - 2010 for OVIDE, with the trends generated representing the frequency of observational sampling, which is limited in both time and space.

- **Model truth from sections** All model output on a section at 5 day mean resolution. The section truth represents observations limited spatially but not in time. The same time period is used for each section: 1990 - 2010.

Since both pseudo sections and model truth from sections use data from the model output which is identically sub-sampled spatially, the difference between trends derived from these data represents temporal uncertainty associated with occupying hydrographic sections infrequently.

To calculate the trend for each grid box along the section and at each vertical model level, a straight line of best fit is applied through the potential temperature, θ (°C) at each timepoint, t (years). The gradient is the linear temperature trend, β:

$$\theta(t) = \beta \cdot t + c + \varepsilon(t) \ . \tag{2.1}$$

c is the interception point with the y-axis, while $\varepsilon(t)$ is an error term derived from the residuals of the linear fit. Following the observational methodology, $\varepsilon(t)$ is not being used to calculate uncertainties. Though $\varepsilon(t)$ is neglected here, it is noted that this error is likely to be large for trends derived from hydrographic style sampling. In a model, it is very likely that the trend fit will have a magnitude of error lower than that for real-world observations, since temperature variability in model-world is less than observed in the real oceans. Therefore, even if this error was calculated using the model results here, it is particularly important that observational oceanographers calculate this independently using real world data to generate this uncertainty as it is likely to be larger than calculated in a model.

At each point along the section, the pseudo section trend is compared to that for the section truth at each vertical level. Then both the pseudo section trends and the section truth trends are horizontally averaged along each section to give a section average trend, $\hat{\beta}_{\text{section}}$, from (a) pseudo sections and (b) from the truth on section.

2.4.3 Estimating the basin trend

By computing the length of each section (using distances in the model), relative weightings for each section are created which enable calculation of a basin trend estimate composed from the weighted average temperature trends from the three sections. The same weightings are assumed for each model simulation estimate to allow comparability between the results from different simulations.

The process of estimating the basin temperature trends, generated by calculating the weighted average temperature trends of each section (A16N, A05E and OVIDE), is performed twice, for both the pseudo sections and truth on section data. At each vertical model level, the calculation of the estimated basin trend from sections, $\hat{\beta}_{\text{basin estimate}}$, can be expressed by the following equation, where w represents the weighting (length) of the section:

$$\hat{\beta}_{\text{basin estimate}} = \frac{(w_{A16N} \cdot \hat{\beta}_{\text{A16N}}) + (w_{A05E} \cdot \hat{\beta}_{\text{A05E}}) + (w_{OVIDE} \cdot \hat{\beta}_{\text{OVIDE}})}{w_{A16N} + w_{A05E} + w_{OVIDE}} \quad . \qquad (2.2)$$

These estimates of the basin trend from the pseudo sections and section model truth are then compared to:

- **Basin model truth** The basin average temperature, neither limited in time nor space, calculated at each vertical pressure level, and representing the model truth for the basin. The linear trend through time is calculated from the basin average temperature using all temporal data (5 day means) between 1990 - 2010. This time interval was chosen because it is the same as that used for previous estimates of decadal variability in deep ocean heat content (Purkey and Johnson, 2010).

The basin model truth is directly compared to the basin estimate from the model truth on sections to assess the uncertainty associated with the limited spatial sampling across a basin that hydrography provides (both trends are calculated over the same period of time). As explained earlier, the difference between basin estimates using the model truth from sections and the pseudo sections represents the temporal uncertainty.

Although focus on model validation is not a principal aim of this study, for interest the model output for each section are compared to the equivalent observational analysis calculated by Desbruyères *et al.* (2014).

2.4.4 Generating uncertainty estimates due to the timings of pseudo sections

The analysis was repeated using different timings for the pseudo sections to examine the sensitivity of the results to specific timings of hydrographic sections. Section data were collected at roughly pentadal (5 yearly) intervals, but hydrographic sections could easily have been sampled at least a year before or after their actual time due to factors such as ship time constraints or weather

conditions.

The temporal uncertainty associated with the hydrographic data timings is quantified to address whether there are significant variations in trends obtained using slightly different timings. This analysis highlights whether potential agreement/disagreement between the basin estimates from pseudo sections and that from section truth or the basin truth is particularly sensitive to exact timings of hydrography sections.

A Monte Carlo method is used in which the analysis of trends is repeated 1000 times using alternative cruise dates that are randomly selected between ±12 months of the actual cruise timings. This timescale of around a year seems a reasonable amount of time that each occupation in an ideal sampling frequency of 5 - 10 years might be shifted by, and provides a range of trends that are not biased towards any particular season.

Repeated analysis producing 1000 alternative pseudo sections enables calculation of an uncertainty estimate, which is calculated as ±2 standard deviations of the trends from the 1000 alternative pseudo sections. When compared to the trend derived from the initial pseudo sections, this interval represents the temporal uncertainty that could arise due to scheduling of cruises.

2.4.5 Calculating thermosteric sea level rise from temperature trends

The thermosteric component of sea level rise (SLR) between shallower depth z_1 and seafloor z_{max} was calculated from the temperature trend at each depth z as follows:

$$SLR = \frac{1}{A(z_1)} \int_{z_{max}}^{z_1} \alpha(z) \; . \; \hat{\beta}_{\text{basin}}(z) \; . \; A(z) \;\; dz \;\; . \tag{2.3}$$

$\hat{\beta}_{\text{basin}}$ is the basin average temperature trend at each vertical model level (converted to Cs^{-1}), A is the area of the basin (m^2), and α is the depth varying

thermal expansion coefficient of seawater (K^{-1}) from the Hydrobase 3 climatology (www.whoi.edu/science/PO/hydrobase/php/index.php).

2.5 Results & Discussion

2.5.1 Temperature on sections

The average potential temperature (hereafter 'temperature') on sections and in the basin for the NEMO083 simulation (Figure 2.2) suggest a warming trend in the NEA at 2000 m of 0.02°C/decade since 1990, although section A05E is cooling. The basin average warms at 3000 m by a similar amount to 2000 m, but at 4000 m and 5000 m all sections and the basin average estimate from sections are cooling (Figure 2.3). The spatial uncertainty of the average basin temperature from sections is the difference between the basin average estimate (brown line) and the basin truth temperature (black line). At each depth the basin estimate from sections is consistently warmer than the basin truth (Figures 2.2 and 2.3).

The basin estimate has a warm bias of 0.05°C at 2000 m, 0.02°C at 3000 m and 4000 m, and 0.03°C at 5000 m (Figures 2.2 and 2.3) for the NEMO083 simulation. At 2000 m and 3000 m the bias is very small at the year 1980 but increases during the next few years of the simulation; this might be related to an initially rapid warming of some sections (e.g. OVIDE and A16N at 2000 m) and a subsequent reduction of rate of warming after the year 1990. The change in rate of warming might be related to initial spin up of the model rather than replicating real-ocean variability. Otherwise, the bias remains relatively consistent during the simulation and this might suggest that observation derived estimates of ocean heat content in the NEA may be improved by using a bias adjustment to reduce spatial uncertainty.

To gain some idea of the realism of the model simulations, observational data from each cruise averaged along each section is plotted on the same graph (data from Desbruyères *et al.*, 2014). Figures 2.2 and 2.3 show that the model output temperature is of a similar magnitude to observations at the depths considered,

generally being within about 0.1°C. An exception is the A16N section at 2000 m which is warmer than observed by 0.1 - 0.3°C and OVIDE at 2000 m on which the model overestimates the observed temperature by 0.5°C.

Some of these differences are not insignificant considering the low temperatures of the deep ocean. There is generally a consistent offset between the magnitude of the model average and the observational average, although there is large variability within the observations on A16N not seen in the model average (Figure 2.2). Observations are an average of a couple of months measurements along a section, whereas 5 day mean data from the model are presented here, however it seems unlikely in some cases that temperature variability between observations would encapsulate the modelled temperatures. Where the offset is relatively consistent this suggests that the model may represent the vertical pattern of decadal trends with some degree of skill. Reasons for differences may include uncertainties in the surface forcing in the model, the model not resolving certain processes or uncertainty in parameterisations.

Figures 2.4 and 2.5 show the same figures for the NEMO025 (ERA Interim) simulation. There is slight (0.02°C/decade) warming at 3000 m, and very slight cooling at 4000 m and 5000 m throughout 1990 - 2010. Figures 2.6 and 2.7 illustrate that NEMO025 (CORE2) possesses similar trends to NEMO025 (ERA Interim), and the behaviour of individual sections is similar in both of the NEMO025 runs. The trends observed are similar in each of the three model runs, despite the different forcing products used.

To quantify uncertainty in the basin temperature estimate in the NEMO025 simulations, consider the difference between the brown line (basin estimate) and the black line (basin truth). Figure 2.4 shows that for NEMO025 (ERA Interim) the basin estimate is slightly biased warm by 0.02°C but that at 3000 m the basin estimate is biased cool by 0.03°C. At 4000 m the warm bias is 0.01°C and 0.04°C at 5000 m (Figure 2.7). These biases are similar in magnitude to those in the NEMO083 (DFS) simulation.

For NEMO025 (CORE2) the warm bias at 2000 m is up to 0.03°C although the magnitude of bias changes through the simulation more than for the previous simulations (Figure 2.6). At 3000 m there is a cold bias of 0.04°C (Figure 2.6), at

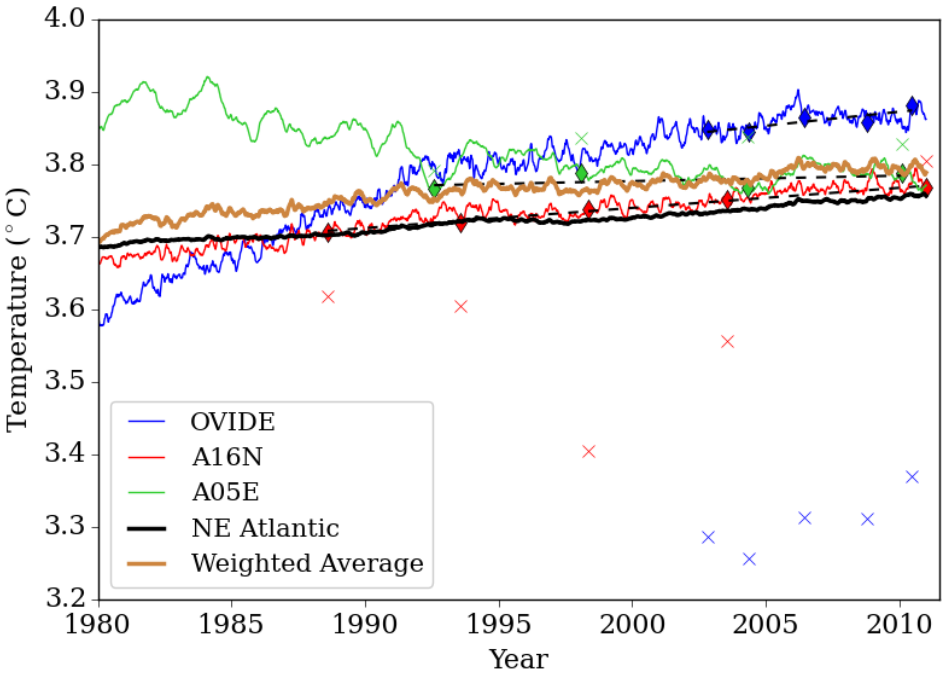
4000 m a warm bias of 0.01°C which increases through the simulation from 0 at 1980, and at 5000 m a warm bias of 0.05°C that is fairly consistent through time (Figure 2.7). At 5000 m there is a step change at the end of 2007 which is very likely an adjustment that occurred when the simulation was restarted.

For NEMO025 (ERA Interim) the model estimates are similar to the observations, though at 2000 m the model estimate is warmer than the observations by up to 0.5°C for A16N and OVIDE (Figure 2.4). At 3000 m and 4000 m the model estimation of the observed average temperature along OVIDE is within 0.1°C, with some slightly larger differences between model and observations at 5000 m (up to 0.2°C for 5000 m, Figure 2.5). The NEMO025 (CORE2) averages have similar relationships to the observational data (Figures 2.6 and 2.7).

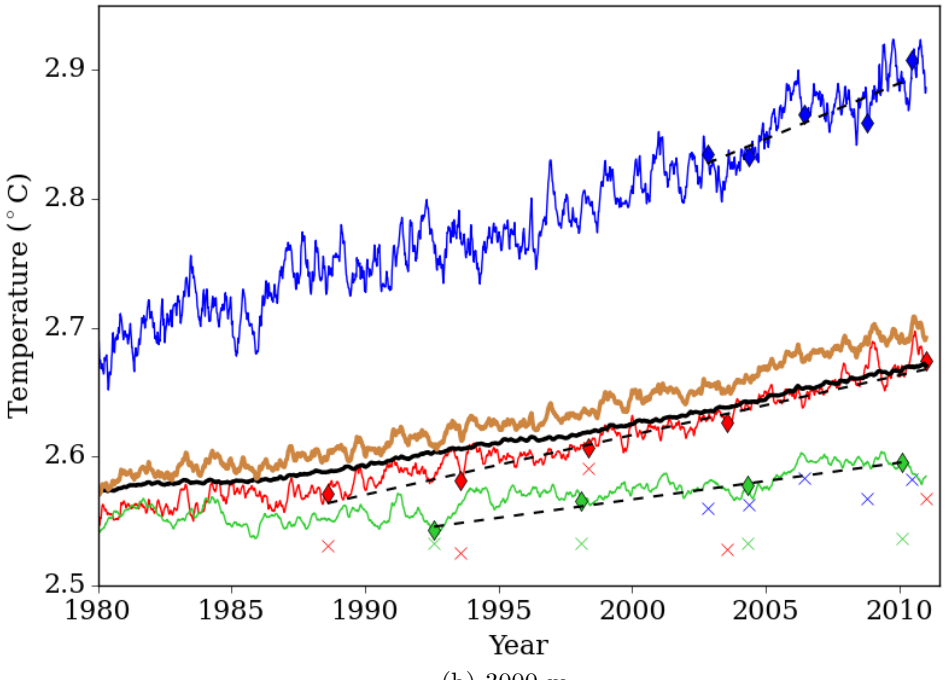
Generally between the period 1990 and 2010 (the period over which the temperature trend in the basin is calculated later in this work), the warming/cooling trends in the model are approximated relatively well by a linear trend. Therefore the linear fit assumed later to calculate the trend (following the method used for the observational data) is judged to be an appropriate technique. On the temperature figures described in this section the linear trend from pseudo sections is highlighted by black dashed lines between the times of the first and last pseudo sections. In this work, the trend over the pseudo section period (using only pseudo section data) is compared to the trend over the period 1990 - 2010 (using 5 day mean data) as a measure of the temporal uncertainty that is inherent in observational sampling because pseudo sections are separated by 2-8 years and in some cases (namely OVIDE) do not cover the whole period of interest (1990 - 2010). This differs from the spatial bias in observational estimates of deep OHC which exists because (in this basin) there are only 3 hydrographic sections considered. This spatial bias is what has been quantified so far in this work by considering the offset between temperature calculated from the weighted average of the three sections and from all grid points in the basin.

2.5.2 Temperature trends along A16N

Temperature trends along sections are explored by considering the linear trend at each point along sections and at each depth level. The trends along A16N using

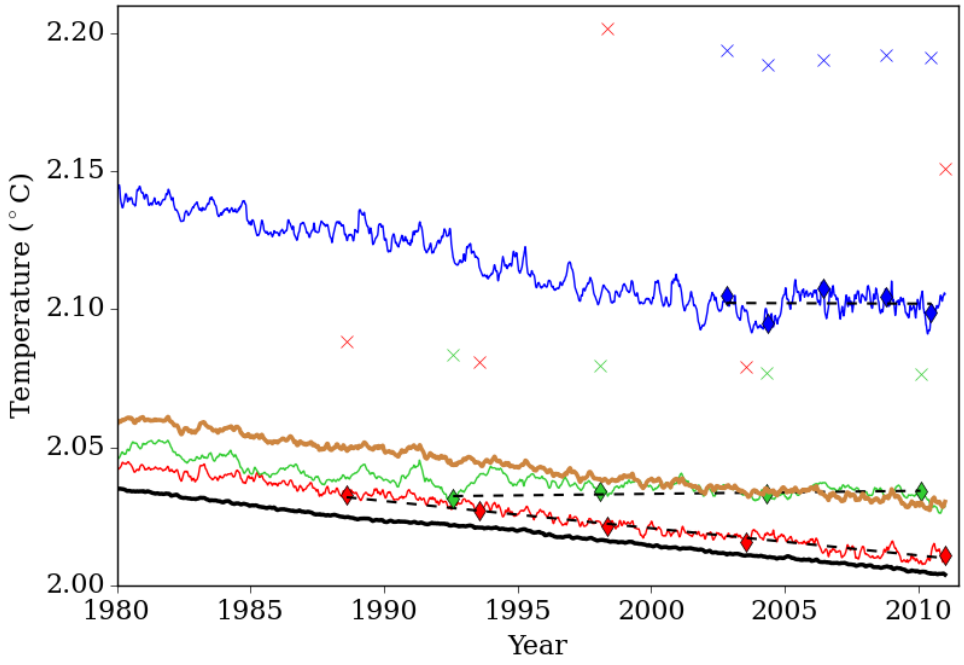


(a) 2000 m

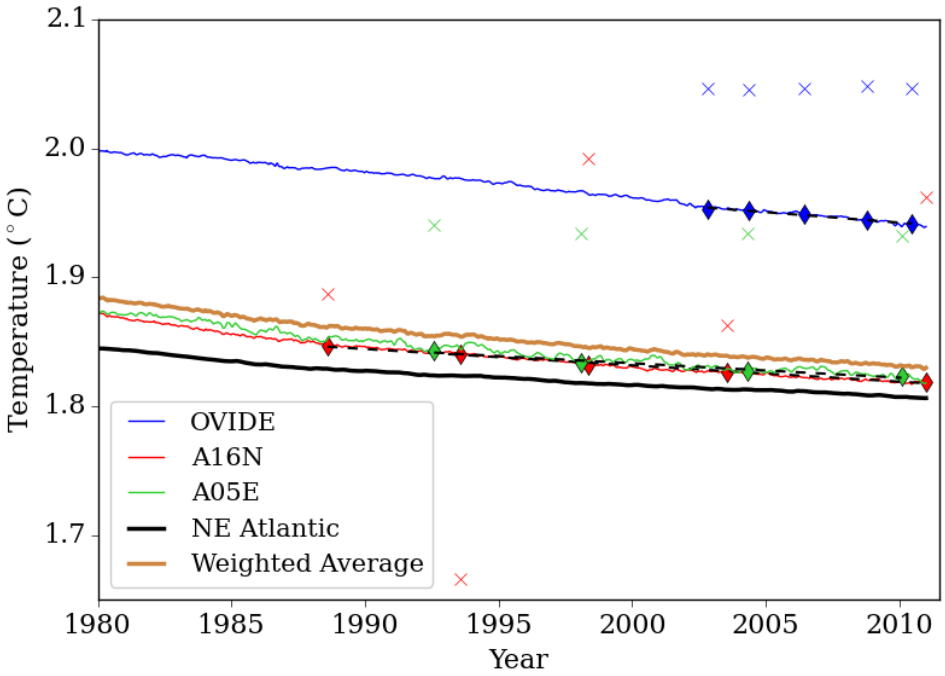


(b) 3000 m

Figure 2.2: Temperature between 1980 and 2010 in NEMO083 at (a) 2000 m and (b) 3000 m. Red/green/blue lines denote the average temperature (Θ) along each hydrographic section (A16N/A05E/OVIDE respectively) with the black dashed line the linear temperature trend from pseudo sections timings (diamonds). Estimated basin average temperature from sections is the brown line, whilst the black line represents the basin average temperature (the model truth). Crosses denote the observed average temperature on each section (Desbruyères *et al.*, 2014).

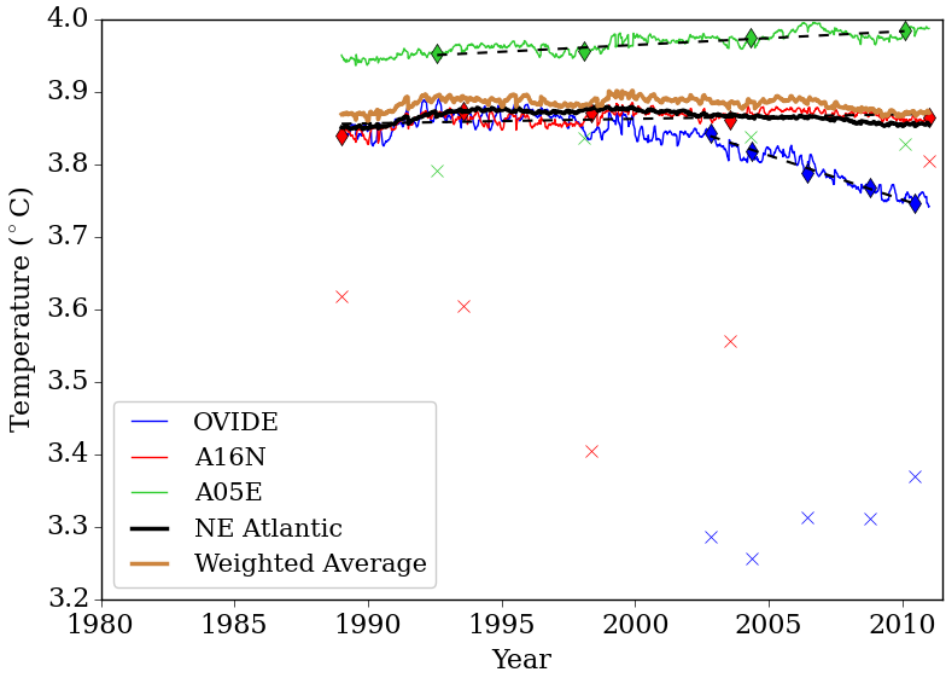


(a) 4000 m

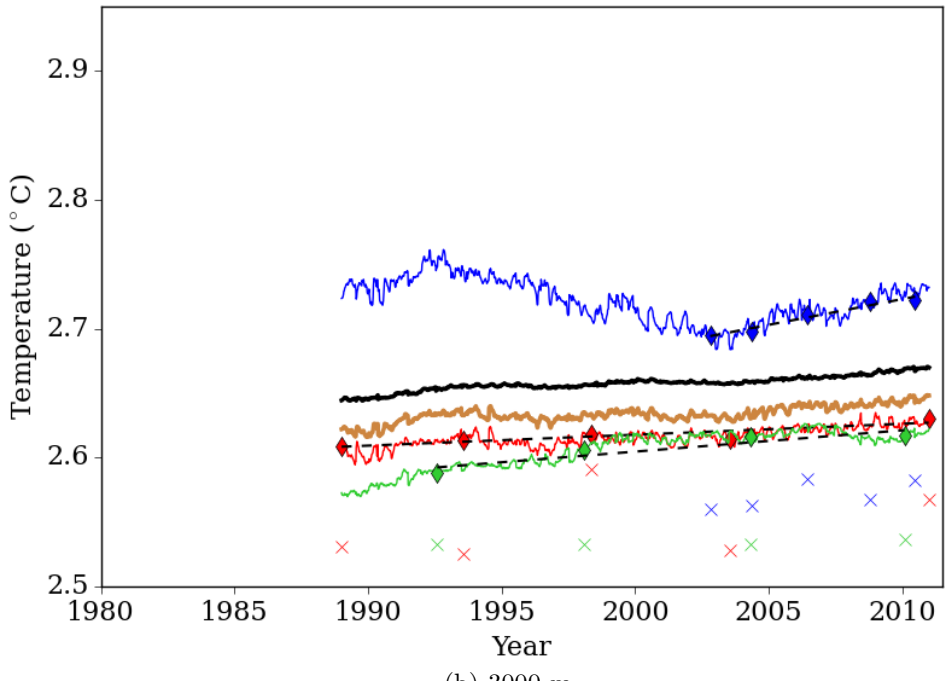


(b) 5000 m

Figure 2.3: Temperature between 1980 and 2010 in NEMO083 at (a) 4000 m and (b) 5000 m. Otherwise as described in Figure 2.2.

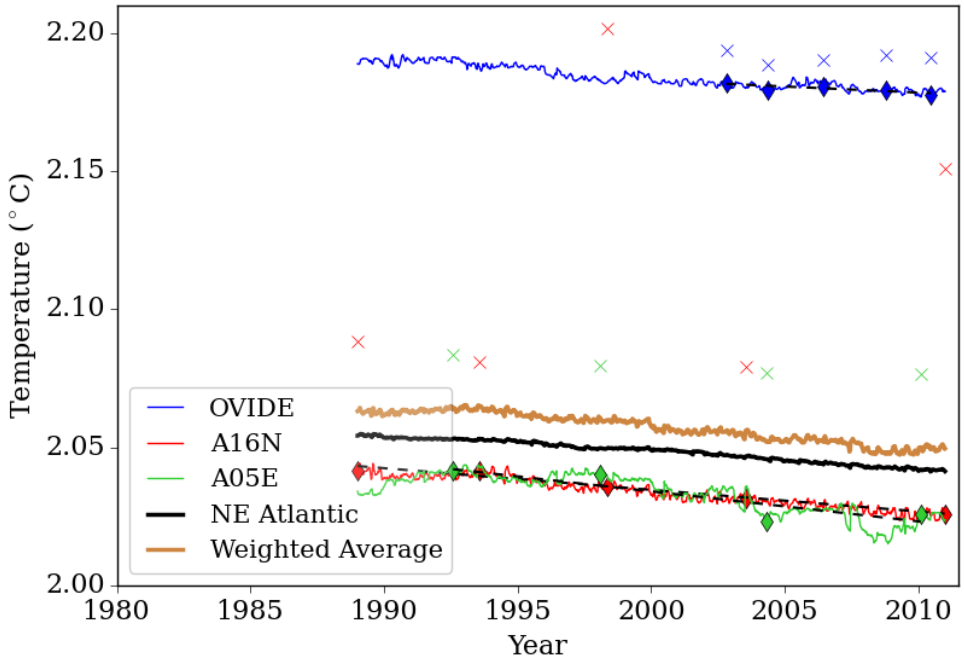


(a) 2000 m

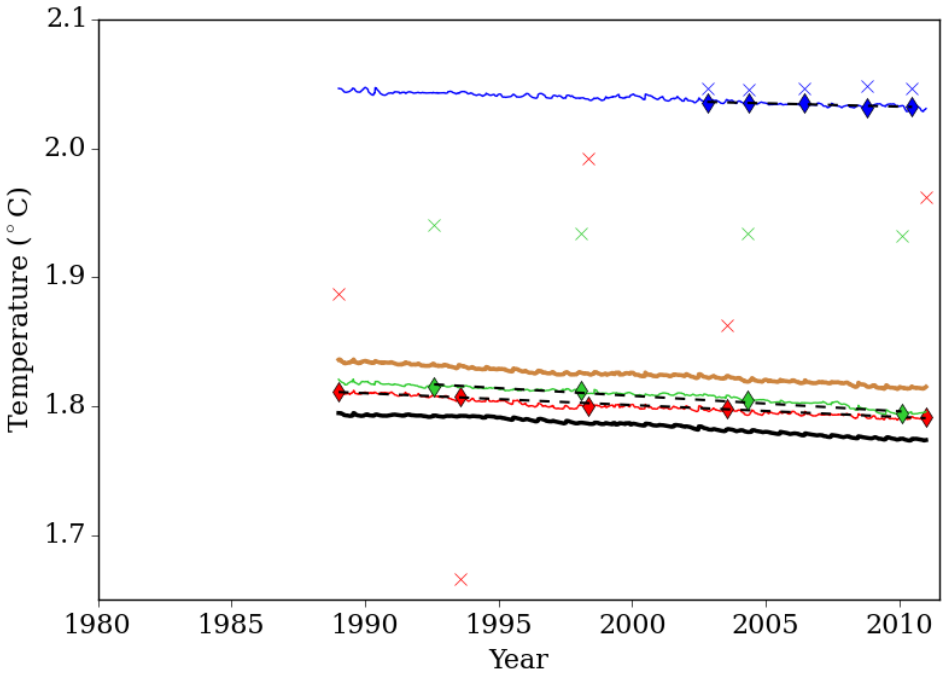


(b) 3000 m

Figure 2.4: Temperature between 1989 and 2010 in NEMO025 (ERA Interim) at (a) 2000 m and (b) 3000 m. Otherwise as described in Figure 2.2.

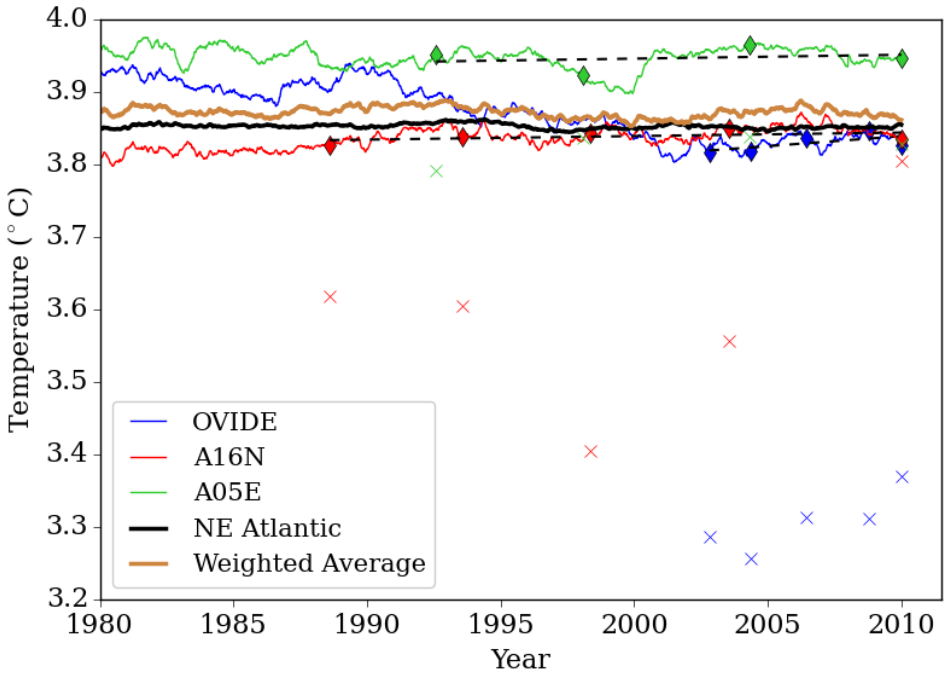


(a) 4000 m

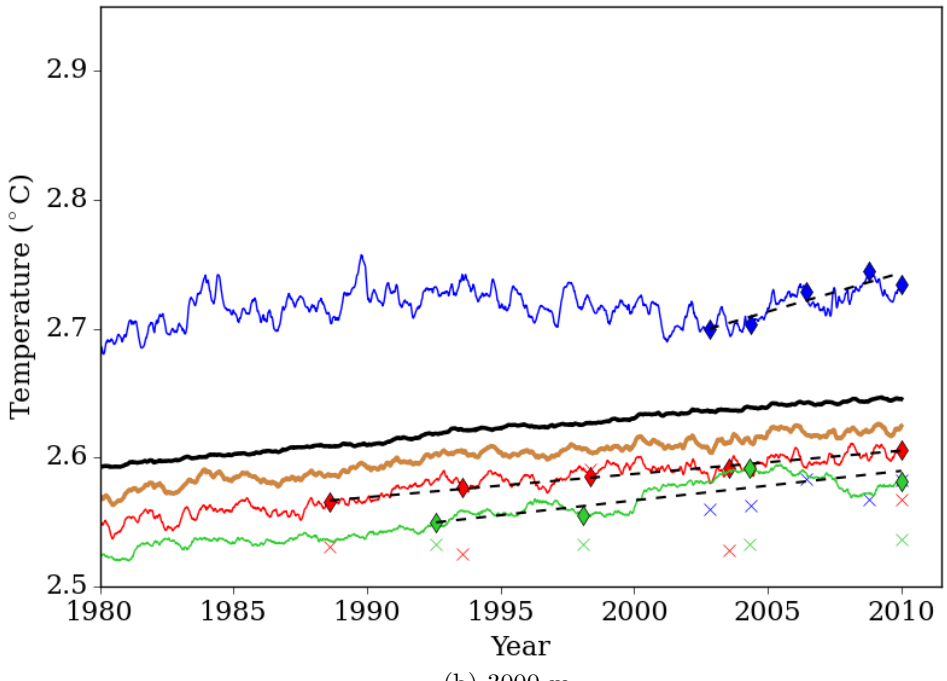


(b) 5000 m

Figure 2.5: Temperature between 1980 and 2010 in NEMO025 (ERA Interim) at (a) 4000 m and (b) 5000 m. Otherwise as described in Figure 2.2.

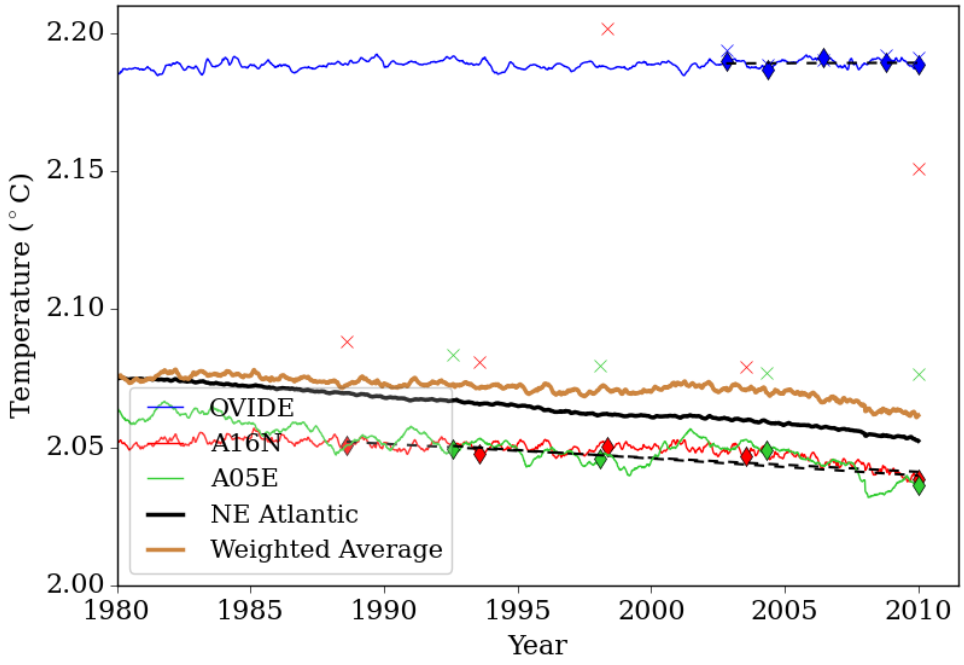


(a) 2000 m

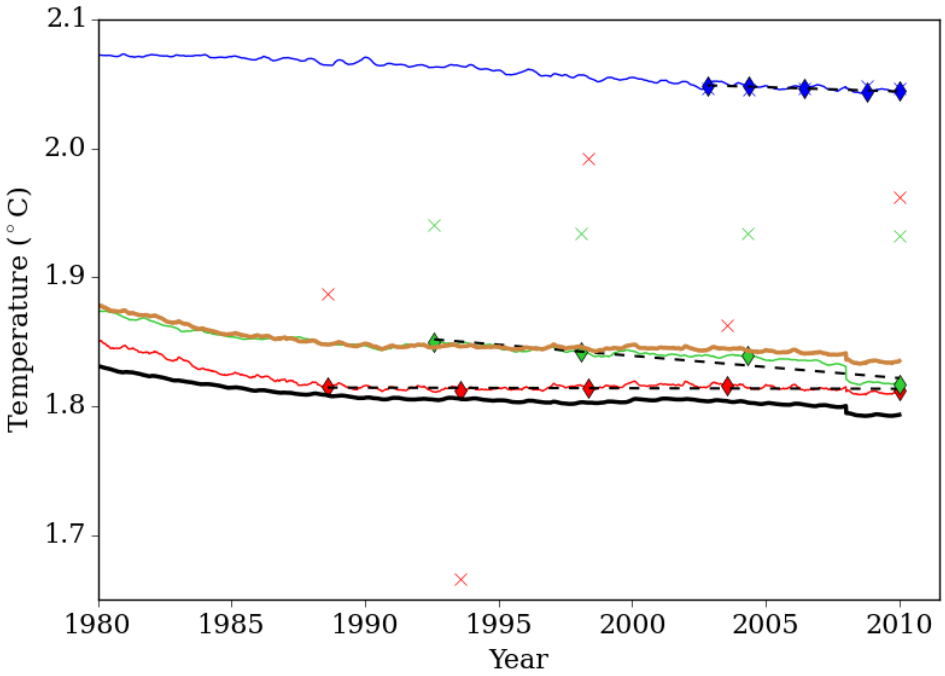


(b) 3000 m

Figure 2.6: Temperature between 1989 and 2010 in NEMO025 (CORE2) at (a) 2000 m and (b) 3000 m. Otherwise as described in Figure 2.2.



(a) 4000 m



(b) 5000 m

Figure 2.7: Temperature between 1980 and 2009 in NEMO025 (CORE2) at (a) 4000 m and (b) 5000 m. Otherwise as described in Figure 2.2.

the model truth on section data (5 day means) are calculated using all temperature data between 1990 and 2010, while the pseudo section trends are calculated using data from 5 pseudo sections between the years 1988 - 2010. Figure 2.8 (a) shows that along section A16N in NEMO083 there is strong warming above 3000 m near Iceland and down to 4000 m in the subtropics/tropics, whilst some cooling is observed at mid-latitudes at around 2000 m and there is very slight cooling in the abyssal basin. In both NEMO025 simulations (Figures 2.8 c and e) the cooling trend is strongest to the north of the section (near Iceland), whereas the mid and southern parts of the basin warm at 2000-3000 m and the abyssal basin cools very slightly.

The trends are calculated in Figures 2.8 (b), (d) and (f) from the temperature data at the timings of pseudo sections only. The differences between the trends from this temporally limited data and the trends from temporally complete (section truth) data in (a), (c) and (e) reveal temporal uncertainty. However, the patterns of the trend along A16N look similar when using the section truth or pseudo sections for each model so temporally limited data captures the pattern of trends along A16N well, suggesting that bias due to temporal uncertainty is not large on this section. As the sampling frequency of the observed hydrographic sections is sufficient to capture the trends along this section in each model (Figures 2.8 b, d and f), temporal uncertainty may not have a large impact on our ability to observe decadal trends with infrequent repeat hydrographic sampling along A16N.

Averaging the trends (for temporally limited pseudo sections) horizontally along the section A16N (Figure 2.9, dashed lines) achieves a very close representation of the model truth on A16N (coloured solid lines), with the biggest differences in the depth range 2000 - 3000 m for each model. On average, the section A16N warms above 3500 m and cools below 4000 m in each model, regardless of whether temporally complete or limited data are used (Figure 2.9). A16N is the longest section and due to its north-south location (of the sections considered) it is arguably the best sole representative of the whole basin, but the warming trend above 3000 m in the models is not found in the hydrographic observations (Figure 2.9, black line), querying the accuracy of the models with respect to this trend. There are clear differences between the magnitude of the trends in each model with the warming in NEMO083 at 3000 m being three times greater than

that in the two NEMO025 simulations, however temporal uncertainty when using pseudo section trends is low regardless because for each model there is a close match between the section truth and pseudo section trends.

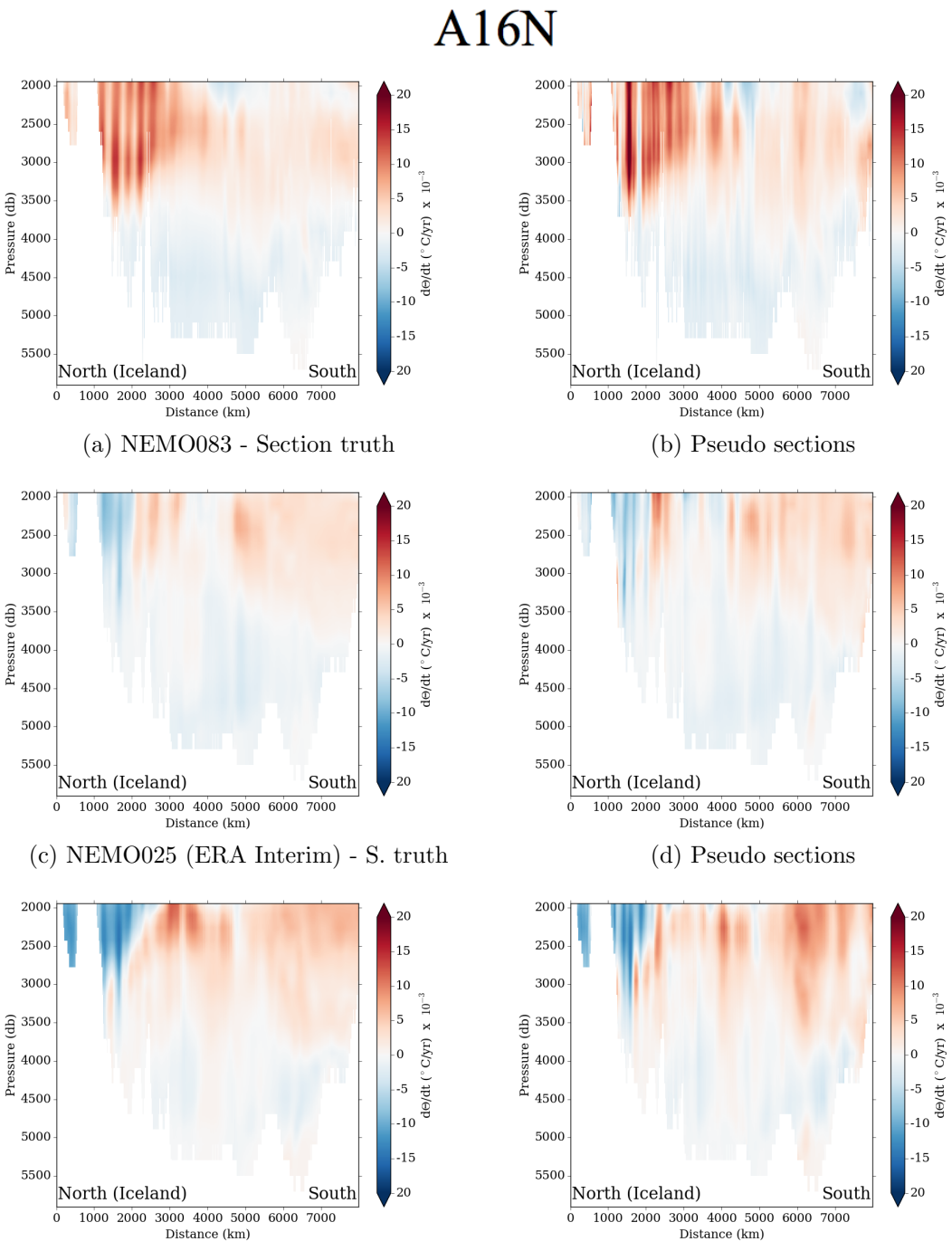
2.5.3 Temperature trends along A05E

In NEMO083 a warming trend is found between the depths 2500 m and 4000 m along the section when the section truth data are used (Figure 2.10 a), with cooling above 2000 m and below 4000 m. When data are limited temporally (use only 4 pseudo sections during the period 1992 and 2010) the pattern is slightly different with a stronger warming trend observed between 2000 - 4500 m in the western half of the section (Figure 2.10 b). This suggests that temporal uncertainty leads to a different trend using pseudo section data than using the section truth on A05E.

In the NEMO025 (ERA Interim) and NEMO025 (CORE2) simulations the differences between the trends from the section truth and from pseudo sections are not as large, with warming intensified in the central part of the section between 2000 - 3000 m in NEMO025 (ERA Interim) (Figures 2.10 c/d) and to the west and eastern parts of the section in NEMO025 (CORE2) (Figures 2.10 e/f). There is a strong cooling as the depth shoals on the eastern side of the basin when using temporally limited data from pseudo sections from NEMO025 (CORE2) (Figure 2.10 f); this feature does not exist when the trend is calculated using the section truth (Figure 2.10 e).

On the whole across models the pattern of trends is captured to some extent using temporally limited model output (patterns in Figure 2.10 b, d and f bear similarities to those in Figure 2.10 a, c and e). However there are clear differences in the patterns of warming across the basin in the different model simulations.

When the trends along the section are averaged (Figure 2.11) the differences between the coloured dashed and solid lines illustrate that although the broad features of the vertical change in temperature trend (warming between 2000 - 3500 m, cooling below 4000 m) are captured when temperature data are restricted in time, temporally limited data does not always yield the correct



(a) NEMO083 - Section truth

(b) Pseudo sections

(c) NEMO025 (ERA Interim) - S. truth

(d) Pseudo sections

(e) NEMO025 (CORE2) - S. truth

(f) Pseudo sections

Figure 2.8: Distance pressure sections below 2000 m of dΘ/dt derived from 5 day mean model output along hydrographic section A16N in time period 1990-2010 from simulation NEMO083 (DFS) (a) and subsampled using the dates of research cruises within period 1988-2010 (b). (c)/(d) NEMO025 (ERA Interim). (e)/(f) NEMOO25 (CORE2).

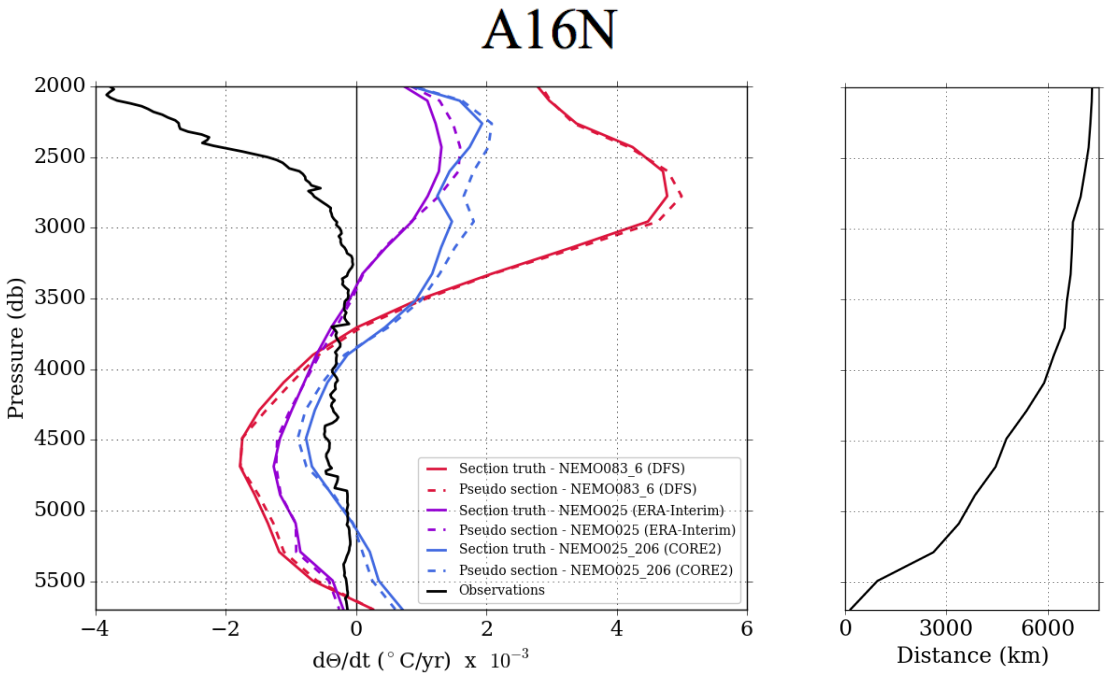


Figure 2.9: Left panel: Horizontally averaged depth profiles of dΘ/dt below 2000 m for hydrographic section A16N (both model truth on section and pseudo sections) for all models. Observational data from Desbruyères *et al.* (2014). Right panel: Distance of the section in the NEMO083 model.

magnitude of warming. The temporal uncertainty (difference between coloured dashed and solid lines) is clearly largest in the NEMO083 simulation (red lines in Figure 2.11), except in the abyssal ocean below 4500 m where it is largest in NEMO025 (CORE2). In NEMO083, at the depth of peak magnitude warming (around 2700 m) the trend from pseudo sections is 50% greater in magnitude than the section truth trend on A05E during 1990 - 2010.

However, unlike for A16N (Figure 2.9), there is relatively close agreement between the three simulations as to the vertical structure of warming/cooling, and the simulations all represent the pattern of the observations (from Desbruyères *et al.*, 2014) reasonably well (warming above 3000 m and cooling below 4000 m), although the maximum magnitude of warming/cooling is greater in the models than in observations and the warming penetrates deeper in the models than in observations (Figure 2.11). That the vertical structure is similar for each model when the section data are averaged (Figure 2.11) while the horizontal structure of trends along the section differed (Figure 2.10) highlights that when averaging along the section important information about the

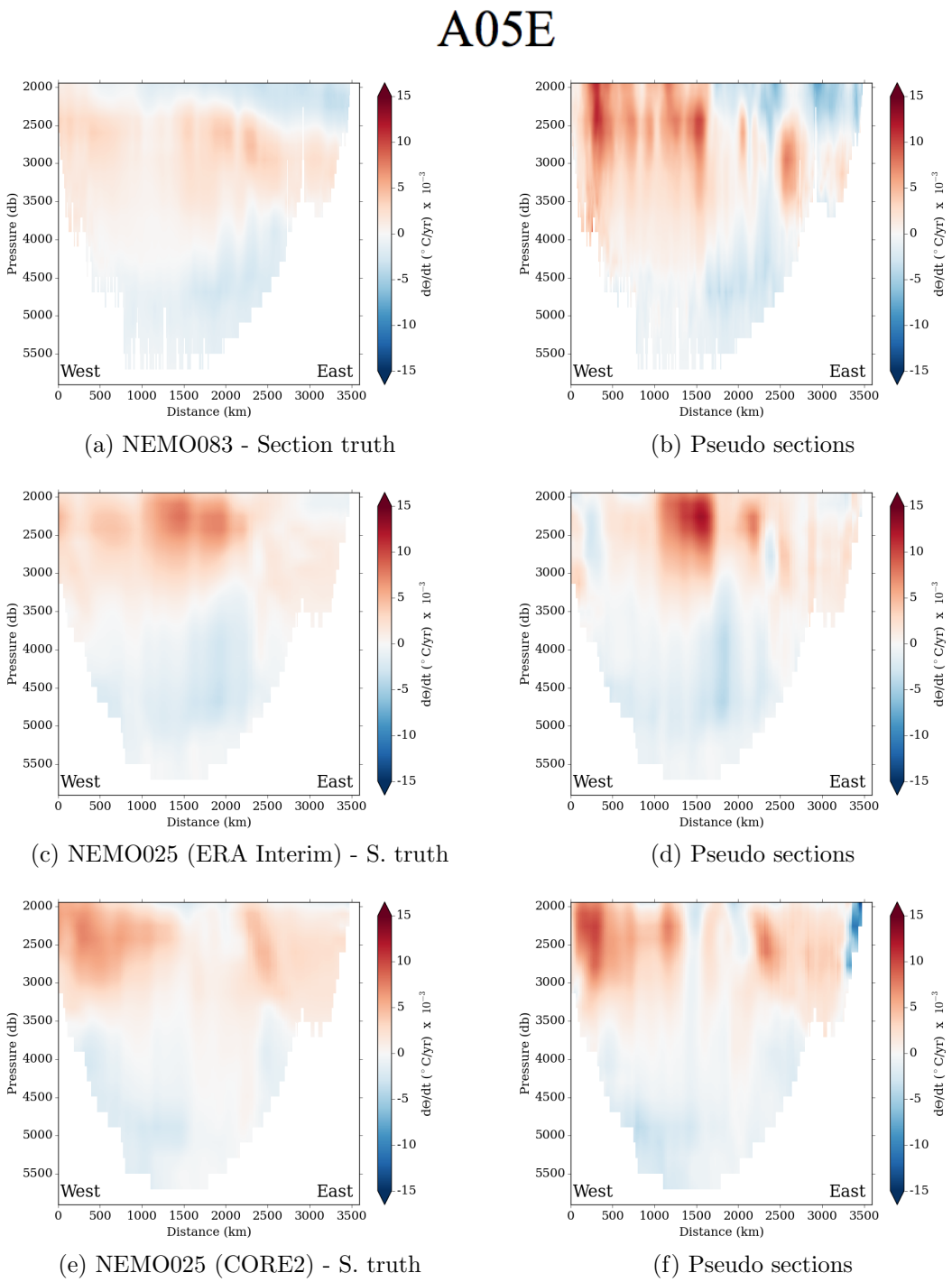
distribution of trends along the section is lost, and implies that the agreement of the vertical structure between models may be fortuitous. In addition, the differences between models and the differences between model truth and pseudo section data suggests that additional deep ocean sampling would be of benefit to provide higher temporal resolution data that reduces the uncertainty in the magnitude and horizontal structure of trends along this section.

2.5.4 Temperature trends along OVIDE

In NEMO083, OVIDE is warming between 2000 - 3000 m (Figure 2.12 a) when the trends are calculated from section truth data between 1990 and 2010. When the trends are calculated from temporally limited data (4 pseudo sections during the period 2002 - 2010), the warming between 2000 - 3500 m is still apparent but restricted to certain locations with cooling in between (especially towards Greenland), resulting in a vertical stripy pattern of trends along the section. The pseudo section data do not give an accurate representation of the trends along the section over 1990 - 2010, and this will introduce temporal uncertainty into estimates of the basin trend using pseudo section trends on OVIDE.

The model truth on OVIDE in NEMO025 (both ERA Interim and CORE2 forced experiments) warms further to the south-east (Figures 2.12 c and e), but again the spatial patterns of trends using pseudo section data do not capture this introducing uncertainty in basin estimates using the OVIDE section (Figures 2.12 d and f). Pseudo sections along OVIDE are not as good at representing the model truth as for A05E and A16N (Figures 2.12 b, d and f). When considering the average temperatures on OVIDE in the NEMO025 simulations there are contrasting trends between 1990-2000 and 2000-2010 particularly at 3000 m (Figures 2.4 and 2.6). As OVIDE has been sampled only since 2002, pseudo sections do not capture trends as well as they do for A16N/A05E, whose repeat sections better span the period of study (1990-2010). For this reason, estimates of the basin temperature trends from sections included later in this chapter are calculated with and without OVIDE to consider how OVIDE contributes to spatial and temporal uncertainty.

For each simulation, the trends on OVIDE below 4000 m reveal slight cooling



(a) NEMO083 - Section truth

(b) Pseudo sections

(c) NEMO025 (ERA Interim) - S. truth

(d) Pseudo sections

(e) NEMO025 (CORE2) - S. truth

(f) Pseudo sections

Figure 2.10: Distance pressure sections below 2000 m of dΘ/dt derived from 5 day mean model output in period 1990-2010 along hydrographic section A05E from simulation NEMO083 (DFS) (a) and subsampled using the dates of research cruises which span period 2002-2010 (b). (c)/(d) NEMO025 (ERA Interim). (e)/(f) NEMOO25 (CORE2).

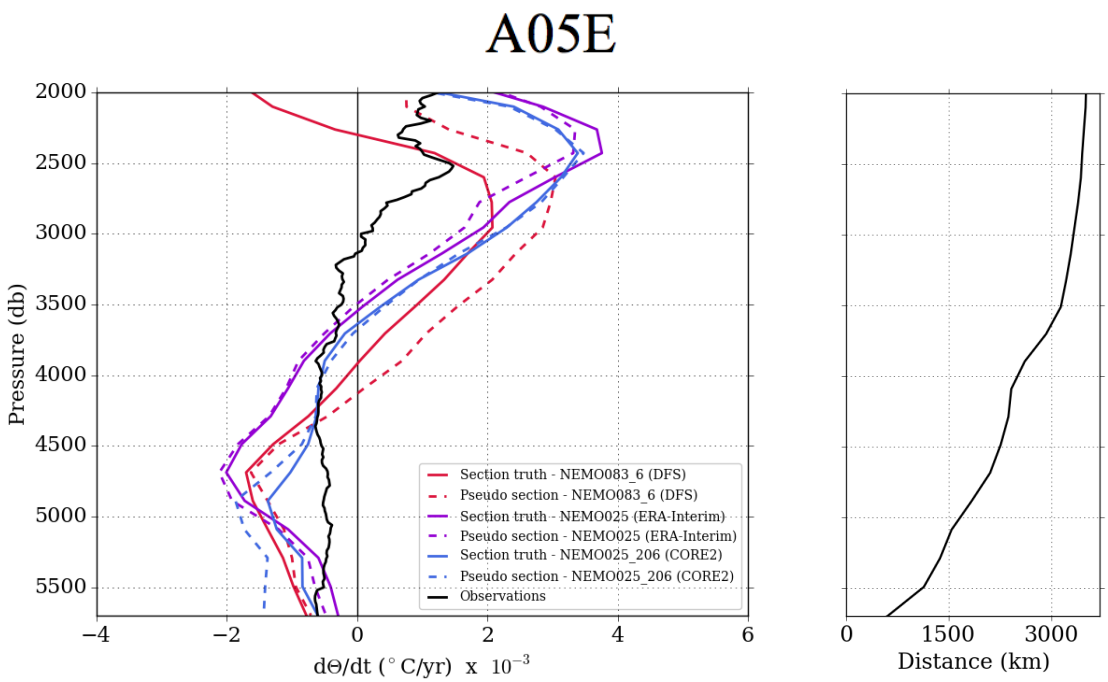


Figure 2.11: Left panel: Horizontally averaged depth profiles of dΘ/dt below 2000 m for hydrographic section A05E (both model truth on section and pseudo sections) for all models. Observational data from Desbruyères *et al.* (2014). Right panel: Distance of the section in the NEMO083 model.

which is stronger to the eastern side of the basin in NEMO083 and NEMO025 (ERA Interim) and weakest in the NEM025 (CORE2) simulation (Figures 2.12). This is confirmed by considering the vertical average trend on the section in each model in Figure 2.13, where each model shows a slight cooling which is not found in observations (Desbruyères *et al.*, 2014). When considering the model truth, there is on average a warming above 4000 m in NEMO083 between 1990 - 2010 (contrasting with observations that show shallower warming above 2500 m) whilst the two NEMO025 simulations cool.

Figure 2.13 shows that there are large differences between the pseudo section and section truth trends in each simulation (especially above 4000 m). The large differences between the coloured solid and dashed lines mean that pseudo sections can lead to a misleading interpretation of the trend on the section, especially for NEMO025 (CORE2) for which pseudo sections suggest a warming trend above 4000 m but the model truth on sections show that in reality the section cools above 3000 m between 1990 - 2010. Temporal uncertainty is clearly important in limiting our ability to capture trends along OVIDE accurately over the period

1990 - 2010 (OVIDE was sampled every two years between 2002 - 2010, which is not a good representation of the whole 1990 - 2010 period).

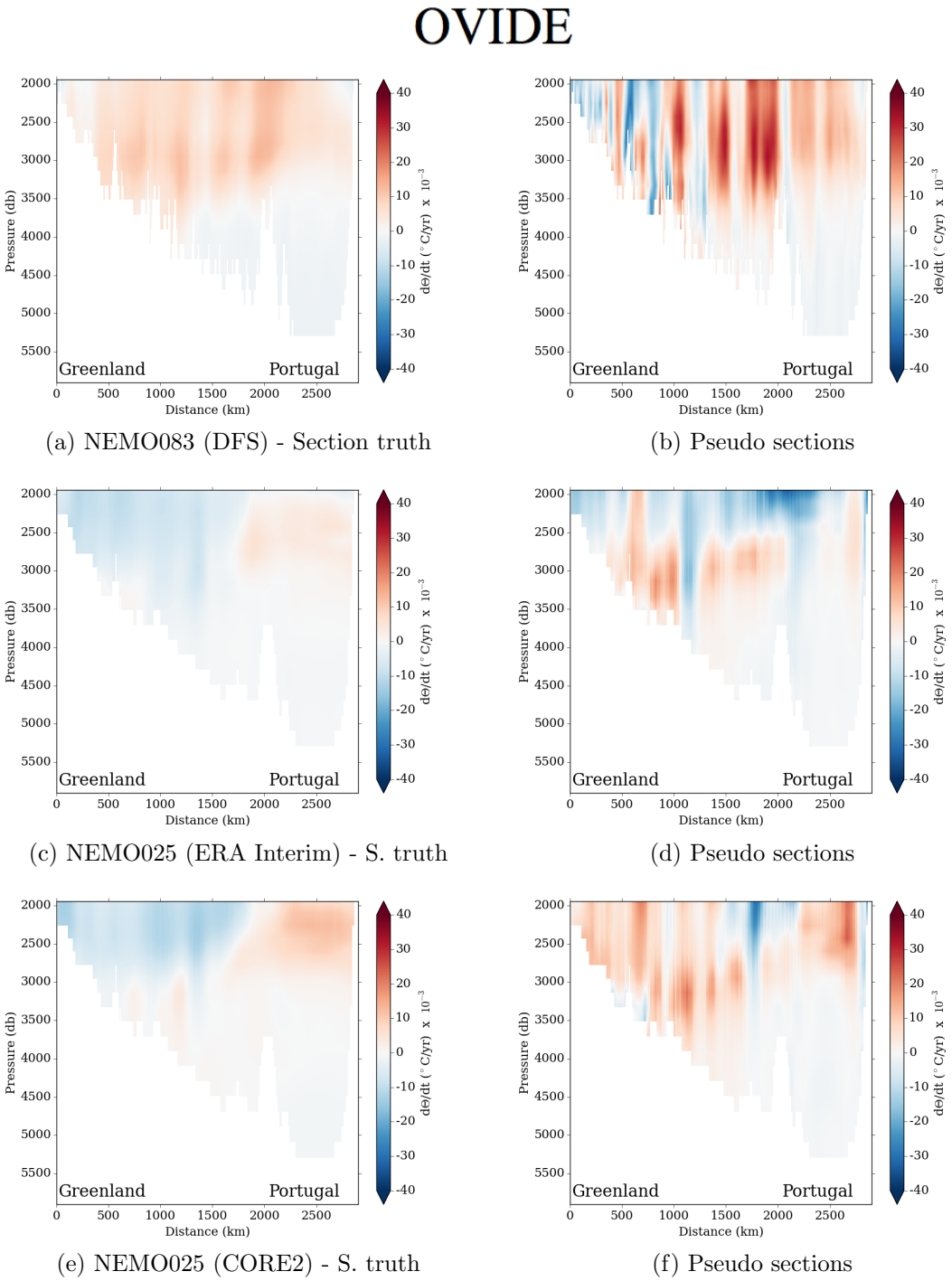
2.5.5 Estimating basin trends from hydrographic sections

The horizontally averaged trends for each section were averaged to produce a estimate of the basin trend from hydrographic sections (Figure 2.14) using both the section truth and pseudo section trend data. These two estimates (dashed lines) are compared to the basin average trend (solid line) to reveal if trends from sections can be combined to give a basin estimate that is representative of the model truth basin average trend.

Figure 2.14 (a) demonstrates that for NEMO083 both the basin estimates from the model truth on sections and pseudo sections show the same overall vertical structure as the basin average (basin truth) when all three sections are used. However there is a clear difference between the basin estimate from the section truth (red dashed line) and the basin truth (black solid line) below 4500 m, and this represents spatial uncertainty in the abyssal ocean due to not enough sections spanning the basin. The basin estimate from section truth suggests a greater cooling than the basin truth (Figure 2.14 a). Since at this depth spatial uncertainty is of greater magnitude than temporal uncertainty, this suggests that there might be a greater improvement in the accuracy of hydrographic sampling in the abyssal ocean below 4500 m if more sections were sampled rather than sampling existing sections more frequently.

Above 4500 m there are differences between all three lines (Figure 2.14 a), with the difference between the two dashed lines (black and red) indicating the temporal uncertainty, whereas the difference between the solid black line and the dashed red line indicates the magnitude of spatial uncertainty. Between 2500 m and 3000 m both temporal and spatial uncertainty are of a similar magnitude and at approximately 2800 m the total uncertainty would result in a basin estimate of warming from pseudo sections which is 25% greater than that from the basin truth (Figure 2.14 a).

To consider if the magnitude of temporal uncertainty might change if different



(a) NEMO083 (DFS) - Section truth

(b) Pseudo sections

(c) NEMO025 (ERA Interim) - S. truth

(d) Pseudo sections

(e) NEMO025 (CORE2) - S. truth

(f) Pseudo sections

Figure 2.12: Distance pressure sections below 2000 m of dΘ/dt derived from 5 day mean model output in period 1990-2010 along hydrographic section OVIDE from simulation NEMO083 (DFS) (a) and subsampled using the dates of research cruises within period 1988-2010 (b). (c)/(d) NEMO025-ERA. (e)/(f) NEMOO25-206.

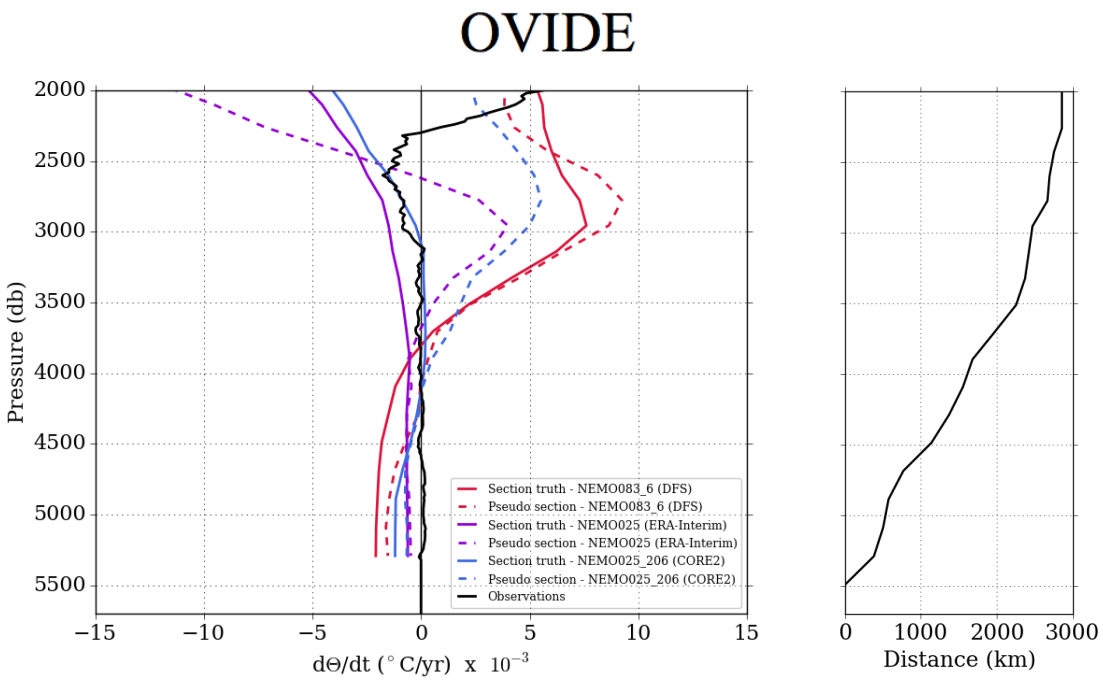
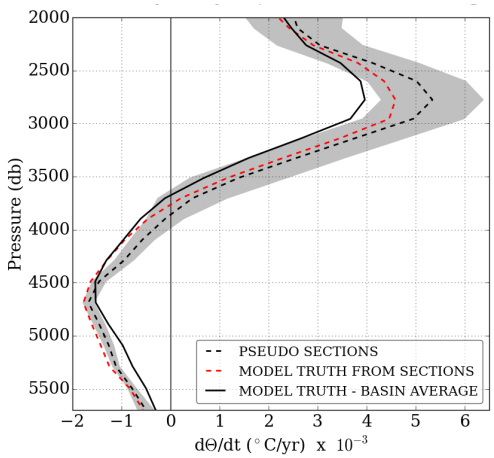


Figure 2.13: Left panel: Horizontally averaged depth profiles of dΘ/dt below 2000 m for hydrographic section OVIDE (both model truth on section and pseudo sections) for all models. Observational data from Desbruyères *et al.* (2014). Right panel: Distance of the section in the NEMO083 model.

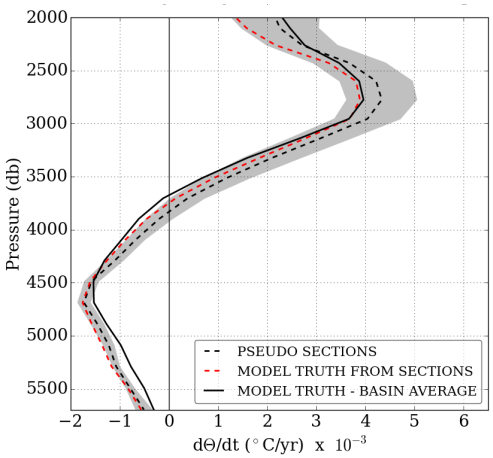
cruise dates were used, the analysis of basin trends using pseudo section data is repeated one thousand times using alternative pseudo sections dates (each date randomly selected between ±1 year of the real timing for each cruise). The spread of results is demonstrated by the shaded areas in Figure 2.14 which reflect ± 2 standard deviations of the 1000 alternative pseudo sections around the trend of the actual pseudo sections. The magnitude of the shaded area indicates that temporal uncertainty could be larger or smaller depending on the cruise dates, which are in reality heavily influenced by ship availability, availability of key scientists or severe weather.

For NEMO083 (Figure 2.14 a) the range of temporal uncertainty is broadest between 2000 - 3000 m. The range indicated by shaded areas suggests that, if certain alternative cruise dates had been used, the combination of larger temporal uncertainty on top of the spatial uncertainty (constant) could result in an overestimation of the warming trend at 2800 m of up to 50%.

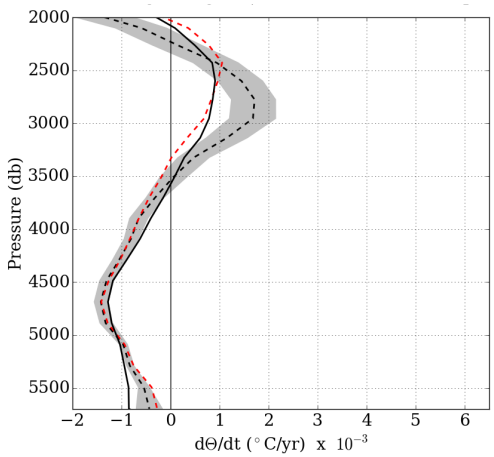
When excluding OVIDE from the trend analysis because of its large temporal



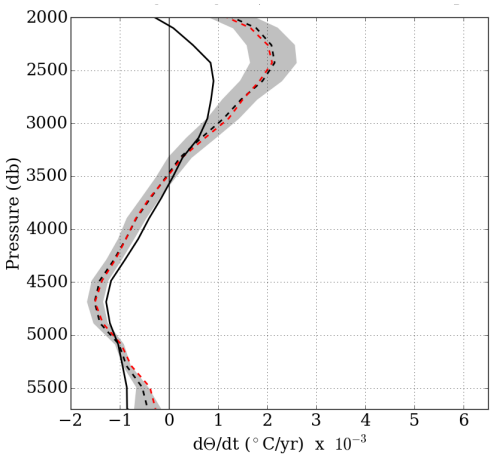
(a) NEMO083 - including OVIDE



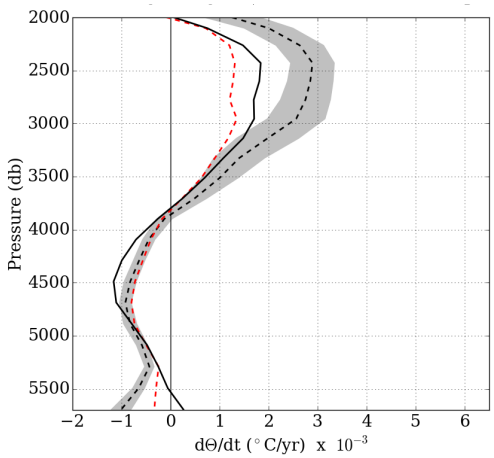
(b) NEMO083 - excluding OVIDE



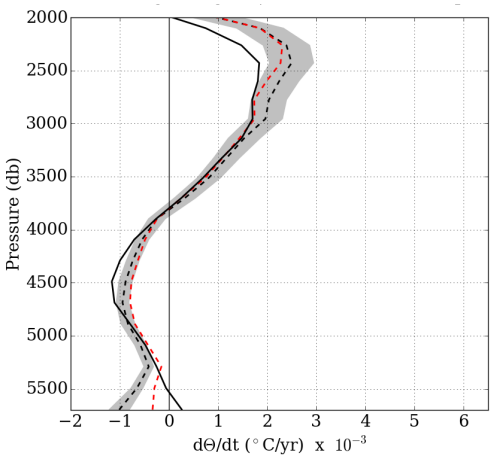
(c) NEMO025 (ERA Interim) - incl. OVIDE



(d) NEMO025 (ERA Interim) - excl. OVIDE



(e) NEMO025 (CORE2) - incl. OVIDE



(f) NEMO025 (CORE2) - excl. OVIDE

Figure 2.14: For each model, depth profiles of $d\Theta/dt$ calculated from the NE Atlantic basin average temperature and from the weighted average $d\Theta/dt$ of the sections (a)/(c)/(e) A16N,A05E,OVIDE and (b)/(d)/(f) A16N and A05E (OVIDE removed). The shaded area indicates the region between $\pm 2\sigma$ of 1000 alternative pseudo sections using randomly selected dates between ± 1 year of the actual cruises.

uncertainty (Figure 2.14 b), there is large reduction in the difference between the pseudo section estimate and the model truth from sections above 3500 m, so temporal uncertainty does reduce in this simulation. There is a close match between the trend estimated by pseudo sections and the basin truth, however, the section truth estimate provides a warming trend of less than a third of the magnitude of the basin truth trend at 2000 m. Since the basin truth and pseudo section trends are very similar at 2000 m, this means that temporal and spatial uncertainty are of approximately the same magnitude but acting in opposite directions. However, the width of the shaded area indicates that temporal uncertainty could be smaller or larger if different dates were used. Compared to including OVIDE, the spatial uncertainty without OVIDE decreases in the 2500 - 3000 m range but increases in the 2000 - 2500 m range (Figure 2.14 b).

For the NEMO025 (ERA Interim) simulation (Figure 2.14 c), estimates from pseudo sections, the model truth from sections and the basin model truth are in reasonable agreement below 3500 m. Above 3500 m pseudo sections strongly overestimate (between 2500-3000 m) or underestimate (above 2500m) the model truth on sections and basin truth (Figure 2.14 c). This indicates large temporal uncertainty (which could not be much reduced by small alterations to the cruise timings) compared to a smaller but not insignificant spatial uncertainty above 3500 m (Figure 2.14 c).

When OVIDE is removed (Figure 2.14 d), the trends from pseudo sections and model truth from sections are very similar for the NEMO025 (ERA Interim) simulation, indicating that the temporal uncertainty exhibited when all sections are included is largely due to OVIDE (and the lack of sampling along it between 1990-2002). Instead there is a consistent offset due to spatial undersampling which becomes very large above 3000 m; at 2500 m the trend estimated from sections is twice the magnitude of the basin truth trend (Figure 2.14 d).

The NEMO025 (CORE2) simulation (Figure 2.14 e) has similar trends to the ERA Interim simulation, but with a larger uncertainty relating to the limited spatial sampling of the basin between 2000-3500 m. When OVIDE is excluded there still exists larger temporal related uncertainty between 2000-3000 m than in the ERA Interim simulation (Figure 2.14 f). There are also strong deviations in the trend estimates below 5000 m (Figure 2.14 e / f).

The analysis here can be summarised for each simulation by volume weighting the trends at each depth to get an overall average trend below 2000 m (Table 2.2), and comparing the volume weighted average of the pseudo section trend with the basin average trend to evaluate how observational style sampling under- or over-estimates the trend in the basin. Due to the increased volume weighting in the shallower layers, overall each simulation shows a warming below 2000 m, and the basin truth is overestimated when using pseudo section data (Table 2.2).

When OVIDE is included, the pseudo section trend can be up to twice the magnitude of the basin truth trend (NEMO025 (CORE2)), and when OVIDE is excluded, up to ten times (for NEMO025 (ERA Interim) since the basin truth trend is near zero) (Table 2.2). Using the difference between the two trends (Table 2.2) shows that for NEMO083 and NEMO025 (CORE2) the pseudo section overestimate is larger when OVIDE is included (increased temporal uncertainty), but for NEMO025 (ERA Interim) the overestimate is much larger without OVIDE (spatial uncertainty increases). These differences may suggest that warming/cooling is occurring in different parts of the basin in the different simulations so that it is captured to different extents by hydrographic sections, and OVIDE in particular.

Table 2.2: Overestimation of average trend below 2000 m by pseudo sections

Model	Model truth trend °C/year * 10^{-3}	With OVIDE	Pseudo section trend °C/year * 10^{-3}	Overestimate by pseudo sections °C/year * 10^{-3}
NEMO083	1.50	With	2.09	+ 0.59
		Without	1.70	+ 0.20
NEMO025	0.04	With	0.05	+ 0.01
(ERA Interim)		Without	0.50	+ 0.46
NEMO025	0.62	With	1.25	+ 0.63
(CORE2)		Without	0.96	+ 0.34

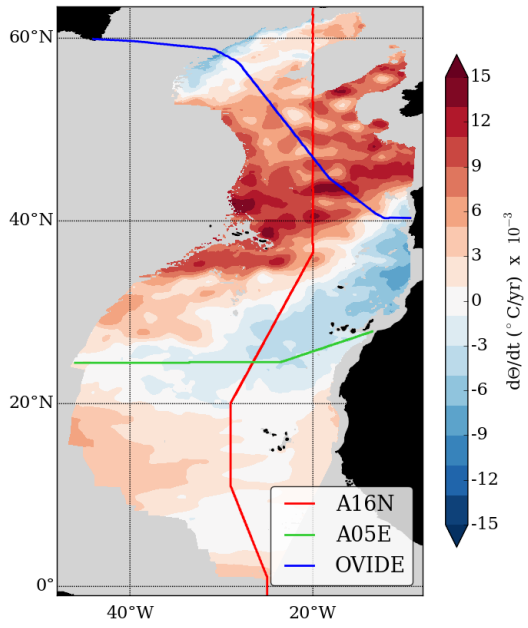
2.5.6 Further analysis of spatial variability

To investigate the spatial variability of trends in the basin, spatial maps of the temperature trends are presented at one thousand metre intervals in the deep and

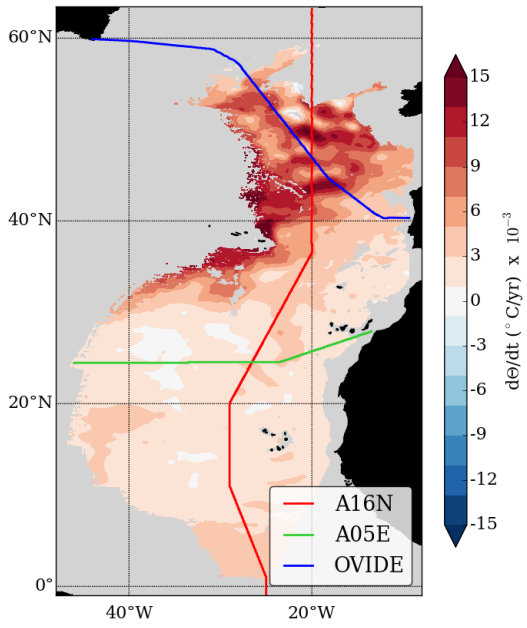
abyssal ocean. A map of the temperature trend during the period 1990 - 2010 at 2000 m in NEMO083 (Figure 2.15 a) shows that the northern part of the basin exhibits the strongest warming whilst the strongest cooling exists to the north-east of the A05E section, perhaps linked to changes in formation of Mediterranean Water. From eye the sections appear to capture spatial variability reasonably; Figure 2.14 (a) shows that the pseudo sections undersample the overall warming trend in the basin only very slightly at this depth. At 3000 m (Figure 2.15 b), there is warming throughout the basin, intensified in the north and at 4000 m and 5000 m (Figures 2.15 c and d) most of the basin is cooling.

In contrast to the NEMO083 simulation, in the NEMO025 (ERA Interim) simulation there is cooling in the far north at 2000 m and 3000 m mostly between 50°N and 60°N, with warming water further south (Figures 2.16 a / b). The patterns of temperature trends are very different across the basin in the NEMO025 (ERA Interim) simulation compared to NEMO083, explaining why including OVIDE has different consequences in different simulations (increases overestimate by pseudo sections in NEMO083, reduces in NEMO025 (ERA Interim)). As for NEMO083, Figures 2.16 (c) and (d) show that there is cooling at 4000 m/5000 m in NEMO025 (ERA Interim), although the locations where there is warming differ between the simulations.

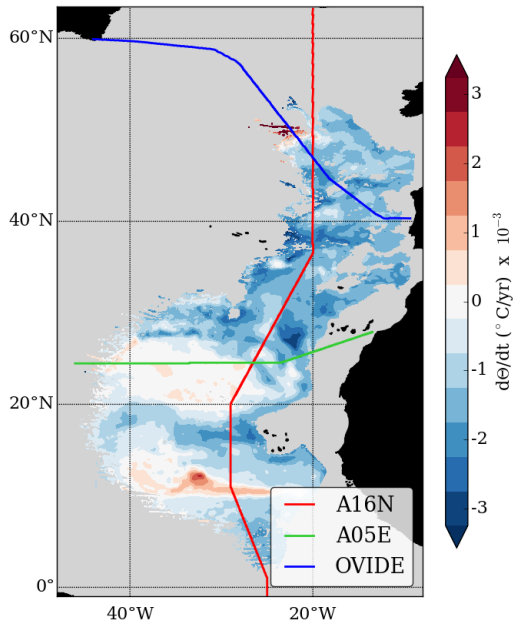
The observational study of Desbruyères *et al.* (2014) identified a warming between 1000 - 3000 m in the subpolar gyre; the NEMO083 simulation is consistent with the observational study but the NEMO025 simulations are not. The trends across the subpolar gyre are spatially consistent in each simulation so this discrepancy between the NEMO025 simulations and the observations is not due to spatial sampling limitations of the observations, although temporal uncertainty could play a role, particularly since the subpolar gyre is sampled by OVIDE (and A16N). However, the warming and cooling patterns in the respective simulations were found on section A16N when analysed alone, and this work has shown temporally subsampled data along A16N to give a good representation of the trends from 5 day mean data. In addition, the higher spatial and temporal resolution Argo data shows that there was a warming trend at 2000 m in the subpolar gyre (Desbruyères *et al.*, 2014). Together, this suggests that the trends in the NEMO083 simulation are more likely to be representative of the real world than those in the NEMO025 simulations.



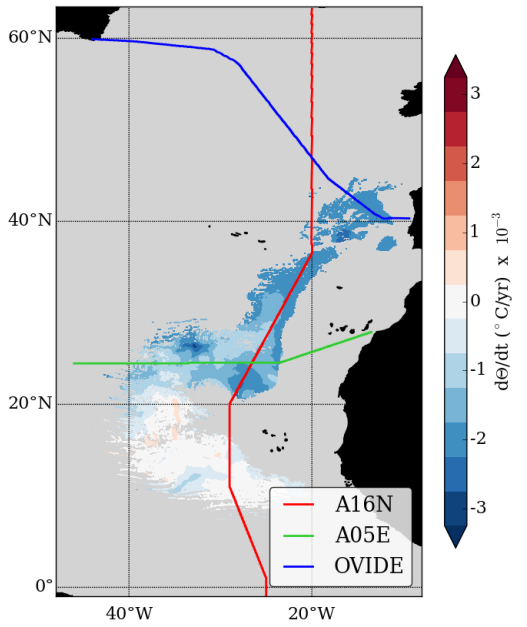
(a) dΘ/dt at 2000 m



(b) dΘ/dt at 3000 m

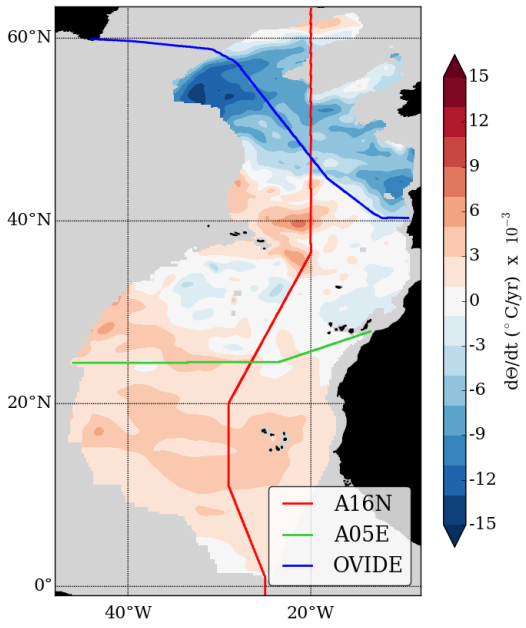


(c) dΘ/dt at 4000 m

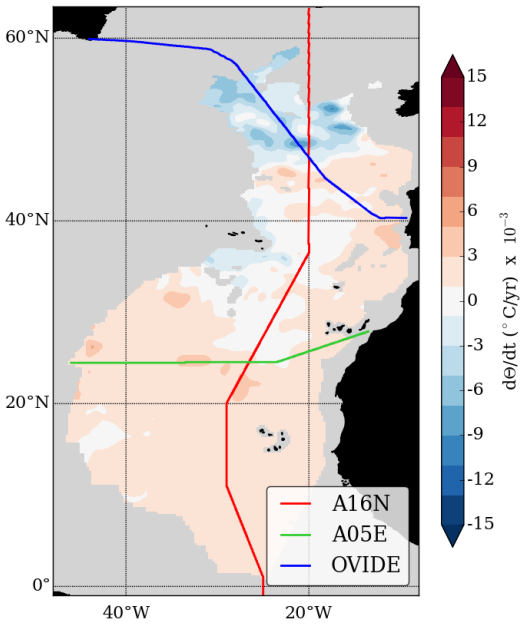


(d) dΘ/dt at 5000 m

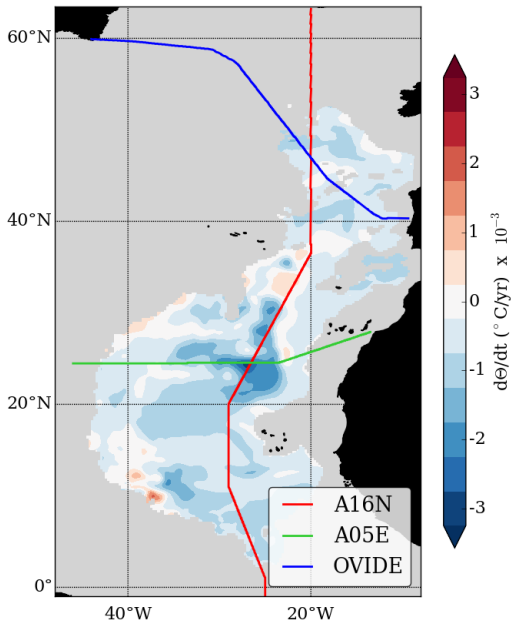
Figure 2.15: Spatial maps of linear temperature trend (dΘ/dt) between 1990-2010 at 2000 m, 3000 m, 4000 m and 5000 m (NEMO083).



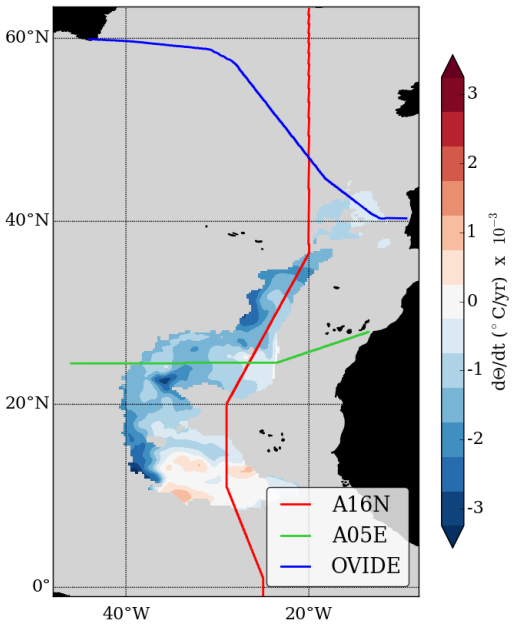
(a) dΘ/dt at 2000 m



(b) dΘ/dt at 3000 m



(c) dΘ/dt at 4000 m



(d) dΘ/dt at 5000 m

Figure 2.16: Spatial maps of linear temperature trend (dΘ/dt) between 1990-2010 at 2000 m, 3000 m, 4000 m and 5000 m (NEMO025 (ERA Interim)).

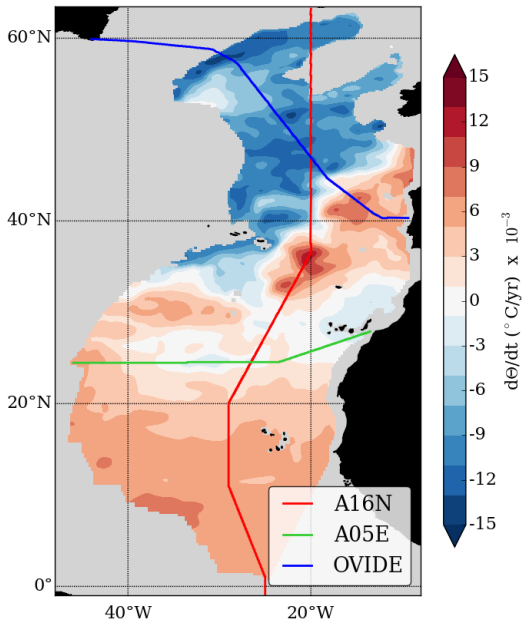
Broadly similar magnitudes and spatial patterns are observed in the NEMO025 (CORE2) simulation (Figure 2.17) to the NEMO025 (ERA Interim) simulation, though there are some differences in magnitude e.g., areas of stronger warming at 2000 m and regions of intensified cooling around the boundaries in NEMO025 (CORE2) (Figure 2.16).

Overall the differences in spatial patterns of trends across the basin influence the spatial uncertainty calculated in this analysis. The large differences in spatial patterns between models may mean that the real world scenario will look something like one of the possibilities outlined here, although as previously discussed, the NEMO025 simulations are less likely to be realistic than the NEMO083 simulation in the subpolar gyre. However, this uncertainty in the spatial uncertainty provides a case for additional spatial coverage of the deep ocean in this basin.

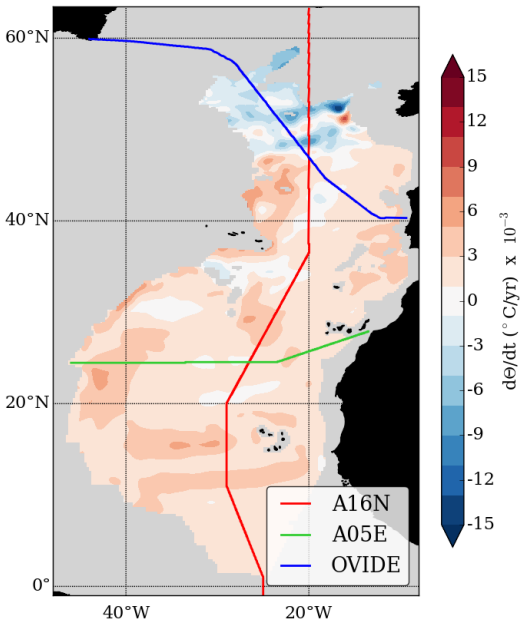
When considering the locations of the sections, it is possible there is oversampling in the northern-most part of the basin. There may be merit to trend analysis excluding the OVIDE section as a simple solution to both reduce temporal uncertainty and potential spatial oversampling in the north of the basin, but the differences in response in the different simulations show that excluding OVIDE may increase or reduce the accuracy of basin estimates from hydrographic sections depending on the pattern of temperature change across the basin. Another scheme for developing a section based estimation could include adjusting the distance weighting for sections by reducing the weighting for parts of the sections that are deemed oversampled (that is, utilising optimal interpolation techniques to integrate the section output into a potentially improved estimate).

2.5.7 Sea level rise

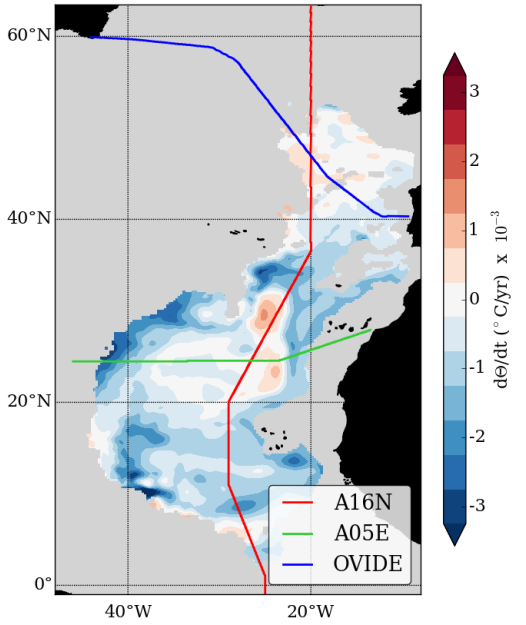
To further consider the need for additional sampling in the deep ocean, Figure 2.18 shows the cumulative (from the seafloor) thermosteric component of sea level rise that results from the change in temperature. This is calculated from the temperature trends from pseudo sections, section truth and basin truth data, with temporal uncertainty relating to the precise timing of pseudo sections highlighted by grey shading representing 2σ of the results from one thousand sets



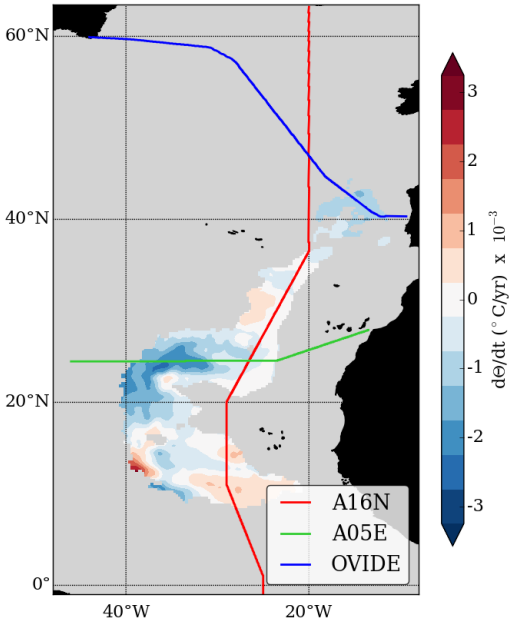
(a) dΘ/dt at 2000 m



(b) dΘ/dt at 3000 m



(c) dΘ/dt at 4000 m



(d) dΘ/dt at 5000 m

Figure 2.17: Spatial maps of linear temperature trend (dΘ/dt) between 1990-2010 at 2000 m, 3000 m, 4000 m and 5000 m (NEMO025 (CORE2)).

of alternative pseudo sections.

For NEMO083 the cumulative thermosteric contribution to sea level rise goes from negative below 3500 m to positive above 3000 m; the increased warming trend calculated from pseudo sections means the crossover from a sea level fall to a rise occurs at a deeper depth than when using section or basin truth data (Figure 2.18 a). When OVIDE is excluded, the same general pattern occurs but the sea level rise goes positive at a shallower depth, which reflects the slightly weaker warming trend captured without OVIDE at around 3000 m (Figure 2.18 b). At 2000 m the cumulative sea level rise from pseudo sections is just under 0.75 mm yr^{-1} but the truth from sections is 0.6 mm yr^{-1}, so the temporal uncertainty using the true dates of hydrographic sampling can be expressed by a positive bias of around 0.15 mm yr^{-1} (Figure 2.18 a). However, the magnitude of the shaded area shows that the temporal variability itself could be subject to change with different sampling times. The difference between the truth from section and basin truth (the spatial uncertainty) is slightly smaller than the magnitude of the temporal uncertainty at 2000 m (Figure 2.18 a). When OVIDE is excluded the magnitude of temporal uncertainty remains about the same, but the spatial uncertainty at 2000 m halves, to around 0.05 mm yr^{-1} (Figure 2.18 b). For NEMO083 the uncertainties tend to grow as the sea level rise is cumulatively summed up the water column.

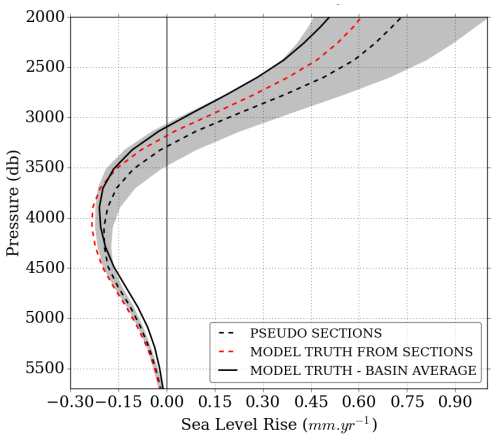
In the NEMO025 (ERA Interim) simulation the sea level rise is negative until close to 2000 m, depending on which data type is used (Figure 2.18 c). In Figure 2.18 (c) (including OVIDE) the uncertainties are larger between 3000 m and 2000 m than at 2000 m, due to biases for each uncertainty changing sign. Spatial uncertainty is around twice the magnitude of temporal uncertainty (Figure 2.18 c). For this simulation cumulative sea level rise from the deep ocean is around 0, but spatially limited data (section truth) suggest it is slightly negative, whilst spatially and temporally limited data (pseudo sections) suggest it is slightly positive (Figure 2.18 c). The effect of excluding OVIDE in this simulation is to increase the spatial uncertainty such that there is a positive bias of 0.15 mm yr^{-1} at 2000 m when using section truth or pseudo section data compared to the basin truth (Figure 2.18 d). Although there is uncertainty due to the timing of pseudo sections (grey shading), the pseudo sections using the actual hydrography timings result in virtually no bias due to temporal uncertainty at

2000 m in this simulation when OVIDE is excluded (Figure 2.18 d).

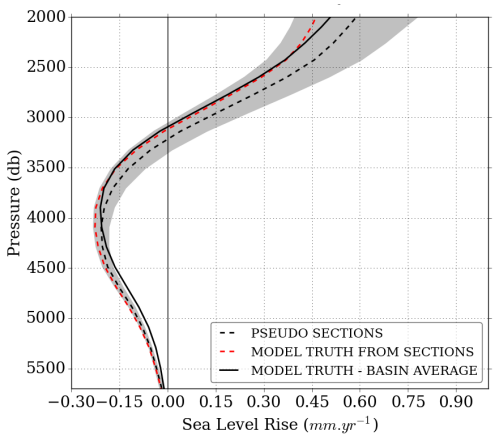
The greater warming in the NEMO025 (CORE2) simulation results in a positive value of the deep ocean component of thermosteric sea level rise (just over 0.2 mm yr^{-1}) (Figure 2.18 e). Spatial uncertainty is low compared to the temporal uncertainty when OVIDE is included; the bias due to spatial uncertainty is around -0.05 mm yr^{-1} but the bias due to temporal uncertainty is nearly +0.3 mm yr^{-1} (Figure 2.18 e). The red line deviates significantly out of the shading indicating that significant spatial uncertainty would exist in the upper deep ocean regardless of any slight adjustment of pseudo section timings. When OVIDE is excluded the bias due to spatial uncertainty increases to nearly 0.1 mm yr^{-1} but the bias due to temporal uncertainty reduces to less than a tenth of what it was when OVIDE is included (Figure 2.18 f).

The uncertainties quantified here highlight the need for additional spatial and temporal sampling to reduce the magnitude of these uncertainties and better constrain estimates of future thermosteric sea level rise. As the models disagree on which uncertainty is most important (due to different spatial patterns of temperature trends across the basin), it is necessary to observe the deep ocean with increased spatial and temporal frequency. With more information about real-world ocean variability it may then be possible to identify which uncertainty is a priority to reduce further.

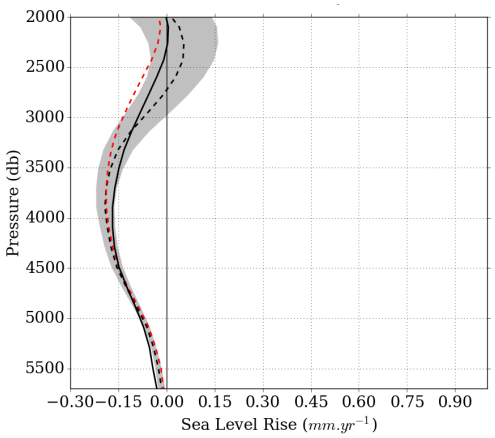
Purkey and Johnson (2010) calculated thermosteric sea level rise in the NEA of -0.07 $\pm$0.06 mm yr^{-1} below 4000 m, which is slightly less than the magnitude estimated in the NEMO025 model simulations, and somewhat less than that calculated for the NEMO083 simulation (-0.20 mm yr^{-1}) (Figure 2.18). The thermosteric component of sea level rise can often be partially compensated by the halosteric component (which has not been calculated here) so the magnitude of the total of the thermosteric and halosteric components of sea level rise may vary from (and may be less than) the thermosteric component of sea level rise.



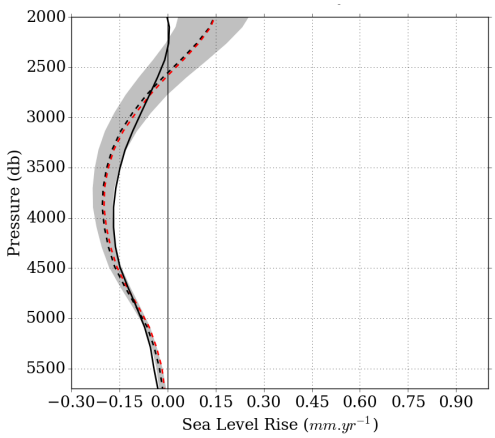
(a) NEMO083 - including OVIDE



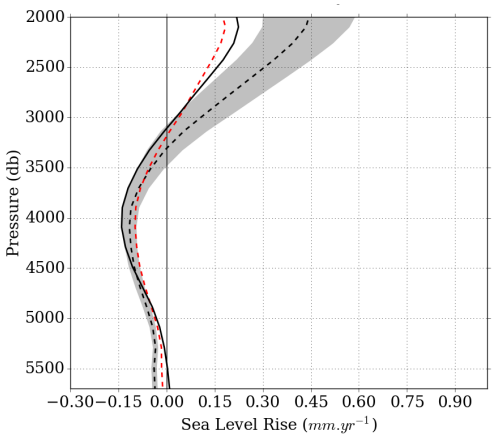
(b) NEMO083 - excluding OVIDE



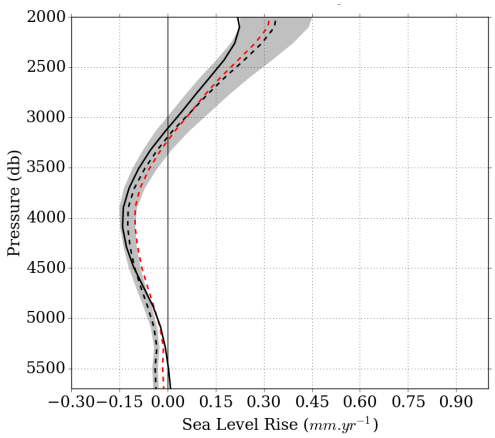
(c) NEMO025 (ERA Interim) - incl. OVIDE



(d) NEMO025 (ERA Interim) - excl. OVIDE



(e) NEMO025 (CORE2) - incl. OVIDE



(f) NEMO025 (CORE2) - excl. OVIDE

Figure 2.18: For each model, vertical profiles of the thermosteric sea level change (mm yr^{-1}) (integral from seafloor) that would occur in the NE Atlantic basin as a consequence of the temperature trends between 1990-2010 (Figure 2.14). Shading indicates temporal uncertainty in the pseudo section estimate ($\pm 2\sigma$ of 1000 alternative pseudo sections using randomly selected dates between ± 1 year of the actual cruises).

2.6 Conclusions

Ocean temperature variability is an important component of the Earth's energy balance and surface climate variability including the thermosteric component of sea level rise. Scientific understanding of the upper and intermediate ocean (to 2000 m) variability is aided by regular sampling by the upper ocean Argo array. In contrast the ocean below 2000 m is currently only sampled by scientific research ships repeating hydrographic sections with a frequency typically between 2 - 8 years. This chapter provides an analysis of how representative basin estimates of deep ocean temperature trends from temporally and spatially limited measurements might be of the whole NEA basin. Observations from the year 1990 show cooling in the abyssal NEA (below 4000 m) but warming in the intermediate and upper deep ocean. This cooling is interesting and significant when considering global climate variability given the proximity of the NEA to Northern Hemisphere deep water formation regions and the Atlantic Meridional Overturning Circulation.

In this work, temperatures and their trends for the complete model output in the NEA basin are compared to those from pseudo sections, a representation of limited frequency observational sampling in numerical model simulations. Temperature itself along the three sections varies over time, with each section representing a different part of an oceanographically variable basin. However, when the temperature data from the three pseudo sections are averaged, a good estimate of the variability of temperature change in the basin is made, with an offset which is often fairly consistent through time, potentially providing opportunity to reduce spatial bias in temperature sampling through a bias adjustment.

Linear temperature trends derived from pseudo sections largely capture the main features of temperature trends along sections. The three NEMO model simulations used do not always agree on the magnitude of trends along sections, nor do they exactly reflect the trends generated from observations. Examination of the temporal uncertainty on each section finds that it is relatively small for section A16N (greatest in magnitude between 2000 m - 3000 m) compared to A05E, where it is sufficiently low to enable the right features (but wrong

magnitude) of temperature change on the section to be captured. Temporal uncertainty is largest on OVIDE, resulting in completely different trends obtained when sampling is limited in time, but sampling on OVIDE was restricted to the last 9 years of the 21 year study period.

The magnitude of temporal and spatial uncertainties associated with restricted sampling across the basin (both in space and time) varies between simulations but in general when all three sections are used to estimate the basin average, the magnitude of temporal uncertainty is greater than the magnitude of spatial uncertainty. However this temporal uncertainty is subject to a change in magnitude depending on the exact timings of the pseudo sections, demonstrating that the uncertainty is sensitive to the precise timings of the occupations of each section. When the section OVIDE is excluded (based on the temporal uncertainty known to be introduced by this section) each simulation has a different response. Total uncertainty reduces in two simulations when OVIDE is excluded, but increases in one due to the increase in spatial uncertainty in that simulation. The different responses between simulations reflect the different spatial patterns of temperature trends across the basins, and motivates increased spatial resolution sampling to determine the true spatial patterns of temperature change across the deep ocean basin. This would also help provide further guidance to model developers about the abilities of models to capture deep ocean temperature trends and their changes.

The total uncertainty when using the three sections in the NEA to estimate the basin trends leads to an overestimate of the temperature trend in the basin between 2000 m - seafloor that is between 25% and 100% depending on the simulation. Therefore although the main vertical features of the spatially averaged basin temperature trends may be captured by pseudo sections, there is a clear need for additional sampling in the basin at both higher spatial and temporal resolution to accurately capture the magnitude of temperature trends. The error in the temperature trends is equivalent to a overestimate of the sea level rise contribution from the deep basin of up to 0.3 mm yr^{-1}.

This chapter has demonstrated that in the NEA different model simulations, using different forcing products and grid resolutions but the same underlying code, provide a good indication of ocean heat content trends. However additional

sampling is likely required to reduce temporal and spatial uncertainties to ensure that the correct magnitude of ocean heat content change is captured. This immediately generates questions regarding whether the same is true of other basins in the global ocean, some of which (e.g., the Southern Ocean) are much less well sampled spatially than the NEA. For climate system understanding and prediction of future change, it is important to know whether both the global integral and regional variation in ocean heat content change are captured well, so the next chapter extends the methodologies developed here into a global analysis.

CHAPTER 3

Did hydrographic sampling capture global and regional deep ocean heat content trends accurately between 1990-2010?

3.1 Abstract

Estimates of heat content change in the deep oceans (below 2000 m) for the last thirty years are obtained from temperature measurements made by hydrographic survey ships with specialist equipment occupying the same tracks across an ocean basin approximately every 5+ years. Measurements may not be sufficiently frequent in time or space to allow accurate evaluation of total ocean heat content (OHC) and its rate of change. It is widely thought that additional deep ocean sampling will also aid understanding of the mechanisms for OHC change on annual to decadal timescales, including how OHC varies regionally under natural and anthropogenically forced climate change. Here a 0.25° ocean model is used to investigate the magnitude of uncertainties and biases that exist in estimates of deep ocean temperature change from hydrographic sections due to their infrequent timing and sparse spatial distribution during 1990 - 2010. Biases in the observational data may be due to lack of spatial coverage (not enough sections covering the basin), lack of data between occupations (typically 5-10 years apart) and due to occupations not closely spanning the time period of interest. Between 1990 - 2010, the modelled biases globally are comparatively small in the abyssal

ocean below 3500 m although regionally certain biases in heat flux into the 4000 - 6000 m layer can be up to 0.05 Wm^{-2}. Biases in the heat flux into the deep 2000 - 4000 m layer due to either temporal or spatial sampling uncertainties are typically much larger and can be over 0.1 Wm^{-2} across an ocean. Overall, 82% of the warming trend below 2000 m is captured by observational-style sampling in the model. However, at 2500 m (too deep for additional temperature information to be inferred from upper ocean Argo) less than two thirds of the magnitude of the global warming trend is obtained, and regionally large biases exist in the Atlantic, Southern and Indian Oceans, highlighting the need for widespread improved deep ocean temperature sampling. In addition to bias due to infrequent sampling, moving the timings of occupations by a few months generates relatively large uncertainty due to intra-annual variability in deep ocean model temperature, further strengthening the case for high temporal frequency observations in the deep ocean (as could be achieved using deep ocean autonomous float technologies). Biases due to different uncertainties can have opposing signs and differ in relative importance both regionally and with depth revealing the importance of reducing all uncertainties (both spatial and temporal) simultaneously in future deep ocean observing design.

3.2 Introduction

The deep (2000 - 4000 m) and abyssal (4000 - 6000 m) oceans are vast reservoirs for heat in the climate system, but they are relatively under-sampled both in time and in space when compared to the atmosphere and upper ocean. Monitoring of ocean heat content (OHC) change in each of Earth's heat stores is required to properly evaluate the status and predict the future of the climate system, and on climate relevant timescales the oceans are by far the dominant sink for heat accumulating in the climate system (Von Schuckmann *et al.*, 2016). Limited knowledge of deep ocean temperature and its behaviour hinders solving of open questions about the pathways of heat through the oceans and how they change over time (e.g. Xie, 2016). Rapid climate change in response to sustained anthropogenic greenhouse gas emissions may lead to potentially severe societal consequences, with OHC change particularly important as a contributor for sea level rise and its regional variations (e.g., Church and White, 2011; Church *et al.*,

2013b) and surface heat fluxes that can lead to extreme weather events (e.g., Grist *et al.*, 2015).

Below 2000 m temperature observations from repeated hydrographic sections enable the estimation of global and regional trends in deep OHC. Purkey and Johnson (2010) analysed hydrographic section data from the time period 1990-2010 providing the first assessment of decadal deep ocean temperature change. Purkey and Johnson (2010) identified statistically significant warming through many regions of the deep oceans, particularly the Southern Ocean.

Hydrographic sections provide sparse spatial coverage of the global oceans and typically sections are repeated every few years, so temporal and spatial sampling frequencies are low. Systematic biases arising from infrequent or inhomogeneous sampling in time and space may potentially exist. This work estimates the magnitude of biases and uncertainties associated with trends from hydrographic sampling by repeating the analysis of Purkey and Johnson (2010) using state-of-the-art global ocean model output. The same temporal and spatial restrictions on data are not present when using model data, and so historical global and regional trend estimates using data subsampled from the model along hydrographic sections can be compared to global and regional trends using spatially and temporally complete model data.

The objective of the work is to assess the adequacy of hydrography for constraining global and regional trends in deep OHC during 1990 - 2010, and to provide the observational community with a guide as to the level of confidence that can be attributed to observational studies of deep OHC change from hydrographic sections. The following specific questions are addressed: (a) Do estimates of deep OHC change from hydrographic sections during 1990 - 2010 represent global and regional trends accurately? (b) If not, what are the reasons for inaccuracy and do they relate to temporal or spatial biases in the sampling? (c) What magnitude do these biases have? (d) Do temporal and spatial biases vary regionally? (e) What implications do the results have for future deep ocean observing?

3.3 Methodology

This chapter extends and further develops and extends the methodology laid out in the previous chapter, applying the principles globally. Instead of using a multi-model study approach, this work provides a more detailed study of the different uncertainties in hydrography sampling across the whole globe, but is provided for one model simulation. This limits the confidence with which the magnitude of uncertainties can be estimated; the previous chapter showed that in the North-East Atlantic the magnitude of estimates varied between simulations although the same overall message was obtained from each simulation. However, regional differences in uncertainties and an estimate of total uncertainty using hydrographic style subsampling in a model field is novel work, and one estimate alone provides a useful scientific contribution.

3.3.1 Choice of model

An introduction to the NEMO model and ORCA grid is provided in Section 2.3. In this chapter, a simulation of NEMO ORCA025 is used, N025_DFS_01, which has a resolution of 1/4 ° and was surface forced with the Drakkar Forcing Set (DFS) version 5.2, which dictates surface air temperature, winds, humidity, surface radiative heat fluxes and precipitation (Brodeau *et al.*, 2010). A 1/4° simulation was selected rather than a higher resolution ORCA 1/12° simulation in part due to the computational advantages for a global analysis. In addition the available 1/12° simulation at commencement of this work (N001) had clear biases in deep ocean temperature change associated with unrealistic circulation in the Southern Ocean and a consequently unrealistically large northward transport of Antarctic Bottom Water.

Compared to other available 1/4° simulations, this simulation was chosen because it used a consistent forcing product from initial spin up and covered the 1990-2010 period of interest. The simulation was initialised in 1958 so the ocean had over 30 years prior to the analysis period to spin up. Optimal spin up length for ocean model output is a compromise between being long enough to reduce the effect of large initial adjustments but not so long that ocean properties drift too

far from observed climatological values. Computing power availability often restricts the length of simulations, as well as the availability of realistic historical forcing products.

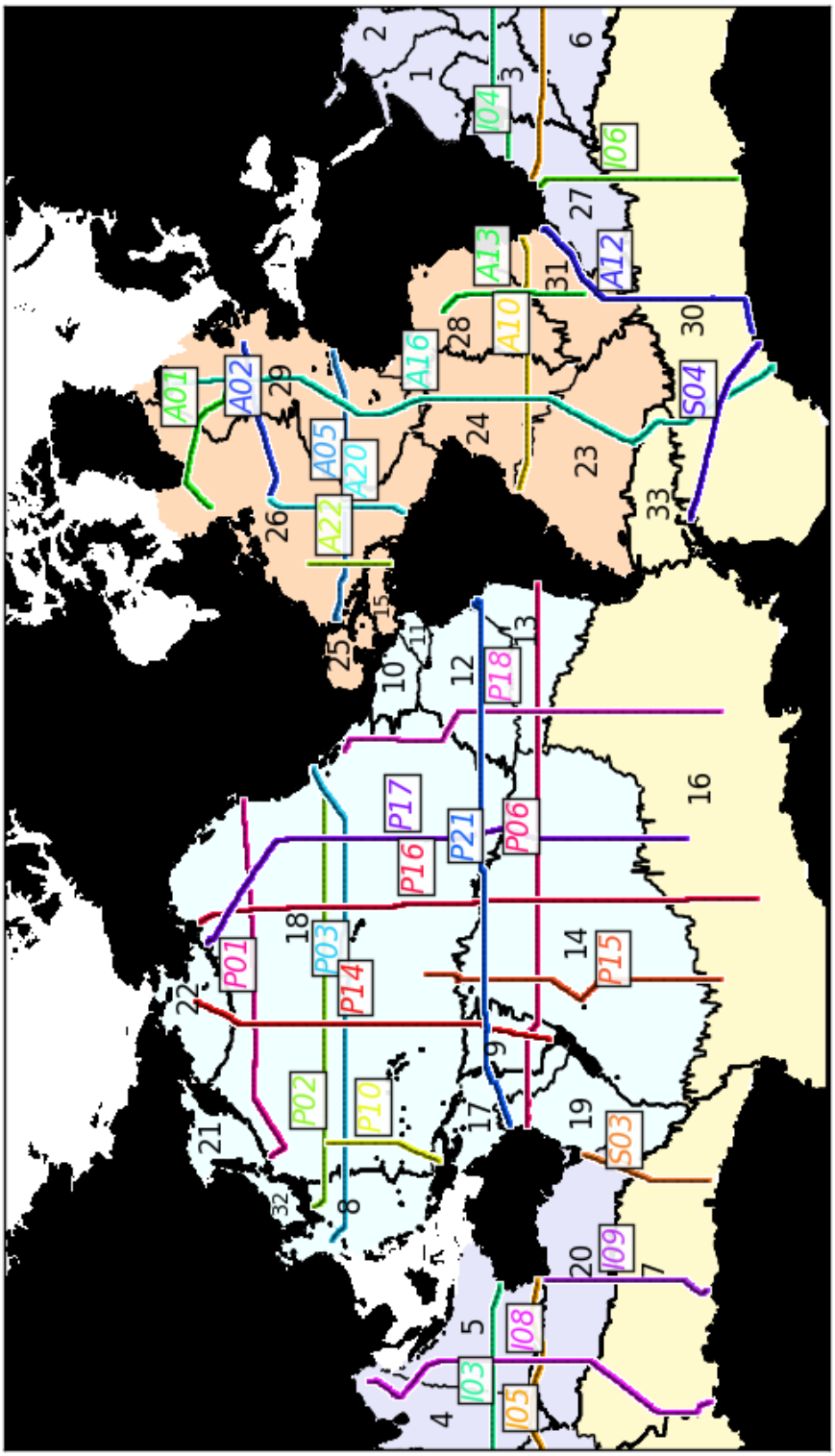


Figure 3.1: Map of hydrographic basins in the global oceans, as defined and numbered by Sarah Purkey, and hydrographic sections occupied at least twice during the period 1980 - 2010 and used in an observational analysis of deep OHC change by Purkey and Johnson (2010). Block colours indicate which ocean (Atlantic, Southern, Indian or Pacific) each numbered basin is part of.

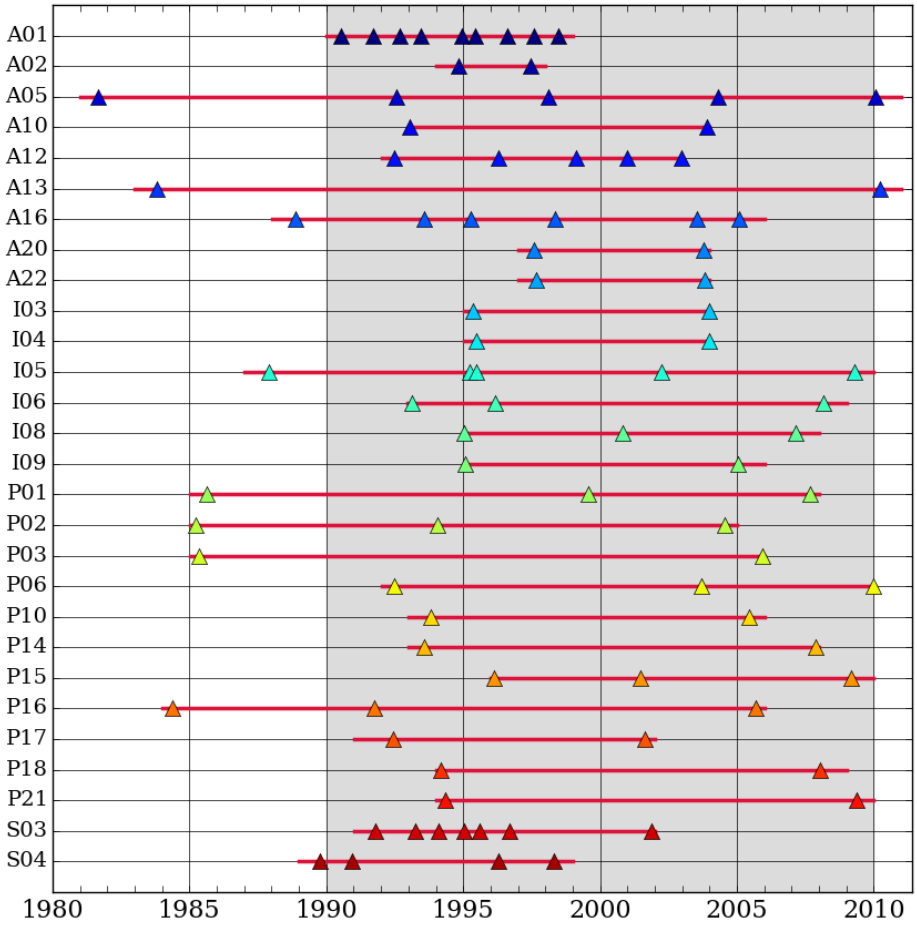


Figure 3.2: Hydrographic sections occupied at least twice during the period 1980 - 2010. Triangles denote the times of 'pseudo sections': model output along the section is restricted to the times when sections were actually occupied by hydrographic survey ships. Red lines indicate the period of 'frequent sections': 5 day mean model output along the section is extracted from the start of the year of the first pseudo section to the end of the year of the last pseudo section. The grey shading represents the time interval over which 'section truth' and 'basin truth' data are extracted (1990 - 2010 inclusive). 'Section truth' restricts model output along a section to all 5 day mean temperature data between 1990 and 2010, while the 'basin truth' uses the 5 day mean basin average temperature data for the same time period.

3.3.2 Hydrographic section choice

To ensure the method followed replicates that of the existing observational analysis of decadal heat content change below 2000 m as closely as possible, the same repeat hydrographic sections (locations and timings) as Purkey and Johnson (2010) are used. The rationale used when selecting sections is outlined extensively in Purkey and Johnson (2010). A map showing the locations of the hydrographic sections is provided in Figure 3.1 and the observational timings are denoted by triangles in Figure 3.2. Note that although the aim is to evaluate how well sections represent heat content change between 1990-2010, in following the observational methodology, section occupations between 1980-1990 are included and used to inform the early part of the temperature change before the first occupation in the 1990's. The temperature change generated by these (and the later) occupations is considered to be representative of the 1990-2010 period. The full extent of each section is assumed to be measured during every occupation; in reality not all of a section is sampled completely every occupation due to constraints on ship availability or delay (for example due to adverse weather). In line with the assumptions made by the observational analysis, every model gridbox on a section is sampled simultaneously at the midpoint in time of a ship's occupation. In reality, occupations of a section can take months.

3.3.3 Hydrographic section coordinates

The coordinate files generated in the observational analysis by Purkey and Johnson (2010) are derived from the locations of the CTD casts by interpolation onto 'an evenly spaced standard 2' longitudinal or latitudinal grid along the section trackline using a space-preserving piecewise cubic Hermite interpolant' (Purkey and Johnson, 2010). A netCDF tool developed by the NEMO consortium, find_coords, was used to find the nearest neighbour of each coordinate pair, with the output assessed to check that the distances between the original coordinates and the model grid coordinates are reasonably small, and that the new coordinate output represents the section correctly. The coordinates typically include duplicates of grid boxes because of their fine resolution relative to a 1/4° model grid. Duplicate coordinates are removed, so that each gridbox is

retained only once.

3.3.4 Masking sections by basin

Purkey and Johnson (2010) defined 33 geographic basins below 2000 m in the global oceans, and their locations are shown in Figure 3.1. Using the geographic basin coordinate information determined by Sarah Purkey, a mask was created for each of the basins on the ORCA025 grid. Using this mask, each whole section is split into partial sections that are entirely contained in one basin. From here on, each partial section entirely contained within a basin is referred to as a section, while the combined sections that represent the initial hydrographic section are referred to as whole sections.

3.3.5 Extracting sections from the model

The highest temporal frequency model data are used to ensure high accuracy in the temperature trends calculated, with 5 day mean data the highest frequency model output provided. For each coordinate pair along a section, every 5 day mean potential temperature model output between 1980 and 2010 (inclusive) is extracted at each vertical model level. Henceforth the potential temperature data output from the model is referred to simply as temperature data.

3.3.6 Section weightings

The distances along sections are calculated by finding the distances in the model between each x and y coordinate pair in the x and y directions, and then the Pythagorean theorem was used to calculate a single distance between each pair. This assumes x and y distances are perpendicular, which over the short distances between coordinate pairs should be approximately correct. As a method check, distances between the coordinate pairs are calculated using the Haversine formula, which gives great-circle distances between points on a sphere.

The distances are then cumulatively summed to give the total length of the section. A weighting for each gridbox along section (later used to combine the data along the section into a section average) is calculated using the distances between the points on the section. The first and last points along the section are given the weighting of half the distance to the closest point, while each intermediary point is given a weighting composed of the sum of the two half distances to each adjacent point. This ensures that the sum of the weightings remains equal to the total distance along the section.

The total distance along each section is converted into a weighting for the section at each vertical model level, which removes any distance along the section that is not ocean. This weighting is later used to combine data for the section with others in the same basin at each vertical model level to generate a temperature trend estimate for the basin.

3.3.7 Calculating trends on sections

Using the section data extracted from the model, three types of temperature data are generated by sub-sampling the section data in time at varying frequencies (as illustrated in Figure 3.2):

- Section truth - 5 day mean temperature data between 1990 and 2010 (inclusive). Trends calculated from this data are denoted by $\beta^{\text{section}}_{\text{truth}}$.
- Frequent sections - 5 day mean temperature data between the beginning of the first year in which a section was occupied and the end of the last year in which a section was occupied. Trends calculated from this data are denoted by $\beta^{\text{frequent}}_{\text{section}}$.
- Pseudo sections - the data only at the mid-point in time of each occupation, so that the data are representative of the temporal frequency of observations. Trends calculated from this data are denoted by $\beta^{\text{pseudo}}_{\text{section}}$.

The remainder of the analysis is repeated for each type of data. The temperature trends are calculated at each location on a section. In line with the analysis of

Purkey and Johnson (2010), if there are only two data points (that is, a pseudo section with only two occupations) the trend is calculated by dividing the difference in temperature by the difference in time between the occupations. Otherwise, for each grid box along the section and at each vertical model level, a straight line of best fit is applied through the potential temperature, θ (°C) at each timepoint, t (years). The gradient is the linear temperature trend, β:

$$\theta(t) = \beta \cdot t + c + \varepsilon(t) \quad . \tag{3.1}$$

c is the interception point with the y-axis, reflecting θ at $t = 0$, while $\varepsilon(t)$ is an error term derived from the residuals of the linear fit. Following the observational methodology, $\varepsilon(t)$ is not being used to calculate uncertainties.

As previously mentioned in Section 2.4.2, neglecting $\varepsilon(t)$ means a potentially large uncertainty is not considered, and this uncertainty is likely to be larger in the real world than model world. Although we do not calculate this error here, if desired, a standard error from residuals (SEr) at each point along the section could be calculated from the residuals of the trend fitting, R, and the number of data points used to calculate the trend, n:

$$SEr = \sqrt{\frac{\Sigma R^2}{n-2}} \quad . \tag{3.2}$$

This could be weighted by the along section distances between the gridbox midpoints, w_d, to calculate an error for the section (SEr_{section}) associated with the trend fitting:

$$SEr_{\text{section}} = \frac{\Sigma(w_d \cdot SEr)}{\Sigma(w_d)} \quad . \tag{3.3}$$

To return to the methodology used, for each vertical model level on a section, the average trend, $\hat{\beta}_{\text{section}}$ is calculated using weightings derived from the along section distances between the gridbox midpoints, w_g:

$$\hat{\beta}_{\text{section}} = \frac{\Sigma(w_g \cdot \beta)}{\Sigma(w_g)} \quad . \tag{3.4}$$

Since there are three types of section data, the following three average trends are calculated for each section at each vertical level: $\hat{\beta}^{\text{section}}_{\text{truth}}$, $\hat{\beta}^{\text{frequent}}_{\text{section}}$, $\hat{\beta}^{\text{pseudo}}_{\text{section}}$.

3.3.8 Calculating the error of the average trend on a section

Although this step of calculating the error of the average trend on a section can be done for each type of data, it is only calculated and presented for the pseudo section data so that the uncertainty calculation replicates the observational analysis as closely as possible.

Standard error, $\text{SE}_{\text{section}}$, at each vertical model level is calculated in the same way as for the observational analysis by Purkey and Johnson (2010). This is calculated by dividing the standard deviation of trends along the section, σ_{section}, by the square root of the degrees of freedom (DOF) for the section, $\text{DOF}_{\text{section}}$:

$$\text{SE}_{\text{section}} = \frac{\sigma_{\text{section}}}{\sqrt{\text{DOF}_{\text{section}}}} \quad . \tag{3.5}$$

The standard deviation is calculated by taking the square root of the variance of trends on a vertical model level along the section. The squared differences between the temperature trends, β, from the section mean temperature trend, $\hat{\beta}_{\text{section}}$, weighted by distances between the gridbox midpoints, w_g, gives the variance:

$$\sigma^2_{\text{section}} = \frac{\sum(w_g \cdot (\beta - \hat{\beta}_{\text{section}})^2)}{\sum w_g} \quad . \tag{3.6}$$

The effective degrees of freedom for a section, $\text{DOF}_{\text{section}}$, are calculated by dividing the length of the section, L_{section} (km), by 163 km (the mean horizontal

decorrelation length scale from observational data identified in Purkey and Johnson (2010)):

$$\mathrm{DOF}_{\mathrm{section}} = \frac{L_{\mathrm{section}}}{163} \quad . \tag{3.7}$$

This decorrelation length scale is an estimate of how far apart two temperature measurements must be in the horizontal before they are assumed to be independent. Following the observational methodology, when sampled regions become separated by seafloor topography at depth to the extent that the sampled regions are more than the decorrelation length scale apart, the data are assumed to be independent and contribute one DOF to the estimate.

Assuming a Student's t distribution (assumes data are normally distributed), the two sided 95% confidence intervals are estimated, with the endpoints calculated by:

$$\hat{\beta}_{\mathrm{section}} \pm \mathrm{SE}_{\mathrm{section}} \cdot t(0.975, \mathrm{DOF}_{\mathrm{section}}) \quad . \tag{3.8}$$

$\mathrm{SE}_{\mathrm{section}}$ is the standard error from averaging the trends along the section (Equation 3.5), whilst t is a function which gives the desired quantile (here 0.975) from the probability density function for a Student's t continuous random variable using the degrees of freedom for the section, $\mathrm{DOF}_{\mathrm{section}}$ (Von Storch and Zwiers, 2001).

3.3.9 Basin trends as estimated from sections and associated errors

Following the observational methodology of Purkey and Johnson (2010), the section data are used to estimate the temperature trend for a specific basin, $\hat{\beta}_{\text{basin estimate}}$. As there are three types of data for each section, the following analysis is repeated to obtain three basin estimates: $\hat{\beta}^{\text{section truth}}_{\text{basin estimate}}$, $\hat{\beta}^{\text{frequent section}}_{\text{basin estimate}}$ and $\hat{\beta}^{\text{pseudo section}}_{\text{basin estimate}}$.

At each vertical model level, the section trend, $\hat{\beta}_{\text{section}}$, for each section contained within the basin is weighted by the total distance along the section, w_s to contribute to the basin estimate:

$$\hat{\beta}_{\text{basin estimate}} = \frac{\sum(w_s \cdot \hat{\beta}_{\text{section}})}{\sum w_s} \quad . \tag{3.9}$$

To calculate the standard error for a basin, the standard deviation for the basin estimate, σ_{basin}, is divided by the square root of the degrees of freedom for the basin, $\text{DOF}_{\text{basin}}$:

$$\text{SE}_{\text{basin}} = \frac{\sigma_{\text{basin}}}{\sqrt{\text{DOF}_{\text{basin}}}} \quad . \tag{3.10}$$

The standard deviation for the basin estimate, σ_{basin}, is calculated by weighting the standard deviation for each section, σ_{section}, contained within the basin, by the total distances along each section, w_s:

$$\sigma_{\text{basin}} = \frac{\sum(w_s \cdot \sigma_{\text{section}})}{\sum w_s} \quad . \tag{3.11}$$

The DOF for the basin is determined by summing the DOF for each section:

$$\text{DOF}_{\text{basin}} = \sum \text{DOF}_{\text{section}} \quad . \tag{3.12}$$

95% confidence intervals are calculated from the standard error in the same way as they were for a section (Equation 3.8 in Section 3.3.8).

3.3.10 Average trend in basin - model truth

To validate the section based estimates, the basin truth data are extracted from the model output (data location described for the section data in Section 3.3.5).

The average temperature in a basin, $\hat{\theta}_{\text{basin}}$, is determined from the model output potential temperature, θ, at each 5 day mean, by weighting temperature in each grid box by the volume of the grid box, w_g at each vertical model level:

$$\hat{\theta}_{\text{basin}} = \frac{\sum(w_g \cdot \theta)}{\sum(w_g)} \quad . \tag{3.13}$$

A linear fit is applied to $\hat{\theta}_{\text{basin}}$ through time, t, and the gradient of that fit gives the model truth basin temperature trend, $\hat{\beta}_{\text{basin truth}}$:

$$\hat{\theta}_{\text{basin}}(t) = \hat{\beta}_{\text{basin truth}} \cdot t + c + \varepsilon(t) \quad . \tag{3.14}$$

c is the interception point with the y-axis, reflecting $\hat{\theta}_{\text{basin}}$ at $t = 0$, while $\varepsilon(t)$ is an error term derived from the residuals of the linear fit.

3.3.11 Ocean trends from section estimates and basin truth and associated errors

The data from each basin for the three section-based estimates and the true basin temperature trend are then weighted by the volume of each component basin at each vertical model level to get an average temperature trend for the ocean basins (that is, the Atlantic/Southern/Indian/Pacific Oceans) and for the total global ocean. The volume of the basin is calculated at each vertical model level by calculating the volume of each ocean gridbox contained in the basin and then summing them.

The standard error and confidence intervals for the ocean trends are calculated from the basin standard deviations (σ_{basin}) and DOFs ($\text{DOF}_{\text{basin}}$) in the same way that the basin values were determined from the section values (the methodology follows that in Section 3.3.9).

3.3.12 Heat fluxes

For each basin estimate or the basin truth, a heat flux, HF, into the deep and abyssal (2000 - 6000 m) or abyssal ocean (4000 - 6000 m) is calculated from the temperature change in the layer. This assumes that the heat change in the layer is a function only of the heat fluxes into the layer from the upper ocean rather than due to horizontal transport from adjacent basins or due to any geothermal influences within the basin. Note that the seafloor is always less than 6000 m in the model, which does not contain deep ocean trenches; therefore reference to 6000 m is equivalent to measuring to the bottom of the ocean. First the heat gain as a function of depth, $H(z)$ (Wm^{-1}), is calculated:

$$H(z) = \hat{\beta}_{\text{basin}} \,.\, A \,.\, \rho \,.\, c_p \;\; . \tag{3.15}$$

where $\hat{\beta}_{\text{basin}}$ is the basin average temperature trend at each vertical model level (converted to Cs^{-1}), A is the area of the model layer in that basin (m^2), ρ is the density of seawater (kgm^{-3}) and c_p is the specific heat capacity of seawater ($Jkg^{-1}C^{-1}$). The quantities used for ρ and c_p are depth varying estimates from the World Ocean Atlas climatology (www.nodc.noaa.gov/OC5/indprod.html, data for the basins provided by Damien Debruyères). β, A, ρ and c_p are therefore all functions of depth.

The downward heat flux through depth z_1 is calculated as follows:

$$HF = \frac{1}{A(z_1)} \int_{z_{max}}^{z_1} H(z) \;\; dz \;\; . \tag{3.16}$$

To calculate the heat flux into an intermediate layer such as the deep ocean layer from 2000 - 4000 m, the heat flux through the lower bound of the desired layer (4000 - 6000 m layer) is subtracted from the heat flux through the upper bound of the desired layer (2000 - 6000 m layer).

3.3.13 Uncertainties in the estimation of basin heat content change from hydrographic sections

There are many uncertainties surrounding the estimation of basin heat content change on decadal timescales using this method of averaging linear trends along hydrographic occupations. They include:

- **Uncertainty in the linear fit applied to data at each location through time to give a temperature trend.**

  A linear fit is the methodology used to assess the temperature change on decadal timescales in the observational methodology of Purkey and Johnson (2010). The total temperature change between two time points can be calculated by assuming a linear fit (as in the observational analysis) but when using high resolution data in time from the model is it useful to know if the temperature change over a period is likely to be accurately represented by a linear fit. In addition, when converting heat content change into a heat flux into a layer, the temperature change is then assumed to be linear in time. The appropriateness of a linear fit methodology to the temperature change in the model can be visually assessed by plotting temperature through time at certain depths at point locations, as an average on a section or in a basin. It is determined that linear trends are a reasonable approximation for basin temperature change in this model simulation (figures not shown).

  Uncertainty in the linear fit itself can be calculated from the residuals of the linear fit (the spread of the data either side of the line of best fit). However the observational analyses of decadal deep OHC do not consider this term when calculating their uncertainties. This uncertainty is not addressed further in this work.

- **Uncertainty in the averaging of trends along sections**

  There is variability in the trends along a section, but the observational analyses (whose methodology is closely followed) average the trends to give a single trend on the section. The variability is expressed by the standard

deviation, which can be converted into a standard error and then used to generate a confidence interval for a section, which reflects how well the mean on the section can be predicted. This is the only uncertainty that observational deep ocean analysis provides. Here the same uncertainty is quantified by the method described earlier in Sections 3.3.8, 3.3.9 and 3.3.11.

- **Uncertainty due to spatially limited sampling**

  Hydrographic sections cover a small geographic proportion of the basin. To quantify this spatial uncertainty, bias due to limited spatial sampling is calculated from the difference between the section truth temperature trend or heat flux into a basin (using 5 day mean data between 1990 - 2010 along sections) and that from the basin truth over the same time period. The calculation of the bias in trends obtained due to this uncertainty, along with the other calculated uncertainties, are summarised in Table 3.1.

- **Temporal uncertainty due to lack of sampling between occupations**

  Hydrographic sections are usually sampled at a frequency of approximately 5-10 years. The large periods of time between samples may mean that trends calculated from this temporally limited data may not be reflective of the trend between occupations. This uncertainty is quantified by subtracting the temperature trend or heat flux in a basin calculated from the pseudo sections in a model (same limited times and spatial coverage as observations) from that calculated from the frequent data (5 day mean data from the beginning of the first year of occupation to the end of the last year of occupation of a section, with the same limited spatial coverage as observations). This gives a bias in the pseudo section estimate that is directly attributable to this uncertainty (Table 3.1). The pseudo and frequent data do not span exactly the same period of time because the frequent data goes from the beginning of the first year of occupation to the end of the last year of occupation, so comparing them directly involves the assumption that a small extrapolation of the trend from pseudo-sections to the beginning/end of the first and final years of occupation does not

introduce further uncertainty. Choosing whole years only for the frequent data type also removes any uncertainty due to seasonal sampling bias that might be influencing the pseudo section trend. Therefore, the section truth and frequent section trends will not have any seasonal bias, though trends from pseudo sections might. The effect of potential seasonal biases is also evaluated in this work, with the methodology described later in Section 3.3.14.

- **Temporal uncertainty due to extrapolating trends**

  Very few hydrographic sections have occupations spanning 1990 - 2010. Therefore when comparing basin estimates from sections to the basin truth, making the assumption that they are representative of 1990 - 2010 effectively extrapolates the calculated trend from the time between hydrographic sections to the same time period as the basin truth. The trend could also be calculated over a longer period by including part of the 1980s, as some sections use occupations earlier than 1990, and that error is also considered part of this uncertainty. So although it is referred to in this work as uncertainty due to extrapolation, it simply includes any error due to the trend being calculated over a different time period to 1990 - 2010. As summarised in Table 3.1, this uncertainty can be quantified by a bias which is calculated by taking the difference between frequent section data estimates (using 5 day mean data on a section between the first and last year of occupation) and the section truth data estimates (5 day mean data on a section during the whole time period, 1990 - 2010).

A summary of differences that can be calculated between data types and the uncertainties that can be inferred from them is provided in Table 3.1. Isolating biases due to only one type of uncertainty enables comparison of the relative importance of the different uncertainties.

Table 3.1: Isolating specific uncertainties by subtracting data calculated by different methods

Uncertainty	Bias obtained by:
Spatially and temporally limited sampling	PSEUDO SECTIONS - BASIN TRUTH
Limited spatial sampling	SECTION TRUTH - BASIN TRUTH
No data between occupations	PSEUDO SECTIONS - FREQUENT SECTIONS
Extrapolation of trend to 1990 - 2010	FREQUENT SECTIONS - SECTION TRUTH

3.3.14 Monte Carlo analysis of sensitivity to pseudo section timing

The timing of hydrographic sections is dictated by a number of constraints around the nominal sampling frequencies intended by observational scientists, such as ship time and availability, weather and the availability of principal investigators. Therefore sampling frequencies vary and may, particularly in polar regions, be biased towards sampling at certain times of year. The possible effects of varying the timings of pseudo sections slightly are considered by re-running the analysis and analysing the differences in basin estimate that might be obtained using different times for the occupations. This indicates how sensitive the pseudo section basin estimate is to the precise timing of the occupations. This gives some understanding whether the relationship between the pseudo section estimate and basin truth calculation is fortuitously good or bad, or representative of the sampling frequency.

For each occupation the timing test pseudo section date varies from the original by choosing a date randomly from within a specified time interval (e.g. 1 year prior to the original date to 1 year past the original date). This random selection is repeated many times (e.g., 1000). This gives new sets of timings for each section that have broadly the same time intervals as the original dates, but with

variations of up to a couple of years. Then the analysis of trends is repeated for each of these timing tests. The different timing tests performed are described in Table 3.2.

Table 3.2: Repeating pseudo section analysis with differing dates

New dates chosen from period	Number of repeats
± 1 year	1000
± 2 months	500
- 7 months to - 5 months	500
+ 5 months to + 7 months	500

Moving the dates in this timing test approach aims to expose the uncertainty in pseudo section trends due to the precise timing of the dates. The initial choice of 1 year prior to 1 year past the original date seems a sensible choice for likely movement from an ideal sampling strategy to a realistic timetable for observations. It also provides coverage of the whole year and thus provides a range of possibilities of pseudo section trends that could emerge due to sampling at different times of the year.

Additional experiments are designed to expose the effect of sampling at different times of the year. Five hundred timing tests where the dates range from 2 months prior to 2 months past the original date will provide the range of possibilities for pseudo sections sampling within the same season as the original dates. The difference between this experiment and the first experiment where dates range from 1 year prior to 1 year past will highlight whether sampling in the same seasons to the actual observational dates is particularly important in determining large scale decadal temperature trends. In the same way, experiments with dates ranging from 7 months prior to 5 months prior and from 5 months past to 7 months past will reveal what trends occur when occupation dates are moved to opposing seasons, and comparison with the initial two experiments will highlight if these are significantly different to trends from the same season or the whole year.

The uncertainty related to the pseudo section timings is quantified by calculating two standard deviations of the timing tests. Where the temperature trends are presented against depth, the uncertainty is presented as two standard deviations of the trends from the timing tests around the mean value of the timing tests. When the uncertainty is presented as a heat flux, the heat flux is calculated for

all the timing tests and then the uncertainty reflects two standard deviations of those data.

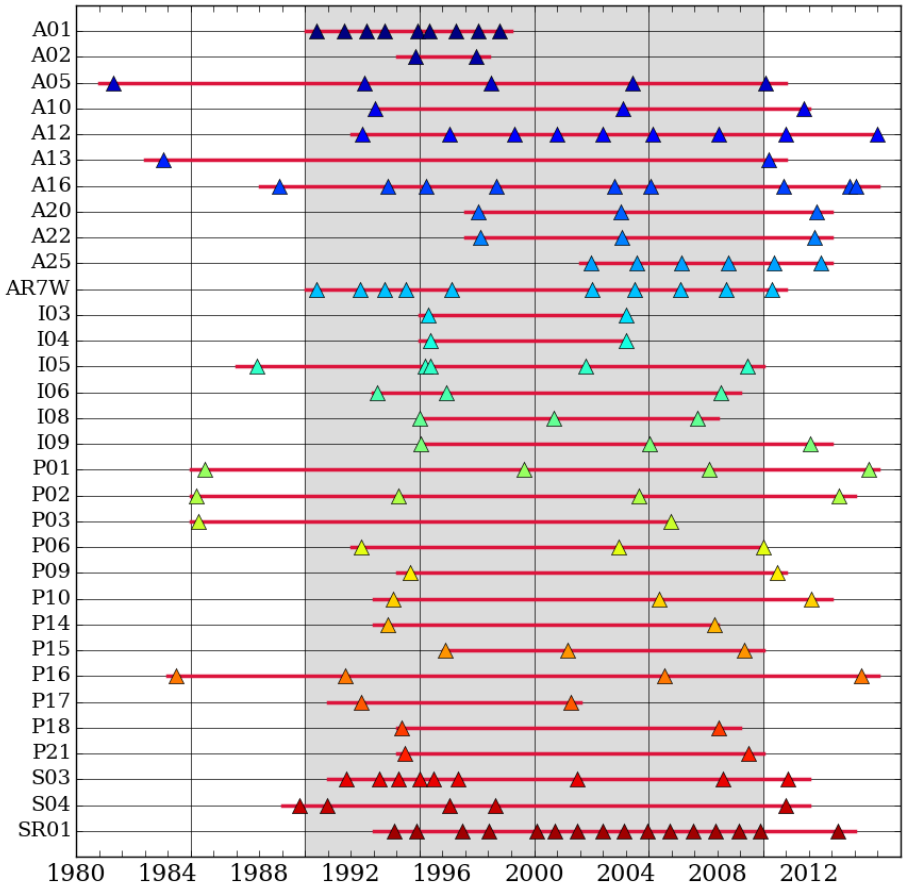


Figure 3.3: Hydrographic sections occupied at least twice during the period 1980 - 2015 as used in an observational analysis of deep OHC by Desbruyères *et al.* (2016b). Triangles denote the times of 'pseudo sections': model output along the section is restricted to the times when sections were actually occupied by hydrographic survey ships. Red lines indicate the period of 'frequent sections': 5 day mean model output along the section is extracted from the start of the year of the first pseudo section to the end of the year of the last pseudo section. The grey shading represents the time interval over which 'section truth' and 'basin truth' data are extracted (1990 - 2015 inclusive). 'Section truth' restricts model output along a section to all 5 day mean temperature data between 1990 and 2015, while the 'basin truth' uses the 5 day mean basin average temperature data for the same time period.

3.3.15 Assessing whether the same results are gained from an updated analysis of observations to 2015

A recent update to Purkey and Johnson (2010) by Desbruyères *et al.* (2016b) uses additional hydrographic sections and timings up to 2015 to provide an up to date assessment of OHC trends in the deep oceans over the last 35 years. The sections and timings used are presented in Figure 3.3 and the key differences between those used previously are the inclusion of sections AR7W and A25 in the Atlantic, P09 in the Pacific and SR01 in the Southern Oceans, as well as the addition of occupations for any sections which occurred between 2011 and 2015. The main analysis is repeated in Section 3.4.5 using this new set of timings to identify if there are large changes in the main uncertainties (excluding Section 3.3.14). The global and oceanic temperature trends are presented at each vertical depth level along with the uncertainty due to averaging along sections. Comparison of these results to those in the main analysis following the methodology of Purkey and Johnson (2010) give some idea of whether the main uncertainties change much for the new analysis by Desbruyères *et al.* (2016b). Since the differences would be attributable not just to observations spanning a longer time period, but also due to the addition of extra sections, there is the potential for the reduction of both temporal and spatial uncertainties. Providing this analysis helps support conclusions from the main analysis which consider what deficiencies may exist in the current deep ocean observing system.

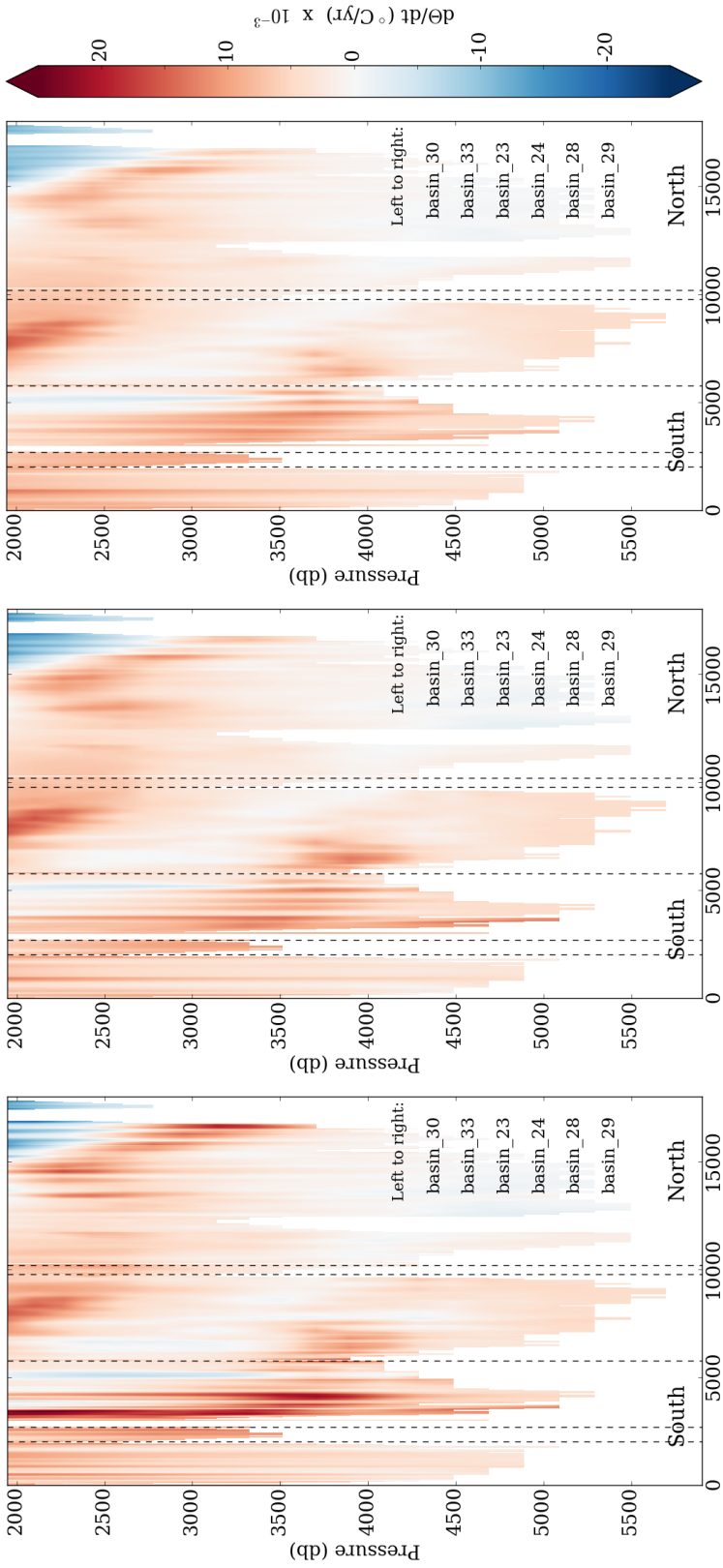
3.4 Results

3.4.1 Trends on sections

As shown in Figure 3.2, pseudo sections (triangles) reflect temperature trends along sections calculated using model output from the timings of actual occupations, frequent sections (red lines) represent temperature trends calculated from 5 day mean model output between the beginning of the year of first occupation and the end of the year of final occupation, whilst section truth trends

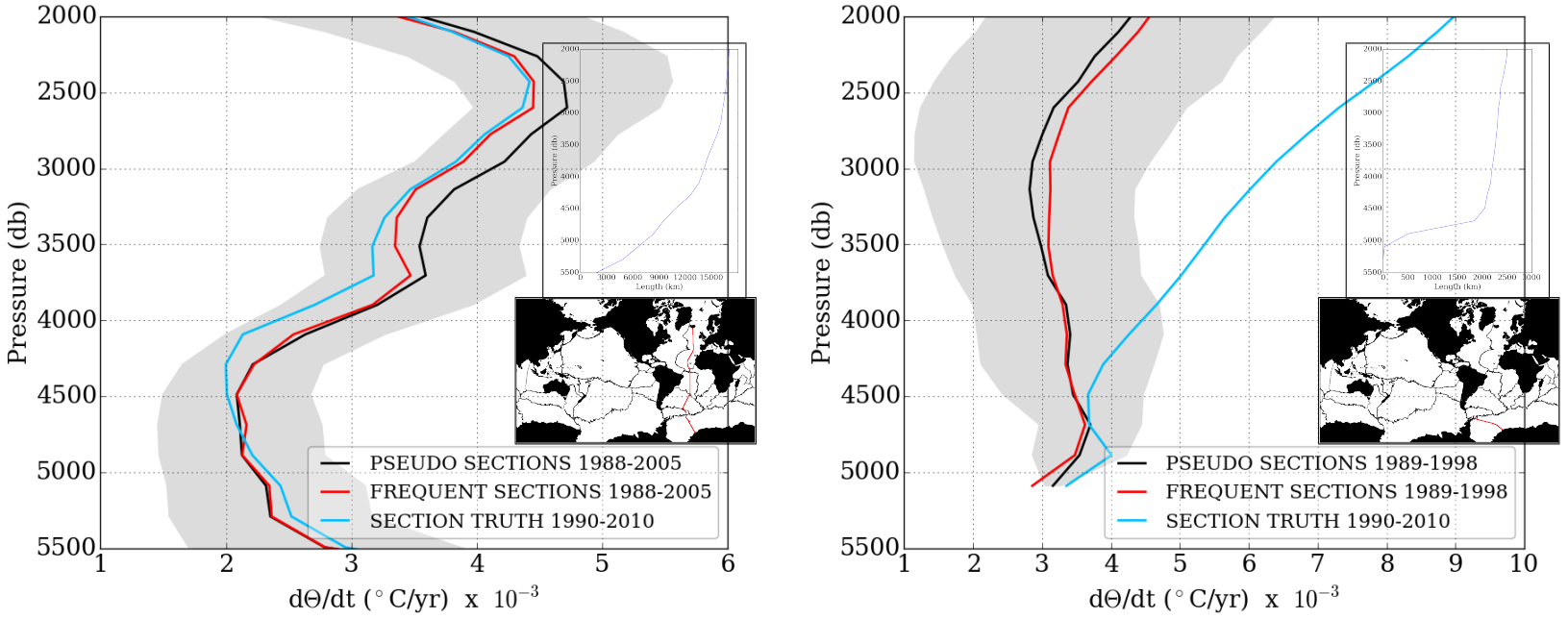
are calculated from 5 day mean model output between 1990 - 2010 (grey shading). For the pseudo sections, frequent sections and section truth data, the trends calculated at each point along a whole section can be visually assessed to give some understanding of the variability of trends along a section, and the pattern differences when trends are calculated from all temporal data, or limited temporal data (Figure 3.4).

There are 28 whole sections (Figure 3.2) but Figure 3.4 presents the data for the longest section, A16, which spans the Atlantic Ocean and Atlantic sector of the Southern Ocean. The patterns of how the trends change along A16 are similar irrespective of the temporal frequency used for the trend calculation. For many of the whole sections, the temporally limited pseudo sections or frequent sections capture the patterns of the section truth well. However the pattern of trends for 5 of the 28 sections (S04, A01, A02, A12 and A22) is notably poor at replicating the section truth patterns with pseudo or frequent sections. These sections are relatively short, and four of them are in the Atlantic. Figure 3.2 highlights that each of these sections are only sampled for half (or less) of the time period 1990 - 2010.



(a) Pseudo sections (b) Frequent sections (c) Section truth 1990 - 2010

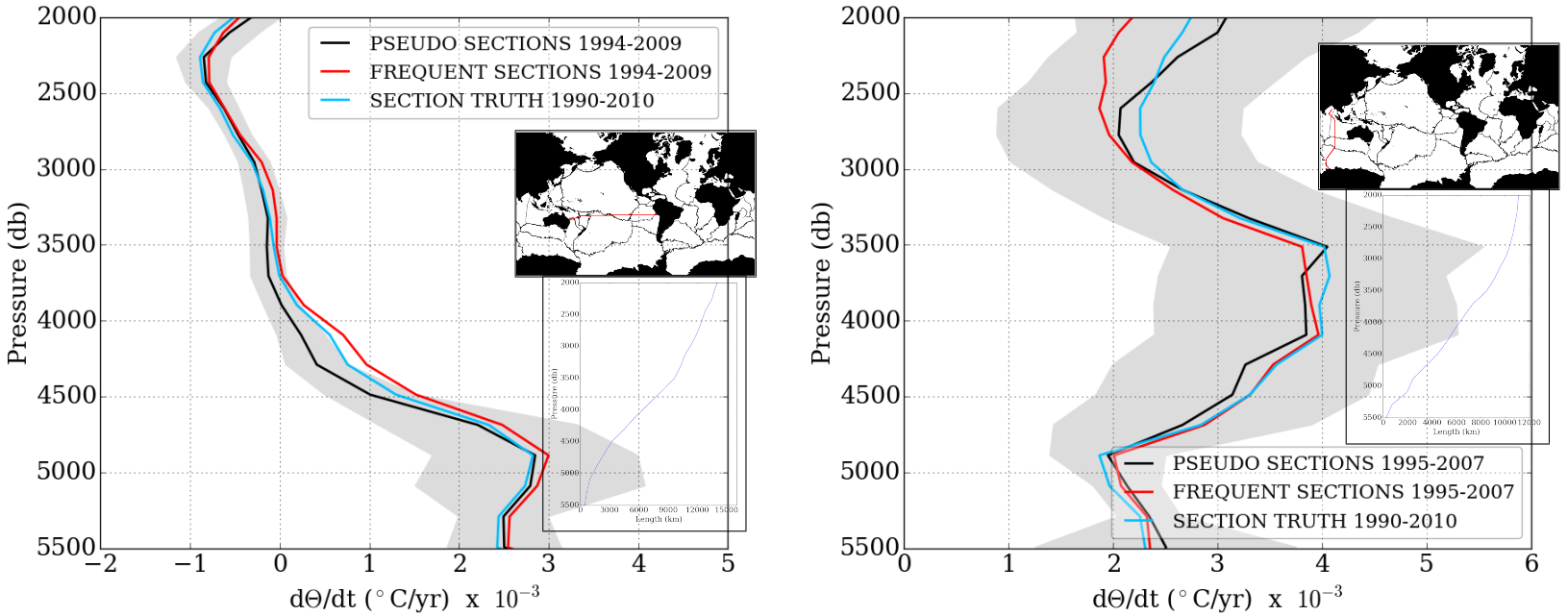
Figure 3.4: Distance/pressure section of $d\Theta/dt$ using model output along hydrographic section A16 calculated from pseudo sections (model output limited to timings of real occupations), frequent sections (5 day mean model output between the years of the first and final occupations) and from section truth (5 day mean model output between 1990 and 2010 inclusive).



(a) A16 (Atlantic Ocean))

(b) S04 (Southern Ocean)

Figure 3.5: Horizontally averaged depth profile of $d\Theta/dt$ along the longest hydrographic sections (a) in the Atlantic (A16) and (b) in the Southern (S04) Oceans calculated from pseudo sections (model output limited to timings of real occupations), frequent sections (5 day mean model output between the years of the first and final occupations) and from section truth (5 day mean model output between 1990 and 2010 inclusive). The grey shading represents uncertainty associated with averaging trends derived from pseudo section data along the section. The inset graph denotes the length of the section (km) at each vertical level in the model.



(a) P21 (Pacific Ocean)

(b) I08 (Indian Ocean)

Figure 3.6: As Figure 3.5 but for the longest hydrographic sections (a) in the Pacific Ocean (P21) and (b) in the Indian Ocean (I08).

Figure 3.5 (a) provides the trends on the whole section A16 when the data in Figure 3.4 is averaged horizontally along the section. For A16, the pseudo section and frequent section estimates and the section truth are similar in pattern and magnitude, though temporally limited data consistently overestimates the section truth trend between 2000 and 5000 m (Figure 3.5 a). The longest whole section solely in the Southern Ocean is S04 (note that a number of North/South hydrographic sections sample significant portions of the Southern Ocean too). Figure 3.5 (b) demonstrates that for SO4, pseudo and frequent section estimates are very close but that the section truth estimate is very different in pattern and magnitude between 2000 - 4000 m, showing a much stronger warming trend. This section was only sampled between 1989 - 1998, so the difference between the section truth and frequent truth is due to the lack of sampling over the final 12 years of the study period. This highlights that calculating long term trends from two occupations close in time and extrapolating them over a much longer period may not be appropriate.

In contrast, the longest whole section in the Pacific, P21 (Figure 3.6 a) was sampled over 1994 - 2009 (the majority of the 1990 - 2010 study period) and the pseudo, frequent and section truth estimates are in excellent agreement. Figure 3.6 (b) reveals that for I08, the longest whole section in the Indian Ocean, the pseudo sections and frequent sections are a reasonably good representation of the section truth except above 3000 m. Between 2000 and 3000 m, the pseudo section and section truth both give a greater trend than the frequent section estimate. This highlights that the lack of sampling between occupations causes an overestimate of the trend relative to 5 day sampling between occupations, but that considering the trend to be representative of the 1990 - 2010 period (uncertainty due to extrapolation) results in an underestimate of the true trend on this whole section. Each uncertainty is of similar magnitude resulting in the pseudo sections giving a good representation of the trend from the section truth.

Considering the other whole sections (not shown) shows that there is no systematic offset of pseudo sections over- or under- estimating the section truth, with at least a third of the sections showing no consistent pattern of over or underestimating through any significant depth range. It is not clear that sections in certain oceans are more or less likely to over or underestimate the section truth from pseudo sections. Since the data for one section covers the same spatial area,

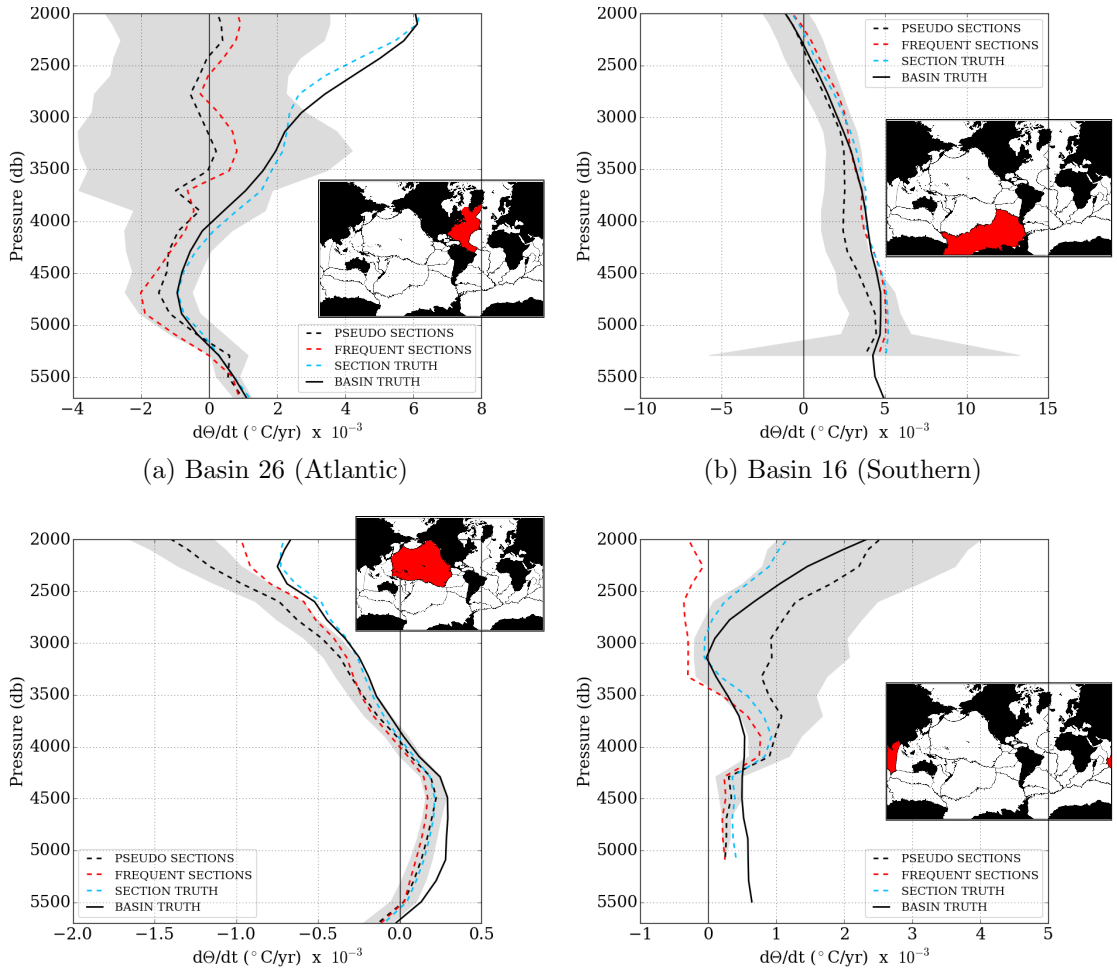
differences between pseudo sections and the section truth are purely due to uncertainty associated with infrequent sampling on the section (no data between limited occupations) and sampling not covering the time period 1990-2010.

The shading in Figures 3.5 and 3.6 is calculated using Equation 3.8, and represents the uncertainty that arises from averaging the trends along a section. This uncertainty is calculated for the pseudo sections and provides potential to account for differences in the different trends through most of the water column, but the deviation of the frequent section estimate and/or section truth trend out of the shaded area for some of the sections shown indicates the importance of other sources of error. Typically this metric of uncertainty is fairly broad for most whole sections, especially in the upper part of the deep ocean where sections are much longer than in the abyssal ocean (e.g. see inset figure on Figure 3.5 a). The length of many whole sections typically covers more than one basin, so variability in trends along the section is not unexpected, and it is worth noting that as these whole sections are split into partial sections constrained in a basin for the rest of the analysis, uncertainty for a partial section may not be as large. As the section truth goes outside these bounds of uncertainty for some sections, it is not a satisfactory way to estimate total uncertainty on the section.

3.4.2 Basin trends and section based estimates

Figure 3.7 presents basin estimates (calculated by averaging the data on sections within a basin) for the largest basins in each ocean alongside the basin truth. There are 33 deep ocean basins in the global ocean, although not all basins are intersected by hydrographic sections. Figure 3.7 (a) shows a close match between the section truth and basin truth data, indicating that basin 26 in the north west Atlantic could be well represented by section data if sampled frequently in time through the study period. The frequent and pseudo section estimates suggest there is a negative bias (warming trends are much weaker and cooling trends are stronger) when sampling is limited in time. Since the difference between the section truth and frequent section estimate is much greater than the difference between the frequent estimate and pseudo section estimate, the temporal bias is almost all due to the bias associated with extrapolating measured trends to the

study period.



(a) Basin 26 (Atlantic)

(b) Basin 16 (Southern)

(c) Basin 18 (Pacific)

(d) Basin 4 (Indian)

Figure 3.7: Largest individual basins in (a) Atlantic Ocean (b) Southern Ocean (c) Pacific Ocean (d) Indian Ocean. Horizontally averaged depth profiles of $d\Theta/dt$ calculated from the basin truth (basin average temperature data between 1990 - 2010) and from three section based estimates. The first section based estimate is calculated from pseudo sections (model output limited to section locations and timings of real occupations), the second from frequent sections (5 day mean model output along sections between the years of the first and final occupations) and the third from the section truth (5 day mean model output along sections between 1990 and 2010 inclusive). The grey shading represents uncertainty associated with averaging the trends derived from pseudo section data along the sections.

In contrast, basin 16 (Figure 3.7 b) in the Southern Ocean reveals a close match between the frequent section estimate, section truth and basin truth, but the trend suggested by pseudo-sections is biased towards a lower trend between 3000

and 5000 m. This indicates the major source of uncertainty in this basin is the lack of sampling between the occupations, which could also include uncertainty due to sampling at a particular time of year (seasonal bias). Although trends are much smaller in magnitude, Basin 18 in the Pacific Ocean (Figure 3.7 c) shows a good fit between all data types apart from above 2500 m where temporal uncertainty (difference between pseudo sections and section truth) is much larger than spatial uncertainty (difference between section truth and basin truth). Approximately half of the temporal bias is due to extrapolating trends to the study period and half due to having no data between occupations. In contrast below 4000 m, spatial uncertainty is much larger than temporal uncertainty.

When considering all basins (not shown), some section estimates for the basins overestimate at some depths, some underestimate, and some, like basin 18, show a reasonable fit at some depth ranges. Generally the vertical pattern of the basin truth is captured by section estimates, but some basins such as basin 4 in the central Indian Ocean (Figure 3.7 d) do not show this. For basin 4 the differences between section estimates reflect a large role for temporal uncertainty whilst the difference between the section truth and basin truth in pattern and magnitude of trends reflects error due to limited spatial sampling. So in some cases there can be markedly large differences at some depths between the basin truth and section estimates, and between the section estimates themselves (another example is basin 31 in the south east Atlantic Ocean, not shown). Overall there is a lack of consistency observed between major sources of uncertainty in different basins, and at different depths even within the same basin. The equivalent plots for the major oceans and global ocean are shown later in Section 3.4.4.

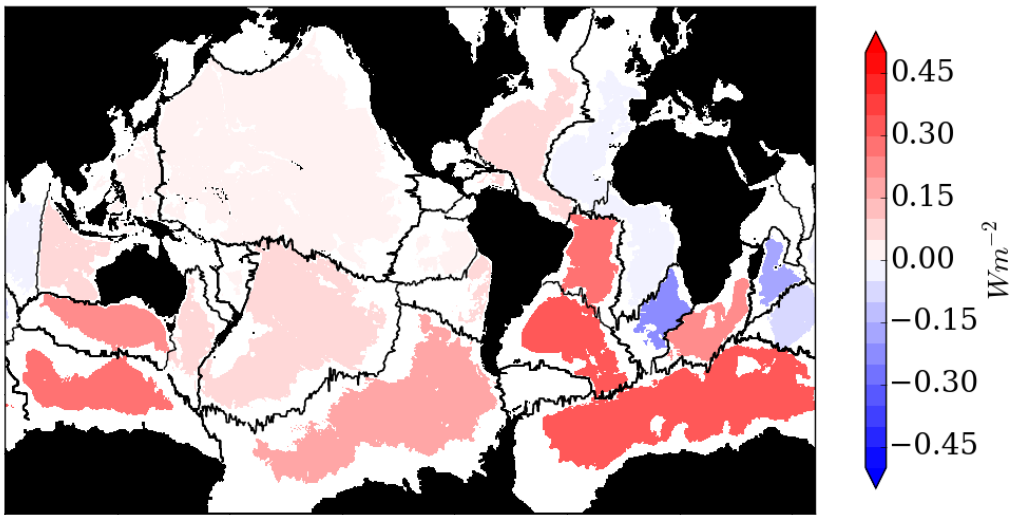
3.4.3 Heat fluxes and estimating biases

Despite the vertical variation in the importance of the uncertainties, to give a more comprehensive overview of the regional variation in uncertainties, the basin estimates and basin truth data are converted into the equivalent heat flux into two layers, 2000 - 4000 m (the deep ocean) and 4000 - 6000 m (the abyssal ocean). First a direct comparison between the observational data for the abyssal ocean (below 4000 m) as calculated from hydrographic sections by Purkey and

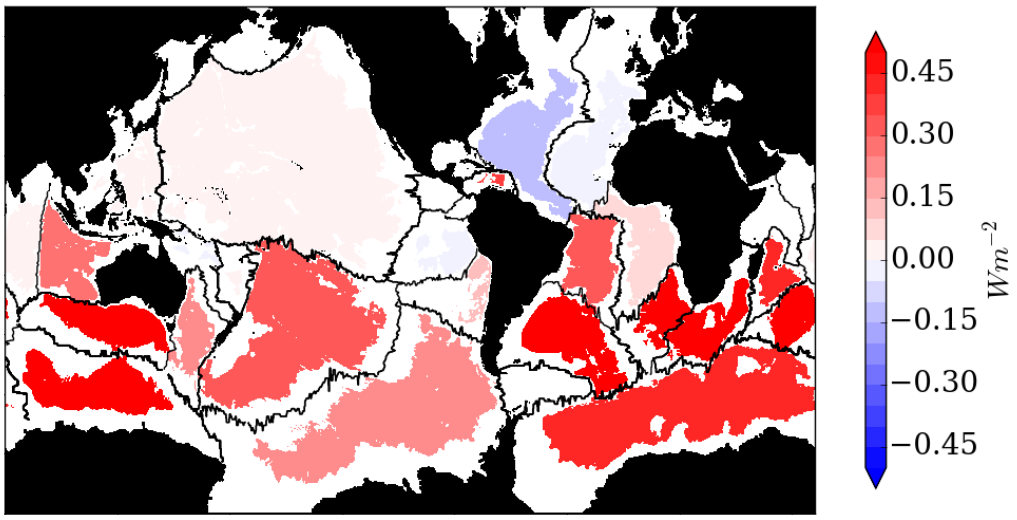
Johnson (2010) and the pseudo section estimate of the basin heat fluxes is presented in Figure 3.8. The equivalent analysis in the model (b) shows many similarities to the observations (a). For both, the strongest warming is present through the south west Atlantic and the Atlantic and Indian sectors of the Southern Ocean, while Pacific Ocean warming is relatively weak in the northern hemisphere and stronger in the southern hemisphere. In both observations and model, there is cooling in the North Atlantic, but this exists down the eastern Atlantic in the observations while the model has a stronger cooling in the north west. In contrast to the observations, the model shows warming through the south east Atlantic and through the western Indian Ocean (Figure 3.8), but the cooling in the observations is not significant (Purkey and Johnson, 2010).

Figure 3.9 shows both the pseudo section estimate and basin truth fluxes for the 2000 - 4000 m and 4000 - 6000 m layers. In addition, by presenting the difference between them, the total bias in sampling the deep oceans using temporally limited hydrographic sections relative to the basin truth is quantified. Between 2000 - 4000 m, Figure 3.9 (a) and (c) show that the pseudo section estimates broadly match the pattern of the basin truth, but (e) highlights that pseudo sections seem to slightly tend toward under-estimating the trend in many basins. Large areas of the Atlantic and Southern Ocean sector of Atlantic have trends underestimated by 0.3 - 0.6 Wm^{-2} by pseudo sections between 2000 - 4000 m. The main features between 2000 - 4000 m in the model deep ocean are warming through the Atlantic, Indian and Southern Oceans with strongest Southern Ocean warming in the Indian and Atlantic sectors, while the northern Pacific Ocean cools (Figure 3.9).

Figure 3.9 (b) and (d) demonstrate that in the abyssal oceans (4000 - 6000 m) pseudo sections and basin truth both show warming across the Southern Hemisphere, with Northern Hemisphere basins showing a much smaller warming trend, or cooling in some basins. The differences (Figure 3.9 f) reveal that pseudo sections in this layer may either over or under estimate the basin truth (with more over-estimating basins than for 2000 - 4000 m). The over-estimates span the subpolar Indian and Atlantic basins, but total error is generally less than 0.12 Wm^{-2} between 4000 - 6000 m. The exceptions are basins 20, 27 and 31, where the over-estimates are higher, at 0.20, 0.22 and 0.20 Wm^{-2} respectively (Figure 3.9).



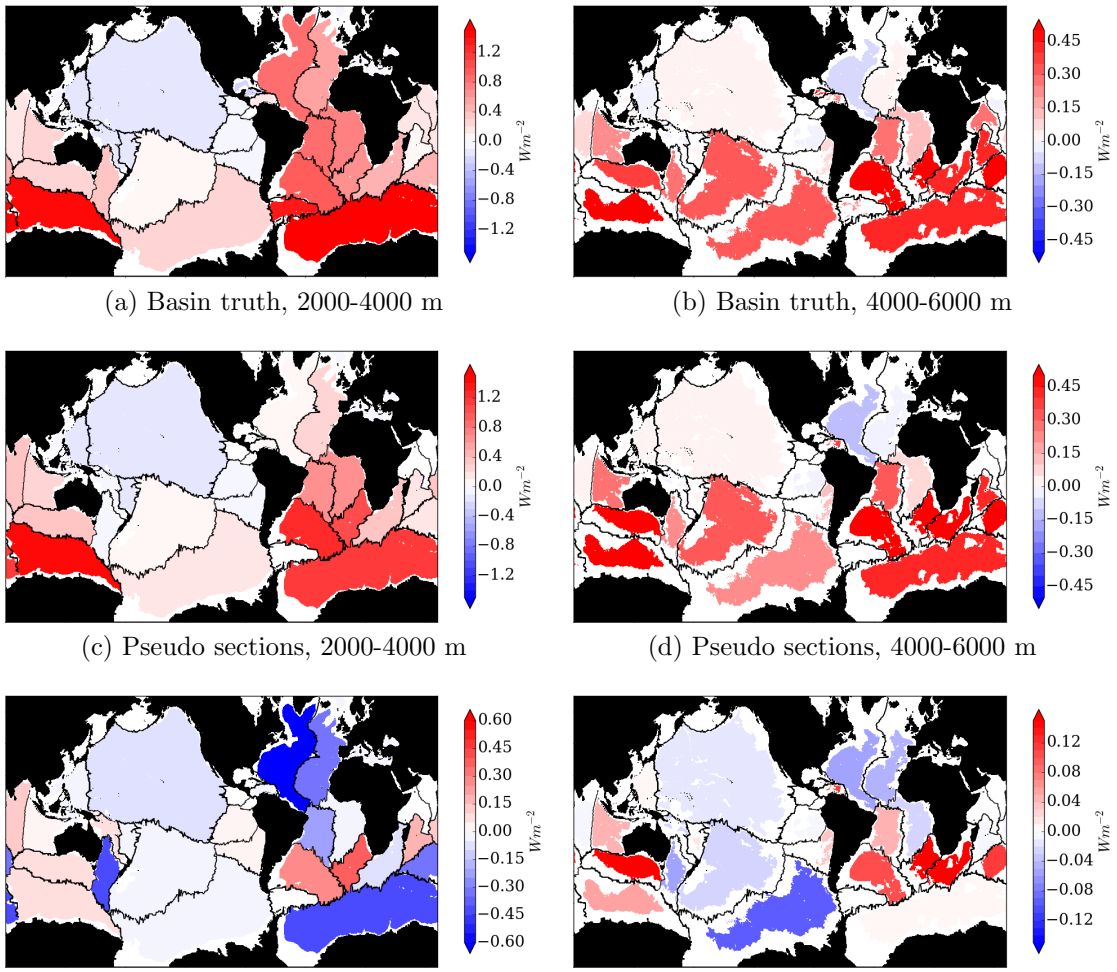
(a) Purkey & Johnson, 2010



(b) N025_DFS_01 Equivalent

Figure 3.8: Local heat fluxes in each basin through 4000 m calculated from abyssal warming below 4000 m during the period 1990 - 2010 from (a) observational hydrographic sections (data from Purkey and Johnson, 2010) and (b) calculated in a NEMO 0.25° simulation from temperature trends along the same sections and at the same timings as the observations.

For the 4000 - 6000 m layer, the total error in Figure 3.9 (f) can be compared to the uncertainty estimates made by Purkey and Johnson (2010) in Figure 3.8 (a) to consider whether the method of estimating uncertainty used in observational analysis is likely to give error bars that would encompass the magnitude of calculated total bias in the model due to spatial and temporal uncertainty. The uncertainty estimates made by Purkey and Johnson (2010) exceed the total bias



(a) Basin truth, 2000-4000 m

(b) Basin truth, 4000-6000 m

(c) Pseudo sections, 2000-4000 m

(d) Pseudo sections, 4000-6000 m

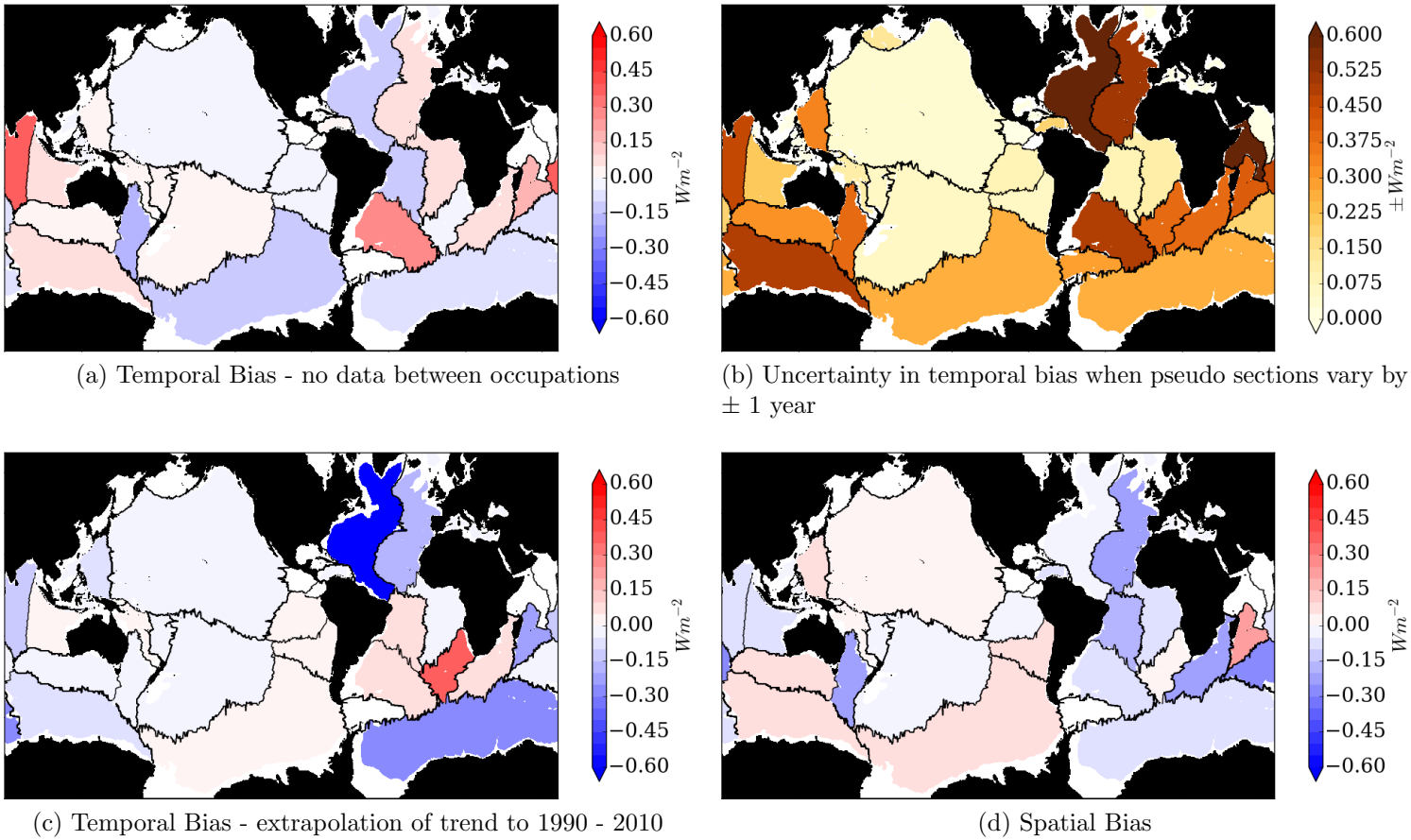
(e) Bias in pseudo section trends, 2000-4000 m

(f) Bias in pseudo section trends, 4000-6000 m

Figure 3.9: Local heat fluxes in each basin into the 2000 - 4000 m and 4000 - 6000 m layers, calculated from the warming in the deep and abyssal layers respectively during the period 1990 - 2010. (a) and (b) show the heat fluxes equivalent to basin truth warming trends (calculated from basin average temperatures), whilst (c) and (d) the heat fluxes equivalent to pseudo section warming trends (model output limited to section locations and timings of real occupations). (e) and (f) show for each layer, the basin truth heat fluxes subtracted from the pseudo section heat fluxes, illustrating whether the pseudo section estimate is biased positive or negative relative to the basin truth.

calculated in the model in all but four basins. In three of these four, the difference is very small; the magnitude of the total bias (Figure 3.9 f) exceeds the Purkey and Johnson (2010) uncertainty estimate by 0.0018 Wm^{-2} in basin 18 (North Pacific), by 0.0065 Wm^{-2} in basin 20 (south west of Australia), and by 0.00071 Wm^{-2} in basin 19 (south east of Australia). In basin 15, the relatively

small Caribbean Sea, the magnitude of the total bias is 0.08 Wm^{-2} greater (Figure 3.9 f) than the magnitude of the uncertainty in Purkey and Johnson (2010). The uncertainty estimates made in the observational analysis by Purkey and Johnson (2010) are generally sufficient to encompass the magnitude of modelled spatial and temporal biases in the abyssal ocean (4000 - 6000 m).



(a) Temporal Bias - no data between occupations

(b) Uncertainty in temporal bias when pseudo sections vary by ± 1 year

(c) Temporal Bias - extrapolation of trend to 1990 - 2010

(d) Spatial Bias

Figure 3.10: Heat fluxes calculated from the basin truth, section truth, frequent sections and pseudo sections are compared in different pairings to separate the different biases in 2000 - 4000 m ocean sampling. The effect of collecting temporally limited data along sections can be separated into two parts: there is no data between the occasional occupations (a) and the trend calculated from limited sections in the period is assumed to be representative of the trend between 1990 and 2010 (c). Uncertainty in relation to the exact timing of pseudo sections is tested by moving the pseudo section dates about randomly in a two year interval from one year before to one year after the real timings (b). The spatial bias (d) illustrates where the estimate from the section truth is not representative of the basin truth.

Total bias can be decomposed into the components representing the spatial uncertainty and the temporal uncertainty, with the temporal uncertainty split between uncertainty due to observations not spanning 1990-2010 (extrapolation) and uncertainty due to no data between the limited occupations. Figure 3.10 illustrates the biases for each basin. In addition, another uncertainty is presented that addresses the importance of the precise timing of pseudo sections, using a Monte Carlo approach of repeating the pseudo section analysis 1000 times with dates that randomly vary in the interval from 1 year prior to 1 year past the true occupation dates. This ensures that the approximate frequency of sampling is preserved. The results provide an uncertainty around the pseudo section estimate and thus the temporal bias due to no data between occupations.

Between 2000 - 4000 m, biases due to no data between occupations (Figure 3.10 a) are positive through the east Atlantic and Indian Oceans (trends overestimated there). The western Atlantic and much of the Southern Ocean show negative biases suggesting that having no data between occupations results in an underestimate of the trends in these regions. However, Figure 3.10 (b) indicates that uncertainties in the timing of the pseudo sections can be relatively large and might result in changes to the biases in (a) of at least the same magnitude as the calculated biases. Therefore the patterns observed in Figure 3.10(a) may be largely a function of the very precise timings of the pseudo sections.

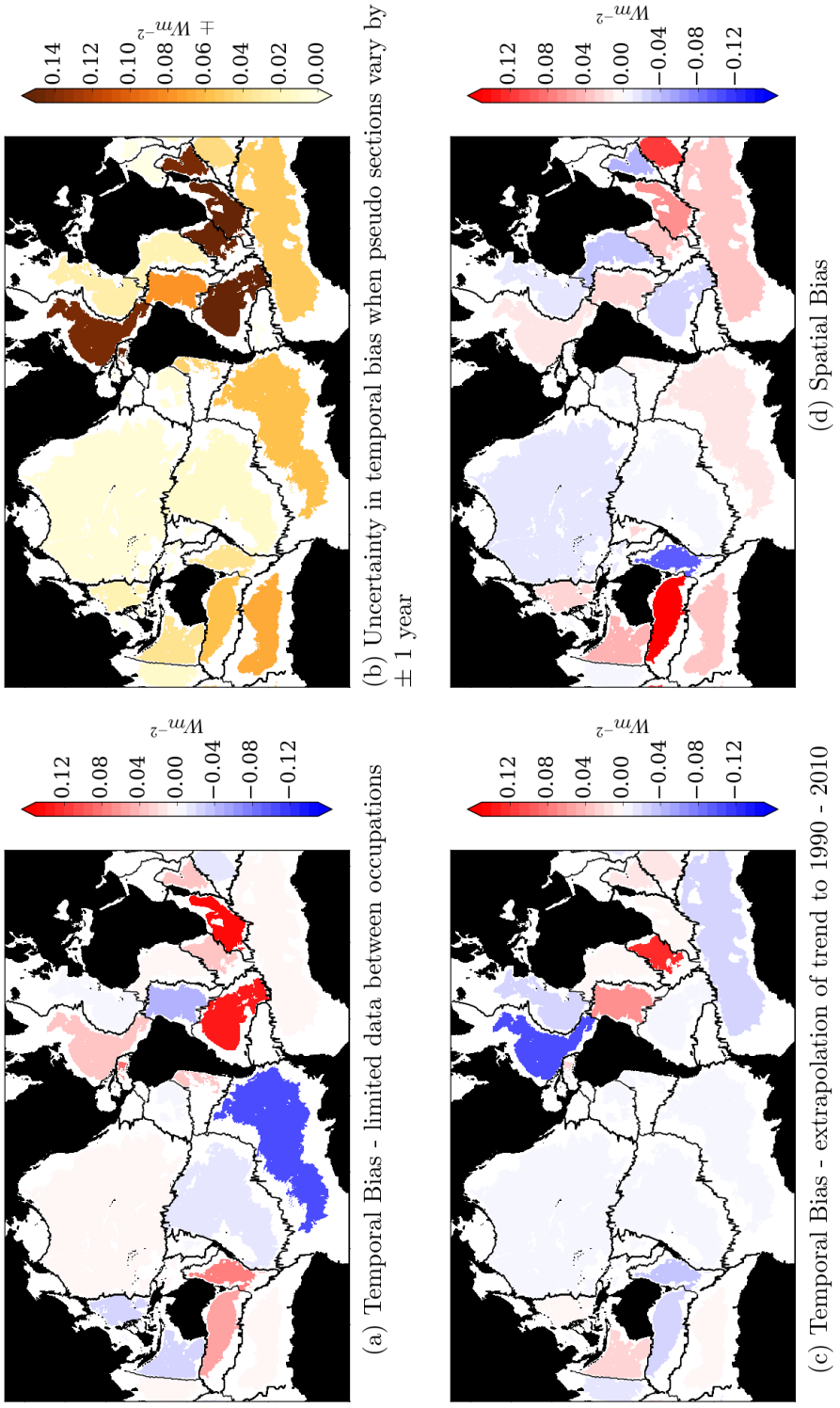
Figure 3.10 (c) indicates that extrapolating trends from sections to assume representativity for the period 1990-2010 is relatively unimportant across the Pacific basins, but can lead to sizeable biases in some of Atlantic and Indian basins (including their Southern Ocean sectors). The largest bias due to extrapolation is in the north west Atlantic (Figure 3.10 c). Lack of spatial coverage leads to an under-estimation of warming through the Atlantic and some Indian Ocean basins (Figure 3.10 d), but a relatively minor over-estimation in other Indian basins, the Pacific and the Indian and Pacific sectors of the Southern Ocean. Figure 3.10 illustrates that some basins have positive values for one bias and negative for another (e.g. the Pacific sector of the Southern Ocean); different uncertainties can result in opposing biases in the same basin.

Biases in heat fluxes into the 4000 - 6000 m layer (Figure 3.11) do not appear to exhibit spatially coherent patterns across oceans. For each bias there are some

basins where the biases are comparatively strong compared to the rest of the global oceans, but these are different basins for each bias (Figure 3.11). Uncertainty in the timing of pseudo sections (Figure 3.11 b) is also particularly large in several basins across the Atlantic and Indian Oceans, with moderate biases also in the Southern Ocean.

The uncertainties for the 4000 - 6000 m layer calculated in Figure 3.11 (b) due to the precise timings of pseudo sections can be compared to the uncertainty estimates by Purkey and Johnson (2010) in Figure 3.8 (a). This metric of uncertainty is also smaller than the observational estimates of uncertainty made by Purkey and Johnson (2010) in all basins except the relatively small basin 15 (the Caribbean Sea) where the uncertainty due to timing is 0.4 Wm^{-2} greater than that estimated by Purkey and Johnson (2010) in Figure 3.8 (a). Again, this illustrates that the observational uncertainty estimates made by Purkey and Johnson (2010) in the abyssal ocean may be large enough to encompass modelled uncertainty associated with changes to the sampling times.

Figure 3.11 (b) reveals that as between 2000 - 4000 m, the magnitude of the uncertainty due to pseudo section timings between 4000 - 6000 m is of a similar order of magnitude to the calculated biases, in particular the temporal bias due to no data between occupations (Figure 3.11 a), which could be significantly affected by relatively minor adjustments to the dates of the pseudo sections. Lack of spatial coverage leads to a positive bias across the Southern Ocean between 4000 - 6000 m (Figure 3.11 d) .



(a) Temporal Bias - limited data between occupations

(b) Uncertainty in temporal bias when pseudo sections vary by ± 1 year

(c) Temporal Bias - extrapolation of trend to 1990 - 2010

(d) Spatial Bias

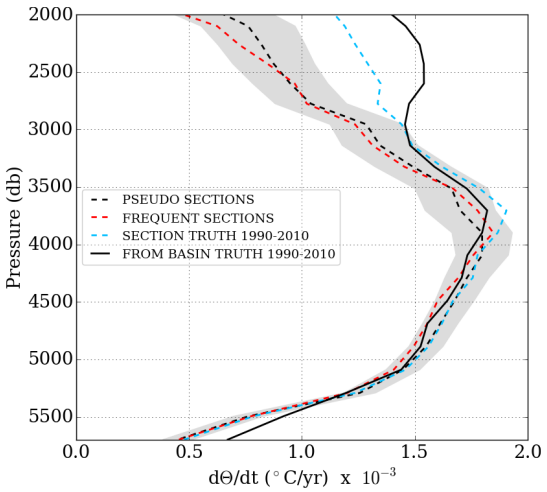
Figure 3.11: As Figure 3.10 but for 4000 - 6000 m.

3.4.4 Global and major ocean trends

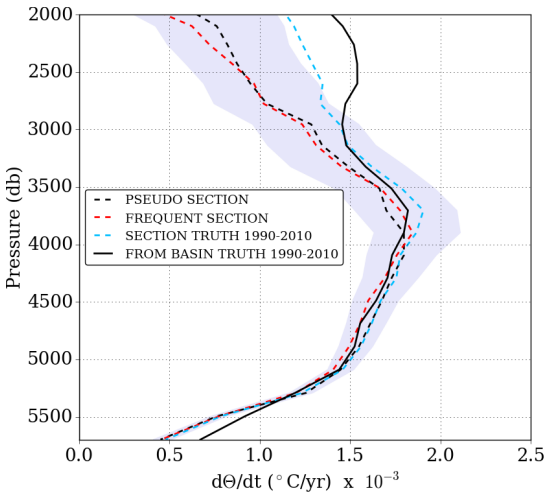
To consider if pseudo sections can represent global temperature trends accurately, the basin temperature trends are combined to get section based estimates of the global trend together with the global truth from the basins. Figure 3.12 provides the temperature trends vertically from 2000 m with a different uncertainty shown in each panel. Globally the section based estimates under-estimate the truth at whichever temporal frequency sampled between 2000 m and 3000 m, highlighting that spatial coverage is partly responsible for hindering capture of the warming trend (Figure 3.12). In this 2000 - 3000 m layer there is also a large difference between the section truth estimate and those from frequent and pseudo sections, revealing that the role of extrapolation of section trends across the study period introduces a large bias; extrapolation bias accounts for approximately two thirds of the total bias compared to spatial bias comprising one third. Below 3000 m the section truth is similar to the truth from basins, indicating that the spatial bias becomes small. The offset of the pseudo and frequent section from the section truth estimates until 3500 m indicates that temporal bias due to extrapolating section trends to cover the study period is an important source of error between 3000 - 4000 m. Total bias is relatively small at greater depths. Below 5300 m, spatial bias again results in an underestimation of the global trend by section estimates.

The pseudo section data in Figure 3.12 is volume weighted during averaging to give an estimation of the overall temperature trend below 2000 m of 0.00128 °C year^{-1}, compared to the basin truth trend of 0.00156 °C year^{-1}. The total bias in the pseudo sections estimate is therefore -0.00028 °C year^{-1}, which is equivalent to pseudo sections underestimating the trend by 18% / capturing only 82% of the trend. The underestimate is greater in the upper deep ocean, especially between 2000 - 2500 m, where less than 60% of the trend is captured.

The uncertainty associated with averaging the pseudo section trends along sections (Figure 3.12 a) is largest between 2000 - 3000 m and decreases with depth, becoming almost non-existent below 5000 m. The uncertainty interval encompasses the frequent section estimate but not the section truth or the truth from basins above 3500 m, highlighting the importance of other uncertainties.



(a) Shading indicates uncertainty due to averaging along sections



(b) Shading indicates uncertainty when varying pseudo section timings by ± 1 year

Figure 3.12: For the global oceans, horizontally averaged depth profiles of $d\Theta/dt$ calculated from basin truth (basin average data between 1990 - 2010), pseudo sections (model output along sections limited to timings of true observations), frequent sections (5 day mean model output along sections between the first and final years of section occupation) and from section truth (5 day mean model output along sections between 1990 and 2010 inclusive). In addition, in (a) the uncertainty associated with averaging temperature trends horizontally along the sections (calculated for pseudo sections) is shown with grey shading reflecting the 95% confidence interval. In (b) light blue shading indicates ± 2 standard deviations from the mean of the trends from 1000 timing test pseudo sections. They differ from the original pseudo section estimate because their dates have been slightly shifted by randomly selecting dates between -1 year and 1 year of the actual occupations.

Shifting the dates of pseudo sections randomly between 1 year prior to 1 year past the true dates, and repeating the analysis 1000 times (Figure 3.12 b), results in a timing uncertainty that is generally larger than that associated with averaging the pseudo section trends along sections (Figure 3.12 a). The timing uncertainty (Figure 3.12 b) decreases with depth but the width of the shading exceeds 0.0002 °C above 5000 m. It does not encompass the truth from basins or section truth between 2000 - 2700 m, where spatial bias results in a significant offset regardless of the section timings (Figure 3.12 b). In addition, the extrapolation of section based estimates from the period of sampling to 1990 - 2010 remains a major uncertainty between 2000 - 3500 m.

To quantify the contribution of different oceans to the global uncertainties the

biases and uncertainties are calculated from the equivalent heat fluxes into the deep and abyssal layers and are presented in Tables 3.3 and 3.4 respectively. Globally the temporal bias due to extrapolation dominates in the deep layer, whilst being marginally the largest bias in the abyssal layer. In the abyssal layer, the biases are all of approximately the same magnitude globally. The temporal bias due to limited data almost compensates the bias due to extrapolation (Table 3.4), so the pseudo sections at their observed timings actually provide a fairly close fit to the truth (Figure 3.12).

Table 3.3: Uncertainties in measuring global and oceanic trends, 2000 - 4000 m

Uncertainty	Global	Atlantic	Southern	Pacific	Indian
			10^{-3} Wm^2		
Total bias	-92.5	-226.0	-143.7	-49.5	-21.6
Spatial bias	-25.5	-99.4	+26.4	-8.8	-37.9
Temporal bias due to extrapolation	-74.5	-144.2	-117.0	-20.0	-106.4
Temporal bias due to limited data	+7.5	+17.7	-53.2	-20.7	+122.7
Uncertainties when pseudo sections differ in timings by ± 1 year	± 62.1	± 191.5	± 180.8	± 32.7	± 191.9
Uncertainties due to averaging along sections	± 38.3	± 154.7	± 157.1	± 26.3	± 108.8

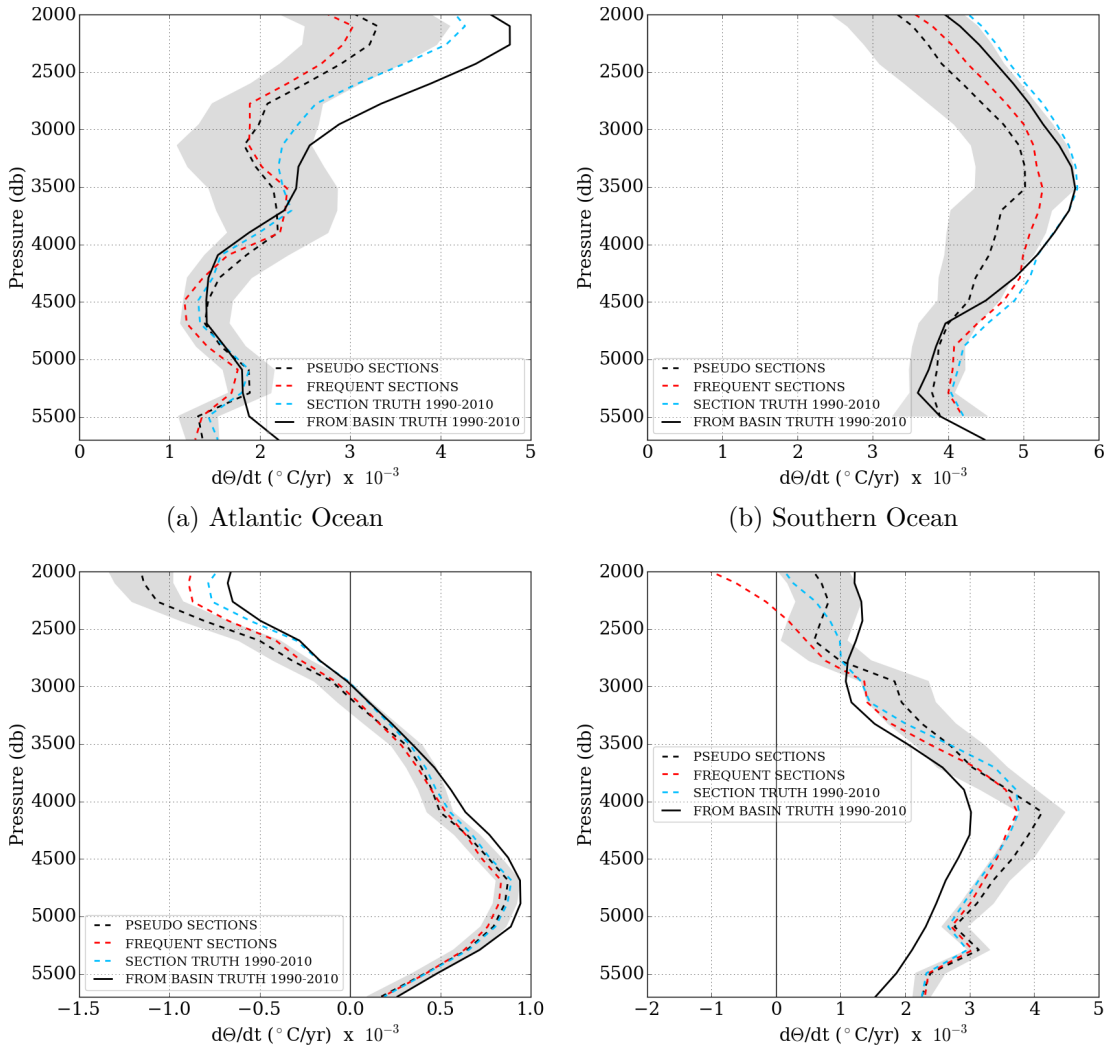
Reconstructing the same analysis by major ocean basins shows the vertical trends in each ocean behave quite differently, and are captured by section based estimates to varying extents (Figure 3.13). The Atlantic basin trends are not well captured by section estimates with sizeable biases above 3500 m and below 5300 m (Figure 3.13 a). The separation of each of the lines between 2000 - 3500 m in Figure 3.13 (a) emphasises the role of spatial and both types of temporal uncertainty, and Table 3.3 reveals relatively high values for the biases in heat flux into the deep ocean layer: - 0.14 Wm^{-2} for the temporal bias due to extrapolation and - 0.1 Wm^{-2} for the spatial bias. However in the abyssal ocean the greatest Atlantic bias is due to no data between occupations; this is twice the magnitude of the other temporal bias (due to extrapolation), while spatial bias appears negligible in the abyssal Atlantic Ocean (Table 3.4).

Table 3.4: Uncertainties in measuring global and oceanic trends, 4000 - 6000 m

Uncertainty	Global	Atlantic	Southern	Pacific	Indian
			10^{-3} Wm^2		
Total bias	+4.0	+10.7	-28.6	-9.8	83.6
Spatial bias	+4.3	-0.3	+21.8	-7.5	+56.8
Temporal bias due to extrapolation	-5.5	-9.6	-12.6	-3.6	-1.6
Temporal bias due to limited data	+5.2	+20.6	-37.7	+1.3	+25.2
Uncertainties when pseudo sections differ in timings by ± 1 year	± 12.3	± 42.2	± 35.0	± 4.7	± 34.5
Uncertainties due to averaging along sections	± 9.6	± 35.3	± 40.1	± 6.2	± 29.4

The Southern Ocean (Figure 3.13 b) trends are under-estimated by pseudo sections due to temporal biases down to 4500 m (since the section truth agrees well with truth from basins over the same depth range); Table 3.3 reveals that the biases in heat flux due to temporally limited sampling are composed primarily from extrapolation (0.12 Wm^{-2}), with a contribution from limited data between occupations (0.05 Wm^{-2}). The spatial sampling bias for the Southern Ocean is much smaller in magnitude than for the temporal uncertainties between 2000 - 4000 m (Table 3.3), suggesting that increased spatial coverage would not be as useful as higher frequency sampling to capture the decadal trends accurately between 2000 - 4000 m. As in the Atlantic, the 4000 - 6000 m Southern Ocean bias due to extrapolation is much less important in the abyssal ocean than for the deep layer, but together the temporal biases are over twice as important as the the spatial bias (Table 3.4).

Trends in the Pacific Ocean (Figure 3.13 c) are generally well captured by section based estimates but with a systematic underestimate due to spatial sampling, with temporal biases also playing a role especially between 2000 - 3000 m. In the Indian Ocean (Figure 3.13 d) spatial biases are clearly significant shallower than 3000 m and deeper than 4000 m; in the 4000 - 6000 m layer the spatial bias is twice as big as the timing bias due to limited data and many times larger than that due to extrapolation (the latter being relatively unimportant) (Table 3.4).



(a) Atlantic Ocean (b) Southern Ocean

(c) Pacific Ocean (d) Indian Ocean

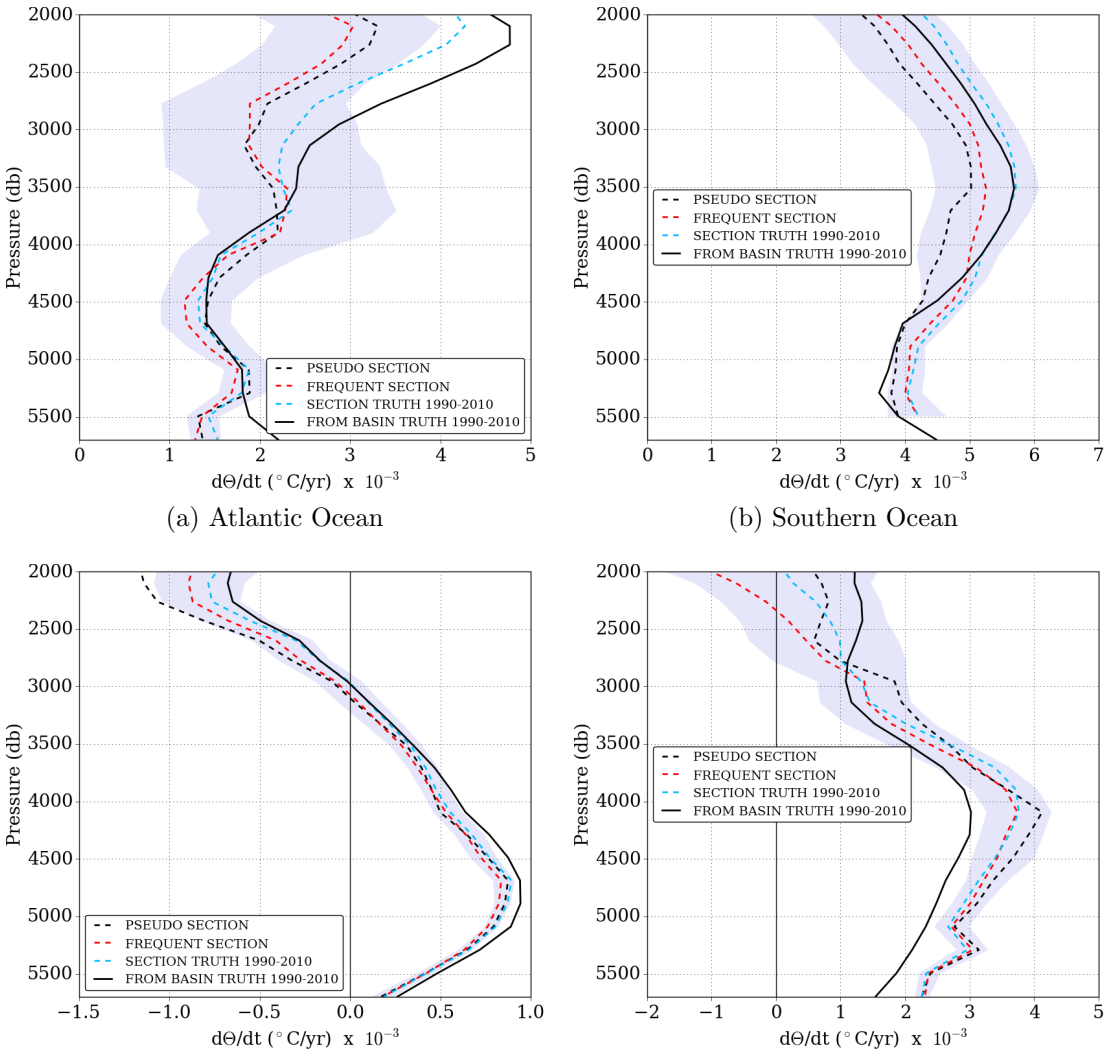
Figure 3.13: For each ocean, horizontally averaged depth profiles of $d\Theta/dt$ calculated from basin truth (basin average data between 1990 - 2010), pseudo sections (model output along sections limited to timings of true observations), frequent sections (5 day mean model output along sections between the first and final years of section occupation) and from the section truth (5 day mean model output along sections between 1990 and 2010 inclusive). In addition, the uncertainty associated with averaging temperature trends horizontally along the sections (calculated for pseudo sections) is shown with grey shading reflecting the 95% confidence interval.

Table 3.3 shows that over the 2000 - 4000 m layer in the Indian Ocean the timing biases are much larger than the spatial bias, but Figure 3.13 (d) reveals that the negative spatial bias above 3000 m is compensated by a positive spatial bias below 3000m.

Comparing Figure 3.13 to Figure 3.12 (a) reveals that spatial and temporal biases that exist when estimating global trends between 2000 - 3500 m are largely due to a combination of spatial biases in the Atlantic (above 3500 m) and Indian Oceans and temporal biases in the Atlantic, Southern and Indian Oceans.

The shading in Figure 3.13 reflects the uncertainty due to averaging pseudo section trends along sections, and is broadest in the Southern and Atlantic Oceans. Figure 3.14 shows the same data but with different uncertainty shading, reflecting uncertainty associated with sensitivity to specific pseudo section dates. This was calculated by re-running the analysis 1000 times with pseudo section dates shifted randomly within 1 year prior to 1 year past the real pseudo section dates. The uncertainties are relatively large for the Atlantic Ocean and 2000 - 3000 m in the Indian Ocean (Figure 3.14), compared to the uncertainties in Figure 3.13. This reinforces the point that using uncertainty due to averaging along sections may not be a sufficiently broad uncertainty for deep ocean temperature trends. However neither uncertainty encompasses the truth from basins in the 2000 - 3000 m Atlantic and below 4000 m Indian Oceans, highlighting the critical importance of the spatial bias in these regions.

Timing tests involve performing the pseudo section analysis for timings randomly shifted in a certain interval, so to test the importance of sampling at different times or year the interval is shifted from one year prior to one year past the true dates, to: 2 months either side of the true dates, 7 to 5 months prior to the true dates and 5 to 7 months past (see Table 3.2). In Figure 3.15 the results are shown for the Southern Ocean, which is a region where hydrography usually takes place in the summer months due to adverse weather conditions and sea ice during winter. The range of uncertainty that is generated when sampling at any time of year (Figure 3.15 a) differs slightly from that when sampling in the same season (Figure 3.15 b) and differs even more when sampling in the opposite season to the true dates (Figure 3.15 c and d). The uncertainty interval is slightly reduced in width when sampling in the same season (Figure 3.15 b) than over a two year period (Figure 3.15 a), but (b) does not include the section truth line at some depths when (a) does. However since the shading represents 2σ of the timing tests, one in twenty of the timing tests may lie outside the shading. However, it may also suggest that sampling in the same seasons that Southern Ocean observations have been made in over recent decades may be less likely to yield a

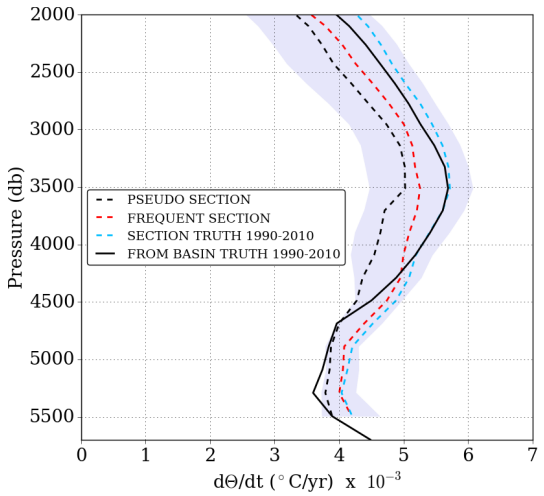


(a) Atlantic Ocean (b) Southern Ocean

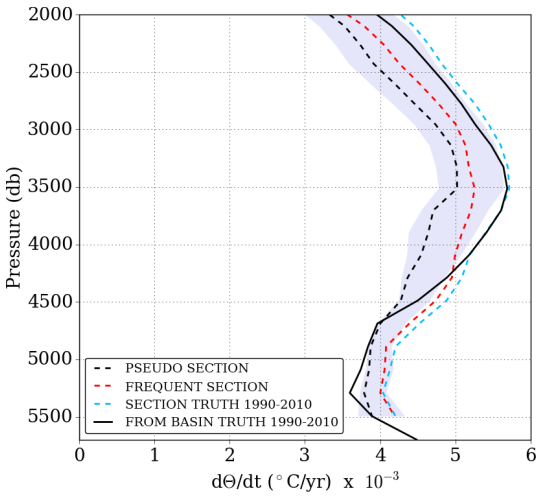
(c) Pacific Ocean (d) Indian Ocean

Figure 3.14: For each ocean, horizontally averaged depth profiles of $d\Theta/dt$ calculated from basin truth (basin average data between 1990 - 2010), pseudo sections (model output along sections limited to timings of true observations), frequent sections (5 day mean model output along sections between the first and final years of section occupation) and from section truth data (5 day mean model output along sections between 1990 and 2010 inclusive). In addition, light blue shading indicates $\pm$ 2 standard deviations from the mean of the trends from 1000 timing test pseudo sections. They differ from the original pseudo section estimate because their dates have been slightly shifted by randomly selecting dates between -1 year and 1 year of the actual occupations.

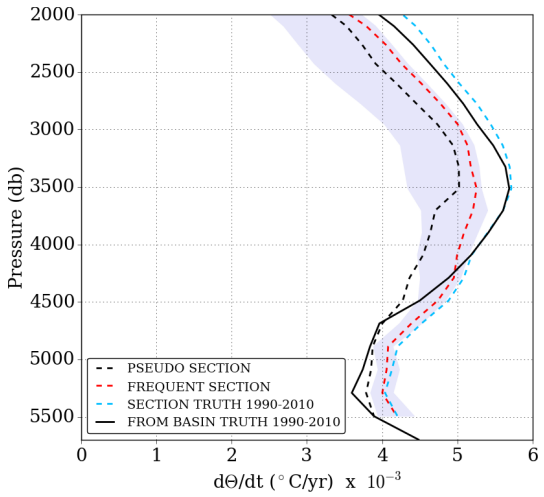
representative trend than sampling throughout the year. The trend from any adjusted timing pseudo section is likely to be an underestimate of the section and basin truth, because the shading is biased negative compared to the section truth



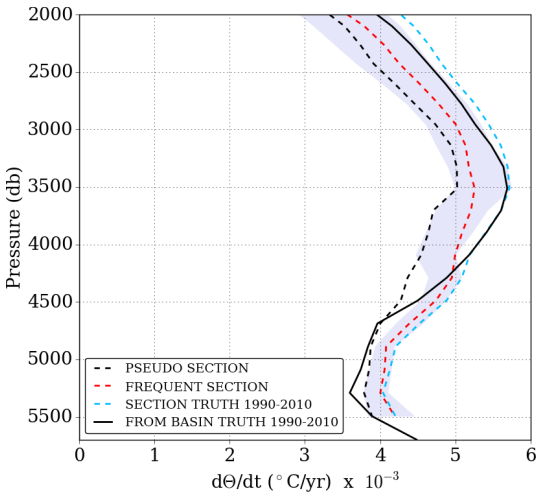
(a) ± 1 year (Any time of year)



(b) ± 2 months (Same season as original)



(c) - 7 mos. to - 5 mos. (Opposite season to original)



(d) +5 mos. to +7 mos. (Opposite season to original)

Figure 3.15: For the Southern Ocean, horizontally averaged depth profiles of $d\Theta/dt$ calculated from basin truth (basin average data between 1990 - 2010), pseudo sections (model output along sections limited to timings of true observations), frequent sections (5 day mean model output along sections between the first and final years of section occupation) and from section truth data (5 day mean model output along sections between 1990 and 2010 inclusive). In addition, shading indicates ± 2 standard deviations from the mean of the trends from 1000 timing test pseudo sections. They differ from the original pseudo section estimate because their dates have been slightly shifted by randomly selecting dates in a time interval relative to the actual occupations. Time intervals chosen are (a) ± 1 year from actual pseudo section timings (b) ± 2 months (c) 7 to 5 months before actual timings and (d) 5 to 7 months after actual timings.

and basin truth lines.

The uncertainty interval is approximately the same width when sampling in the opposite season (Figure 3.15 c and d) compared to the same season (Figure 3.15 b), although it varies in shape at some depths. However the section truth line is further outside the uncertainty interval more often when sampling in the opposite season (c) to sampling in the same season (b). A second opposing season interval is proposed in (d) by shifting timings forward in time as opposed to backwards (c), but the differences between (b) and (d) are less pronounced than those between (b) and (c). Sampling is more likely to get a representative trend on sections when timings are similar to those used at the moment (i.e. in summer) compared to measuring in winter only, and the best representation would be obtained when measuring all year round (a). This reveals that seasonal variability exists in the model deep Southern Ocean.

The pseudo section estimate from actual dates tends to present a cooler trend than that of the mean of the pseudo sections, which is the middle of the shaded area in Figure 3.15 when sampling in the same season and or any time of year, so this is something of an outlier result particularly between 4500 - 5500 m, where it is very much on the lower end of the predicted range of trends. This suggests that the Southern Ocean bias due to no data between occupations, which is directly affected by a change to the trend from pseudo-sections, is likely to be larger in the earlier analysis using original pseudo section dates than that for a randomly selected set of dates from the timing tests. This changes when sampling at a different time of year, as the pseudo section trend leaves the bounds of the uncertainty between 4000 - 5500 m (Figure 3.15 c and d), suggesting that the same trends from pseudo sections (as well as the section truth) could not be achieved by similar sampling frequency but in the opposing seasons, and further highlighting the seasonality in deep ocean temperatures.

The Indian, Atlantic, Pacific and Global Oceans all show similar patterns for the timing tests (not shown) whereby the uncertainty range is much smaller when measuring either in the same season as actual timings or in opposing seasons, than sampling at any time of year. When sampling in the same or opposing season to the actual measurements the section truth often leaves the uncertainty interval, suggesting a representative trend on sections is unlikely to be gained.

However spreading sampling randomly across the year removes any seasonal biases to the sampling and often the section trend is near the midpoint of the uncertainty interval. The 2000 m to 3000 m Atlantic is an exception as the section truth does not fall in the uncertainty interval highlighting the role of uncertainty due to extrapolating trends there; this also occurs for the global ocean analysis (not shown).

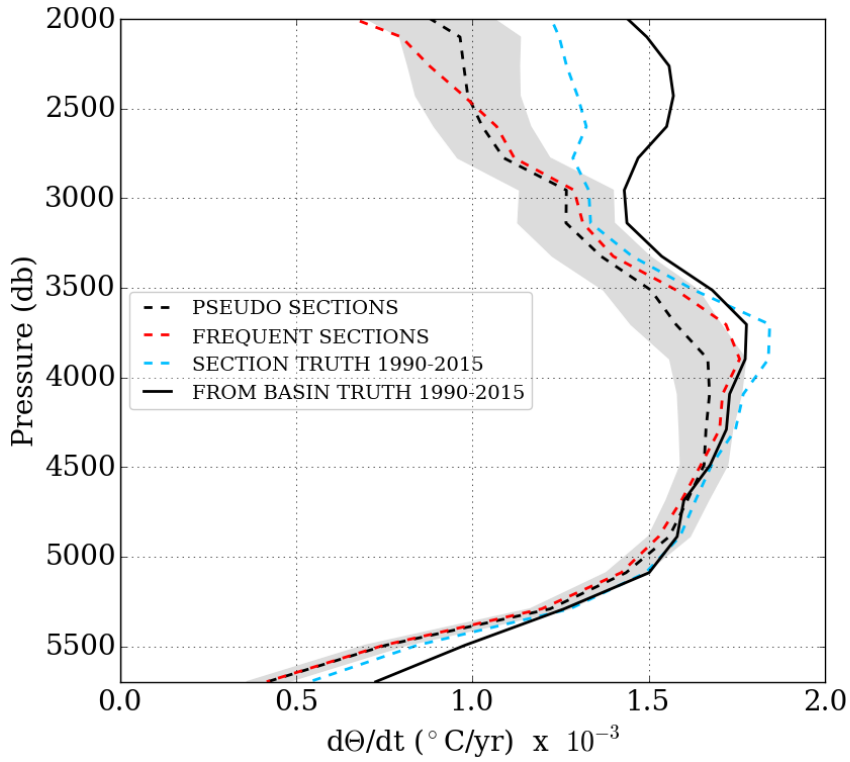
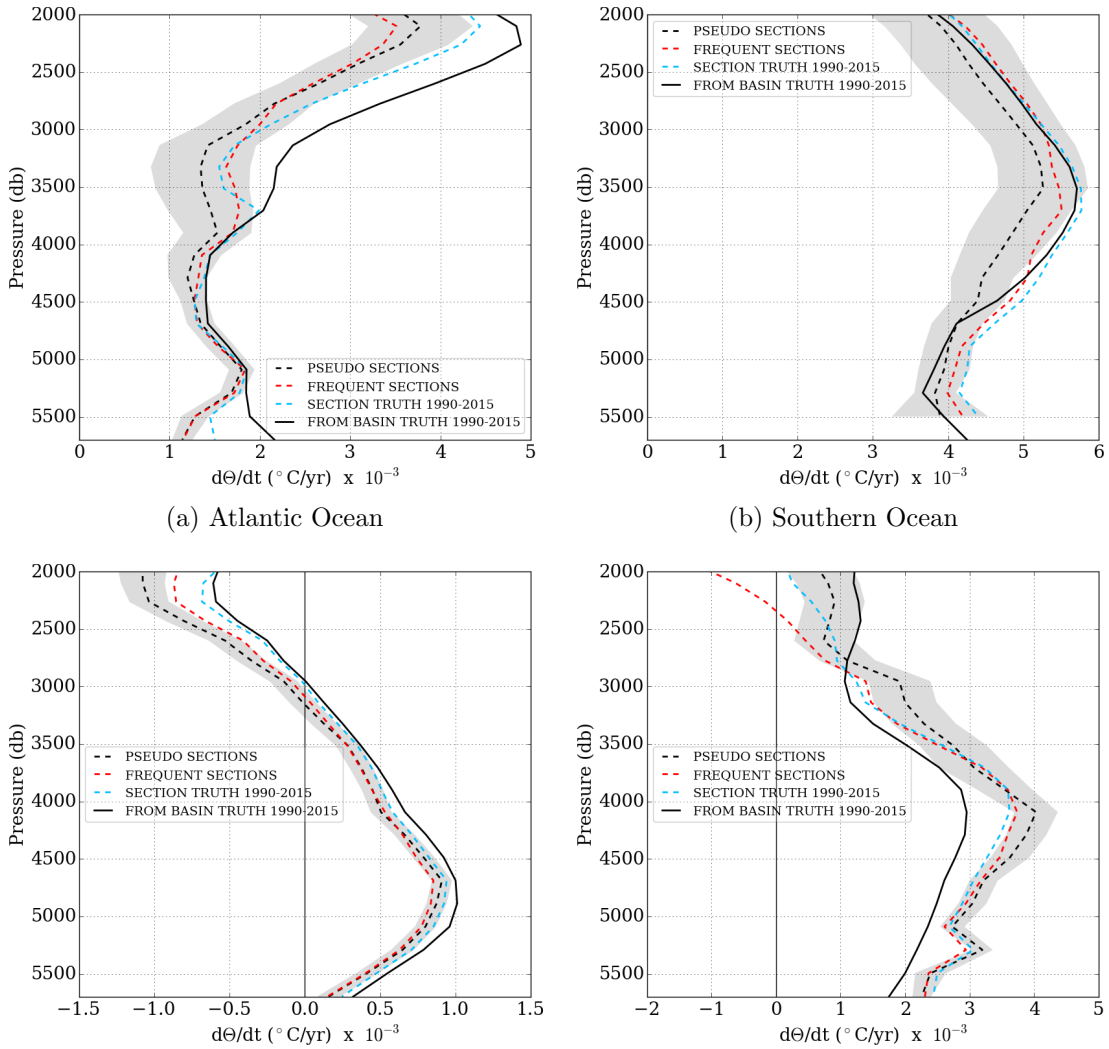


Figure 3.16: For the global oceans, horizontally averaged depth profiles of $d\Theta/dt$ calculated from basin truth (basin average data between 1990 - 2015), pseudo sections (model output along sections limited to timings of true observations as used in an updated analysis by Desbruyères *et al.* (2016b)), frequent sections (5 day mean model output along sections between the first and final years of section occupation) and from section truth (5 day mean model output along sections between 1990 and 2015 inclusive). In addition, in the uncertainty associated with averaging temperature trends horizontally along the sections (calculated for pseudo sections) is shown with grey shading reflecting the 95% confidence interval.



(a) Atlantic Ocean

(b) Southern Ocean

(c) Pacific Ocean

(d) Indian Ocean

Figure 3.17: For each ocean, horizontally averaged depth profiles of $d\Theta/dt$ calculated from basin truth (basin average data between 1990 - 2015), pseudo sections (model output along sections limited to timings of true observations as used in an updated analysis by Desbruyères *et al.* (2016b)), frequent sections (5 day mean model output along sections between the first and final years of section occupation) and from the section truth (5 day mean model output along sections between 1990 and 2015 inclusive). In addition, the uncertainty associated with averaging temperature trends horizontally along the sections (calculated for pseudo sections) is shown with grey shading reflecting the 95% confidence interval.

3.4.5 Using additional sections and timings up to 2015

Figures 3.16 and 3.17 provide the same analysis as Figures 3.12 (a) and 3.13, but using four additional sections and additional timings for any occupations that

occurred up to 2015, which replicates an updated observational analysis of deep OHC change (Desbruyères *et al.*, 2016b) in the model output. Globally changes to the trends from basin truth and section truth are very small (Figure 3.16). This indicates that globally spatial uncertainty has not been reduced significantly by the additional sections. However in the 2000 to 3000 m range, the basin trends estimated from frequent sections and pseudo sections are larger when using the additional sections and timings (Figure 3.16). This means that the temporal bias, and in particular that due to extrapolation, has been reduced by the addition of occupations up to 2015. However the pseudo section trend at 2000 m is around two thirds of that for the basin truth, so there is still a large error in the estimate of global heat content change from hydrographic sections.

When looking at the differences for each ocean by comparing Figure 3.17 to Figure 3.13, the results for the Indian Ocean are very similar and the results for the Southern Ocean and Pacific Ocean also have relatively small differences when additional sections and timings are used. The largest differences are seen in the 2000 - 3000 m Atlantic Ocean, in which there have been additional sections added to the analysis and additional occupations of sections during 2011 - 2015. The ocean trend estimated by pseudo sections at 2000 m is 0.0005 °C / year greater, and is therefore somewhat closer to the basin trend at this depth (just over 0.0045 °C/ year) than that for analysis following the methodology of Purkey and Johnson (2010) in Figure 3.13.

3.5 Discussion

3.5.1 Capturing trends on individual sections or in individual basins

For an individual section, temporally restricted data can represent the trends along the section very well, especially if data are collected on the section over the whole time period represented; sections where the pseudo section trend is calculated over just a few years and then extrapolated are often poorly represented. This suggests that inclusion of sections into basin estimates which

require sizeable extrapolation may introduce significant error. As only a few whole sections (mostly in the Atlantic) are really poorly represented by pseudo sections, parts of those sections will be particular culprits for skewing basin estimates. Some sections only cover a few years and hence do not represent the study period well, indicating a role of inter-decadal variability in the deep oceans; this could be due to natural variability in deep ocean temperatures and/or non-uniformly changing temperatures due to increased heat forced into the oceans under anthropogenic climate change.

Decadal variability in the deep ocean has been identified (for example) by Desbruyères *et al.* (2014), who find cooling in the North East Atlantic between 1988 and 2003 preceding warming between 2003 and 2013. The period of study here (1990 - 2010) is a strong candidate for decadal variability in heat content because the oceans played a key role in varying heat uptake in the 2000s such that surface temperature warming slowed considerably, although it is still not clear to what extent deep ocean changes are responsible for this 'hiatus'; Liu *et al.* (2016) argue deep ocean changes simply reflect the continued downward penetration of anthropogenic heat. To clarify deep ocean behaviour, higher temporal frequency sampling in the ocean is of clear value, particularly on sections covering only a few years of the study period.

There is no regional pattern or consistency as to whether section trends are over or under estimated by pseudo sections, and equally when section data are amalgamated into basins, no pattern has been identified whereby basin trends are systematically over or under estimated through the deep and abyssal oceans. However, there is no reason to think that there should be any systematic pattern here, so the result is not surprising. However, the element of randomness in whether trends are over or under estimated highlights the difficulty in identifying what additional sections or timings might improve the sampling. In addition, vertical changes in which biases are most important exist within the same basin.

It is also not uncommon for biases in a basin to be acting in opposite directions. Of the biases due to temporal uncertainty due to no data between occupations, temporal uncertainty due to extrapolation and spatial uncertainty, one might cause a cooling bias and another a strong warming bias. The opposing nature of biases and their cancellation or addition might result in a small overall bias with

the perception that the basin is well sampled by pseudo sections. However, in some cases 'solving' one of the biases, for example by providing high resolution temporal sampling on a section, may result in the estimated trend becoming less representative of the basin truth.

A global comparison of heat fluxes through 4000 m measured by observations and calculated in the model is made and although the pattern and relative strength of trends is broadly similar, there are distinct differences in the North Atlantic and the Indian oceans. An updated observational analysis to 2015, and using additional sections following Desbruyères *et al.* (2016b), finds some changes in regional trends to those in Purkey and Johnson (2010), specifically in the deep Pacific and Atlantic Oceans, and consider this reversal to be consistent with ideas of heat distribution change around 'hiatus' periods. Despite real world surface forcing the representation of OHC trends in present day hindcast simulations is not expected to be an exact match to the real world, as numerous factors which limit model development including limited understanding of heat transport pathways and ocean variability are unlikely to lead to models providing an exact manifestation of the oceans. However, model based results are a critical tool because only by using both observations with very large uncertainties and model data which is not a perfect representation of the real world, is the scientific community provided with the best chance of increasing our understanding of how best to measure the oceanic part of the climate system and improving the future design of both observations and models.

In the deep and abyssal layers, biases in heat flux may be of opposite signs even within the same basin, and some basins with a strong bias in one layer will show a small bias in the other layer. This highlights the structural variation vertically in temperature trends in a basin and that limiting data either spatially or temporally restricts our ability to capture that variation accurately.

Overall the picture that emerges when considering major uncertainties on sections or basins is one of randomness, where different uncertainties have different levels of importance in different places. Therefore the model indicates there is no clear way that global deep ocean trends could have been better measured with hydrography using a globally uniform approach. However, certainly in places where there are large biases, a basin-by-basin approach could be employed to

amend basin sampling strategies and thus improve the likelihood of capturing the basin truth.

3.5.2 Capturing trends globally or across oceans

When considering how well sections have captured the global trends below 3500 m the model suggests that hydrography does a very good job of capturing the trends in deep ocean temperature, though regionally it has already been established that biases can still be important below 4000 m. Regional biases in measured deep ocean temperature estimates may result in erroneous calculations for the thermal expansion contribution to sea level rise, skewing regional sea level predictions.

Above 3500 m biases result in a severe under-estimation of the trend and highlight that deep ocean temperature sampling should be improved substantially at this depth range (ideally with sampling improvements in every basin) to ensure trends are accurately captured. Globally the bias in pseudo sections could be considerably reduced by less reliance on extrapolation (making sure section occupations span the time period studied closely) and through reducing spatial bias.

In the 2000 - 4000 m layer, where the bias due to extrapolation is the largest bias, there must be significant inter-decadal variability on some sections and as already discussed, this is not always captured by occupations. In the Atlantic Ocean between 2000 - 4000 m temporal and spatial biases are sizeable, suggesting that Atlantic deep ocean temperature is highly variable in time and space, requiring higher frequency sampling both in time and space than can be provided using hydrography. In the Atlantic Ocean, section based estimates deviate from the vertical structure of the basin truth mostly above 3500 m and below 5000 m suggesting that it is in these depth ranges where more spatial sampling would be most beneficial.

In the 2000 - 4000 m layer the bias due to extrapolation is very large in the Atlantic, Southern and Indian Oceans, suggesting inter-decadal variability not captured by hydrography, but this bias becomes (relative to the other biases) much less important in the 4000 - 6000 m layer. This indicates that more

consistent monitoring of temperature in time across the whole study period is required in the 2000 - 4000 m layer, but is not so critical in the abyssal layer.

In the 4000 - 6000 m layer many of the biases are much smaller than in the 2000 - 4000 m layer, and especially globally, where biases are very small, however regionally some biases are still very large and therefore should not be considered unimportant (particularly for regional sea level rise prediction). In the abyssal ocean, insufficient data between occupations is relatively important in the Atlantic and Southern Oceans. This highlights variability in the abyssal Atlantic and Southern Oceans that is higher frequency than the time between occupations, so sampling limited in time (with no data between occupations) can alias the true trend. In the Indian Ocean the bias due to spatial sampling in the 4000 - 6000 m layer (larger than that in the 2000 - 4000 m layer) may be partly a result of limited hydrography in the deep Indian Ocean due to safety issues with sampling in the West Indian ocean. Yet more frequent sampling in space in that depth range could alone considerably improve temperature trend estimates. Relatively small biases in the Pacific indicate sampling there is representative compared to the other oceans, but there remain biases and improved sampling could improve the measurement of trends there. It is also worth noting that parts of the deep oceans with relatively small recent OHC changes or low variability may not remain that way as anthropogenic heat increasingly penetrates the deep oceans.

That the biases due to limited sampling can exceed the uncertainties from averaging along sections suggests that the uncertainty measure currently adopted for observational analyses of the deep ocean is not sufficient to account for the uncertainties in trends due to restricted sampling both in time and space. Another uncertainty developed here indicates that high frequency changes in deep ocean temperature mean that slight changes to pseudo section occupation timings can significantly change the calculated trends. It suggests that trends measured by hydrographic sampling could be a function of the exact timing of occupations and so it may not be possible to accurately capture trends with hydrographic sampling, which by its nature is temporally infrequent. The short timescale temporal variability in temperature identified here may be due to short term heave, where isopycnals move relative to depth, potentially oscillating up and down over time; this could be due to eddies or internal waves.

Possibly due to the model not capturing the variability of deep ocean temperatures sufficiently well, the uncertainties actually estimated by Purkey and Johnson (2010) have been found in this work to exceed the uncertainties (either total bias due to limited sampling, or sensitivity to precise timings) in the model for most basins in the 4000 - 6000 m layer. Therefore it may be that the observational uncertainties from averaging along sections are sufficiently large to capture the uncertainties considered in this work in the abyssal ocean. However, a caveat of this is that real world uncertainties relating to spatial and temporal variability may be larger than uncertainties that exist in model world.

The relative importance of each bias changes vertically and not just by basin, so trying to 'solve' a bias in a basin or ocean by an additional section or occupations is not necessarily going to improve the estimation of the trend at all depths. Instead the results here suggest that future sampling strategies would have most success by addressing all biases at all depths simultaneously. Deep ocean autonomous floats have the potential to simultaneously increase both the spatial and temporal sampling frequency.

In addition, the relatively low biases in the Pacific Ocean compared to the other basins suggest that there is merit in considering regional focus in the deployment of future deep ocean resources rather than providing a globally uniform distribution. This is because any future deep ocean temperature sampling strategy will be additional to the hydrographic section strategy examined here; continuation of hydrographic surveys is vital for the future of oceanographic research. Hydrography provides a comprehensive suite of oceanographic properties many of which are not able to be measured using autonomous float technology, and continuation of present sampling methods is vital to ensure comparability of temperature measurements through time. Critically, the very high accuracy hydrographic observations are the highest quality observations available, so are required for making high accuracy estimates of ocean temperature change as well as to ground-truth temperature data from other observing strategies (e.g. autonomous floats). Since globally the biases in deep ocean sampling using hydrography are much greater between 2000 - 4000 m than 4000 - 6000 m, deep Argo floats which sample to 4000 m may be largely sufficient for ensuring accurate monitoring of the global OHC change in the below 2000 m oceans. However, the inclusion of some floats measuring to 6000 m may be

prudent to address regional abyssal ocean sampling biases which may still be large.

3.5.3 Using additional sections and timings up to 2015

The analysis using the sections and timings up to 2015 (replicating the observational analysis by Desbruyères *et al.*, 2016b) does not change the conclusions made in this work based on the methodology of Purkey and Johnson (2010). Although there is some improvement to the global estimate of temperature trends from pseudo sections in the upper part of the deep ocean when additional sections and occupations to 2015 are included, there still remain large biases in the estimation of temperature trends from hydrographic style sampling, especially between 2000 and 3000 m, and large regional biases still remain at other depths. The improvement appears to be largely due to additional occupations in the Atlantic Ocean. Increased frequency temperature measurements both in time and space should be employed to reduce biases in temperature trends in the deep ocean.

3.6 Conclusions

The model replication of deep ocean observational sampling in this work indicates that temperature data from hydrographic sections between 1990 - 2010 give a reasonable representation of global temperature trends below 3500 m, but that large biases due to restricted temporal and spatial sampling may exist between 2000 - 3500 m. Overall, 82% of the temperature trend from 2000 m to the seafloor is captured from pseudo hydrographic sections in the model output, with hydrography capturing less than 60% of the trend at depths above 2500 m, but closely estimating trends below 3500 m. The model indicates that observational estimates of deep ocean temperature trends between 2000 - 3500 m may significantly underestimate real world warming in the deep oceans. In both the deep and abyssal oceans there can be strong regional spatial or temporal biases and therefore regional estimates of heat content change may not be dependable even in the abyssal ocean. Relatively large biases exist in the Atlantic, Indian and Southern Oceans, and therefore widespread amendment to deep ocean sampling is required. This is key for correctly measuring the thermosteric component of sea level rise and improving future sea level predictions.

Between 2000 - 3500 m biases due to temporal and spatial sampling remain important regardless of the precise timings of occupations; with current observational sampling frequency large biases exist globally, largely due to inadequate sampling in the deep Atlantic Ocean and abyssal Indian Ocean. The lack of spatial coverage in the deep oceans is a major barrier to sampling deep ocean temperature changes accurately in these locations, especially the abyssal Indian Ocean. Error due to trends on sections not being sampled throughout the whole time period and thus giving unrepresentative change signals due to inter-decadal variability is a major source of error globally in the deep oceans (2000 - 4000 m).

Removing seasonal biases in hydrographic sampling in the model improves the likelihood of capturing the section truth accurately, but the spatial biases and temporal bias due to extrapolating data to the study period still remain. The relative importance of different biases (due to either restricted temporal or spatial sampling) differ regionally and with depth, and therefore addressing all biases

simultaneously (for example, by using deep ocean autonomous floats) is likely required to ensure accurate future deep and abyssal ocean temperature observation.

Uncertainty estimates made by observational studies and in this chapter indicate that additional measurements of deep ocean temperature would increase the accuracy of ocean heat content change estimates over multi-decadal (20 - 30 year) timescales. However, when considering how to observe deep ocean heat content change it is also important to consider the internal variability of ocean heat content over longer timescales and how ocean heat content change might vary under projected 21st-century forcing. The next chapter provides a contribution to the scientific understanding of ocean heat content change and its spatial variability over longer timescales (over 100 years).

CHAPTER 4

The evolution of spatial patterns of ocean heat content in CMIP5 models and implications for deep ocean observing

4.1 Abstract

The oceans absorb the vast majority of heat accumulating in the climate system and so changes in ocean heat content (OHC) reflect global climate change. OHC changes influence global and regional sea level rise as well as ocean and atmosphere dynamics. The sparseness and infrequency of current ocean temperature measurements below 2000 m may hinder precise estimation of full depth OHC and change. This work uses global climate simulations from the Coupled Model Intercomparison Project 5 (CMIP5) to reveal the regions where below 2000 m temperature observations are necessary to accurately estimate OHC and adequately resolve the planetary energy imbalance over the twenty-first century (under the RCP 8.5 forcing scenario). First, a comparison is made between the locations and magnitudes of forced and internal variability in the CMIP5 models and observations of heat content trends between 1990 - 2010, revealing that CMIP5 models are a useful tool for predicting broad scale changes in OHC. Consistent with available observations, OHC variability and the emergent patterns of climate change in the deep oceans (below 2000 m) are shown to be dominated by signals in the Southern Ocean and Atlantic Ocean.

Variability captured by sampling to specific depths compared to full depth OHC is presented. The biases in the OHC estimate when sampling to specific depths and in certain deep ocean regions compared to full depth OHC are calculated for the twenty-first century, revealing the requirements of an OHC observing system capable of resolving the planetary energy imbalance. The results suggest that current and future observing strategies must include sampling below 2000 m as heat penetrates deep into the oceans, but that for the purposes of constraining the planetary energy imbalance over the coming decades, a key priority should be 2000 to 4000 m ocean observations in the Southern and Atlantic Oceans.

4.2 Introduction

The oceans are Earth's dominant reservoir of the heat accumulating in the climate system (Hansen *et al.*, 2011) and have absorbed at least 93% of Earth's energy increase since the 1970s (Rhein *et al.*, 2013). The rate of energy accumulation in the Earth system is a fundamental measure of the state of climate change. Modelling studies demonstrate that the global oceans also dominate the planetary energy budget with respect to internal variability on timescales of longer than a year (Palmer *et al.*, 2011; Palmer and McNeall, 2014). Changes in how the oceans take up and store heat are at least partly responsible for regional sea level changes and decadal surface temperature changes such as global surface warming 'hiatus' periods (e.g., the early twenty-first century) although the mechanisms and locations of heat uptake are still under debate (e.g., Chen and Tung, 2014; Drijfhout *et al.*, 2014; England *et al.*, 2014; Trenberth *et al.*, 2014; Lee *et al.*, 2015; Liu *et al.*, 2016).

Current satellite observations of incoming short-wave radiation and outgoing short- and long- wave radiation do not provide sufficient accuracy to generate absolute estimates of the net top of the atmosphere (TOA) energy imbalance (Hansen *et al.*, 2011). Accurate in-situ observations of ocean heat content (OHC) are therefore essential for constraining the evolution of the TOA energy imbalance due to external climate forcing and internal variability (Loeb *et al.*, 2012). It is widely considered that OHC measurements are too sparse in parts of the global oceans for accurate estimation of OHC, especially below 2000 m. This work examines the spatial patterns of OHC change and how they vary with depth using climate model simulations, and analyses how this can inform ocean observing system design.

In recent decades ocean warming has been observed through repeat hydrographic ship sections (full-depth), by expendable bathythermographs (XBTs, 0-700m) and most recently by the 0-2000 m Argo array. Estimated rates of recent warming between 0 and 700 m range from 0.27 Wm^{-2} over the period 1955 - 2010 (Levitus *et al.*, 2012) to 0.64 $\pm$ 0.29 Wm^{-2} between 1993 - 2008 (Lyman *et al.*, 2010). In the 0 - 2000 m layer, estimates include 0.39 Wm^{-2} over the period 1955 - 2010 (Levitus *et al.*, 2012) and 0.4 - 0.6 Wm^{-2} between 2006 -

2013, with half the warming below 700 m (Roemmich *et al.*, 2015). Wijffels *et al.* (2016) note that the heat anomaly in the 0 - 2000 m layer is not uniform globally; the southern hemisphere warms 4 times faster than the northern hemisphere with the majority of heat accumulating between 30°- 50 °S.

Despite the rapid expansion of OHC observing during the last 20 years, parts of the upper ocean (here defined as 0 - 2000 m) as well as much of the deep ocean (below 2000 m) may not be sufficiently well sampled to accurately estimate heat content change for climatic purposes such as regional thermosteric sea level prediction or constraining the planetary energy imbalance, now or in the future. Durack *et al.* (2014) suggest that 0 to 700 m observational estimates are likely biased low because of poor sampling in data sparse regions such as the Southern Ocean. Gleckler *et al.* (2016) find more than a third of heat going into the ocean is accumulating below 700 m. The recent deployment (during 2000 - 2004) of the 0-2000 m Argo array vastly improved temporal and spatial sampling resolution above 2000 m (e.g., Abraham *et al.*, 2013; Riser *et al.*, 2016) although its use is limited for sampling sea ice covered regions, continental shelves or where swift western boundary currents exist. Although OHC estimates above 2000 m over the last decade rely heavily on the Argo array, Lyman and Johnson (2014) reveal that assumptions regarding areas poorly sampled by 0-2000m Argo can significantly affect global OHC estimates and their trends.

Uncertainties in upper ocean observing systems currently prevent inference of temperature and salinity changes very far beneath 2000 m (Von Schuckmann *et al.*, 2014). Instead, estimates of deep OHC rely on repeated hydrographic (ship-based) observations; the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP; www.go-ship.org) provides a well-planned distribution of repeated hydrographic sections. Deep ocean observations indicate global warming between 1990 - 2010 of 0.068 ± 0.061 Wm^{-2} below 2000 m, 0.053 ± 0.031 Wm^{-2} below 3000 m and 0.027 ± 0.009 Wm^{-2} below 4000 m (Purkey and Johnson, 2010). Desbruyères *et al.* (2016b) also calculated warming of 0.065 ± 0.040 Wm^{-2} below 2000 m during 2004 - 2014. In contrast, between 2005 - 2012, Llovel *et al.* (2014) infer a cooling (OHC change of - 0.08 ± 0.43 Wm^{-2}) below 2000 m using upper ocean and satellite measurements, but uncertainties in this method are large (Johnson and Lyman, 2014). A full depth estimate of ocean heat uptake during 2003 - 2012 by

Desbruyères *et al.* (2016b) using Argo data and hydrography is 0.71 ± 0.12 Wm^{-2}.

These deep estimates of OHC are obtained using temperature data from repeated ship based hydrographic sections that are limited in time and space (there are few repeated sections in each ocean and they are typically repeated only every 2-10 years). Earlier in this thesis it has been demonstrated that estimates of OHC change from hydrographic sections may not be representative of the true trends in recent decades. Prior to this, the scientific community has identified a necessity to improve the resolution of deep ocean measurements to accurately capture full-depth changes in OHC and quantify steric sea level changes (e.g., Von Schuckmann *et al.*, 2014; Wunsch and Heimbach, 2014). Accurate deep OHC estimates will enable sea level and planetary energy budgets to be verified; currently budgets are estimated from upper ocean and satellite measurements but confidence in budgets can be increased through cross checking of ocean mass, ocean warming and sea level measurements (Johnson and Lyman, 2014). Deep ocean Argo floats capable of sampling to 4000 m or 6000 m exist in small numbers and are under continued development. Carefully considered deployment strategies will allow efficient and effective detection of subtle changes in the deep oceans and maximise value during the early stages of deployment.

In an ideal world a uniform high resolution deep ocean observing system would be deployed (a uniform design for capturing decadal change with unprecedented accuracy is presented in Johnson *et al.*, 2015). However achieving funding for both the implementation and maintenance of ocean observing systems is not easy; political difficulties are summarised in Cristini *et al.* (2016). The current observing systems of GO-SHIP and upper ocean Argo are projected to decline in size over the period 2016-2018 (Durack *et al.*, 2016). If funding is limited, choices about locations for maximum value for money must be made based on current scientific understanding.

Here, global climate simulations are used to investigate the spatial patterns of OHC under internal variability and forced climate change scenarios. From an understanding of the regions where OHC changes due to internal variability and/or anthropogenic forcing, and how the patterns of OHC are likely to evolve over time, this work evaluates whether and where deep ocean observations are

needed to accurately estimate and capture future changes in full-depth OHC.

4.3 Models

The fifth Coupled Model Intercomparison Project (CMIP5, Taylor *et al.*, 2012) includes simulations from a range of climate models from over 20 modelling centres. Although not as high resolution as some state-of-the-art ocean and climate model simulations, at the present time CMIP5 provides the best freely available model simulations for answering questions about long term variability in the climate system and for predicting future changes to the climate system. CMIP5 models are high resolution climate models with long control simulations, hindcasts over the industrial period, and with future scenario runs over the coming century (and in some cases beyond). The range of simulations enables investigation of variability in the coupled climate system on multi-decadal time scales and the prediction of future changes during the rest of this century.

Cheng *et al.* (2016) evaluate global OHC in many of the CMIP5 models, finding that although the CMIP5 models have a large spread in OHC change, the ensemble median is in excellent agreement with observational estimates. Gleckler *et al.* (2016) separately also determined that there is very good agreement between the mean global deep OHC change in CMIP5 models and that from the observations made by Purkey and Johnson (2010). Climate models are an imperfect representation of the real world (in part due to our limited understanding of the Earth system) but they are an appropriate tool for predicting how the climate and oceans will evolve in the coming decades and to estimate how observational requirements will change.

This work focusses on four models (HadGEM2-ES, MPI-ESM-MR, GISS-E2-R and GFDL-CM3, see Table 4.1) from the CMIP5 archive. By selecting models that encompass a range of behaviours in the CMIP5 archive, the intention is to provide a spread of results that encompasses the behaviour of the majority of the archive, whilst presenting an in-depth study of several models. The models selected were developed at different modelling centres to reduce the likelihood of shared model features.

The decadal to centennial variability of global OHC in the control simulations of 22 CMIP5 models was examined (Figure 4.1). Figure 4.1 illustrates that the

Table 4.1: CMIP5 models analysed in this work

Model	Institute/Country	References
HadGEM2-ES	Met Office Hadley Centre, UK	Collins *et al.* (2011) Jones *et al.* (2011)
MPI-ESM-MR	Max Planck Institute for Meteorology, Germany	Jungclaus *et al.* (2013)
GISS-E2-R	Goddard Institute for Space Studies, USA	Schmidt *et al.* (2014)
GFDL-CM3	Geophysical Fluid Dynamics Laboratory, USA	Griffies *et al.* (2011)

global full-depth OHC variability of the four models chosen (highlighted in red) differ in their characteristics, with GFDL-CM3 being one extreme outlier in the archive. Cheng *et al.* (2016) analyse full-depth OHC change in CMIP5 models during the periods 1970-1991 and 1992-2005, and of 24 models examined, GISS-E2-R and MPI-ESM-MR are included in the top 7 greatest OHC change during 1970-1991, while GFDL-CM3 and HadGEM2-ES are both in the four models with the smallest OHC change over the same period. The selected models encompass most of the range between the maximum and minimum values from all models.

The heat uptake efficiency was examined for a range of CMIP5 and CMIP3 (the precursor to CMIP5) models by Kuhlbrodt and Gregory (2012). Heat uptake efficiency is a measure of heat taken up by the oceans from the atmosphere and so a realistic value of this may indicate a realistic coupled climate system in a model. GISS-E2-R is not included in the study but of CMIP5 models considered, both HadGEM2-ES and MPI-ESM-MR have relatively small heat uptake efficiencies (both in the four smallest) (Kuhlbrodt and Gregory, 2012). In contrast the ocean heat uptake efficiency for GFDL-CM3 is toward the upper end of the CMIP5 models. The strength and variability of the Atlantic Multi-decadal Oscillation (AMO) and Atlantic Meridional Overturning Circulation (AMOC) in CMIP5 models was considered by Zhang and Wang (2013); the AMO index for GISS-E2-R was on the lower end of the 28 models considered, whilst HadGEM2-ES and GFDL-CM3 were at the upper end (MPI-ESM-MR not included in the study).

Mean transports during 1986-2005 by the AMOC and Southern Meridional Overturning Circulation (SMOC) were calculated by Heuzé *et al.* (2015). GFDL-CM3 and GISS-E2-R have the third and fourth largest transports of the AMOC (defined as the maximum streamfunction at 30°N) of the 25 models

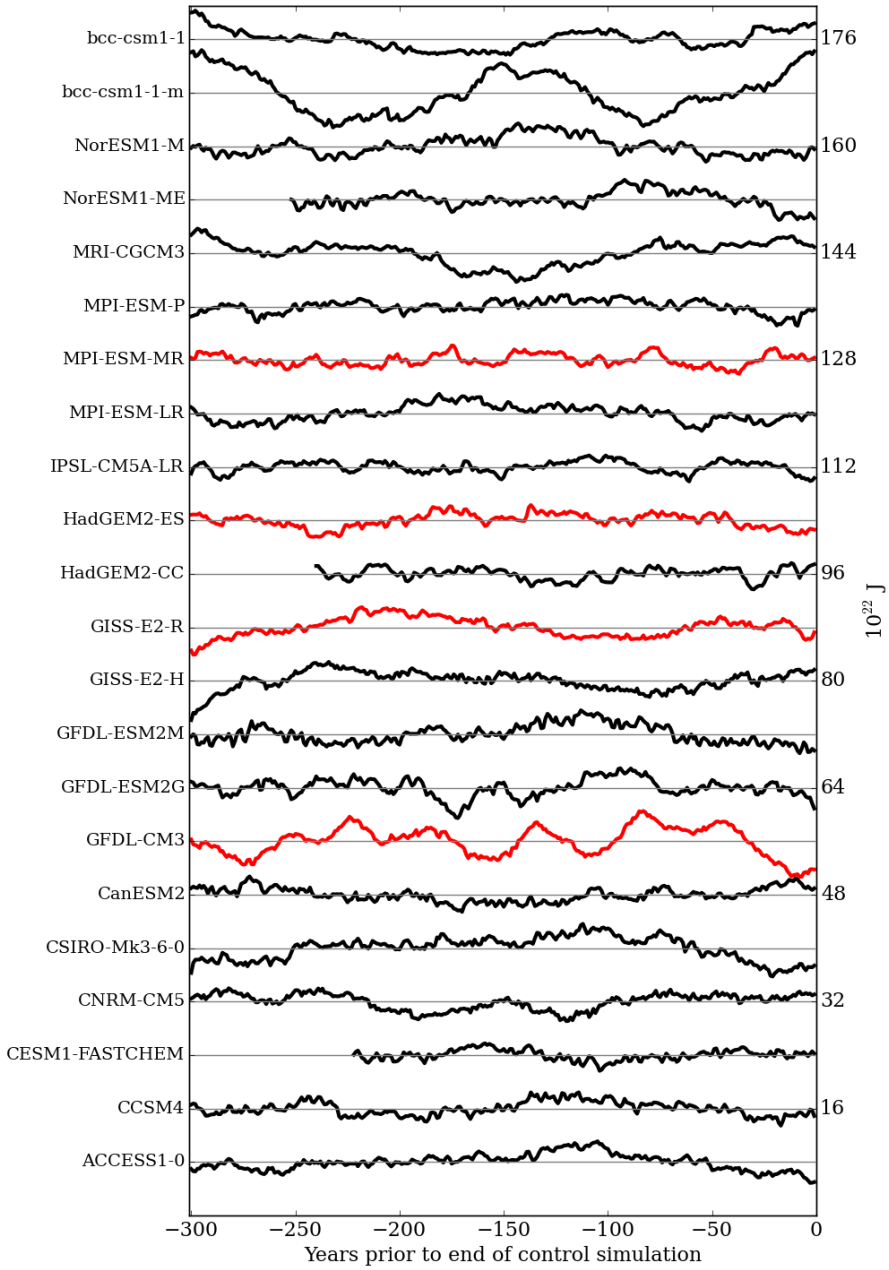


Figure 4.1: The relative variability of global OHC in the last 300 years of the control simulation of 22 CMIP5 models. Each model has an offset of some multiple of 8 $\times 10^{22}$J to best illustrate the patterns of OHC change. The data are detrended linearly. The models used for the remainder of this work are plotted in red.

studied; both have an AMOC of 21 Sv, compared to a multi-model mean of 18 Sv ($\pm$ 5 Sv). The transport in HadGEM2-ES is 17 Sv, close to the multi-model mean. MPI-ESM-MR exhibits the lowest mean transport of all the studied models at 10 Sv. For the SMOC at 30°S, the multi-model mean is 10 Sv ($\pm$ 7

Sv). GFDL-CM3 has the fifth smallest transport (3 Sv), that for MPI-ESM-MR (9 Sv) is close to the multi-model mean, and the transports in GISS-E2-R and HadGEM2-ES (both 17 Sv) are both in the five greatest of all the models studied.

Heuzé *et al.* (2013) evaluated Southern Ocean bottom water characteristics in 15 CMIP5 models and although the study overlaps with only two of the selected models, it shows that the HadGEM2-ES model has bottom temperature and densities that are close to those observed; HadGEM2-ES mean bottom temperature is between 0-1°C warmer than climatology, whilst GISS-E2-R is fairly poor at representing observations (particularly for temperature). Heuzé *et al.* (2013) conclude that HadGEM2-ES is one of the best performing models in terms of Antarctic Bottom Water properties, despite deep water formation mechanisms in CMIP5 models depending on open ocean convection (though most modern deep water is formed on continental shelves).

The models selected have known biases, some of which are common biases amongst the current generation of climate models. Some of the specific known biases for the models used in this work are outlined below. All state-of-the-art complex models, whether coupled ocean-atmosphere-sea ice or Earth System models, neglect some Earth System processes (some of the reasons for which include lack of understanding of the key processes and computational power limitations).

Biases in the Southern Ocean in CMIP5 models (highlighted by Heuzé *et al.*, 2013) are likely to be partly due to the way deep water formation occurs in models. For example, deep mixed layers and active convection in parts of the Southern Ocean such as the Weddell Sea often feature in models, but are scarcely observed. Models are thought to overestimate the frequency of modern open ocean convection while underestimating (or not having the capacity to represent) the frequently observed on-shelf dense water formation and dense water cascades that form modern deep water. Like many climate models, HadGEM2-ES has a warm Southern Ocean bias resulting from atmospheric biases (Sallée *et al.*, 2013; Hyder *et al.*, 2016).

Known biases in MPI-ESM-MR include the ocean being too cool and fresh in the upper layers and too warm and saline in the intermediate and deep ocean

(Jungclaus *et al.*, 2013). The North Atlantic Current is too zonally orientated resulting in a cold and fresh bias in the North Atlantic upper ocean in the model (Jungclaus *et al.*, 2013). There is an equatorial cold bias in the MPI-ESM-MR Pacific and upwelling there is not represented correctly, while the maximum Atlantic Meridional Overturning Circulation streamfunction is slightly weak, but with maximal meriodional heat transport close to observations (Jungclaus *et al.*, 2013).

Griffies *et al.* (2011) highlights that GFDL-CM3 has warmer than observed oceans (including the abyssal oceans). The Atlantic basin is too salty, the North Pacific gyre is slightly salty, while there are large fresh biases near high latitude coasts and inland seas in GFDL-CM3 (Griffies *et al.*, 2011). The deep counterclockwise convective overturning cells in the Indian and Pacific Oceans of GFDL-CM3 are much weaker than observed (Griffies *et al.*, 2011).

Schmidt *et al.* (2014) identify a warm bias, excessive Antarctic Circumpolar Current and large sea surface temperature and sea ice biases in the Southern Ocean of GISS-E2-R. The Gulf Stream separation point is incorrect, but the Atlantic Meridional Overturning Circulation streamfunction is close to observational estimates, although its maximum may be a little high (Schmidt *et al.*, 2014). The GISS-E2-R Arctic is too fresh and there are other high salinity and coastal temperature biases linked to insufficient marine stratus clouds (Schmidt *et al.*, 2014).

Wang *et al.* (2014) show that regional biases in CMIP5 models may be linked to other remote biases (for example, weakening of the Atlantic Meridional Overturning Circulation may be related to warm sea surface temperature biases in the Southern Ocean and excess Antarctic Bottom Water Formation). Solving regional biases locally may not, therefore, be especially useful (Wang *et al.*, 2014), emphasizing the complexity involved in reducing model bias. In addition, even models that do not appear to exhibit large biases in certain respects have been found to be a good representation of the climate system for the wrong reasons.

4.4 Method

4.4.1 Aim and clarification of associated issues in OHC sampling that are not covered

The aim here is to test the extent of vertical sampling required now and in the future, and the broad basin areas where sampling in the deep ocean should be prioritised. Throughout this work model data interpolated onto a 1° by 1° resolution grid are used, and so the method does not test adequacy of horizontal sampling of OHC. The method does not consider importance of ice-cover or shallow seas, rather assuming sampling above 2000 m in the modern observing system to be perfect; in reality 0 - 2000 m observing would be improved with additional sampling in marginal seas, under sea-ice and in hard to access parts of the ocean (e.g., McKinnon and Huybers, 2016). Neither does this work present details of coverage or realistic deployment strategies for future observing systems.

4.4.2 Calculating OHC

For each selected model, this work uses output from the CMIP5 control (typically longer than 400 years), historical (years 1850 to 2005) and representative concentration pathway (RCP) 8.5 (years 2005 to 2100) simulations. Annual means had been calculated from monthly mean ocean potential temperature (θ, henceforth 'temperature') model output at the Met Office. To facilitate inter-model comparison, temperature data had been regridded at the Met Office to a regular 1° latitude by 1° longitude horizontal grid. To reduce any influence of model drift, data from the historical and RCP 8.5 simulations were detrended (grid-point by grid-point) by subtracting the linear trend in the companion control simulation. A recent study by Cheng *et al.* (2016) detrend global OHC data using both a linear and quadratic fit and determine that there is little difference in using either. Consideration of global OHC in a range of models including those selected (Figure 4.1) also suggests that a linear fit is appropriate.

OHC between two depths (z_{upper} and z_{lower}) is calculated from depth-weighted

integrated temperature in each model layer (z_n to z_{n+1}), the density of seawater (ρ, assumed constant 1025 kg m^{-3}), the specific heat capacity of seawater (cp, assumed constant 4000 J $kg^{-1}K^{-1}$) and the horizontal area, A, of the water column (m^2), as follows:

$$\mathrm{OHC}_{z_{upper}-z_{lower}} = \int_{z_{upper}}^{z_{lower}} \theta(z)\ \rho\ c_p\ A\ dz \quad . \tag{4.1}$$

$\mathrm{OHC}_{z_{upper}-z_{lower}}$ reflects OHC in an ocean layer (e.g., OHC_{0-700m} reflects OHC integrated from the surface to 700 m). OHC in three distinct ocean layers, OHC_{0-700m}, $\mathrm{OHC}_{700-2000m}$ and $\mathrm{OHC}_{2000m-seafloor}$ are frequently referred to in this work. In calculating $\mathrm{OHC}_{z_{upper}-z_{lower}}$, the nearest depth layer in the model to the desired depth is used. Across all the models, the maximum difference between the desired depth and the actual depth layer used is 6% of the magnitude of the desired depth.

4.4.3 Variability in models compared to observations during 1990 - 2010

For each model, the linear trend in OHC_{0-700m}, $\mathrm{OHC}_{700-2000m}$ and $\mathrm{OHC}_{2000m-seafloor}$ during the period 1990 - 2010 was calculated from the relevant years of the historical and RCP 8.5 scenarios. For each layer the linear trend in observational data over the same time period is also presented. For 0-2000 m observations are taken from the EN4 dataset provided by the Met Office Hadley Centre, which comprises global gridded fields of potential temperature at 1° resolution with 42 vertical levels (Good *et al.*, 2013). Below 2000 m the observational trends in OHC are those calculated by Purkey and Johnson (2010). The time period 1990 - 2010 is chosen because it is the time period during which sufficient data were collected from below 2000 m for Purkey and Johnson (2010) to calculate regional trends in deep and abyssal OHC.

To identify regions of internal variability in the models, all overlapping 21 year linear trends in OHC in the control simulation are calculated, and the standard deviation of these trends is presented for OHC_{0-700m}, $\mathrm{OHC}_{700-2000m}$ and

$OHC_{2000m-seafloor}$. This highlights regions where there exists high internal variability without anthropogenic forcing; control simulations represent only the variability that naturally emerges from the coupled atmosphere-ocean nature of the climate system.

For each model, regions below 2000 m where the magnitude of the linear trend during 1990 - 2010 exceeds 2 standard deviations of the overlapping 21 year trends from the control simulation are highlighted using hatching. This identifies regions which may be subject to anthropogenic forcing in the deep oceans during the period 1990 - 2010.

4.4.4 Using EOF analysis to identify modes of variability

For each model key regions of spatial variability in deep OHC are identified using empirical orthogonal function ('EOF') analysis. The matrix of OHC beneath 2000 m through time in the control scenario is partitioned into mathematically orthogonal modes. This technique may yield modes which are interpretable using oceanographic understanding, although as a statistical technique, modes may be generated from noise in the data rather than due to multi-decadal oceanographic phenomena.

In EOF analysis, the detrended matrix of OHC data, D, is decomposed (using singular value decomposition) into three matrices:

$$D = USV^t \tag{4.2}$$

U and V^t are orthonormal matrices (where U is a matrix of principal component weights and V is a matrix of standardized-rescaled principal component scores) and S is a diagonal matrix of singular values. The matrix, V, comprises the spatial patterns of each mode of variability, whilst the principal component time series are found by multiplying U and S.

Typically most of the variance associated with atmospheric and oceanographic processes is contained within the first few modes calculated by EOF analysis, but modes should not be considered independently, and conclusions should not be

drawn without examining variability in a dataset using other methods (Monahan *et al.*, 2009). Examining principal component time series alongside spatial patterns generated from the analysis reveals the period over which a mode of variability occurs. A comprehensive overview of EOF analysis and its use in climate science can be found in Chapter 13 of Von Storch and Zwiers (2001).

For this work, the leading EOFs are first calculated globally, identifying the North Atlantic and Southern Ocean as the key regions. However, EOFs calculated globally might result in local patterns of variability generated by remote forcing (due to variability in another part of the global ocean). To constrain the patterns of variability to reflect sources of variability within the local region, OHC in the local region is restricted and then the EOFs are calculated for the local region. The leading two EOFs are presented for each model in the North Atlantic and Southern Oceans (where the latter represents all ocean below 40°S).

4.4.5 Emergent signals of ocean climate change

To consider potential future signals in OHC_{0-700m}, $\mathrm{OHC}_{700-2000m}$ and $\mathrm{OHC}_{2000m-seafloor}$ the RCP 8.5 simulation is considered. This simulation includes forcing from rising levels of greenhouse gases in the atmosphere. By considering linear trends in $\mathrm{OHC}_{z_{upper}-z_{lower}}$ during the period 2010 - 2100 at each grid-point, the regions where the largest signals of twenty-first century climate change are likely to emerge are identified.

4.4.6 How deep do we need to measure to resolve full depth OHC?

To estimate the depth at which full-depth OHC, OHC_{total}, can be accurately estimated from partial-depth OHC, OHC_{0-z}, a fraction of variability metric is used. This is presented for both the control and RCP 8.5 simulations. Fraction of variability, F, is calculated from the standard deviations, σ, of the difference between OHC_{total} and OHC_{0-z}, and of OHC_{total} as follows:

$$\mathrm{F} = 1 - \frac{\sigma(OHC_{total} - OHC_{0-z})}{\sigma(OHC_{total})} \quad . \tag{4.3}$$

Calculated for each depth, z, this metric produces a number between 0 and 1 for each grid-point. Locations where the full-depth OHC is well represented by partial-depth OHC generate a fraction of variability close to 1, whilst poorly represented locations give numbers closer to 0. A particular water column is considered sufficiently well observed by partial-depth OHC when fraction of variability > 0.95.

4.4.7 Estimating planetary energy imbalance from OHC

Simple observing scenarios are used to evaluate the large-scale regions and depths that are most important from a budget perspective for estimating the planetary energy imbalance from OHC. The ocean is sub-divided by depth range and ocean basins to generate the following measurement scenarios: to OHC_{0-700m} globally, to $\mathrm{OHC}_{0-2000m}$ globally, to $\mathrm{OHC}_{0-2000m}$ globally with additional $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$ in the Southern Ocean (defined by the deep basins of Purkey and Johnson, 2010, and in Figure 3.1 of this thesis), to $\mathrm{OHC}_{0-2000m}$ globally with $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$ south of 40°S and in the Atlantic Ocean and to $\mathrm{OHC}_{0-4000m}$ globally.

The bias in planetary energy imbalance is estimated for each scenario by calculating the difference between each 10 year trend in OHC measured by each of these scenarios and the equivalent 10 year trend in OHC_{total}. How biases evolve with greenhouse gas emissions forcing is determined by calculating biases for both the historical and RCP 8.5 simulations. This highlights if measurement in certain regions of the deep oceans can effectively provide sufficient additional OHC information to enable accurate estimation of the TOA energy imbalance by considering in which scenarios biases do not exceed 0.1 Wm^{-2}.

4.5 Results

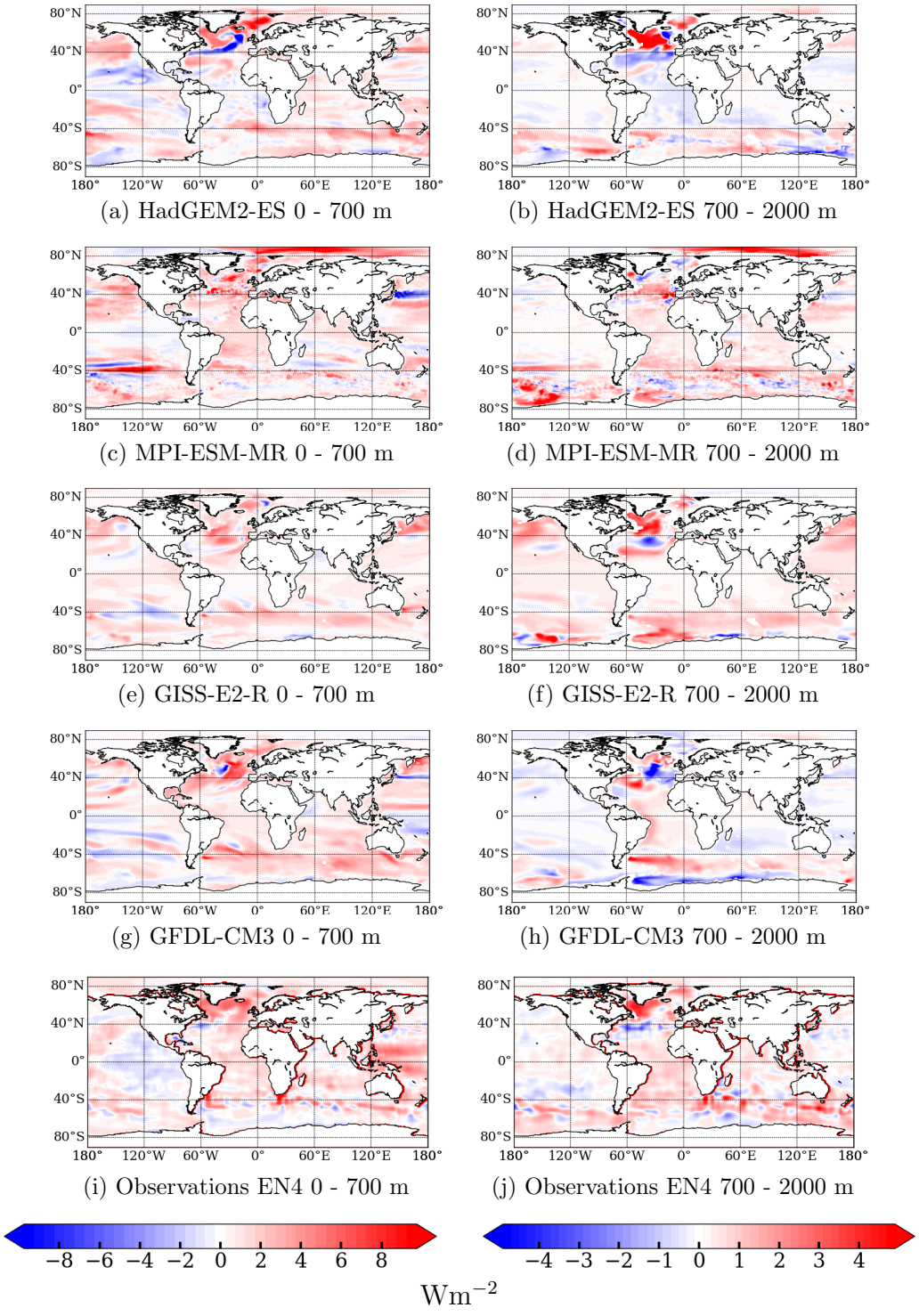
4.5.1 21 year trends in OHC in models and observations

Firstly the linear trend in OHC_{0-700m}, $OHC_{700-2000m}$ and $OHC_{2000m-\text{seafloor}}$ during the time period 1990 - 2010 is presented for each model and from observations. For OHC_{0-700m} (Figure 4.2, left column), some of the greatest trends are focussed in the subpolar regions, along regions of strong currents or frontal regions of the oceans, e.g., the Gulf Stream and North Atlantic Current at 45-50°N, 20-50°W and the Kuroshio at 35-45°N, 140-170°E. Patterns of warming and cooling across each ocean vary between models and are less spatially consistent than in observations, except in the Southern Ocean where MPI-ESM-MR reflects the variability across the oceans in observations relatively well, but the other models show greater spatial coherence.

However, the models generally exhibit some trend of similar magnitude to those in observations in most locations between 0-700m; an exception is the western tropical Pacific, where the strong observed warming trend is not adequately reproduced by any model (Figure 4.2). Decadal trends in OHC in areas such as the upper ocean Pacific are likely to be heavily influenced by internal climate oscillations such as El Niño, which do not occur in climate model simulations at the observed timings, as internal variability in these models is not constrained by observed temperatures.

Alongside this analysis, the standard deviations of all overlapping 21 year trends in OHC_{0-700m} in the control simulation are presented in Figure 4.3 (left column). Figure 4.3 highlights regions in each model which are subject to large internal variations in OHC. It suggests that a given 21 year trend in OHC_{0-700m} is likely to include internal forcing across most of the oceans, with relatively small values in the tropics and the largest values in subpolar regions (particularly the North Atlantic).

For $OHC_{700-2000m}$ (Figure 4.2, right column) strong trends are observed across models and observations in the North Atlantic, Southern Ocean and to a lesser



(a) HadGEM2-ES 0 - 700 m

(b) HadGEM2-ES 700 - 2000 m

(c) MPI-ESM-MR 0 - 700 m

(d) MPI-ESM-MR 700 - 2000 m

(e) GISS-E2-R 0 - 700 m

(f) GISS-E2-R 700 - 2000 m

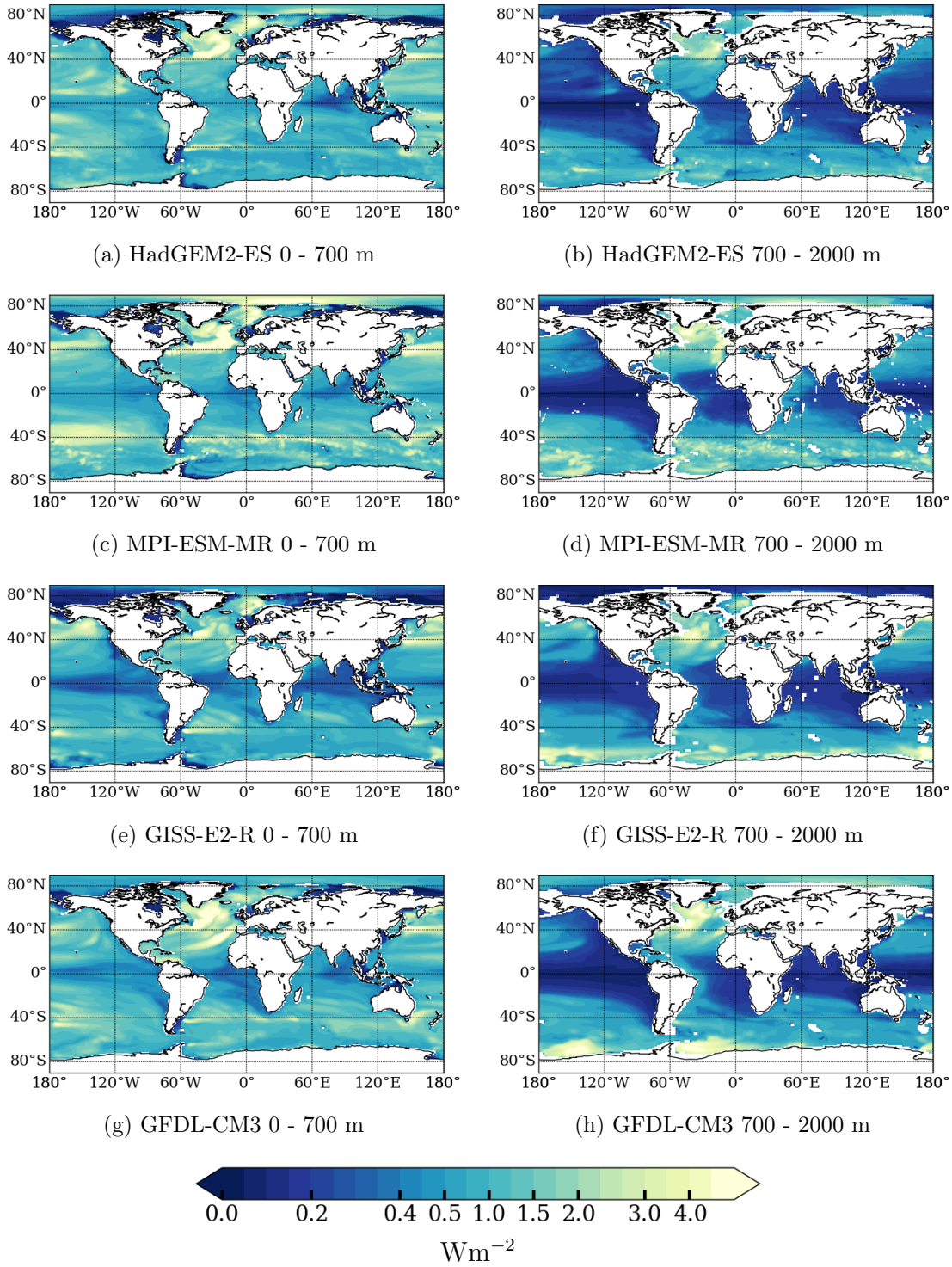
(g) GFDL-CM3 0 - 700 m

(h) GFDL-CM3 700 - 2000 m

(i) Observations EN4 0 - 700 m

(j) Observations EN4 700 - 2000 m

Figure 4.2: Linear trend in OHC in the 0 - 700 m and 700 - 2000 m layers during the period 1990 - 2010 inclusive, presented as a heat flux (Wm^{-2}) into the layer. Note the different colour scales for OHC_{0-700m} and $\mathrm{OHC}_{700-2000m}$.



(a) HadGEM2-ES 0 - 700 m

(b) HadGEM2-ES 700 - 2000 m

(c) MPI-ESM-MR 0 - 700 m

(d) MPI-ESM-MR 700 - 2000 m

(e) GISS-E2-R 0 - 700 m

(f) GISS-E2-R 700 - 2000 m

(g) GFDL-CM3 0 - 700 m

(h) GFDL-CM3 700 - 2000 m

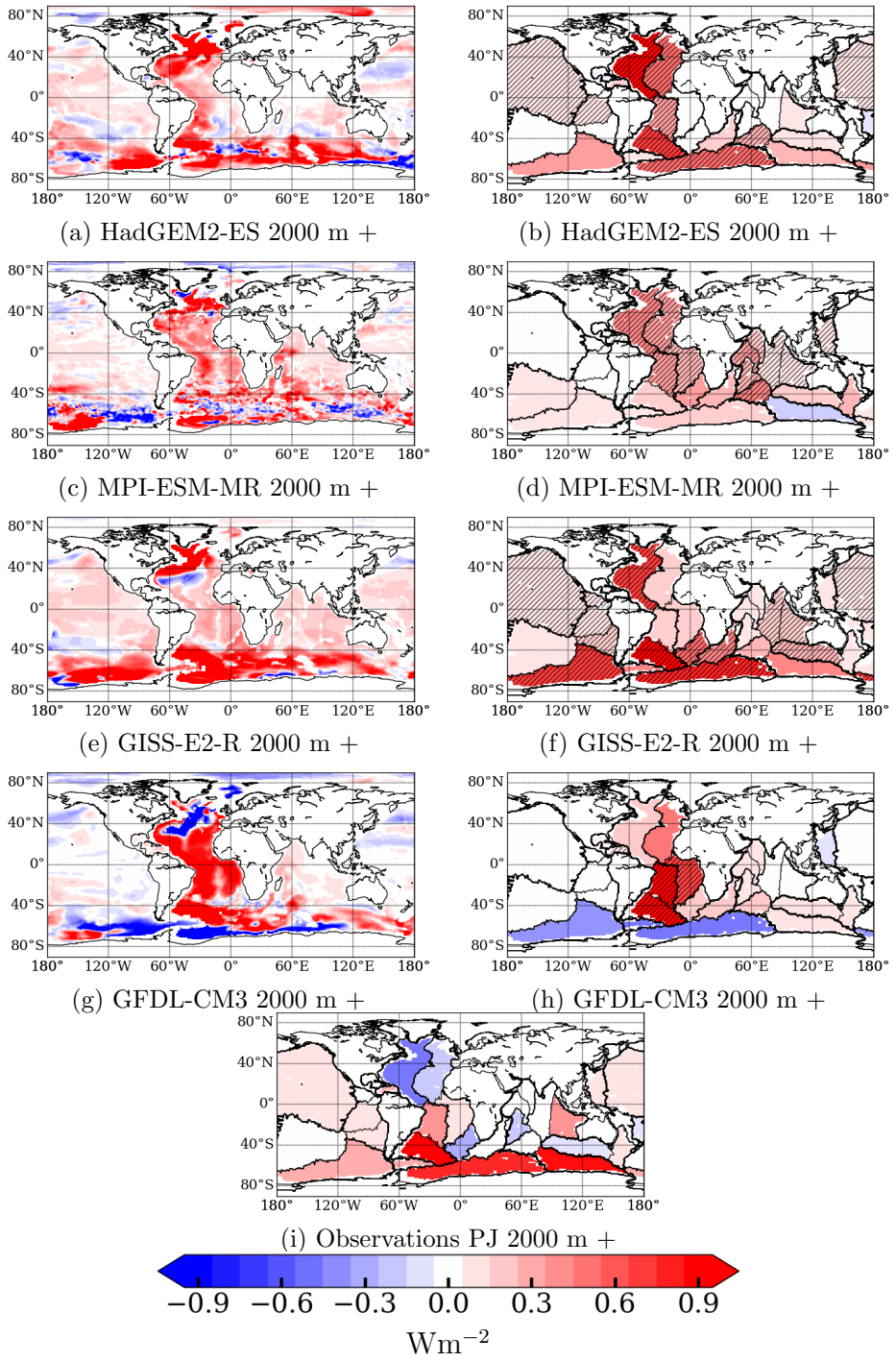
Figure 4.3: σ of the linear trends in OHC for all overlapping 21 year periods in the control simulations, calculated for both the 0 - 700 m (left column) and 700 - 2000 m (right column) layers. The same colour scale is used for each OHC layer.

extent the South Atlantic and Indian Oceans. The strong warming / cooling in the North Atlantic is reproduced very well in some models (HadGEM2-ES and GISS-E2-R) and is reproduced to some extent by all the models. The main region in which there are major differences between the magnitude of trend in a model compared to observations is the Pacific in GISS-E2-R, which shows warming across the basin which does not exist in observations (Figure 4.2, right column). This may be related to an unrealistic Pacific overturning in the GISS-E2-R model.

Considering the standard deviation of all 21 year overlapping trends in $\mathrm{OHC}_{700-2000m}$ (Figure 4.3, right column) in GISS-E2-R, values are low in the Pacific, which may suggest that the warming is due to anthropogenic forcing in that model. Otherwise, the regions of large trends in $\mathrm{OHC}_{700-2000m}$ (Figure 4.2, right column) are associated with regions of large internal forcing, that is, where the standard deviation in all overlapping 21 years trends of $\mathrm{OHC}_{700-2000m}$ in the control simulation is high (Figure 4.3, right column).

For the temperature trend in $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$ during the time period 1990 - 2010 (Figure 4.4), the largest magnitude temperature trends in each model occur across the Southern Ocean and Atlantic Ocean (particularly North Atlantic). In the left column of Figure 4.4 the model trends are shown for each gridpoint, and in the right column the model trends are averaged across the deep ocean basins defined by Purkey and Johnson (2010). Although the gridpoint trends (Figure 4.4, left column) generally show warming, there are coherent regions of cooling, and this information is sometimes lost when looking at an averaged basin (e.g., regions of cooling in North Atlantic).

For $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$, the estimates of basin averaged trends from hydrographic sections are the only available observations (Figure 4.4, bottom row). Regions of greatest trends in the models are consistent with regions of greatest trends in the observations (Atlantic and Southern Oceans) although the trends do not always agree in direction. The observations of Purkey and Johnson (2010) show cooling in the North Atlantic and some other regions, although the only region deemed to show a significant cooling was the north east Atlantic basin (Purkey and Johnson, 2010). A more recent observational analysis including some extra sections in the North Atlantic shows the North-East Atlantic warming over the period 1991 - 2010, but a statistically significantly cooling in the North-West Atlantic



(a) HadGEM2-ES 2000 m +

(b) HadGEM2-ES 2000 m +

(c) MPI-ESM-MR 2000 m +

(d) MPI-ESM-MR 2000 m +

(e) GISS-E2-R 2000 m +

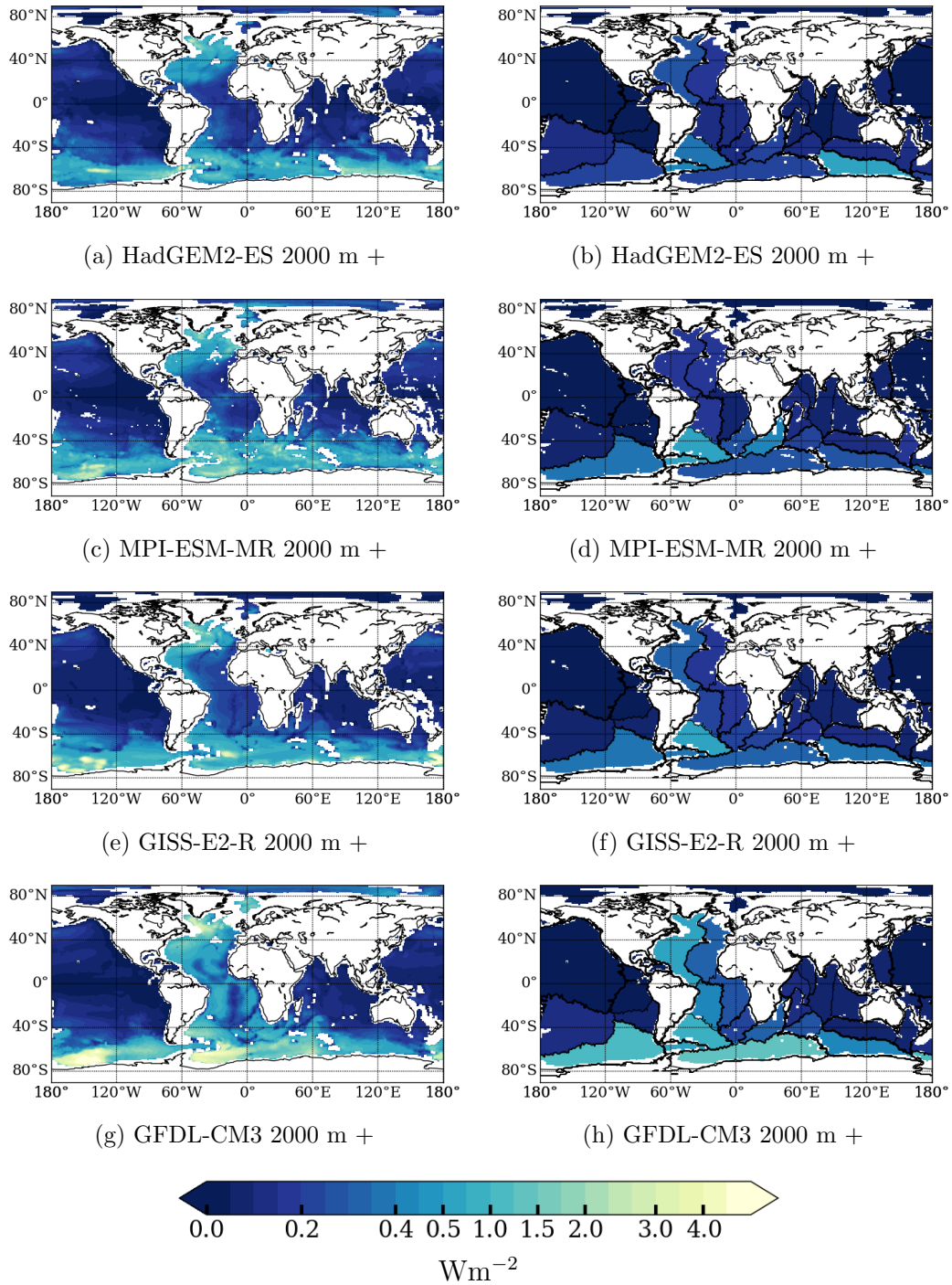
(f) GISS-E2-R 2000 m +

(g) GFDL-CM3 2000 m +

(h) GFDL-CM3 2000 m +

(i) Observations PJ 2000 m +

Figure 4.4: Linear trend in OHC in the 2000 m to seafloor layer during the period 1990 - 2010 inclusive, presented as a heat flux (Wm^{-2}) into the layer. The left column displays the value for each 1° gridbox, whilst the right column shows the trends averaged across deep ocean basins for comparability with the observational record over same period. Hatched basins are where the magnitude of the basin trend between 1990-2010 is larger than two standard deviations of the overlapping 21 year basin average trends in the control (highlights influence of anthropogenic forcing).



(a) HadGEM2-ES 2000 m +

(b) HadGEM2-ES 2000 m +

(c) MPI-ESM-MR 2000 m +

(d) MPI-ESM-MR 2000 m +

(e) GISS-E2-R 2000 m +

(f) GISS-E2-R 2000 m +

(g) GFDL-CM3 2000 m +

(h) GFDL-CM3 2000 m +

Figure 4.5: σ of the linear trends in OHC for all overlapping 21 year periods in the control simulations, calculated for the 2000 m to seafloor layer. The left column displays the value at each 1° gridbox, whilst the right column shows σ calculated after the trends are averaged across deep ocean basins.

(Desbruyères *et al.*, 2016b). The GFDL-CM3 model shows much of the Pacific and Atlantic sectors of the Southern Ocean cooling on average, whilst the same areas warm on average in all other models and observations.

The standard deviation of all 21 year overlapping trends in $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$ (Figure 4.5) in the control simulation shows regions of greatest unforced variability to be across the Southern Ocean and northern Atlantic Oceans. Particularly in the basin averaged trends (Figure 4.5, right column) some differences between the level of unforced variability exhibited in a basin are apparent between models (e.g., compared to other models, much higher variability in Southern Ocean in GFDL-CM3).

In each model, a basic detection/attribution calculation is made to establish in which regions the modelled trend in $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$ during the period 1990-2010 is anthropogenically forced. If the magnitude of the basin trend during 1990-2010 is larger than two standard deviations of the overlapping 21 year basin average trends in the control, then the basin is assumed to be under the influence of anthropogenic forcing, and it is hatched in Figure 4.4. The hatching shows that there is no consistency between models in which basins deep OHC change is influenced by anthropogenic forcing, so no conclusions regarding the attribution of deep ocean warming between 1990-2010 are drawn.

4.5.2 EOF analysis in the North Atlantic and Southern Oceans

To ensure the EOF analysis captures only patterns of variability that represent variability sourced in the same region, EOF analysis is performed on OHC from 2000 m to seafloor in the control simulations in the North Atlantic regions and Southern Ocean below 40°S separately. Figure 4.6 presents the leading three EOFs in the North Atlantic region for each model and the percentage of variance in each model that these modes of variability explain. For each model (except MPI-ESM-MR) over 70% of the total variance is expressed by the first three modes; only 47% of variance is explained by the first three modes in MPI-ESM-MR suggesting that a larger fraction of variance in MPI-ESM-MR represents noise rather than coherent large scale multi-decadal variability (Figure 4.6).

The patterns of variability of the first modes for HadGEM2-ES, GISS-E2-R and GFDL-CM3 look very similar across the Labrador Sea and subpolar gyre suggesting that they represent the same oceanic phenomena (Figure 4.6). The patterns of variability in the second mode of variability for MPI-ESM-MR (note that the first and second modes for MPI-ESM-MR both explain a very similar percentage of variability), second mode for HadGEM2-ES and GFDL-CM3 and third mode for GISS-E2-R all look similar (Figure 4.6); together, these patterns may represent the propagation of ocean heat anomalies that form in the Labrador Sea and move southwards through the deep Atlantic Ocean. However EOFs are typically good at showing large scale pattens in variability, but not regional phenomena, so caution is exercised here and limited oceanographic interpretation is made from these results.

The patterns of variability identified in EOF analysis in the North Atlantic (Figure 4.6) look similar to those identified in the trends in OHC during 1990-2010 in the deep ocean (Figure 4.4), and this suggests that much of the variability in the North Atlantic during 1990-2010 could be due to internally forced variability. The patterns during 1990-2010 are not significantly different to those generated in the EOF analysis, so no distinct influence of anthropogenic forcing can be detected yet (whereas a big difference might suggest that

anthropogenic forcing is having a different effect to that of internal variability).

In the Southern Ocean (Figure 4.7) the first three modes of the EOF analysis show much less coherent spatial patterns than for the North Atlantic (Figure 4.6). In GFDL-CM3, the first three modes represent 79% of the variance, but in the other three models the first three modes represent between 39% - 56%, suggesting that in these models much of the variability is due to noise rather than coherent large scale multi-decadal variability. It seems that EOF analysis is of limited use for identifying large scale modes of variability in the Southern Ocean. However, the 1990-2010 trend in $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$ in the Southern Ocean (Figure 4.4) is much more spatially coherent than the variability identified using EOF analysis, and this may suggest that in the deep ocean, the influence of anthropogenic forcing is apparent in the Southern Ocean during 1990-2010 in CMIP5 models.

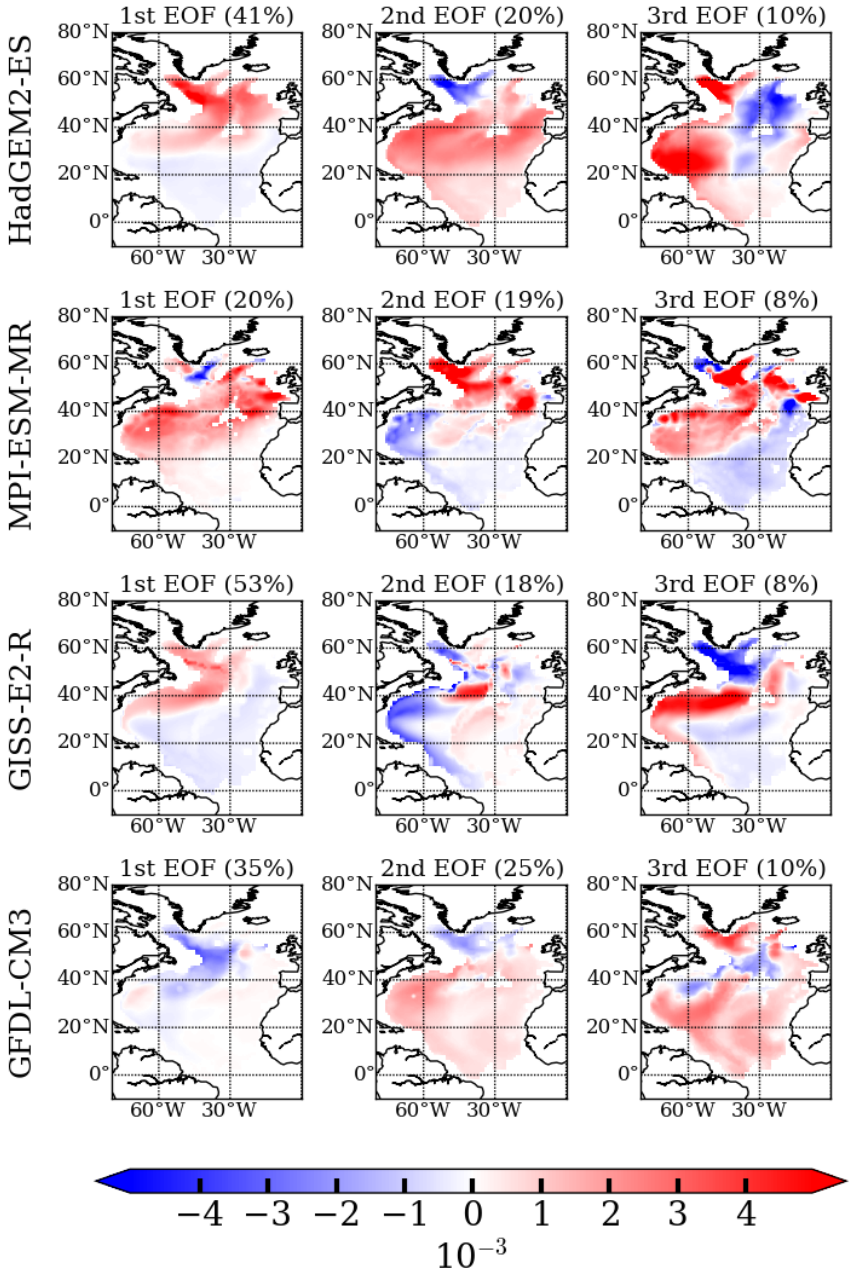


Figure 4.6: The spatial patterns of the three leading modes of variability calculated by EOF analysis of OHC between 2000 m and the seafloor in control simulations. The percentage in brackets is the percentage of total variance explained by each mode.

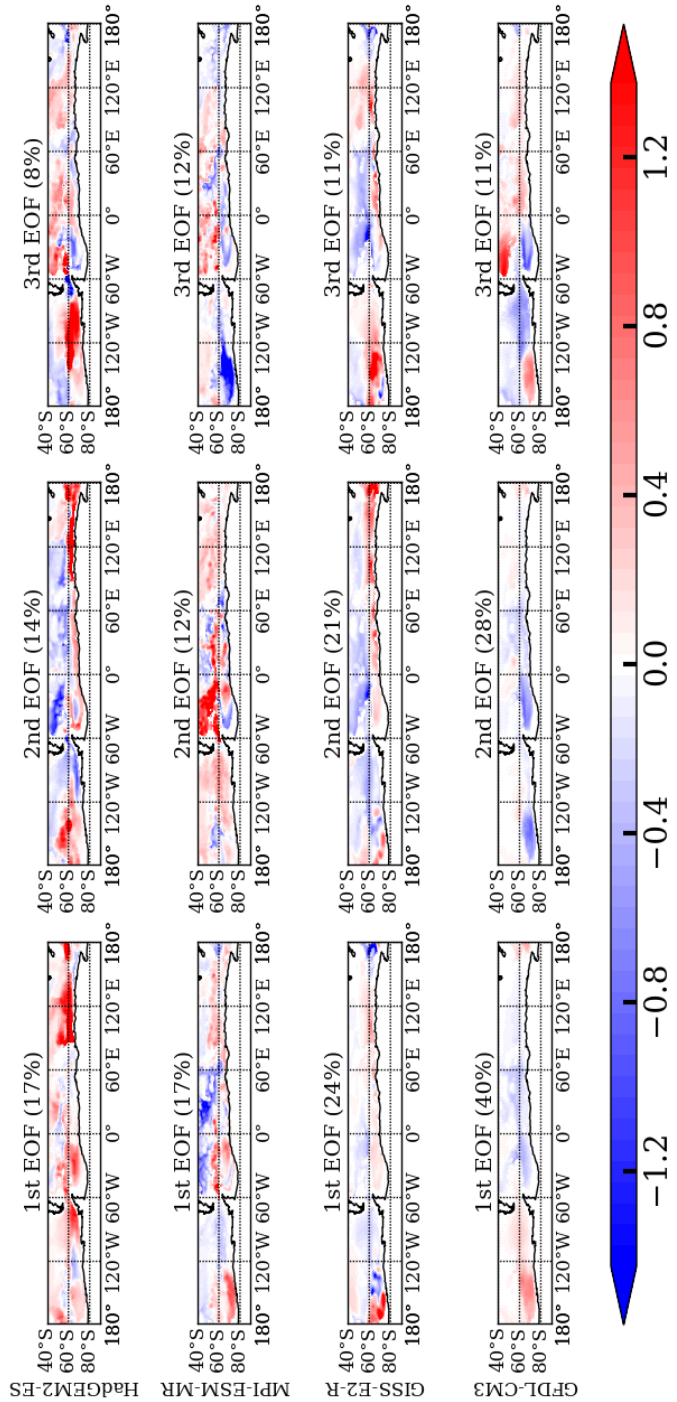


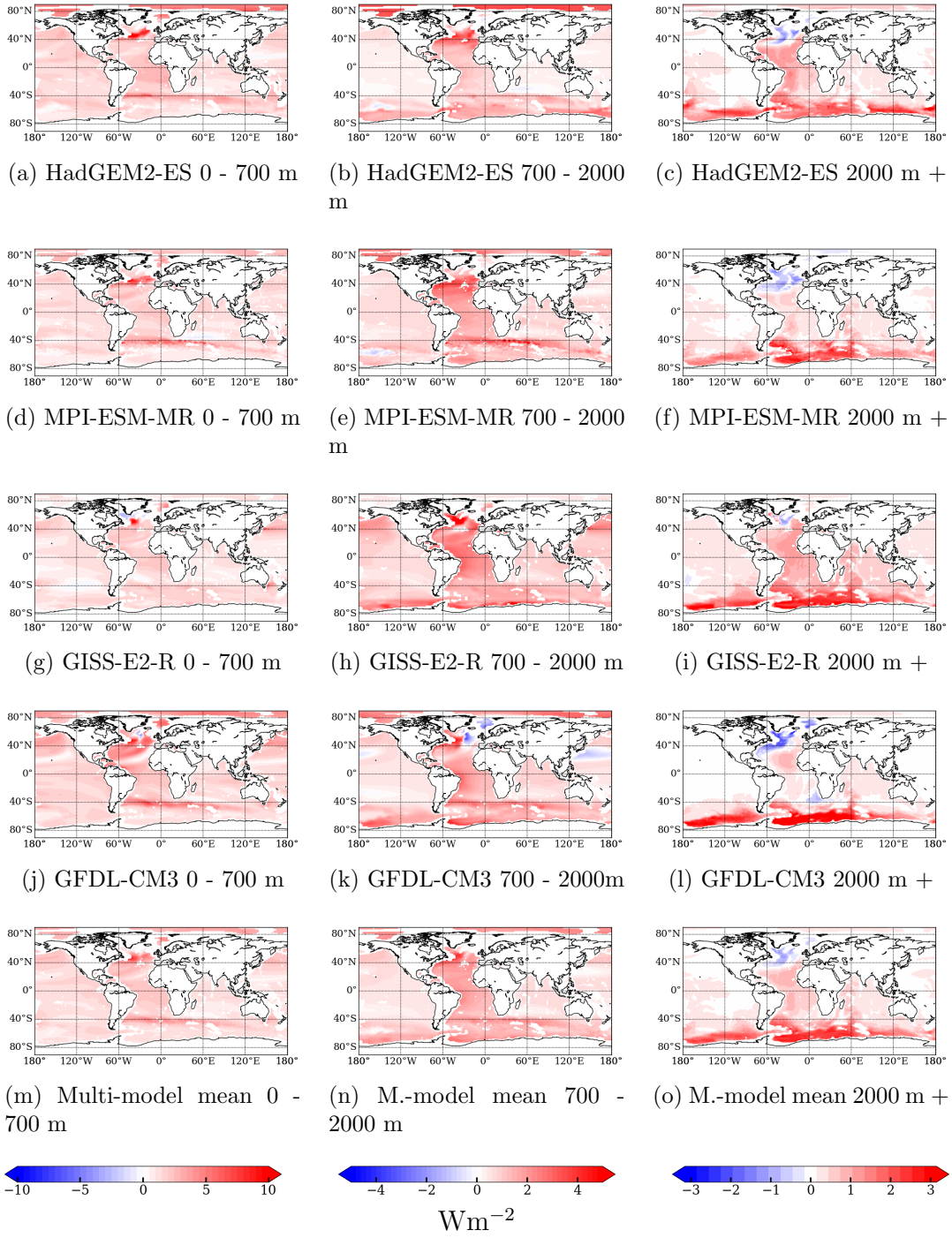
Figure 4.7: As Figure 4.6 but for the Southern Ocean.

4.5.3 Emergent signals of ocean climate change

The climate system is rapidly evolving due to anthropogenic climate forcing and therefore the most important ocean regions to sample to capture future OHC changes may differ from those in an unforced scenario. Using the greenhouse gas emission forced RCP 8.5 simulations, trends through the twenty-first century are calculated to highlight in which regions we expect the largest changes in OHC to occur. The OHC_{0-700m} trend in the RCP 8.5 scenario between 2010 and 2100 is generally positive, reflecting global ocean warming in each model (Figure 4.8, left column). The highest warming trends appear in the Gulf Stream and in standing eddy trains north of the Antarctic Circumpolar Current, most pronounced in GFDL-CM3 (Figure 4.8 j).

Between 700 - 2000 m warming is strongest in the Atlantic (although there is cooling in some parts of the sub-polar gyre in some models, particularly GFDL-CM3 which also shows cooling in the Greenland-Iceland-Norwegian Seas) and warming in the Southern Ocean, including along the standing eddy trains/path of the Antarctic Circumpolar Current (Figure 4.8, middle column).

Between 2000 m - seafloor there is relatively strong warming in the Southern Ocean and Atlantic except in the western subpolar Atlantic and Labrador Sea, which cools in all models (at least partially). The rest of the Atlantic shows stronger warming throughout in HadGEM2-ES (Figure 4.8 c) and GISS-E2-R (Figure 4.8 i) compared with the other models, but overall the main features of warming are relatively consistent between models (Figure 4.8). The deep ocean regions where the largest signals of climate change emerge (Figure 4.8) are broadly the same regions where the largest magnitude decadal internal variability exists (Figure 4.5). This strongly indicates that deep ocean sampling in these regions, namely the Southern Ocean and Atlantic, is vital to measure OHC and its changes accurately.



(a) HadGEM2-ES 0 - 700 m (b) HadGEM2-ES 700 - 2000 m (c) HadGEM2-ES 2000 m +

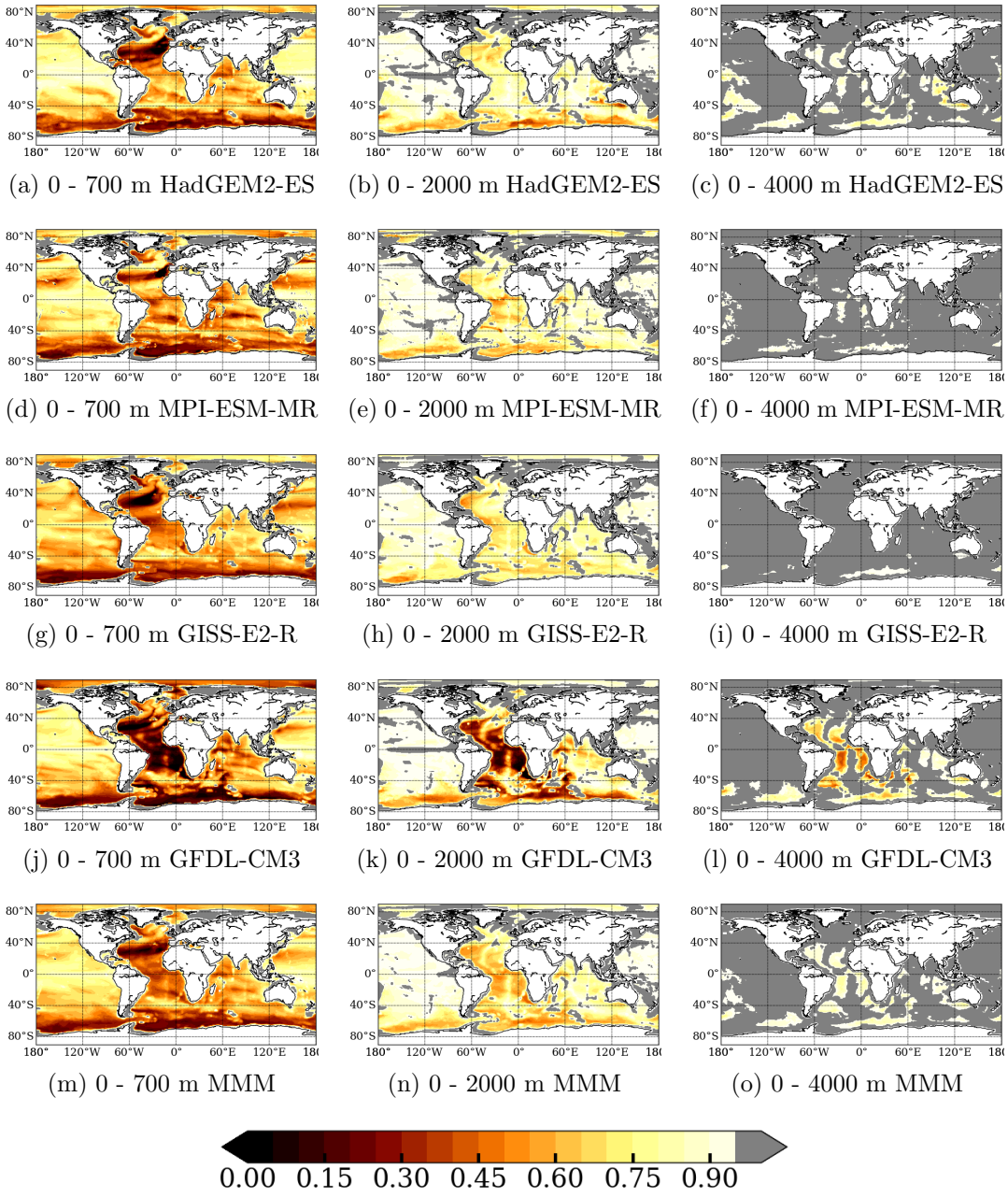
(d) MPI-ESM-MR 0 - 700 m (e) MPI-ESM-MR 700 - 2000 m (f) MPI-ESM-MR 2000 m +

(g) GISS-E2-R 0 - 700 m (h) GISS-E2-R 700 - 2000 m (i) GISS-E2-R 2000 m +

(j) GFDL-CM3 0 - 700 m (k) GFDL-CM3 700 - 2000m (l) GFDL-CM3 2000 m +

(m) Multi-model mean 0 - 700 m (n) M.-model mean 700 - 2000 m (o) M.-model mean 2000 m +

Figure 4.8: Trends in the column integrated OHC between 0 - 700 m, 700 - 2000 m and 2000 m - seafloor, calculated over the period 2010 to 2100 in the RCP 8.5 scenario in CMIP5 models HadGEM2-ES, MPI-ESM-MR, GISS-E2-R and GFDL-CM3, and expressed as a heat flux into the layer (Wm^{-2}). The multi-model mean for these models is also shown.



(a) 0 - 700 m HadGEM2-ES (b) 0 - 2000 m HadGEM2-ES (c) 0 - 4000 m HadGEM2-ES

(d) 0 - 700 m MPI-ESM-MR (e) 0 - 2000 m MPI-ESM-MR (f) 0 - 4000 m MPI-ESM-MR

(g) 0 - 700 m GISS-E2-R (h) 0 - 2000 m GISS-E2-R (i) 0 - 4000 m GISS-E2-R

(j) 0 - 700 m GFDL-CM3 (k) 0 - 2000 m GFDL-CM3 (l) 0 - 4000 m GFDL-CM3

(m) 0 - 700 m MMM (n) 0 - 2000 m MMM (o) 0 - 4000 m MMM

Figure 4.9: For control simulations, the fraction of total variability in full-depth OHC captured by OHC integrated from the surface to a specified depth.

4.5.4 How deep do we need to measure to resolve full depth OHC?

Control simulations

The fraction of variability (F) of OHC_{total} expressed by OHC_{0-700m} in each model and the multi-model mean of the control scenarios is less than 0.8 across most basins (Figure 4.9, left column), with the exception of shallow shelf regions. When F is 0.8 this means that the magnitude of the variability of the unsampled ocean (in this case 700 m to seafloor) is one fifth of the magnitude of the variability of the total water column. F is approximately 0.8 across the tropical Pacific in all models. In some regions, such as the Southern Ocean and subpolar and subtropical gyres of the North Atlantic, F is much smaller at under 0.2. When F is 0.2, the magnitude of variability of the unsampled ocean is four fifths that of the total ocean. Locations are deemed adequately sampled when the fraction of variability > 0.95; these locations are marked in grey on Figure 4.9. Less than 15% of ocean locations are considered adequately sampled in each model if only 0-700 m measurements are considered (Figure 4.9).

Figure 4.9 (middle column) suggests 0-2000 m observations are sufficient to represent OHC in a few parts of the Pacific and Indian Oceans, across most of the Arctic Ocean, and a few locations along Antarctic shelves, but many of these regions are shallower than 2000 m. F is less than 0.8 across most of the Atlantic and Southern Oceans (Figure 4.9, middle column) therefore the variability of OHC_{total} in these regions cannot be accurately estimated with 0-2000 m measurements. With 0-2000 m observations, only around 30% of the global oceans (model spread: 27-31%) are deemed adequately sampled (Figure 4.9, middle column).

Observing to 4000 m (Figure 4.9, right column) gives F greater than 0.95 in a majority of ocean locations across all models. In each model, large parts of all oceans are adequately sampled (in fact observing to 3000 m is sufficient to adequately observe large areas of the Pacific Ocean, not shown). In GISS-E2-R virtually all the oceans (97% of locations) are adequately observed, with the exception of some parts of the Southern Ocean. In MPI-ESM-MR, 93% of

locations are adequately observed, but the Southern Ocean is inadequately sampled along with some parts of the Atlantic and mid Pacific Oceans. 78% of locations are adequately observed in GFDL-CM3 including the Pacific Ocean, but compared to GISS-E2-R and MPI-ESM-MR, larger parts of the Southern, Atlantic and Indian Oceans are not sufficiently well observed with F still less than 0.4 in parts of the Atlantic. In HadGEM2-ES (82% of locations adequately observed) parts of the Pacific, Atlantic, Indian and Southern Oceans are not adequately observed, although the magnitude of F is still generally greater that 0.8 (Figure 4.9).

In cases where sampling to 4000 m yields a low value of F, this indicates that there is significant variability below 4000 m, and so total variability cannot be predicted by variability observed above 4000 m. Figure 4.10 provides scenarios for the depth varying trend in OHC over decadal timescales. Figure 4.10 (a) illustrates a situation where variability in the abyssal ocean is very low and thus the variability in the oceans down to and including the deep ocean is a good metric for total variability (so F would be close to 1). In contrast Figure 4.10 (b) illustrates a scenario where sampling that includes the deep ocean but not the abyssal ocean would result in a low value of F because the variability in the abyssal ocean will be a significant fraction of the total variability of the water column. Examination of temperature change over multi-decadal periods in global OHC at each model level in the control simulations (not shown) illustrate that both scenarios occur in the CMIP5 control runs presented here.

RCP 8.5 simulations

Fraction of variability is calculated in the same way for the RCP 8.5 scenario (Figure 4.11) and the results compared with those for the control simulation (Figure 4.9). For 0-700 m (Figure 4.11, left column) the pattern of F looks similar to that in the control for each model, but there are some locations where F increases compared to the control scenario (e.g., the north Atlantic in all models), and some locations where F reduces (e.g., east of South Africa in MPI-ESM-MR and east and south-west of Brazil in GISS-E2-R). The number of locations for each model deemed adequately observed is similar to that in the control (between 12%-14%, Figure 4.11).

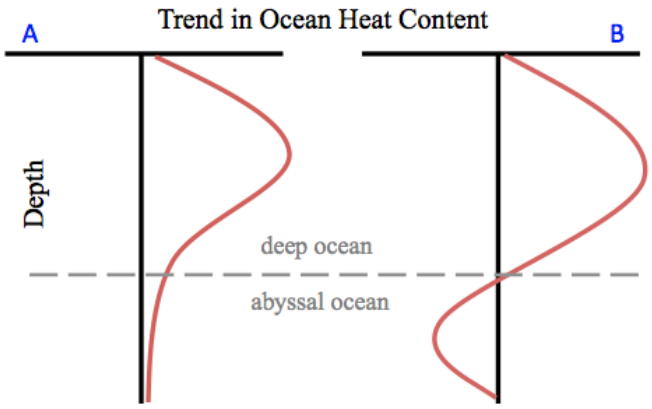


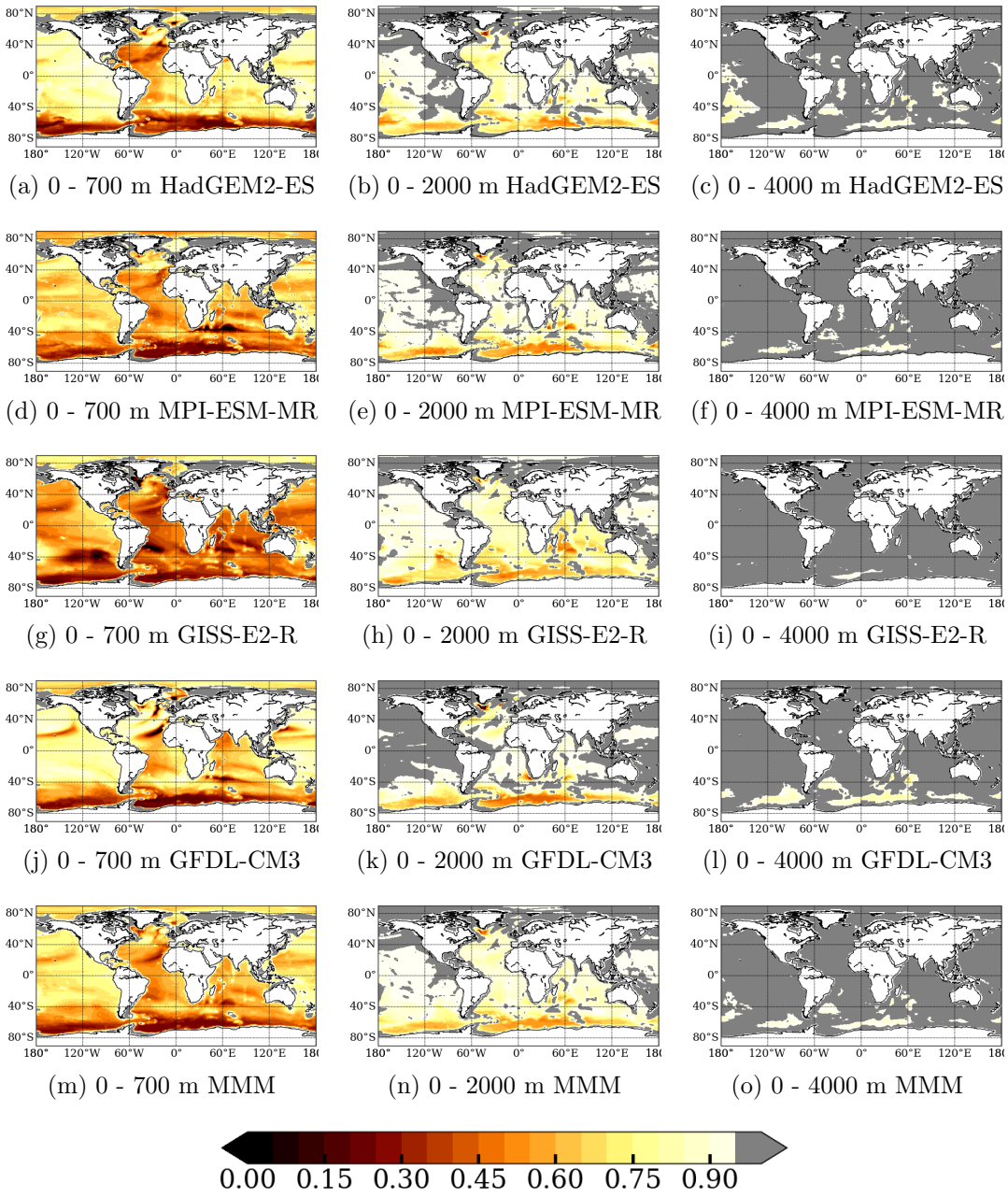
Figure 4.10: Schematic of possible profiles of decadal trends in OHC. When the deep oceans but not the abyssal oceans are sampled, i.e., the oceans are sampled down to the dashed line, (A) would yield a high value of F, whereas (B) would produce a low value of F.

For 0-2000 m (Figure 4.11, middle column) more locations are deemed adequately observed for each model (spread 31%-59%), but there are again regions (for example Southern Ocean in GISS-E2-R model) where F reduces in the RCP 8.5 scenario compared to the control. Observing to 0 - 4000 m in the RCP 8.5 simulation (Figure 4.11, right column) finds a similar (HadGEM-ES) or greater (other models) number of locations to be adequately observed compared to the control simulation (spread 86% - 99%). The influence of heating in RCP 8.5 may increase variability in the upper ocean layers relative to the abyssal ocean, reducing the ratio of the variability in the abyssal ocean to the total variability, and increasing F.

4.5.5 Estimating planetary energy imbalance from OHC

Differences in 10 year running trends of OHC_{0-z} compared to OHC_{total} reflect the bias in TOA radiative imbalance (as estimated from OHC) due to limited depth sampling. Figure 4.12 reveals that biases in each model exceed 0.1 Wm^{-2} for OHC_{0-700m} by the end of the twentieth century and increase through the twenty-first century as the influence of greenhouse gas forcing on the ocean grows (Figure 4.12).

Global $\mathrm{OHC}_{0-2000m}$ observing exhibits a multi-model bias of around 0.2 Wm^{-2}



(a) 0 - 700 m HadGEM2-ES (b) 0 - 2000 m HadGEM2-ES (c) 0 - 4000 m HadGEM2-ES

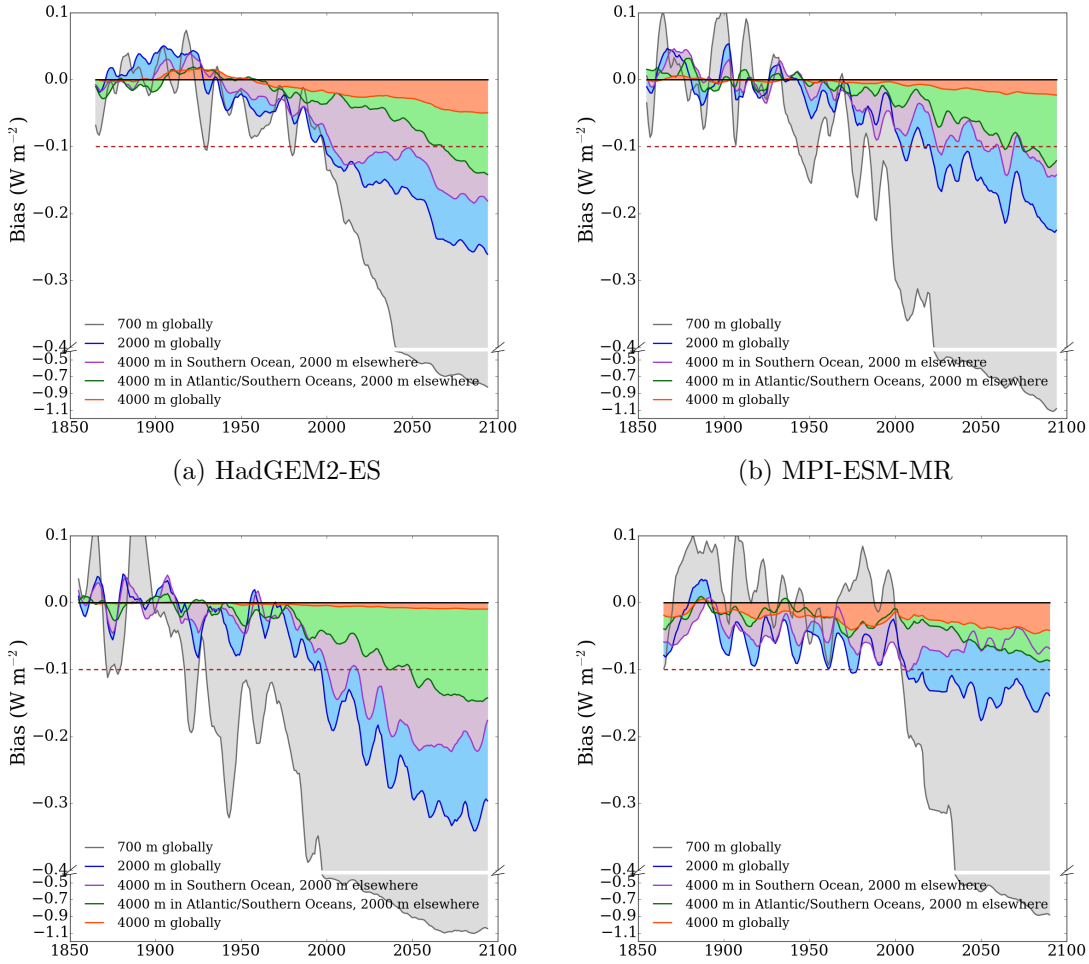
(d) 0 - 700 m MPI-ESM-MR (e) 0 - 2000 m MPI-ESM-MR (f) 0 - 4000 m MPI-ESM-MR

(g) 0 - 700 m GISS-E2-R (h) 0 - 2000 m GISS-E2-R (i) 0 - 4000 m GISS-E2-R

(j) 0 - 700 m GFDL-CM3 (k) 0 - 2000 m GFDL-CM3 (l) 0 - 4000 m GFDL-CM3

(m) 0 - 700 m MMM (n) 0 - 2000 m MMM (o) 0 - 4000 m MMM

Figure 4.11: As Figure 4.9 but for RCP 8.5



(a) HadGEM2-ES

(b) MPI-ESM-MR

(c) GISS-E2-R

(d) GFDL-CM3

Figure 4.12: For each model, observing strategy bias is calculated as the difference in each 10 year trend in OHC captured by a specified strategy to the equivalent 10 year trend in full depth OHC. Bias is presented here as an equivalent surface heat flux. The bias for each overlapping 10 year period in the historical and twenty-first century RCP 8.5 simulations is plotted on the time period midpoint. The idealised observing strategies shown are: perfect (1° resolution) global observing to 700 m (grey line), perfect global observing to 2000 m (blue line), perfect global observing to 2000 m and to 4000 m in Southern Ocean (purple line), perfect global observing to 2000 m and to 4000 m in Southern and Atlantic Oceans (green line), perfect global observing to 4000 m (orange line). The shaded regions represent the differences in biases between strategies.

by the late twenty-first century (Figure 4.12); if the oceans were to be observed only between 0-2000 m it would not be possible to resolve TOA radiative imbalance to within 0.2 Wm^{-2} by the end of the century. Only by supplementing with 2000 - 4000 m Southern Ocean and Atlantic observations can consistent agreement between models of a bias of around 0.1 Wm^{-2} by the late twenty-first century be achieved (Figure 4.12).

Southern Ocean observations provide a large reduction in bias in the period 2000 - 2050 (Figure 4.12). There are notable differences between model projections in the second half of the twenty-first century; the addition of Atlantic 2000 - 4000 m observations reduces the bias below 0.1 Wm^{-2} in the HadGEM2-ES and GISS-E2-R models (Figures 4.12 a & c), but provides little additional benefit over sampling the Southern Ocean in MPI-ESM-MR (Figure 4.12 b). In GFDL-CM3, the Atlantic behaves so differently to the rest of the ocean at 2000 - 4000 m (Figure 4.8) that global sampling to 2000 m with 2000 - 4000 m sampling in the Southern Ocean alone represents the 0-4000 m ocean better than with Atlantic measurements (Figure 4.12 d).

In the first half of the twenty-first century it is shown that a bias in the estimation of the TOA imbalance from OHC may be kept below 0.1 Wm^{-2} with global ocean sampling to 2000 m and 2000 - 4000 m sampling in the Southern Ocean. This suggests that deep OHC measurements should be prioritised in the Southern Ocean.

4.6 Discussion

The spatial patterns of OHC change in CMIP5 models during 1990 - 2010 do not closely match the observations, or each other, but internal variability of climate models is not expected to be in phase with that observed in the real-world. Instead, those areas where the trend is similar across models and observations likely exhibit forced trends. Identifying regions of high internal variability from the standard deviations of all overlapping 21 year trends in the control scenarios, and using this information together with regions where a trend is identified between 1990-2010, and then comparing to observations, gives a better

understanding of the utility of CMIP5 models for assessing decadal change in OHC.

It has been demonstrated that the models generally represent the spatial features of OHC change in upper and deep ocean layers, considering both the forced response between 1990-2010 and variability in trends of the same length from a pre-industrial climate scenario. This analysis supports the utility of climate models for the purpose of considering multi-decadal OHC trends, even in the deep ocean. This is despite known biases in climate models, including their unrepresentative deep water formation mechanisms (Heuzé *et al.*, 2013). This is not to say that climate models do not warrant improvement; the development of models continues and should be fully supported. However, this work indicates useful information about multi-decadal OHC change can be determined from the current generation of climate models and used to inform designers of observational systems. Confidence can be taken in the main results because the models qualititively agree, despite diverse known model deficiencies.

Although commenting on the detection and attribution of deep OHC change is not a primary aim of this work, some analysis towards this has been presented here. There is no consistent agreement between the models as to where the trend in 1990-2010 exceeds the magnitude of variability of 21 year trends in the control, and therefore no particular regions below 2000 m can be suggested to be clearly anthropogenically forced at a magnitude greater than internal variability. EOF analysis also finds that in the North Atlantic patterns of variability in the control look similar to the signals of change there between 1990-2010, and so there is no evidence of anthropogenic forcing behaving differently to internal variability. However in the Southern Ocean strong trends covering broad areas between 1990-2010 differ from the more patchy patterns of variability detected using EOF analysis. This tentatively suggests a possibility of anthropogenic forcing acting on the deep Southern Ocean in a different way to the patterns exhibited by internal variability.

The relatively high variability in multi-decadal deep OHC through the Atlantic and Southern Oceans in control simulations marks these regions as key candidates for high resolution sampling in time and space to improve observational estimates of OHC. It is not necessarily expected that regions where large OHC change

occurs today will be exactly the same as those in the future, as ocean warming and salinity change may adjust ocean circulation and heat storage.

Despite the potential for future change, in these model simulations with anthropogenically forced twenty-first century climate change (RCP 8.5 scenario) the same regions are subject to strong warming as those that experience strong trends in the control and historical simulations, further highlighting the need for observing systems to include frequent deep ocean observations in the Southern Ocean and Atlantic. This concurs with our physical understanding of the oceans: the Southern and Atlantic Oceans are where the deep oceans are ventilated, and so are the most likely deep ocean regions to exhibit a relatively fast response to atmospheric changes. The RCP 8.5 scenario used represents a realistic future: global carbon dioxide emissions are currently slightly exceeding those in RCP 8.5 (Peters *et al.*, 2013). In fact it is likely that the exact choice of emissions scenario is unimportant to broad-scale predictions of deep ocean behaviour over the coming decades since the bulk of anthropogenically forced deep ocean temperature changes will have already been committed to. Although this work focuses on changes up to 2100, it is arguable that observational designers should also be focussing on potential changes beyond the year 2100 (Clark *et al.*, 2016).

The regions of largest warming in the future broadly agree with observational analysis of recent changes to Antarctic Bottom Water, which warms throughout the Southern Ocean but particularly in the Atlantic sector (Purkey and Johnson, 2010). This pattern of warming is consistent with the model simulations shown here, with the exception of HadGEM2-ES, in which Southern Ocean warming in other sectors appears equally strong. Significant warming is projected in the south-west Atlantic in each model (less pronounced in GFDL-CM3) and this is similar in pattern to the relatively strong warming observed during 1990-2010 by Purkey and Johnson (2010) in that region.

The projected cooling in the sub-polar North Atlantic in RCP 8.5 across CMIP5 models was identified and explained by Heuzé *et al.* (2015) as increased quantities of Antarctic Bottom Water filling the basin while North Atlantic Deep Water production reduces. A sub-polar North Atlantic cooling was not observed by the regional analysis of deep OHC trends by Purkey and Johnson (2010), but the time scale of the observational analysis is comparatively short and based on

limited data.

Fraction of variability analysis indicates that past and present global observational systems (relying largely on XBTs measuring to 700 m or Argo floats measuring to 2000 m) are insufficient to accurately capture the magnitude or variability of full-depth global OHC. There is a clear requirement for the maintenance of a 0 - 2000 m Argo array for OHC measurement since shallower sampling does not accurately reflect OHC_{total}. The inadequacy of 0 - 2000 m sampling for capturing OHC in the Southern Ocean and Atlantic and in regions of the Indian and Pacific indicates that below 2000 m observations are required immediately. In some regions the increased fraction of variability in the RCP 8.5 simulation relative to the control simulation reflects the increased heat accumulating in the surface layer of the ocean in RCP 8.5.

More than 90% of global locations are adequately measured by 4000 m in two of the models (80% for the other two) so for broad scale climate predictions, deployment of deep Argo floats that sample to 6000 m is likely to be less critical than deploying the less expensive 4000 m models. However, deep ocean observations do indicate abyssal warming below 4000 m with bottom intensified warming has been observed near the seafloor (e.g., below 5000 m in Atlantic and Southern Oceans, Desbruyères *et al.*, 2016b) which models may not capture (partly due to restricted resolution at depth). Sampling may be required deeper than the models suggest and this should be taken into account in observational design.

The results suggest that the greatest deep ocean sampling need is for 2000 m to 4000 m observations over the coming decades, with priority observing in the Southern and Atlantic Oceans since OHC in these regions is especially poorly captured in all models when measured to only 2000 m. Desbruyères *et al.* (2016b) reinforce this finding with an evaluation of observational OHC estimates from ship sections between 2004 - 2014 which reveals below 2000 m warming was largely due to heave motions in the deep Southern and North Atlantic basins.

Earth's average energy imbalance is generally thought to be around 1 Wm^{-2} (e.g., Hansen *et al.*, 2011) with recent calculations by Johnson *et al.* (2016) calculating it as 0.71 $\pm$ 0.10 Wm^{-2} during 2005 - 2015. Here I use 0.1 Wm^{-2} as

a threshold for acceptable (rather than ideal) bias when estimating planetary energy imbalance from different OHC sampling scenarios. Note that with OHC_{0-700m} sampling the bias is already of the order 1 Wm^{-2} by the end of the twenty-first century; it is vital that the deeper measurements to 2000 m now provided by upper ocean Argo continue.

Assuming complete horizontal coverage, greater than 0.1 Wm^{-2} biases occur in the early twenty-first century using global measurements of $\mathrm{OHC}_{0-2000m}$ revealing the immediate need for observations below 2000 m to constrain the planetary imbalance. By the late twenty-first century biases in the planetary energy imbalance from $\mathrm{OHC}_{0-2000m}$ are as large as 0.2 Wm^{-2}, highlighting that below 2000 m measurements are critical to capture future planetary energy imbalance to sufficient accuracy this century. At least in the first half of the century, adding $\mathrm{OHC}_{2000m-\mathrm{seafloor}}$ measurements just in the Southern Ocean should improve estimation of the planetary energy imbalance from OHC to an acceptable level of error. Southern Ocean sampling using methods such as deep Argo floats can initially provide much of the benefit of global deep ocean observing, though there will be challenges in adequately sampling remote and seasonally or partially-ice covered environments. It is also noteworthy that as the sampling depth increases the inter-annual variability in the bias decreases.

As heat continues to penetrate through the ocean from deep water formation regions and spreads through the Atlantic and into the deep Indian and Pacific oceans, we are likely to need more comprehensive, near-global 2000 - 4000 m sampling to accurately estimate the planetary energy imbalance. By 2100, both Southern Ocean and Atlantic sampling between 2000 - 4000 m is vital to achieve an error of less than 0.1 Wm^{-2}, but global sampling would further reduce the error. However, it is critical that observing systems are deployed before large changes occur to enable accurate measurement of the signal.

It must be noted that despite between-model agreement and similar patterns of variability/warming between models and observations, model deficiencies mean that the real world behaves differently in some important respects. Improvements in climate models are crucial, in particular the representation of deep water formation (Heuzé *et al.*, 2013), to maximise the potential benefit of these tools for informing deep ocean observing strategy. However, state-of-the-art climate

models are our primary tool for predicting future changes to the Earth system and must be considered when designing a future observing strategy.

In reality, OHC estimations below 2000 m are currently obtained from repeat hydrographic ship sections; the maintenance of these sections is vital for measuring a much wider range of oceanic properties and calibrating autonomous floats. While spatial and temporal coverage by hydrographic sections is poor compared to that which can be achieved from continuously profiling autonomous floats, a model-based analysis using a 0.25° resolution forced ocean model determined that in the north-east Atlantic, ship sections between 1990 and 2010 may have recovered deep OHC variability reasonably well (Desbruyères *et al.*, 2014). However, there are regions, e.g., the Southern Ocean, that are widely thought to be poorly sampled by hydrographic sections (either in space or time, and sampling is restricted to the summer months), and earlier in this thesis it is shown that trends from hydrographic sections may not be representative of true trends across many parts of the oceans. Additional deep ocean sampling methods such as deep Argo floats may result in a big improvement in the accuracy of estimated changes in these regions. Indeed in these poorly sampled regions additional deployment of upper 2000 m Argo floats as well as deep measurements may be necessary to observe OHC changes accurately.

4.6.1 Further questions surrounding ocean sampling

The need for further research in aspects of OHC measurement such as developing increasingly accurate sensors for deep ocean measuring and ensuring sufficient sampling in areas of sea ice has already been alluded to. Sampling frequency and coverage may not be adequate in some regions, and estimated full-depth OHC can be dependent on the choice of mapping method (Cheng *et al.*, 2015). Deep OHC change estimates are currently obtained by averaging data across basins according to the length of the hydrographic sections that occupy that basin; this method is relatively straightforward, and with the inclusion of extra deep ocean measurements more sophisticated methods will undoubtably be developed to tease out additional information regarding OHC change. There may be opportunities to optimise measurements through carefully considered

multi-platform deployment (floats, moorings and ships). In addition there remains some uncertainty over the (relatively small) volume of ocean that is too shallow for the upper 2000 m Argo array to sample (marginal seas), and McKinnon and Huybers (2016) find in a model simulation that completely measuring the upper ocean would increase the magnitude and precision of inter-annual OHC trend estimates more than a complete measurement of the deep ocean.

4.7 Conclusions

Through consideration of OHC trends during the period 1990-2010 together with internal variability in OHC in state-of-the-art climate models, the utility of models for evaluating spatial patterns of OHC is confirmed. It is noted that model improvements are also required to increase the accuracy and precision of climate predictions and reduce the large inter-model spread across climate models. The largest internal variability and largest signals of twenty-first century climate change between 2000 - 4000 m are found in the Southern and Atlantic Oceans in CMIP5 model simulations. This work suggests that sampling between 0 - 2000 m is inadequate to capture the magnitude and variability of OHC in the Southern and Atlantic Oceans, and that 0 - 2000 m measurements will not yield accurate estimates of twenty-first century OHC inventory or change in those regions. Even so, the continuation of 0-2000 m Argo is vital; 0 - 700 m observations (e.g., XBT sampling) do not represent heat content and change in any ocean basin. Sampling only between 0 - 2000 m will not allow estimation of twenty-first century planetary energy imbalance to within 0.1 Wm^{-2} so deep ocean measurements are urgently required to supplement the current ocean observing system. This chapter suggests the majority of near-future deep ocean sampling should be focused on the Southern and Atlantic Oceans to capture OHC change and adequately estimate Earth's planetary energy imbalance to within 0.1 Wm^{-2} throughout the twenty-first century.

CHAPTER 5

Conclusions

5.1 Summary

Understanding ocean heat content (OHC) change is important for many areas of climate science (including the planetary energy imbalance, sea level rise, climate sensitivity, carbon storage, ocean biodiversity, initialising climate models and accurate predictions of future global and regional climate change including extreme weather events). Temperatures in the upper oceans (0 - 2000 m) have been sampled frequently over the last decade both in time and space by over 3000 upper ocean Argo floats, providing OHC change estimates that far surpass those of past observational systems in their accuracy. However, present OHC change estimations in the deep oceans (below 2000 m) rely on high quality temperature data along a handful of repeat hydrographic sections across an ocean, occupied at least once a decade. The low spatial and temporal resolution of deep ocean temperature measurements means that these deep OHC estimations are associated with large uncertainties.

The first part of this thesis presents an evaluation of the uncertainties inherent in deep OHC measurements. Using high resolution ocean model simulations, a pseudo hydrographic sampling system is deployed in the model field, which when compared with the complete model field, yields the temporal and spatial uncertainties that arise due to the infrequent sampling in both time and space.

First a regional study of the North-East Atlantic is presented using three different NEMO ocean-sea ice model simulations at two different resolutions (1/4° and 1/12°) and using three different forcing products. Although the main features of the average basin trends below 2000 m are captured in each simulation, there is significant temporal and spatial uncertainty, which cumulates in an overestimate of the total basin trend below 2000 m of 25% to 100% (depending on the simulation) when using a representation of hydrographic sampling in the model field.

The uncertainties found in hydrographic-style sampling in the North-East Atlantic motivate a global study of uncertainties in hydrographic sampling; there are parts of the global ocean which are less well sampled in space than the comparatively well-sampled North-East Atlantic and that exhibit large temperature variability, such as the Southern Ocean. The second science chapter presents uncertainties generated when the hydrographic sections used for an often cited study of global deep OHC change (Purkey and Johnson, 2010) are replicated in a single 1/4° resolution NEMO simulation.

Large biases in temperature trends from hydrographic style sampling were found in the Atlantic, Southern and Indian oceans, and it varies between basins as to whether temporal or spatial uncertainty dominates. In the deep ocean (2000 - 4000 m) these biases can be over 0.1 Wm^{-2} across an ocean. When the global temperature trend is considered the biases are relatively small below 3500 m but shallower than 3500 m the total bias due to limited spatial and temporal sampling grows such that at 2500 m global temperature change is underestimated by a third and at 2000 m only half the total temperature change is captured using hydrographic style sampling in this model simulation.

The last part of the thesis investigates OHC and its spatial patterns over much longer timescales, showing how OHC changes under internal variability in CMIP5 control simulations and under (a now conservative estimation of) projected twenty-first century anthropogenic forcing using the RCP 8.5 scenario. The spatial patterns of OHC change indicate that variability on decadal timescales is likely to be much greater in the Southern and Atlantic Oceans than elsewhere. High resolution deep ocean measurements (such as those that could be achieved with new autonomous float technologies) in the Southern and Atlantic Oceans

may be sufficient to constrain the planetary energy imbalance to $\pm 0.1\ Wm^{-2}$ as estimated from total OHC over the coming decades.

5.2 Implications for deep ocean observing

The first part of this work indicates that over the past thirty years, hydrography is unlikely to have resolved the magnitude of deep ocean temperature trends correctly. This work provides a strong case for increasing both the spatial and temporal sampling resolution of the deep ocean temperature observing system to capture decadal variability in deep OHC trends accurately. This is vital for increasing the accuracy of decadal predictions e.g., for sea level rise. The work shows that improvements to spatial and temporal resolution should be made simultaneously to ensure total uncertainty reduces, and that additional measurements are required across much of the global oceans. Uncertainties are particularly large in the deep ocean (2000 - 4000 m) suggesting that additional measurements in this depth range are critical.

The deep ocean observing system of hydrographic sections is likely to be supplemented by new 'Deep Argo' autonomous float technologies, which are beginning to be deployed in small numbers. Different floats are being developed that can descend to 4000 m or 6000 m. The relatively large uncertainties above 4000 m may mean that cheaper floats that descend to only 4000 m may be sufficient in some regions to reduce deep ocean uncertainties significantly, providing better value for money under a cost-benefit analysis. This work also suggests that key regions for the future deployment of Deep Argo should be the Atlantic and Southern Oceans because these regions exhibit large internal variability in OHC and there are likely to be large changes in OHC in these regions during the twenty-first century relative to other parts of the global ocean.

If the global oceanographic community does not have the resources to deploy a uniform Deep Argo array, due to restricted funding and competing interests from other groups in the Argo community (e.g., bio-Argo floats or under-ice floats), this work suggests that for the purposes of measuring temperature change, the deployment of the majority of Deep Argo floats over the coming decades should

be focussed on key regions in the Atlantic and Southern Ocean, followed by the Indian Ocean. As a caveat to this, of course the long term goal may still be a uniform array globally, and certainly this work suggests that this would be a good idea by the end of the century to be sure to constrain the planetary imbalance to an accuracy of 10-15%. To further increase the accuracy by a given point in time, more widespread deployment of resources is necessary, and it is also important to be measuring with sufficient accuracy in a given region before anthropogenically forced temperature changes occur to be able to identify and attribute them.

5.3 Further work

There are many avenues for potential future research that could improve understanding of deep ocean temperature change, and some of these possibilities are outlined in brief here. The first potential area is the development of more sophisticated methodologies to analyse the data that exists already. Much research already exists into improvement of mapping methods for upper OHC measurements. In the deep oceans, the averaging of data from sections (using their length as a weighting) is a simple approach to generating basin estimates. It potentially introduces error due to the oversampling of some parts of the basin relative to others (for example in regions where two sections cross). There may be more effective ways of averaging, for example using optimal interpolation techniques based on weightings to reduce error, and new methodologies could be evaluated against state-of-the-art model fields and emerging independent data sources such as Deep Argo.

Along similar lines, Desbruyères *et al.* (2014) have already used upper ocean Argo to compare temperature trends at 2000 m to those from hydrography in the North-East Atlantic, and using similar approaches in near deep ocean regions will help identify real-ocean areas where temperature biases from hydrography are likely to be large. In addition, from higher resolution sampling methodologies such as upper ocean or regional pilot arrays of deep ocean Argo floats, information about decorrelation length scales in the deep ocean may be obtained, revealing how far apart particles of water need to be before their temperature properties are no longer related. This will improve new mapping methods and

optimal interpolation techniques for generating global estimates of deep ocean temperature.

As per observational analysis, in this work a constant decorrelation length scale was used. In reality, the horizontal decorrelation length scale is likely to vary regionally, and decrease in areas of high horizontal variability. For the 28 sections examined by Purkey and Johnson (2010), the horizontal length scale is between 25 and 400 km, but the mean and median between 500 and 5000 dbar is 163 km, so they use that as a best estimate of horizontal length scale for all depths and sections. However, it is clear from the analysis of Purkey and Johnson (2010) that in some regions, it would be necessary to sample more frequently to accurately capture spatial patterns (when the decorrelation length scale is smaller), and in some regions, ocean variability is less and the decorrelation length scale is much longer. Further work on deep ocean sampling may consider using regionally varying decorrelation length scales when inferring regional trends from limited data. Decorrelation length scales should also be calculated within ocean model fields as well using the full model field of data, which would serve as a useful comparison to those calculated using limited observations (noting that model variability may underestimate that of the real world and so model derived values may be too long).

Another option for further work is to follow techniques similar to those in this work using additional model simulations. In this work, only the ocean model NEMO is used to evaluate uncertainties in hydrography, and NEMO is thought to mix heat down from the surface to the deep ocean too efficiently. Although no model is perfect, by using other models which will have different deficiencies and thus may have different deep ocean properties, a range of estimates for uncertainties may be obtained. This work already highlights that different simulations using different forcing products can produce different deep ocean trends with different uncertainties, so it would be useful to extend the range of possibilities and find out if there is a consensus between models from different institutions.

To calculate temperature trends in a basin, trends from hydrographic style sampling have been averaged along depth levels in this work and in the observational analyses this work has followed. However, across a basin, the depth

of isopycnals may vary significantly (certainly in some regions, e.g., Southern Ocean), and therefore arguably a more representative average of the basin can be obtained by averaging using an isopycnal framework (that is, average along isopycnals). When averaging across depth levels (as in this work), when isopycnals vary significantly with depth across a basin, information about the magnitude of large trends in some water masses might be lost through the averaging process. Using an isopycnal framework might better highlight how water masses change over time. However long term warming of abyssal waters could cause a water mass to gradually expand, so especially when constrained by the seafloor, might result in a shallowing of isopycnals over time. Changes in density relative to depth can be difficult to interpret, as they may be due to isopycnal changes over time (heave) but there may also be changes to salinity, which in some cases may partially compensate temperature changes. There is certainty scope for further model investigation, not only averaging using an isopycnal framework, but also investigating density changes in more detail, to further understand potential changes in deep and abyssal water masses.

Model runs are now being developed at ultra-high resolution (e.g., NEMO 1/16°, Iovino *et al.*, 2016) and the eddy resolving nature of high resolution simulations may result in more realistic downward mixing of heat. Updated ocean models also tend to include updated forcing products, which may affect deep ocean properties relative to earlier simulations, especially in regions where deep ocean properties are strongly affected by surface water properties (e.g., the Southern Ocean) and where forcing products may change significantly when new data are available in regions where there are few surface flux measurements (again, the Southern Ocean is a prime example).

Climate model simulations used in this work (from the CMIP5 archive) were gridded onto 1° resolution (on average the archive's ocean model resolution is around that) but coupled climate models are now being prepared for CMIP6 that have higher resolution ocean components. It will be worthwhile doing similar analyses of the spatial patterns of OHC change through time in these higher resolution climate model runs as the increased resolution and other model improvements may help generate more realistic deep ocean heat content change.

One of the major criticisms of deep ocean representation in climate models is that

deep water is normally formed by open ocean convection, which we do not today observe as the main mechanism for creating deep water in the Southern Ocean. Instead, dense water forms on shallow shelves around Antarctica due to intense cooling and brine rejection during sea ice formation, before cascading down into the deep ocean in several regions around Antarctica (Rintoul and Garabato, 2013). However, deep ocean convection has been observed in the Southern Ocean (Gordon, 1991), and may have been much more frequent in the past with convection variability thought to be a driver of centennial Antartica climate variability (Martin *et al.*, 2013; Latif *et al.*, 2013; Pedro *et al.*, 2016). Under anthropogenic forcing open ocean convection is likely to become less frequent both in the Southern Ocean (de Lavergne *et al.*, 2014) and northern hemisphere deep water formation regions (Brodeau and Koenigk, 2016). Global models do not have sufficient resolution to resolve shelf water formation, though there are techniques that can be used to artificially source deep water in the correct regions of the deep oceans. Not considering the influence of geothermal heating on abyssal ocean waters may affect model representation of deep ocean circulation substantially, so integration of realistic geothermal fluxes into climate models may also yield improvements in the representation of the deep ocean (Downes *et al.*, 2016).

Limited knowledge of bottom processes, through which energy is dissipated as flows interact with slopes and rough topography, means energy and momentum budgets in models cannot be properly closed (de Lavergne *et al.*, 2016). This restricts our ability to accurately model heat flows through the oceans, as the dissipative processes at the bottom boundary generate turbulent mixing and determine the large scale pattern of ocean properties and the rate of exchange of heat between atmosphere and deep ocean (de Lavergne *et al.*, 2016). In addition to high resolution modelling to analyse flow-topography interactions, instabilities and mixing, in-situ observations are required to properly understand bottom processes, but high pressures challenge current observing systems. Deep Argo, deep-sea gliders and terrain following probes are a few instruments undergoing development that could yield measurements along the bottom of the ocean (de Lavergne *et al.*, 2016).

Despite known model biases, it may be that models with deep ocean properties that are a good representation of observed properties (even if this can only be

tested regionally and at singular timepoints) can provide solutions that are considered more likely than those from unrealistic models. It is difficult to test if the spatial and temporal temperature variability in models is likely to be a good representation of the real ocean, since this is the unknown that models are used to test in this work, so discussions risk becoming somewhat circular. However, whilst the oceanographic community does not have sufficient deep ocean observations to test the deep ocean properties of models against, models known to approximate the real world climate including the upper ocean with some degree of skill can be a useful and important tool for understanding present and future deep ocean temperature change.

Bibliography

Abraham, J., Baringer, M., Bindoff, N., Boyer, T., Cheng, L., Church, J., Conroy, J., Domingues, C., Fasullo, J., Gilson, J., et al. 2013. A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Reviews of Geophysics*, **51**(3),450–483.

Adcroft, A., Scott, J. R., and Marotzke, J. 2001. Impact of geothermal heating on the global ocean circulation. *Geophysical Research Letters*, **28**(9),1735–1738.

Allan, R. P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, M., Smith, D., and Vidale, P.-L. 2014. Changes in global net radiative imbalance 1985–2012. *Geophysical Research Letters*, **41**(15),5588–5597.

Balmaseda, M. A., Trenberth, K. E., and Källén, E. 2013. Distinctive climate signals in reanalysis of global ocean heat content. *Geophysical Research Letters*, **40**(9),1754–1759.

Blaker, A. T., Hirschi, J. J. M., McCarthy, G., Sinha, B., Taws, S., Marsh, R., Coward, A., and de Cuevas, B. 2015. Historical analogues of the recent extreme minima observed in the Atlantic meridional overturning circulation at 26 N. *Climate Dynamics*, **44**(1-2),457–473.

Booth, B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N. 2012. Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**(7393),228–232.

Brodeau, L., Barnier, B., Treguier, A.-M., Penduff, T., and Gulev, S. 2010. An ERA40-based atmospheric forcing for global ocean circulation models. *Ocean Modelling*, **31**(3),88–104.

Brodeau, L. and Koenigk, T. 2016. Extinction of the northern oceanic deep convection in an ensemble of climate model simulations of the 20th and 21st centuries. *Climate Dynamics*, **46**(9-10),2863–2882.

Brown, P. T., Lozier, M. S., Zhang, R., and Li, W. 2016. The necessity of cloud feedback for a basin-scale Atlantic Multidecadal Oscillation. *Geophysical Research Letters*, **43**(8),3955–3963.

Chen, X., Feng, Y., and Huang, N. E. 2014. Global sea level trend during 1993–2012. *Global and Planetary Change*, **112**,26–32.

Chen, X. and Tung, K.-K. 2014. Varying planetary heat sink led to global-warming slowdown and acceleration. *Science*, **345**(6199),897–903.

Cheng, L.-J., Trenberth, K. E., Palmer, M. D., Zhu, J., and Abraham, J. P. 2016. Observed and simulated full-depth ocean heat content changes for 1970-2005. *Ocean Science*, **12**,925–935.

Cheng, L.-J., Zhu, J., and Abraham, J. 2015. Global upper ocean heat content estimation: recent progress and the remaining challenges. *Atmospheric and Oceanic Science Letters*, **8**(6),333–338.

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., Merrifield, M. A., Milne, G. A., Nerem, R. S., Nunn, P. D., et al. 2013a. Chapter 13: Sea level change. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., Merrifield, M. A., Milne, G. A., Nerem, R. S., Nunn, P. D., et al. 2013b. Sea-level rise by 2100. *Science*, **342**(6165),1445–1445.

Church, J. A. and White, N. J. 2011. Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics*, **32**(4-5),585–602.

Clark, P. U., Shakun, J. D., Marcott, S. A., Mix, A. C., Eby, M., Kulp, S., Levermann, A., Milne, G. A., Pfister, P. L., Santer, B. D., et al. 2016. Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nature Climate Change*, **6**,360–369.

Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C., Joshi, M., Liddicoat, S., et al. 2011. Development and evaluation of an Earth-system model–HadGEM2. *Geoscientific Model Development*, **4**(4),1051–1075.

Cristini, L., Lampitt, R. S., Cardin, V., Delory, E., Haugan, P., O'Neill, N., Petihakis, G., and Ruhl, H. A. 2016. Cost and value of multidisciplinary fixed-point ocean observatories. *Marine Policy*, **71**,138–146.

Curry, J. A., Schramm, J. L., and Ebert, E. E. 1995. Sea ice-albedo climate feedback mechanism. *Journal of Climate*, **8**(2),240–247.

Dawson, A., Matthews, A. J., Stevens, D. P., Roberts, M. J., and Vidale, P. L. 2013. Importance of oceanic resolution and mean state on the extra-tropical response to El Niño in a matrix of coupled models. *Climate Dynamics*, **41**(5-6), 1439–1452.

de Lavergne, C., Madec, G., Capet, X., Maze, G., and Roquet, F. 2016. Getting to the bottom of the ocean. *Nature Geoscience*, **9**(12),857–858.

de Lavergne, C., Palter, J. B., Galbraith, E. D., Bernardello, R., and Marinov, I. 2014. Cessation of deep convection in the open Southern Ocean under anthropogenic climate change. *Nature Climate Change*, **4**(4),278–282.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**(656),553–597.

Desbruyères, D., McDonagh, E. L., King, B. A., and Thierry, V. 2016a. Global and Full-depth Ocean Temperature Trends during the early 21st century from Argo and Repeat Hydrography. *Journal of Climate*, **30**,1985–1997.

Desbruyères, D. G., Purkey, S. G., McDonagh, E. L., Johnson, G. C., and King, B. A. 2016b. Deep and abyssal ocean warming from 35 years of repeat hydrography. *Geophysical Research Letters*, **43**(19).

Desbruyères, D., McDonagh, E., King, B., Garry, F., Blaker, A., Moat, B., and Mercier, H. 2014. Full-depth temperature trends in the northeastern Atlantic

through the early 21st century. *Geophysical Research Letters*, **41**(22), 7971–7979.

Domingues, C. M., Church, J. A., White, N. J., Gleckler, P. J., Wijffels, S. E., Barker, P. M., and Dunn, J. R. 2008. Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, **453**(7198),1090–1093.

Downes, S. M., Hogg, A. M., Griffies, S. M., and Samuels, B. L. 2016. The transient response of Southern Ocean circulation to geothermal heating in a global climate model. *Journal of Climate*, **29**,5689–5709.

Drijfhout, S., Blaker, A., Josey, S., Nurser, A., Sinha, B., and Balmaseda, M. 2014. Surface warming hiatus caused by increased heat uptake across multiple ocean basins. *Geophysical Research Letters*, **41**(22),7868–7874.

Duchez, A., Frajka-Williams, E., Josey, S. A., Evans, D. G., Grist, J. P., Marsh, R., McCarthy, G. D., Sinha, B., Berry, D. I., and Hirschi, J. J. 2016. Drivers of exceptionally cold North Atlantic Ocean temperatures and their link to the 2015 European heat wave. *Environmental Research Letters*, **11**(7),074004.

Durack, P. J., Gleckler, P. J., Landerer, F. W., and Taylor, K. E. 2014. Quantifying underestimates of long-term upper-ocean warming. *Nature Climate Change*, **4**(11),999–1005.

Durack, P. J., Lee, T., Vinogradova, N. T., and Stammer, D. 2016. Keeping the lights on for global ocean salinity observation. *Nature Climate Change*, **6**(3), 228–231.

Dussin, R., Barnier, B., Brodeau, L., and Molines, J. M. 2016. The making of Drakkar forcing set DFS5. *DRAKKAR/MyOcean Report 01-04-2016, LGGE, Grenoble, France.*

England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., Gupta, A. S., McPhaden, M. J., Purich, A., and Santoso, A. 2014. Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nature Climate Change*, **4**(3),222–227.

Firing, Y. L., McDonagh, E. L., King, B. A., and Desbruyères, D. G. 2016. Deep temperature variability in Drake passage. *Journal of Geophysical Research: Oceans*, **122**(1),713–725.

Gleckler, P. J., Durack, P. J., Stouffer, R. J., Johnson, G. C., and Forest, C. E. 2016. Industrial-era global ocean heat uptake doubles in recent decades. *Nature Climate Change*, **6**,394–398.

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.

Gordon, A. L. 1991. Two stable modes of Southern Ocean winter stratification. *Elsevier Oceanography Series*, **57**,17–35.

Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., Gnanadesikan, A., Hurlin, W. J., Lee, H.-C., Liang, Z., et al. 2011. The GFDL CM3 coupled climate model: Characteristics of the ocean and sea ice simulations. *Journal of Climate*, **24**(13),3520–3544.

Grist, J. P., Josey, S. A., Jacobs, Z. L., Marsh, R., Sinha, B., and Van Sebille, E. 2015. Extreme air–sea interaction over the North Atlantic subpolar gyre during the winter of 2013–2014 and its sub-surface legacy. *Climate Dynamics*, **46**, 4027–4045.

Häkkinen, S. and Rhines, P. B. 2004. Decline of subpolar North Atlantic circulation during the 1990s. *Science*, **304**(5670),555–559.

Hansen, J., Sato, M., Kharecha, P., and Von Schuckmann, K. 2011. Earth's energy imbalance and implications. *Atmospheric Chemistry and Physics*, **11** (24),13421–13449.

Heuzé, C., Heywood, K. J., Stevens, D. P., and Ridley, J. K. 2013. Southern Ocean bottom water characteristics in CMIP5 models. *Geophysical Research Letters*, **40**(7),1409–1414.

Heuzé, C., Heywood, K. J., Stevens, D. P., and Ridley, J. K. 2015. Changes in Global Ocean Bottom Properties and Volume Transports in CMIP5 Models under Climate Change Scenarios. *Journal of Climate*, **28**(8),2917–2944.

Huber, M. and Knutti, R. 2014. Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nature Geoscience*, **7**(9),651–656.

Hyder, P., Bracegirdle, T. J., Bodas-Salcedo, A., Meijers, A., Gregory, J., Allan, R., Edwards, J., Wood, R., Hewitt, H., Williams, K., Josey, S., Liu, C.-l., Roberts, C., Copsey, D., Mulcahy, J., Field, P., Furtado, K., Sanchez, C., Ridley, J., Thorpe, L., Hardiman, S., Mayer, M., Berry, D., and Belcher, S. 2016. Atmospheric model surface heat flux errors responsible for long-standing Southern Ocean climate model biases. *Under review at Nature Climate Change*.

Ineson, S., Scaife, A. A., Knight, J. R., Manners, J. C., Dunstone, N. J., Gray, L. J., and Haigh, J. D. 2011. Solar forcing of winter climate variability in the Northern Hemisphere. *Nature Geoscience*, **4**(11),753–757.

Iovino, D., Masina, S., Storto, A., Cipollone, A., and Stepanov, V. N. 2016. A 1/16 eddying simulation of the global NEMO sea-ice-ocean system. *Geoscientific Model Development*, **9**(8),2665–2684.

IPCC. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. ISBN ISBN 978-1-107-66182-0. URL `www.climatechange2013.org`.

Ishii, M. and Kimoto, M. 2009. Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *Journal of Oceanography*, **65**(3),287–299.

Johansson, D. J., O'Neill, B. C., Tebaldi, C., and Häggström, O. 2015. Equilibrium climate sensitivity in light of observations over the warming hiatus. *Nature Climate Change*, **5**(5),449–453.

Johnson, G. C., Bullister, J. L., and Gruber, N. 2005. Labrador Sea Water property variations in the northeastern Atlantic Ocean. *Geophysical Research Letters*, **32**(L07602).

Johnson, G. C. and Lyman, J. M. 2014. Oceanography: Where's the heat? *Nature Climate Change*, **4**(11),956–957.

Johnson, G. C., Lyman, J. M., and Loeb, N. G. 2016. Improving estimates of Earth's energy imbalance. *Nature Climate Change*, **6**(7),639–640.

Johnson, G. C., Lyman, J. M., and Purkey, S. G. 2015. Informing Deep Argo Array Design Using Argo and Full-Depth Hydrographic Section Data. *Journal of Atmospheric and Oceanic Technology*, **32**(11),2187–2198.

Jones, C., Hughes, J., Bellouin, N., Hardiman, S., Jones, G., Knight, J., Liddicoat, S., O'Connor, F., Andres, R. J., Bell, C., et al. 2011. The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geoscientific Model Development*, **4**(3),543–570.

Jourdan, D., Balopoulos, E., Dooley, H., Garcia-Fernandez, M., Maillard, C., Baudet, L., Fichaut, M., Hassani, A., et al. 1998. The MEDATLAS climatology: objective analysis of thermal and salinity fields in the Mediterranean basin.

Jungclaus, J., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and Storch, J. 2013. Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, **5**(2),422–446.

Katsman, C. and van Oldenborgh, G. J. 2011. Tracing the upper ocean's missing heat. *Geophysical Research Letters*, **38**(L14610).

Kobayashi, T., Amaike, K.-i., Watanabe, K., Ino, T., Asakawa, K., Suga, T., Kawano, T., Hyakudome, T., and Matsuura, M. Deep NINJA: A new profiling float for deep ocean observation. In *The Proceedings of The Twenty-second (2012) International Offshore and Polar Engineering Conference*, volume 2, pages 454–461, 2012.

Kosaka, Y. and Xie, S.-P. 2013. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**(7467),403–407.

Kouketsu, S., Kawano, T., Masuda, S., Sugiura, N., Sasaki, Y., Toyoda, T., Igarashi, H., Kawai, Y., Katsumata, K., Uchida, H., et al. 2011. Deep ocean heat content changes estimated from observation and reanalysis product and their influence on sea level change. *Journal of Geophysical Research: Oceans (1978–2012)*, **116**(C3).

Kuhlbrodt, T. and Gregory, J. 2012. Ocean heat uptake and its consequences for the magnitude of sea level rise and climate change. *Geophysical Research Letters*, **39**(L18608).

Kuhlbrodt, T., Griesel, A., Montoya, M., Levermann, A., Hofmann, M., and Rahmstorf, S. 2007. On the driving processes of the Atlantic meridional overturning circulation. *Reviews of Geophysics*, **45**(RG2001).

Large, W. G. and Yeager, S. G. 2004. *Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies*. National Center for Atmospheric Research.

Lashof, D. A. 1989. The dynamic greenhouse: feedback processes that may influence future concentrations of atmospheric trace gases and climatic change. *Climatic Change*, **14**(3),213–242.

Latif, M., Martin, T., and Park, W. 2013. Southern Ocean sector centennial climate variability and recent decadal trends. *Journal of Climate*, **26**(19), 7767–7782.

Leadbetter, S., Williams, R., McDonagh, E., and King, B. 2007. A twenty year reversal in water mass trends in the subtropical North Atlantic. *Geophysical Research Letters*, **34**(L12608).

Lee, S.-K., Park, W., Baringer, M. O., Gordon, A. L., Huber, B., and Liu, Y. 2015. Pacific origin of the abrupt increase in Indian Ocean heat content during the warming hiatus. *Nature Geoscience*, **8**(6),445–449.

Levitus, S., Boyer, T., Conkright, M., O'Brien, T., Antonov, J., Stephens, C., Stathoplos, L., Johnson, D., and Gelfeld, R. 1998. World Ocean Database 1998, vol. 1, Introduction, NOAA Atlas NESDIS 18. *NOAA, Silver Spring, Maryland, USA*.

Levitus, S., Antonov, J., Boyer, T., Locarnini, R., Garcia, H., and Mishonov, A. 2009. Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophysical Research Letters*, **36**(L07608).

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J., Seidov, D., Yarosh, E. S., et al.

2012. World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, **39**(L10603).

Lewandowsky, S., Risbey, J. S., and Oreskes, N. 2015. On the definition and identifiability of the alleged 'hiatus' in global warming. *Scientific Reports*, **5**, 16784.

Liu, W., Xie, S.-P., and Lu, J. 2016. Tracking ocean heat uptake during the surface warming hiatus. *Nature Communications*, **7**,10926.

Llovel, W., Willis, J., Landerer, F., and Fukumori, I. 2014. Deep-ocean contribution to sea level and energy budget not detectable over the past decade. *Nature Climate Change*, **4**(11),1031–1035.

Loeb, N. G., Lyman, J. M., Johnson, G. C., Allan, R. P., Doelling, D. R., Wong, T., Soden, B. J., and Stephens, G. L. 2012. Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nature Geoscience*, **5**(2),110–113.

Lyman, J. M., Good, S. A., Gouretski, V. V., Ishii, M., Johnson, G. C., Palmer, M. D., Smith, D. M., and Willis, J. K. 2010. Robust warming of the global upper ocean. *Nature*, **465**(7296),334–337.

Lyman, J. M. and Johnson, G. C. 2014. Estimating global ocean heat content changes in the upper 1800 m since 1950 and the influence of climatology choice. *Journal of Climate*, **27**(5),1945–1957.

Madec, G. et al. 2015. *NEMO ocean engine*. Institut Pierre-Simon Laplace, France.

Madec, G. and Imbard, M. 1996. A global ocean mesh to overcome the North Pole singularity. *Climate Dynamics*, **12**(6),381–388.

Mann, M. E., Steinman, B. A., Miller, S. K., Frankcombe, L. M., England, M. H., and Cheung, A. H. 2016. Predictability of the recent slowdown and subsequent recovery of large-scale surface warming using statistical methods. *Geophysical Research Letters*, **43**(7),3459–3467.

Martin, T., Park, W., and Latif, M. 2013. Multi-centennial variability controlled by Southern Ocean convection in the Kiel Climate Model. *Climate Dynamics*, **40**(7-8),2005–2022.

Marzocchi, A., Hirschi, J. J.-M., Holliday, N. P., Cunningham, S. A., Blaker, A. T., and Coward, A. C. 2015. The North Atlantic subpolar circulation in an eddy-resolving global ocean model. *Journal of Marine Systems*, **142**,126–143.

Mashayek, A., Ferrari, R., Vettoretti, G., and Peltier, W. 2013. The role of the geothermal heat flux in driving the abyssal ocean circulation. *Geophysical Research Letters*, **40**(12),3144–3149.

McCarthy, G., Smeed, D., Johns, W. E., Frajka-Williams, E., Moat, B., Rayner, D., Baringer, M., Meinen, C., Collins, J., and Bryden, H. 2015. Measuring the Atlantic meridional overturning circulation at 26 N. *Progress in Oceanography*, **130**,91–111.

McKinnon, K. A. and Huybers, P. 2016. Seasonal constraints on inferred planetary heat content. *Geophysical Research Letters*, **43**(20),10955–10964.

Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A., and Trenberth, K. E. 2011. Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change*, **1**(7),360–364.

Mercier, H., Lherminier, P., Sarafanov, A., Gaillard, F., Daniault, N., Desbruyères, D., Falina, A., Ferron, B., Gourcuff, C., Huck, T., et al. 2015. Variability of the meridional overturning circulation at the Greenland–Portugal OVIDE section from 1993 to 2010. *Progress in Oceanography*, **132**,250–261.

Meredith, M. P., Jullion, L., Brown, P. J., Garabato, A. C. N., and Couldrey, M. P. 2014. Dense waters of the Weddell and Scotia Seas: recent changes in properties and circulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **372**(2019),20130041.

Moat, B., Josey, S., Sinha, B., Blaker, A., Smeed, D., McCarthy, G., Johns, W., Hirschi, J.-M., Frajka-Williams, E., Rayner, D., et al. 2016. Major variations in subtropical North Atlantic heat transport at short (5 day) timescales and their causes. *Journal of Geophysical Research: Oceans*, **121**(5),3237–3249.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H., Stephenson, D. B., and North, G. R. 2009. Empirical orthogonal functions: The medium is the message. *Journal of Climate*, **22**(24),6501–6514.

NOAA / The COMET Program. Cross section of temperature in the Atlantic Ocean with meridional overturning circulation, 2016. URL `http://www.meted.ucar.edu/tropical/met_topics/media/graphics/moc_atlantic_temp.jpg`. [Online; accessed 8 November 2016].

Palmer, M. D., McNeall, D. J., and Dunstone, N. J. 2011. Importance of the deep ocean for estimating decadal changes in Earth's radiation balance. *Geophysical Research Letters*, **38**(L13707).

Palmer, M. and McNeall, D. 2014. Internal variability of Earth's energy budget simulated by CMIP5 climate models. *Environmental Research Letters*, **9**(3), 034016.

Palmer, M., Roberts, C., Balmaseda, M., Chang, Y.-S., Chepurin, G., Ferry, N., Fujii, Y., Good, S., Guinehut, S., Haines, K., et al. 2015. Ocean heat content variability and change in an ensemble of ocean reanalyses. *Climate Dynamics*, pages 1–22.

Pedro, J., Martin, T., Steig, E. J., Jochum, M., Park, W., and Rasmussen, S. O. 2016. Southern Ocean deep convection as a driver of Antarctic warming events. *Geophysical Research Letters*, **43**(5),2192–2199.

Peters, G. P., Andrew, R. M., Boden, T., Canadell, J. G., Ciais, P., Le Quéré, C., Marland, G., Raupach, M. R., and Wilson, C. 2013. The challenge to keep global warming below 2 C. *Nature Climate Change*, **3**(1),4–6.

Petzrick, E., Truman, J., and Fargher, H. Profiling from 6,000 meter with the APEX-Deep float. In *2013 OCEANS - San Diego*, pages 1–3, Sept 2013.

Purkey, S. G. and Johnson, G. C. 2010. Warming of Global Abyssal and Deep Southern Ocean Waters between the 1990s and 2000s: Contributions to Global Heat and Sea Level Rise Budgets. *Journal of Climate*, **23**(23),6336–6351.

Purkey, S. G. and Johnson, G. C. 2013. Antarctic Bottom Water Warming and Freshening: Contributions to Sea Level Rise, Ocean Freshwater Budgets, and Global Heat Gain. *Journal of Climate*, **26**(16),6105–6122.

Rahmstorf, S. 2002. Ocean circulation and climate during the past 120,000 years. *Nature*, **419**(6903),207–214.

Rhein, M., Rintoul, S., Aoki, S., et al. 2013. Chapter 3: Observations: Ocean. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Rintoul, S. R. and Garabato, A. C. N. 2013. *Ocean Circulation and Climate: Chapter 18. Dynamics of the Southern Ocean Circulation*, volume 103. Elsevier Inc. Chapters.

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.

Roberts, C., Palmer, M., Allan, R., Desbruyeres, D., Hyder, P., Liu, C., and Smith, D. 2016a. Surface flux and ocean heat transport convergence contributions to seasonal and interannual variations of ocean heat content. *In press. Journal of Geophysical Research: Oceans*.

Roberts, C. D., Garry, F. K., and Jackson, L. C. 2013. A multimodel study of sea surface temperature and subsurface density fingerprints of the atlantic meridional overturning circulation. *Journal of Climate*, **26**(22),9155–9174.

Roberts, C. D., Palmer, M. D., McNeall, D., and Collins, M. 2015. Quantifying the likelihood of a continued hiatus in global warming. *Nature Climate Change*, **5**(4),337–342.

Roberts, M. J., Hewitt, H. T., Hyder, P., Ferreira, D., Josey, S. A., Mizielinski, M., and Shelly, A. 2016b. Impact of ocean resolution on coupled air-sea fluxes and large-scale climate. *Geophysical Research Letters*, **43**(19),10430–10438.

Robson, J., Ortega, P., and Sutton, R. 2016. A reversal of climatic trends in the North Atlantic since 2005. *Nature Geoscience*, **9**(7),513–517.

Roemmich, D., Church, J., Gilson, J., Monselesan, D., Sutton, P., and Wijffels, S. 2015. Unabated planetary warming and its ocean structure since 2006. *Nature Climate Change*, **5**(3),240–245.

Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A., Bracegirdle, T., Wang, Z., and Roy, T. 2013. Assessment of Southern Ocean water mass circulation and

characteristics in CMIP5 models: Historical bias and forcing response. *Journal of Geophysical Research: Oceans*, **118**(4),1830–1844.

Sarmiento, J. L., Hughes, T. M., Stouffer, R. J., and Manabe, S. 1998. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature*, **393**(6682),245–249.

Scaife, A. A., Copsey, D., Gordon, C., Harris, C., Hinton, T., Keeley, S., O'Neill, A., Roberts, M., and Williams, K. 2011. Improved Atlantic winter blocking in a climate model. *Geophysical Research Letters*, **38**(L23703).

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., et al. 2014. Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, **6**(1),141–184.

Schurer, A. P., Hegerl, G. C., and Obrochta, S. P. 2015. Determining the likelihood of pauses and surges in global warming. *Geophysical Research Letters*, **42**(14),5974–5982.

Sévellec, F., Sinha, B., and Skliris, N. 2016. The rogue nature of hiatuses in a global warming climate. *Geophysical Research Letters*, **43**(15),8169–8177.

Shaffrey, L. C., Stevens, I., Norton, W., Roberts, M., Vidale, P.-L., Harle, J., Jrrar, A., Stevens, D., Woodage, M. J., Demory, M.-E., et al. 2009. UK HiGEM: The new UK high-resolution global environment model-model description and basic evaluation. *Journal of Climate*, **22**(8),1861–1896.

Sheen, K., Garabato, A. N., Brearley, J., Meredith, M., Polzin, K., Smeed, D., Forryan, A., King, B., Sallée, J.-B., Laurent, L. S., et al. 2014. Eddy-induced variability in Southern Ocean abyssal mixing on climatic timescales. *Nature Geoscience*, **7**(8),577–582.

Simmons, A., Willett, K., Jones, P., Thorne, P., and Dee, D. 2010. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **115**(D1).

Smith, D. M., Allan, R. P., Coward, A. C., Eade, R., Hyder, P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, C. D., and Scaife, A. A. 2015. Earth's energy

imbalance since 1960 in observations and CMIP5 models. *Geophysical Research Letters*, **42**(4),1205–1213.

Somavilla, R., González-Pola, C., Schauer, U., and Budéus, G. 2016. Mid-2000s North Atlantic shift: Heat budget and circulation changes. *Geophysical Research Letters*, **43**,2059–2068.

Steele, M., Morley, R., and Ermold, W. 2001. PHC: A global ocean hydrography with a high-quality Arctic Ocean. *Journal of Climate*, **14**(9),2079–2087.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**(4), 485–498.

Timmermann, R., Goosse, H., Madec, G., Fichefet, T., Ethe, C., and Dulire, V. 2005. On the representation of high latitude processes in the ORCA-LIM global coupled sea ice-ocean model. *Ocean Modelling*, **8**,175–201.

Trenberth, K. E., Fasullo, J. T., and Balmaseda, M. A. 2014. Earth's energy imbalance. *Journal of Climate*, **27**(9),3129–3144.

Trenberth, K. E., Fasullo, J. T., von Schuckmann, K., and Cheng, L. 2016. Insights into Earths Energy Imbalance from Multiple Sources. *Journal of Climate*, **29**(20),7495–7505.

U.S. Department of Commerce. 2006. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Geophysical Data Center: 2-minute Gridded Global Relief Data (ETOPO2v2). *http://www.ngdc.noaa.gov/mgg/global/etopo2.html*.

Von Schuckmann, K., Palmer, M., Trenberth, K., Cazenave, A., Chambers, D., Champollion, N., Hansen, J., Josey, S., Loeb, N., Mathieu, P., et al. 2016. An imperative to monitor Earth's energy imbalance. *Nature Climate Change*, **6**(2), 138–144.

von Schuckmann, K., Sallée, J., Chambers, D., Le Traon, P., Cabanes, C., Gaillard, F., Speich, S., and Hamon, M. 2014. Consistency of the current global ocean observing systems from an Argo perspective. *Ocean Science*, **10** (3),547–557.

Von Schuckmann, K., Gaillard, F., and Le Traon, P. 2009. Global hydrographic variability patterns during 2003–2008. *Journal of Geophysical Research: Oceans (1978–2012)*, **114**(C9).

Von Schuckmann, K., Sallée, J., Chambers, D., Le Traon, P., Cabanes, C., Gaillard, F., Speich, S., and Hamon, M. 2014. Consistency of the current global ocean observing systems from an Argo perspective. *Ocean Science*, **10** (3),923–949.

Von Storch, H. and Zwiers, F. W. 2001. *Statistical analysis in climate research*. Cambridge University Press.

Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R. 2014. A global perspective on CMIP5 climate model biases. *Nature Climate Change*, **4**(3), 201–205.

Watanabe, M., Kamae, Y., Yoshimori, M., Oka, A., Sato, M., Ishii, M., Mochizuki, T., and Kimoto, M. 2013. Strengthening of ocean heat uptake efficiency associated with the recent climate hiatus. *Geophysical Research Letters*, **40**(12),3175–3179.

Wijffels, S., Roemmich, D., Monselesan, D., Church, J., and Gilson, J. 2016. Ocean temperatures chronicle the ongoing warming of Earth. *Nature Climate Change*, **6**(2),116–118.

Wong, A. P. and Riser, S. C. 2013. Modified shelf water on the continental slope north of Mac Robertson Land, East Antarctica. *Geophysical Research Letters*, **40**(23),6186–6190.

Wunsch, C. and Heimbach, P. 2014. Bidecadal thermal changes in the abyssal ocean. *Journal of Physical Oceanography*, **44**(8),2013–2030.

Xie, S.-P. 2016. Oceanography: Leading the hiatus research surge. *Nature Climate Change*, **6**(4),345–346.

Yashayaev, I., Bersch, M., and van Aken, H. M. 2007. Spreading of the Labrador Sea Water to the Irminger and Iceland basins. *Geophysical Research Letters*, **34** (L10602).

Zhang, L. and Wang, C. 2013. Multidecadal North Atlantic sea surface temperature and Atlantic meridional overturning circulation variability in CMIP5 historical simulations. *Journal of Geophysical Research: Oceans*, **118** (10),5772–5791.