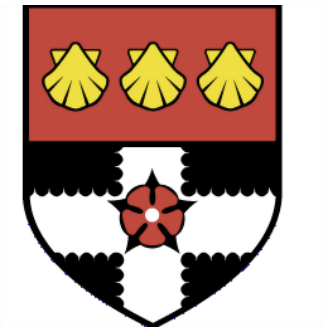
UNIVERSITY OF READING

Department of Meteorology



Understanding the performance of a decadal prediction system

Jonathan Ian Robson

A thesis submitted for the degree of Doctor of Philosophy

February 2010

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Jonathan Robson

Abstract

The UK Met Office has recently developed a Decadal Prediction System (DePreSys), which is initialised from the observed state of the climate system in order to constrain uncertainty in "near term" projections of climate change. This thesis aims to understand the decadal predictions made with DePreSys with a focus on the Atlantic sector.

First, an analysis of the observed decadal variability in the ocean is presented, and a significant warming of the North Atlantic in the 1990s is highlighted. By comparing the observed evolution with ocean model experiments, it is shown that the warming can be understood as a lagged response to the persistently positive North Atlantic Oscillation (NAO) forcing that preceded the warming.

An extended analysis of the mean skill of DePreSys, relative to predictions that do not assimilate information (NoAssim), is presented. It is shown that the mean bias of both DePreSys and NoAssim significantly affects the mean skill, and that the original published comparison of skill was in error. It is further shown that many different choices and processes affect the mean bias, including: sampling uncertainty, model drift, unbalanced initial conditions, and model errors, Moreover, it is demonstrated that much, but not all, of the increased skill of DePreSys at long lead times, relative to NoAssim, is due to a reduction of the bias, rather than prediction of the low frequency variability.

The rapid warming of the North Atlantic in the 1990s is analysed as a specific case study for decadal prediction. This analysis reveals specific limitations of the anomaly assimilation method: hindcasts can be sensitive to the model climatology, and errors can also arise from the non-linear equation of state.

Finally, based on the results, this thesis gives specific recommendations for the future development of DePreSys, and for decadal climate prediction in general.

Acknowledgements

I would like to give a huge thank you to my supervisor Rowan Sutton. Thank you for your guidance, your constant enthusiasm for my results, and above all for allowing me to hunt around at what I found interesting, but still managing to keep me focussed on overall goals. I would also like to thank Doug Smith, my CASE supervisor, for all his help. Thanks also for creating a fascinating decadal prediction system; and for letting me run wild all over it. I would also like to thank Ed Hawkins for all his help, and also the members Monitoring Committee, Maarten Ambaum and Stephen Belcher, for their guidance and advice.

I would also like to thank the members of Not The Tropical Group, and Ocean Group, and the people who kindly discussed my work with me, especially Keith Haines. A massive thank you to Leon Hermanson, Laura Wilcox, Ed Hawkins, Chunlei Liu, Buwen Dong, and Microsoft Word for reading and giving me comments on my thesis. I would like to thank Andy Heaps for all his help with computers, IDL, and especially for my three terabytes (and counting) of disk space. Thanks also to Alan Iwi for helping me with the climatology data in my hour of need.

I have had the pleasure of studying and working in the Meteorology Department for many years. When I think of how much I have grown since I arrived in September 2002, I owe thanks to a great many people; from the support staff, to lectures. Special thanks are due for Keith Shine, who gave me my first taste of research as an undergraduate, and a huge thank you to all the friends and office mates that I have shared my time with; Al H, Nicky C, Bethan H, Lesley A, Dann M, Kim B, Liang G, Yonghui L, and many, many others.

Thank you also to my mum and dad for always supporting and encouraging me to achieve my goals in life. Lastly, and certainly not least-ly, thank you to Kirsty for all the support and for always believing in me; even when I didn't. Thank you for letting me and, most importantly, encouraging me to follow my dreams.

"The same thrill, the same awe and mystery, come again and again when we look at any problem deeply enough. With more knowledge comes deeper, more wonderful mystery, luring one on to penetrate deeper still. Never concerned that the answer may prove disappointing, but with pleasure and confidence we turn over each new stone to find unimagined strangeness leading on to more wonderful questions and mysteries – certainly a grand adventure!"

Richard Feynmann (1918 - 1988)

Contents

1 Introduction **1**

1.1 The motivation for initialised climate projections 1

1.1.1 Uncertainties in projections of climate change 3

1.1.2 The potential to reduce uncertainty using initialised climate predictions 5

1.2 Decadal variability in the climate system and its potential predictability . 6

1.2.1 Mechanisms of decadal variability 9

1.2.2 Is the variability potentially predictable? 12

1.2.3 Summary of the potential to predict low frequency variability . . 13

1.3 Initialised decadal climate forecasts 13

1.4 Motivation of the thesis 17

1.5 Thesis structure 19

2 Observed decadal variability in the Atlantic Ocean: Characteristics and mechanisms **20**

2.1 Introduction 20

2.2 Ocean observations and analyses 21

2.2.1 Description of ocean analyses used in this chapter 22

2.2.2 A cautionary tale of ocean analyses: The global mean heat content 24

2.3 Observed variability of the ocean 26

2.3.1 Global variability 26

2.3.2 North Atlantic variability 29

2.4 A comparison of the evolution of the subpolar gyre in ocean analyses and idealised model experiment 32

2.4.1 Description of model experiments 32

2.4.2 Comparison between observations and model 34

2.4.3 Discussion 44

2.5 Chapter Conclusions 45

3 Description of the DePreSys decadal prediction system and introduction to mean skill scores 48

3.1 Introduction to DePreSys 48

3.1.1 Transient integrations 49

3.1.2 DePreSys 51

3.1.3 Anomaly Assimilation 54

3.1.4 NoAssim 56

3.1.5 DePreSys2 and DePreSys Perturbed Physics Ensemble (PPE) . . 57

3.2 Notation and skill 59

3.2.1 Notation 59

3.2.2 Introduction to skill 62

3.3 Mean Skill Scores of DePreSys 64

3.3.1 Replication of Smith et al, 2007 64

3.3.2 Climatological Difference 68

3.3.3 Sensitivity to the Climatological Period 74

3.3.4 NoAssim's sensitivity to the choice of climatology (C_{NoA}) 78

3.3.5 Summary and discussion 81

3.4 Chapter Conclusions 81

4 Analysis of mean skill and bias in DePreSys and NoAssim 84

4.1 Introduction 84

4.2 Significance testing 85

4.3 Analysis of DePreSys Skill 86

4.4 Mean bias of DePreSys and NoAssim 92

4.4.1 Mean hindcast bias 93

4.4.2 Understanding bias and drift 99

4.5 Understanding the source of improved skill 103

4.5.1 Shifted NoAssim 105

4.5.2 Mean bias correction 109

4.5.3 Summary and discussion 113

4.6 Chapter Conclusions 115

5 Analysis of mean drift in DePreSys and NoAssim 118

5.1 Introduction 118

5.2 The AMOC in the DePreSys and NoAssim hindcasts 120

5.3 Density 127

5.4 Is the drift still present in a updated version of DePreSys? (DePreSys2) . 134

5.5 What is causing the drift? 137

5.5.1 The construction of the climatology 138

5.5.2 The non-linear equation of state 139

5.5.3 The use of Dynamic Assimilation 141

5.6 Chapter Conclusions 142

6 Case study in decadal climate prediction: the rapid warming of the North Atlantic in the mid 1990s 145

6.1 Introduction 145

6.2 Rapid warming case study 146

6.3 Density errors 150

6.3.1 Errors in assimilated density anomalies 151

6.3.2 Did the density error affect the early warming hindcasts? 156

6.4 Chapter Conclusions 160

7 Improving the UK Met Office Decadal Prediction System: Hypothesis testing experiments 163

7.1 Introduction 163

7.2 Experiment description 165

7.2.1 New climatology 166

7.2.2 Perturbed salinity 167

7.2.3 Summary of prediction experiments 172

7.3 Results of prediction experiments 174

7.3.1 New Climatology 174

7.3.2 Perturbed Salinity 182

7.3.3 New climatology and perturbed salinity 185

7.3.4 What is the mechanism of the subpolar gyre warming in the DePreSys hindcasts? 188

7.4 Discussion and implications 197

7.5 Chapter Conclusions 199

8 Conclusions **203**

8.1 Introduction 203

8.2 Thesis overview and main conclusions 204

8.2.1 Observed decadal variability in the Atlantic Ocean: characteristics and mechanisms 204

8.2.2 Mean skill and bias of DePreSys and NoAssim 204

8.2.3 Mean drift of DePreSys and NoAssim 206

8.2.4 Case study in decadal prediction: the rapid warming of the North Atlantic in the mid 1990s 207

8.2.5 Improving the Decadal Prediction System: Hypothesis testing experiments 208

8.3 Recommendations and future work 209

8.3.1 Recommendations for the future development of DePreSys 209

8.3.2 Recommendations for the future development of initialised decadal predictions 211

8.4 Concluding remarks 213

A Description of Ocean Analyses used in this thesis **214**

A.1 Levitus 214

A.2 Original Met Office Ocean analysis 215

A.3 Updated Met Office Ocean analysis 216

A.4 ECMWF Ocean Reanalysis System 3 218

Bibliography **219**

Chapter 1

Introduction

This thesis presents analyses that are aimed at understanding decadal climate predictions, and focuses on the UK Met Office's Decadal Prediction System (DePreSys, *Smith et al.* (2007)). Therefore, this first chapter provides an introduction into the motivation for decadal (or initialised) climate predictions in section 1.1, and a brief review of the potential to predict decadal variability in section 1.2 . A review of the first attempts at initialised climate predictions is presented in section 1.3. Finally, the general motivation for this thesis, and the thesis structure, are presented in sections 1.4 and 1.5 respectively.

1.1 The motivation for initialised climate projections

The Earth's climate is changing. The global mean climate has been warming over the past 100 years, and particularly in the last 50 years (see figure 1.1; *Solomon et al.* (2007)). Changes have also been detected in a vast array of other climate variables; such as oceanic heat content, sea level, precipitation, and snow cover. Many detection and attribution studies have attributed the changes to anthropogenic forcing of the climate, primarily through the emission of radiatively active gases into the atmosphere. The large body of evidence led the Intergovernmental Panel on Climate Change (IPCC) fourth Assessment report (AR4; *Solomon et al.* (2007)) to conclude that it is *very likely* that anthropogenic greenhouse gas increases caused most of the observed warming over the last 50 years.

As anthropogenic emissions of greenhouse gases are likely to continue into the future, the IPCC also perform an array of future projections of climate change under a range of socio-economic scenarios (see figure 1.1). The projections show that the current rate of warming will continue over the 21st century, and could even increase, which may have large impacts on important climate variables such as precipitation and weather extremes. Even if emissions of green house gases were reduced to zero now, future climate change would be inevitable due to the thermal inertia of the global ocean (*Solomon et al.* (2007)). Therefore, even if rigorous mitigation policies were put in place, significant future climate change now seems inevitable.

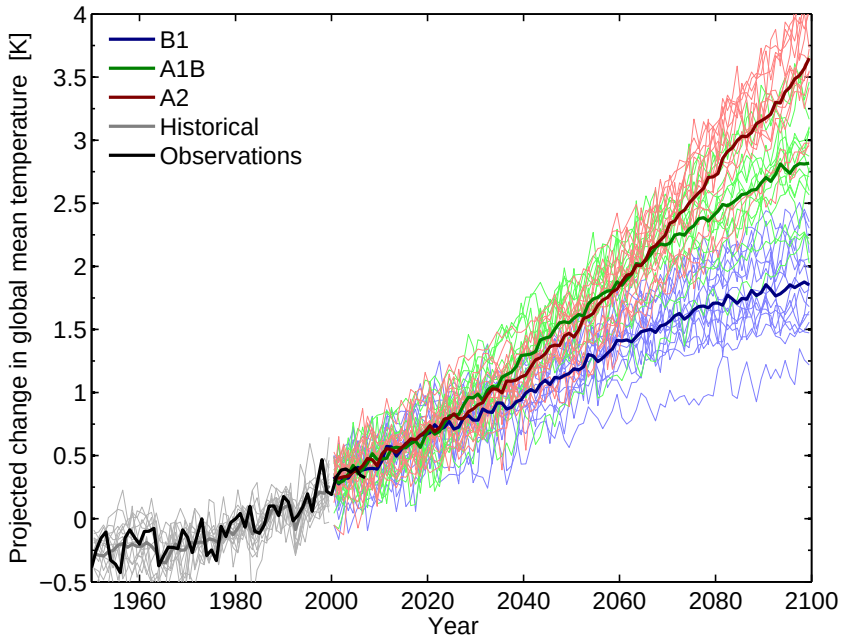


Figure 1.1: Annual average global mean surface air temperature (SAT) observations from 1950 to 2007 taken from the HadCRUT3 (black; *Brohan et al.* (2006)). Also shown is the SAT from 15 climate models that were forced with historical forcings (grey). Predictions of future global mean SAT change from the 15 models forced with three different future emissions scenarios are also shown from 2000 to 2100 (thin coloured lines). The emissions scenarios are the SRES A2 (red), A1B (green), and B1 (blue), and are representative of high, medium and low future emissions. The thick coloured lines represent the multimodel mean of each emission scenario, and represents the uncertainty in future climate projections due the uncertainty in future emissions. All temperatures are expressed as anomalies relative to the 1971-2000 mean.

Figure taken from the *Hawkins and Sutton* (2009a)

As the climate continues to change, the need for society to adapt will become increasingly important. Adaptation decisions will need to be taken at the regional- or local-scale, requiring accurate projections of climate change to be increasingly delivered at increasingly

smaller scales. Unfortunately, there are presently large uncertainties in the predictions of changes in climate variables at these scales. Thus, to be able to deliver predictions to inform adaptation decisions, reducing uncertainties in the projections of climate change is an important priority for the climate modelling community.

1.1.1 Uncertainties in projections of climate change

To reduce the uncertainty in climate projections it is first important to understand the sources of uncertainty. The uncertainty in projections of climate change can be partitioned into three separate contributors, which are:

1. **Internal variability** - The uncertainty that is caused by the natural variability of the climate system. i.e. variability that is internally generated and is, therefore, not externally forced.

2. **Model error** - The models are not perfect representations of the climate system; all models posses errors, and we are uncertain which models are the 'best'. The uncertainty is created because comparable climate models simulate different mean climate states, climate variability, or the climates response to radiative forcing. The different simulations of the models is partly caused by missing processes.

3. **Scenario uncertainty** - The anthropogenic drivers of climate change (e.g. greenhouse gas emissions and land use change) depend on the socio-economic conditions. The exact future socio-economic conditions are dependent on a great many factors, and are, therefore, difficult to predict. The uncertainty in predicting the emissions of greenhouse gases and aerosols creates uncertainty in the future climate. The scenarios used by the IPCC are designed to give a series of different, but plausible, futures.

Recently, *Hawkins and Sutton* (2009a) quantified the uncertainty from all three contributions in projections of the 21st century surface air temperature (SAT) from the World Climate Research Program's (WCRP)'s Coupled Model Intercomparison Project 3

(CMIP3), which were used to inform the IPCC AR4. Figure 1.2 shows the relative contribution of each source of uncertainty for the global mean SAT, and the SAT over a smaller region (in this case the British Isles) from *Hawkins and Sutton* (2009a). Figure 1.2 shows clearly that, although the scenario uncertainty is a large source of uncertainty for predictions of the climate in the latter half of the 21st Century, in the "near term" (i.e. the next few decades) this term is not important.

Hawkins and Sutton (2009a) showed that the model error is a significant source of uncertainty at all lead times for both the global mean and regional climate projections, accounting for up to $\sim$ 70% of the uncertainty (see figure 1.2). However, in the "near term" the internal variability is also a significant source of uncertainty. For the global mean the internal variability accounts for $\sim$ 30% of the uncertainty in the first decade, which quickly reduces. However, at the regional-scales the internal variability contributes $\sim$ 70% of the total uncertainty in the first decade. Thus, the evolution of a particular region's climate, over the next decade, could be completely different to evolution of global mean climate.

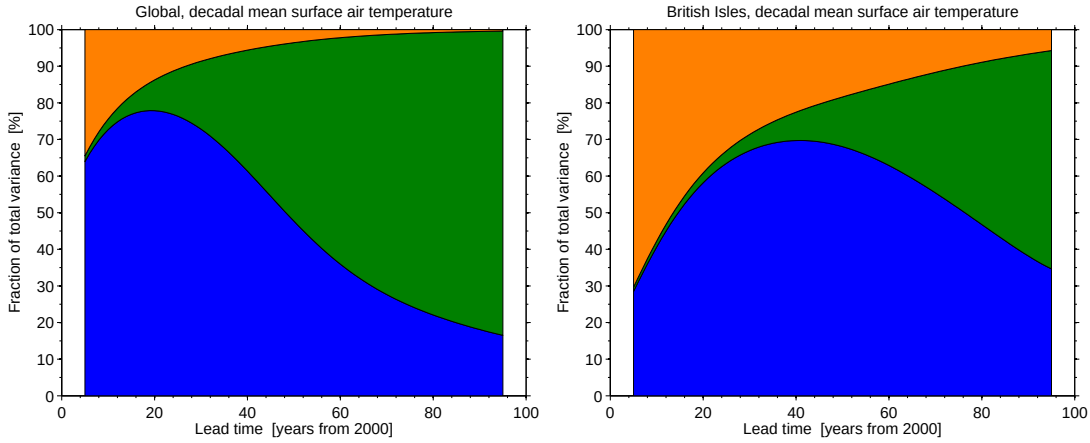


Figure 1.2: The left plot shows the relative contribution, as a percentage, of internal variability (orange), model error (blue) and scenario uncertainty (green) to the *total* uncertainty of **global mean** SAT projections as a function of projection lead time. The plot on the right shows the same but now for the contributions to the uncertainty in SAT projections over a small **region** (the British Isles). Note that the total uncertainty increases with lead time (i.e. see figure 1.1). Both plots are taken from *Hawkins and Sutton* (2009a).

1.1.2 The potential to reduce uncertainty using initialised climate predictions

The internal variability is a source of significant uncertainty for "near term" projections, especially at the regional-scale. In current climate projections such as those used in the IPCC, there is no attempt to predict the internal variability. This is primarily because projections of climate in the year 2100 will not be sensitive to the initial conditions. However, the next decade (where the scenario uncertainty is small) is an important planning time-horizon for decision makers, and for infrastructure planning. Reducing the uncertainty in the "near term" projections is vital for effective decision making, and for efficient adaptation. Therefore, attempting to predict the natural variability of the climate system, by initialising climate models from the observed state, could significantly reduce the uncertainty in "near term" climate projections.

It is also clear from the results of *Hawkins and Sutton* (2009a) that the model error is a huge source of the total uncertainty in climate projections. The model error could be reduced through increased investment into model development. However, initialised climate predictions could provide a powerful new tool to understand climate models and their errors. Validating initialised climate predictions against observations, at the process level, will allow complex climate models to be validated at an unprecedented level of detail. Understanding how the climate models perform could highlight significant deficiencies, or missing processes, therefore, allowing model development to be directed efficiently.

Initialising climate projections from the observed state, and hence attempting to predict the natural variability, could prove extremely useful for decision makers. However, what is the basis for assuming that initialising climate projections has the potential to reduce uncertainty due to internal variability? Is the variability actually predictable, or is it essentially stochastic in nature (e.g. unpredictable)? The next section will give a brief review of decadal variability in the climate system, and its potential to be predicted.

1.2 Decadal variability in the climate system and its potential predictability

Variability is present in the climate system on a range of time and space scales. However, not all the variability is potentially predictable at long time scales. For example, the evolution of the atmosphere is only predictable deterministically for a time horizon of 5-10 days. However, slower parts of the climate system, such as the ocean, evolve much slower than the atmosphere and could potentially offer a source of predictability on time scales longer than a few days. For example, the coupled prediction of ENSO (El Niño-Southern Oscillation) variability enables skillful prediction of climate, particularly in the tropical region, on seasonal time scales (*Latif et al.* (1998)). Are there any climate patterns that could be potentially predictable at interannual to decadal time scales?

Low-frequency variability is observed in the climate system. One example of multi-decadal variability is the Atlantic Multidecadal Oscillation (AMO (*Schlesinger and Ramankutty* (1994); *Kerr* (2000)); which is also referred to as Atlantic Multidecadal Variability (AMV)), which involves multidecadal variability in the Atlantic sea surface temperatures (SSTs), particularly in the North Atlantic. Figure 1.3 shows the observed AMO index that appears to have an oscillatory period of $\sim$ 60-70 years. A positive episode of the AMO is associated with warm surface temperatures across the North Atlantic Ocean (see figure 1.3), but the AMO is also associated with important variability over continents. Research has shown that the AMO is likely to affect the surface temperature of North America and Europe (*Sutton and Hodson* (2005); *Knight et al.* (2005); *Knight et al.* (2006)), hurricane activity (*Webster et al.* (2005); *Trenberth and Shea* (2006)), and rainfall in Brazil and Africa (*Knight et al.* (2006)).

Large scale, low frequency variability is also present in the Pacific. In the North Pacific, covarying fluctuations of surface pressure and the SSTs is known as the Pacific Decadal Oscillation (PDO, *Mantua et al.* (1997); *Mantua and Hare* (2002)). The PDO is associated with changes in the biota of the North Pacific, and temperature over the North American continent (*Mantua et al.* (1997); *Mantua and Hare* (2002)). Linked to the PDO is a similar Pacific wide pattern known as the Interdecadal Pacific Oscillation (IPO;

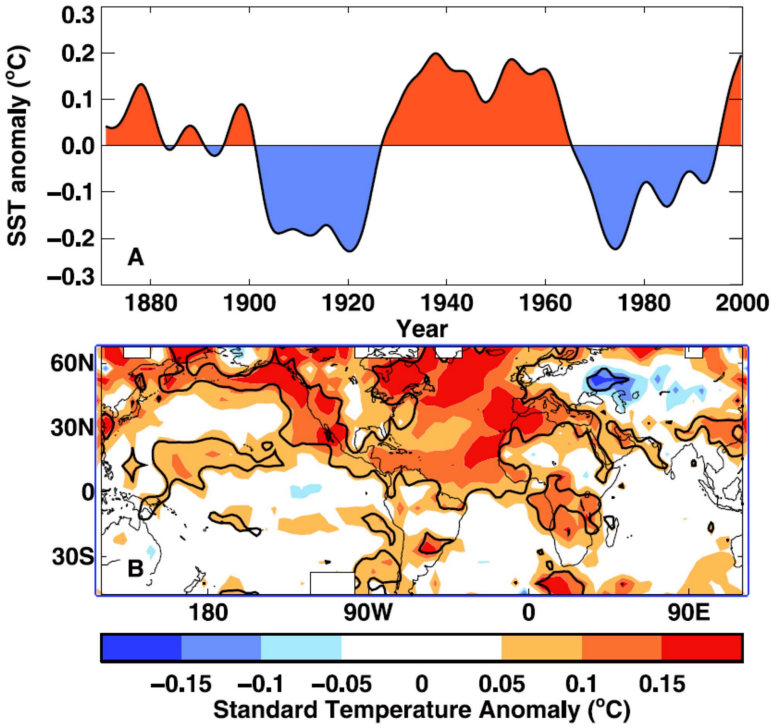


Figure 1.3: a) shows the observed AMO index calculated from detrended area weighted mean North Atlantic SSTs (SST data taken from HadISST (*Rayner et al.* (2003)). b) shows the surface temperature anomaly associated with one positive standard deviation of the AMO index calculated by regressing combined land and sea surface temperatures onto the AMO index. The surface temperature anomalies are taken from the HadCRUT data set (*Jones et al.* (2000))). Figure 3 from *Knight et al.* (2005)

Power et al. (1999)), which has been shown to affect the position of the South Pacific Convergence zone in a way that is separate to ENSO variability (*Folland et al.* (2002)), as well as rainfall and surface temperature variability over Australia (*Power et al.* (1999)).

To place a handle on the relative importance of the low frequency variability, compared to the total variance of the climate system, and where such low frequency exists, *Boer* (2000, 2004) and *Boer and Lambert* (2008) calculated the fraction of the total variance that is low frequency. Figure 1.4, taken from *Boer and Lambert* (2008), shows that a significant fraction of the total variance of SAT in coupled climate models is long time scale variability that could be *potentially* predictable. However, the potential predictability depends on the region, and the climate variable of interest. For example, figure 1.4 shows that, in general, SAT is potentially more predictable than precipitation. The largest fractions of low frequency variability in SAT occur over the ocean: in particular, in the

North Atlantic, North Pacific and the Southern Ocean. It must be noted that although there are encouraging signs of potentially predictable variability over the ocean, the low frequency variability does not seem to account for a large proportion of the variance over land.

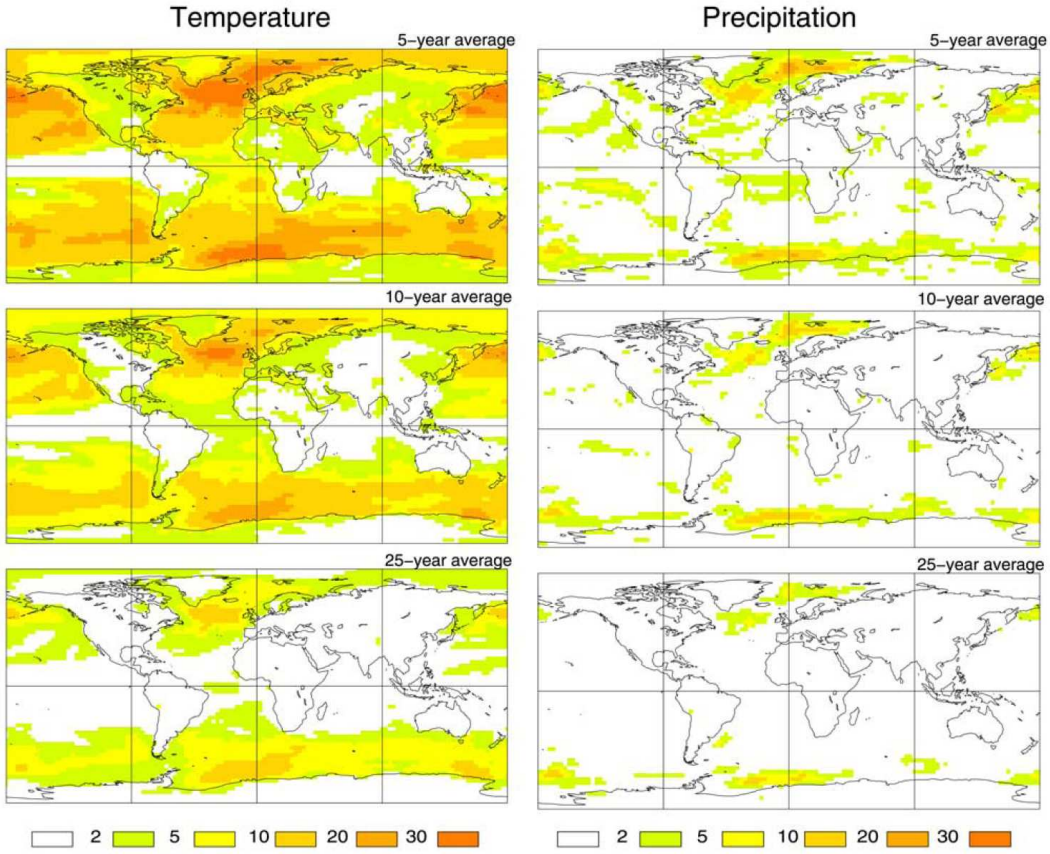


Figure 1.4: The fraction of the total variance from a multimodel ensemble that is associated with low frequency variability, expressed as a percentage, for SAT (left) and precipitation (right), calculated using 5-, 10- and 25-year averages. Taken from *Boer and Lambert* (2008).

The results of *Boer* (2000, 2004) and *Boer and Lambert* (2008) are similar across many coupled climate models, suggesting that the relative importance of the low frequency variability is not just an artifact of one particular model. If the relative importance of the low frequency variability, as suggested in *Boer and Lambert* (2008), is real, what are the mechanisms that control the variability?

1.2.1 Mechanisms of decadal variability

Hasselmann (1976) and *Frankignoul and Hasselmann* (1977) suggested that much of the low frequency variability of temperature over the ocean is simply a thermodynamic response of the ocean's mixed layer to the "white noise" forcing[1] of the atmosphere. Forcing a mixed layer ocean model with a stochastic heat flux produces low frequency variability (*Hasselmann* (1976)). The evolution of the ocean is the response to the sum of the previous forcing; in other words, the ocean reacts to the integral of the previous atmospheric forcing, and not just the instantaneous forcing that is applied. This integrating property of the ocean was later confirmed by *Hall and Manabe* (1997) as an adequate description of the variability in some areas, particularly in the mid latitudes. However, this simple model of climate variability does not describe the variability as well near coasts, or in regions where advection or eddies are important.

Similarly, the ocean also integrates wind forcing, particular the wind stress curl (*Frankignoul et al.* (1997)), which can affect the gyre circulation through Sverdrup balance. Importantly, variability in the wind stress curl can also create westward propagating Rossby waves through vertical displacement of the thermocline. The emission of Rossby waves can contribute significantly to the low frequency variability in the extra tropics due to their slow propagation in the extratropical ocean, compared to the topics.

The variability created because of the wind forcing may be particularly important for the variability in the North Pacific, especially the forcing associated with the Aleutian low (*Schneider et al.* (2002); *Schneider and Cornuelle* (2005)). However, it must be noted that there is some controversy to the exact mechanism for the PDO. The transmission of ENSO variability via atmospheric teleconnections (*Trenberth and Hurrell* (1994); *Newman et al.* (2003)), and via the ocean through coastally trapped Kelvin waves, are also likely to play an important role in the variability of the PDO (*Schneider and Cornuelle* (2005) and of the IPO. Models suggest that it is likely that some PDO type variability can still be produced in the absence of ENSO (*Pierce et al.* (2001)). Therefore, a complex interaction between extratropical and tropical variability seems to drive the PDO

[1] i.e. forcing that has equal power at all frequencies. The variability has no preferred time scales. It is called "white noise" in analogy to white light.

(*Schneider and Cornuelle* (2005)).

The peak in the low frequency contribution to the total variance in the North Atlantic (see figure 1.4) is largely associated with fluctuations in the strength of the Atlantic Meridional Overturning Circulation (AMOC) in the models (*Delworth et al.* (1993); *Timmermann et al.* (1998); *Vellinga and Wu* (2004); *Latif et al.* (2004); *Knight et al.* (2005)). In essence the AMOC is typified by a Northward flow of warm light water at the surface, and a southward return flow of colder dense water at depth, although the actual circulation is much more complicated (*Hawkins and Sutton* (2007)). The AMOC circulation is driven by buoyancy forcing at high latitudes (i.e. primarily through the cooling of the surface water by the atmosphere), which creates North Atlantic Deep Water through convection in the Labrador, Irminger and Greenland-Iceland-Norwegian Seas. Importantly, the AMOC is also driven by mechanical forcing, primarily by wind stress (especially in the Southern Ocean), and also by the tides. The mechanical forcing inputs energy that mixes dense waters from depth to the surface (*Munk and Wunsch* (1998); *Wunsch and Ferrari* (2004); *Kuhlbrodt et al.* (2007)). Significant model evidence suggests that variability in the buoyancy forcing in the North Atlantic is an important driver of the multidecadal variability in the strength of the AMOC (*Timmermann et al.* (1998); *Thorpe et al.* (2001); *Eden and Willebrand* (2001); *Vellinga and Wu* (2004)).

The flow of relatively warm water northwards, and the return flow of cold water southward, results in a net northward transport of heat in the Atlantic. The AMOC's maximum heat transport is approximately 1 PW (10^{15} W; *Trenberth and Caron* (2001)) and is key in regulating the relatively warm climate of Europe (*Rahmstorf* (2000)). In models, fluctuations in the strength of the overturning have been shown to affect the northward heat transport of the ocean, which affects the surface temperatures (*Latif et al.* (2004); *Knight et al.* (2005)). Figure 1.5 shows the decadal mean surface temperature anomalies associated with anomalies in the strength of the AMOC from one particular model (in this case it is the UK Met Office HadCM3 (*Gordon et al.* (2000))). When the AMOC is anomalously strong, much of the Northern Hemisphere is warm (see figure 1.5 a and e). However, when the AMOC is weak, the Northern Hemisphere is relatively cool (see figure 1.5 d and h). Therefore, model evidence suggests that the observed AMO is a manifestation of AMOC variability (*Latif et al.* (2004); *Knight et al.* (2005); *Zhang* (2007)).

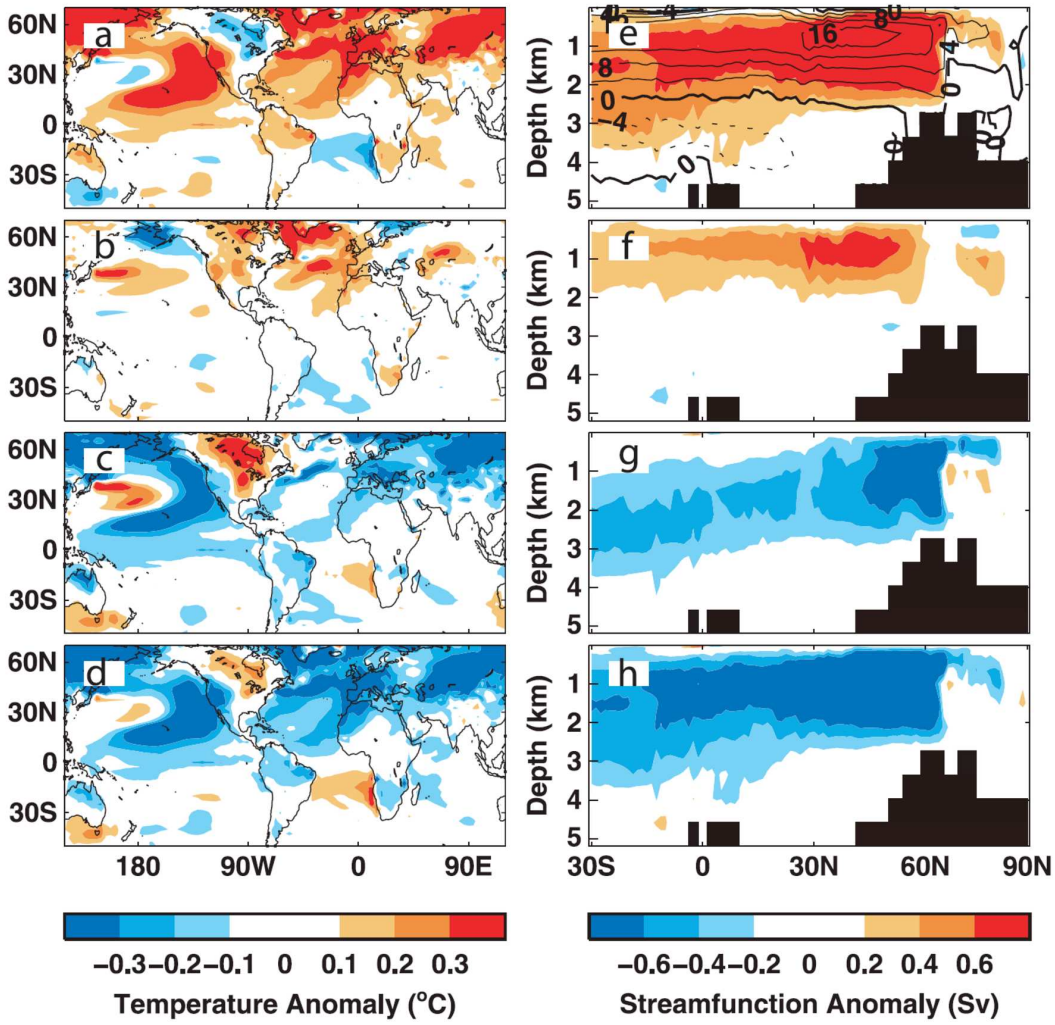


Figure 1.5: a-d shows decadal mean surface temperature anomalies associated with the maximum (a) to the minimum (d) in the overturning circulation in the HadCM3 model. e-h shows the AMOC stream function anomaly associated with the maximum to minimum in the centennial oscillation of AMOC strength. The black contours in figure e) shows the climatological stream function. Figure taken from *Knight et al.* (2005).

Poor observations of the ocean, and especially the subsurface, have meant that an extensive use of coupled climate models has been needed to understand AMOC variability, and its impact. Many climate models agree that the AMOC possesses significant multidecadal variability, but this variability has not been directly observed in the real world. There is also much uncertainty in the AMOC variability in the models: the modes of variability differ from model to model, which makes understanding the exact mechanisms of the variability in the observed AMO difficult. It must also be noted that significant multidecadal variability of Atlantic heat content is likely to be caused by the multidecadal

variability of the North Atlantic Oscillation (NAO, *Hurrell* (1995)), the dominant mode of atmospheric variability in the North Atlantic, even in absence of AMOC variability (*Lozier et al.* (2008)).

1.2.2 Is the variability potentially predictable?

Of course, the presence of low frequent variability in the climate system does not ensure that such variability would be predictable. One way to understand the potential predictability of the climate system is to use "perfect model" experiments. "Perfect model" experiments create an ensemble of model integrations from the same ocean state by perturbing the atmospheric initial condition. Perturbing the atmospheric state, which is not predictable at long time scales, allows a handle on the potential predictability that arises by initialising the slower parts of the climate system. A "perfect model" study essentially examines the ability of a model to predict the climate variability assuming that the ocean was perfectly observed, and so should be taken to represent the upper limit of actual predictability assuming that the model is a perfect representation of the climate system.

Griffies and Bryan (1997a,b) presented the first "perfect model" study of potential predictability in the North Atlantic, which showed that many variables could be potentially predictable for many years. The potential predictability of different climate variables is, again, dependent upon the region and the variable of interest. For example, *Griffies and Bryan* (1997b) showed that sea surface temperature was potentially predictable for a couple of years, and possibly up to 5. The circulation of the ocean, and deeper temperature properties away from the noisy atmosphere, are potentially predictable for longer. Investigating the source of the potential predictability, *Griffies and Bryan* (1997a,b) showed that much originated from the integrating property of the ocean, as postulated by *Hasselmann* (1976). However, a significant proportion of the potential predictability was due to the potential predictability of the AMOC itself.

The high potential predictability of the AMOC was confirmed in later "perfect model" studies, which used different models (*Collins* (2002); *Collins and Sinha* (2003); *Pohlmann et al.* (2004); *Collins et al.* (2006a)). These studies suggest that the AMOC could be po-

tentially predictable for up to a decade in advance, although the potential predictability varies between models (*Collins et al.* (2006a)). Associated with the high potential predictability of the AMOC was potential predictability of SSTs in the North Atlantic for a number of years ahead. The potential predictability of the SSTs in the North Atlantic provides the potential to predict the SAT over North America and Europe. The potential predictability is smaller for SAT, but *Collins* (2002) suggested that the SAT in the North Atlantic could be potentially predictable for 2 years. It must also be noted that the potential predictability is likely to depend on initial conditions (*Collins and Sinha* (2003); *Hermanson and Sutton* (2009)). For example, *Collins and Sinha* (2003) show that for one initial condition the AMOC could be potentially predictable for up to 50 years.

1.2.3 Summary of the potential to predict low frequency variability

Low frequency variability is present in the climate system, and, in models at least, does contribute a significant fraction of the total variance of surface temperature. In particular, the North Atlantic is an area which is associated with significant low frequency variability, which could be potentially predictable. Potential predictability in Atlantic surface temperatures is associated with the potential predictability of the AMOC, which "Perfect model" studies suggest could be predictable for up to a decade in advance. Predictions of the AMOC could improve predictions of temperature in North America and Europe, and so future predictions of the AMOC could bring benefits to society. However, uncertainty still exists in the potential to predict the real ocean, particularly due to the poor observations of the ocean, and because the modes of variability differ from model to model.

1.3 Initialised decadal climate forecasts

The "perfect model" studies showed that initialising climate models could constrain projections of climate change over the next decade in certain regions. However, the actual predictability of the climate on these scales is a very different research question. The

"perfect model" studies assumed that the ocean was perfectly observed, that the assimilation of information was also perfect, and that the model itself is perfect. In reality, the observation of the ocean are historically poor and far from perfect, especially in the deep ocean. There are also significant problems with coupled climate models, such as mean model biases, and the different modes of variability that they posses.

Smith et al. (2007) was the first study to compare initialised climate predictions, from the UK Met Office's Decadal Prediction System (DePreSys)[2], with un-initialised climate predictions which do not assimilate information (called NoAssim). DePreSys and NoAssim are both based upon the UK Met Office's HadCM3 coupled climate model. Predictions from both are forced by a best guess of radiative forcing over the upcoming decade allowing them both to predict the global warming trend. The key difference is that DePreSys is initialised from the observed state of the climate by assimilating information into the atmosphere and, crucially, the ocean. Therefore, the NoAssim hindcasts, which do not assimilate information, are comparable to the climate projections that have been used to inform the IPCC (figure 1.1).

Climate models are not perfect. Therefore, assimilating the observed temperature and salinity values (i.e. the full fields) would move the model away from it's own prefered climatological state. Once initialised, the predictions would then move, or drift, back to the model's prefered state. In seasonal forecasting, the mean drift is calculated from a series of predictions, and then removed to return forecast anomalies (*Stockdale* (1997)). However, the sizable model errors in coupled climate models would likely make the drift over a decade larger than any predictable signals (*Smith et al.* (2007)). To reduce the effects of the climate drift, DePreSys was designed to assimilate *anomalies* onto a model climatology (*Smith et al.* (2007)).

Smith et al. (2007) analysed a series of 10 year DePreSys and NoAssim ensemble hindcasts[3] that were initialised every season from 1982-2001. The hindcasts were validated against the observed variability in order to understand both system's ability to predict the observed changes. Figure 1.6, taken from *Smith et al.* (2007), clearly shows that

[2]The DePreSys system is described in detail in chapter 3.

[3]i.e. predictions that are initialised from an observed state in the past. Hindcasts allow us to understand the prediction system's ability to skillfully predict the evolution of the observed climate.

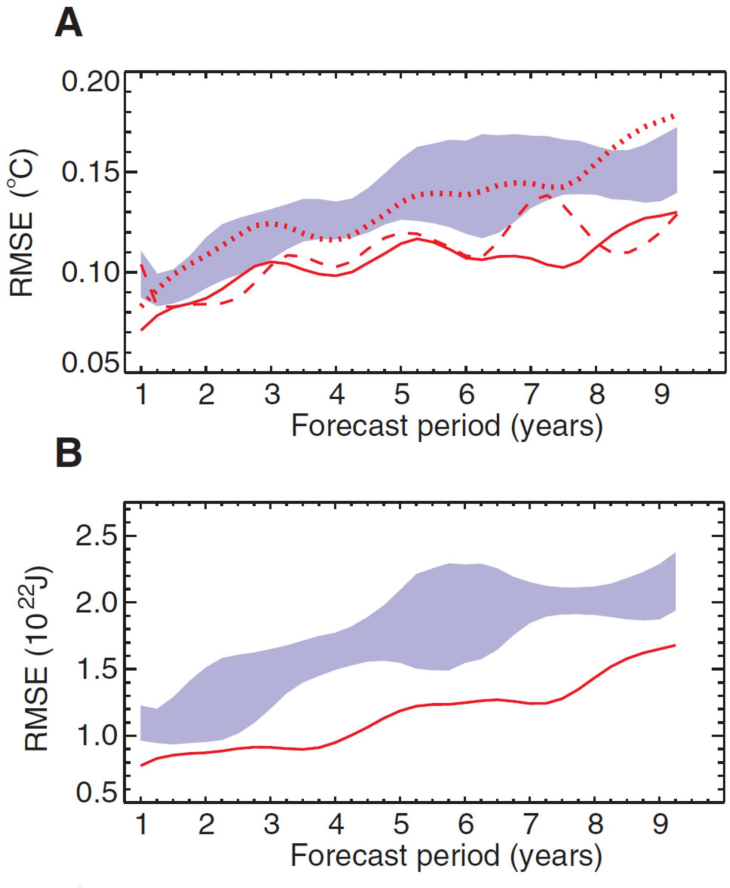


Figure 1.6: Shows the impact of assimilating information on hindcast skill. a) shows the Root Mean Squared Error (RMSE) of global mean SAT anomalies, relative to 1979-2001, as a function of forecast lead time for both DePreSys (solid red line) and NoAssim (the blue area shows the 5 to 95% confidence intervals where the difference between DePreSys and NoAssim are not significant). The hindcast anomalies are verified against HadCRUT2v (*Folland et al.* (2001)). The dashed red curve shows the effect of removing from the DePreSys hindcasts differences between DePreSys and NoAssim that are linearly attributable to the state of ENSO. The dotted red line shows the effect of removing the mean of the difference between DePreSys and NoAssim hindcasts of SAT for coming 9 years. b) shows the same, but now for the upper ocean (0-113m) heat content anomalies , relative to 1941-1996. Observations of upper ocean heat content are taken from the Original Met Office ocean analysis (*Smith and Murphy* (2007)). DePreSys appears to give improved predictions relative to NoAssim for the entire decade. Taken from figure 1 of *Smith et al.* (2007)

DePreSys improves predictions of the global mean SAT and upper ocean (0-113m) heat content, relative to NoAssim, for an entire decade. The improved skill at short lead times was found to be mainly attributable to the ability of DePreSys to predict ENSO, which has a large affect on the global SAT temperatures. However, the prediction of ENSO was not found to improve the predictability beyond 12-15 months. Instead, *Smith et al.*

(2007) show that the improvement in predictions at long lead times was due to a reduced warm bias in DePreSys compared to NoAssim, which *Smith et al.* (2007) argue is due to improved prediction. When the mean difference between the DePreSys hindcasts and the NoAssim hindcasts, over the first 9 years, was removed from the DePreSys hindcasts (see figure 1.6 red dotted line), the significant difference between the two systems is also removed. This led *Smith et al.* (2007) to conclude that the improved predictability was due to the initialisation of low frequency (decadal - multidecadal) variability.

Assimilating information into DePreSys does not improve predictions everywhere. Figure 1.7, also taken from *Smith et al.* (2007), shows spatial maps of Root Mean Squared Error (RMSE) when validating 9-year mean SAT anomalies from DePreSys and NoAssim. The difference between the RMSE's for NoAssim and DePreSys gives a handle on where DePreSys is improving predictions on average. There are large areas where DePreSys is improving prediction of 9-year mean SAT (see figure 1.7 c), with significant improvement over South America, Africa and the North American continent. Much of the improvement in SAT predictions is coincident with improvements in the prediction of upper ocean (0-113m) heat content (see figure 1.7 d). Interestingly, DePreSys does not to seem improve predictions of SAT in the North Atlantic, the area where the potential predictability studies suggest we should expect the largest improvement.

Two further hindcast experiments have also been undertaken by *Keenlyside et al.* (2008) and *Pohlmann et al.* (2009), which are based on separate decadal prediction systems. Both prediction systems also assimilate anomalies to avoid climate drift. The decadal hindcasts created in *Pohlmann et al.* (2009) are similar to DePreSys as the anomalies are assimilated into the full 3D ocean. The predictions of *Keenlyside et al.* (2008) are somewhat different in that they only assimilate SSTs to create the initial conditions of the hindcasts. Due to the large computational expense of decadal climate predictions, both the studies of *Keenlyside et al.* (2008) and *Pohlmann et al.* (2009) use significantly less hindcasts, and so were not as successful in showing improved predictability as *Smith et al.* (2007). Nevertheless, both studies show that the initialisation of the ocean improves predictions of SST in the North Atlantic region for many years ahead, when compared to standard 20th century climate projections (i.e. predictions that do not assimilate information). The improvement for SAT is less clear. *Keenlyside et al.* (2008) show that

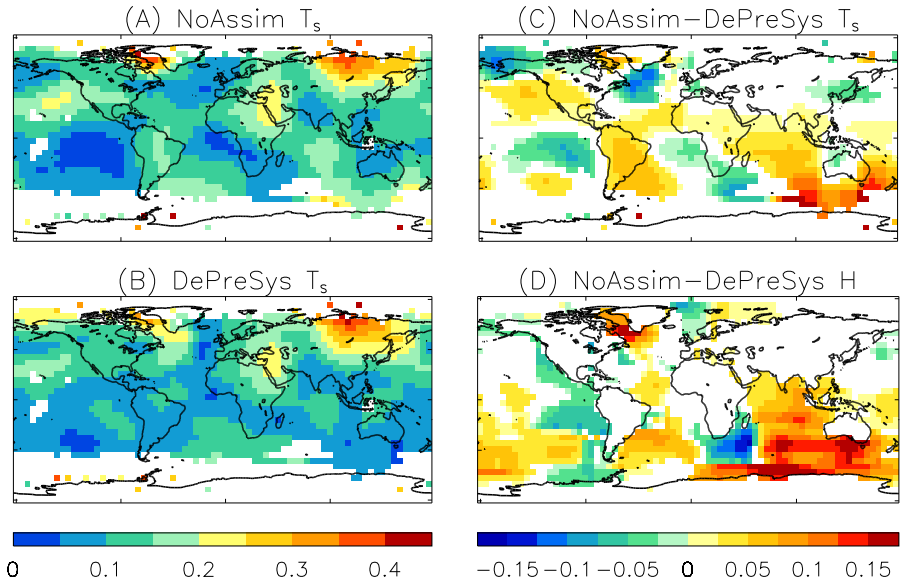


Figure 1.7: a) shows the RMSE of validating 9 year mean SAT anomalies, relative to 1979-2001, from ensemble mean NoAssim hindcasts, verified against HadCRUT2v (*Folland et al.* (2001)). b) as a) but now for DePreSys. c) NoAssim minus DePreSys RMSE for 9-year mean SAT anomalies (i.e. a) minus b)). Differences are shown where they are significant at the 5% level. d) as c) but for mean 0-113m ocean heat content anomalies, relative to 1941-1996. In all panels, each 5° latitude by 5° longitude pixel represents the RMSE for predictions spatially averaged over 35° latitude by 35° longitude box centred on that pixel. Figure 3 from *Smith et al.* (2007)

there is some improved predictability of SAT over Western Europe and North America, but *Pohlmann et al.* (2009) show little improvement at predicting SAT over land. However, *Pohlmann et al.* (2009) do show encouraging signs at predicting the AMOC at high latitudes (48°).

1.4 Motivation of the thesis

Although the assimilation of information seems to improve the predictions of surface temperature in *Smith et al.* (2007), there is still a huge range of questions that remain regarding initialised climate predictions. For example, what mechanisms give rise to the improved predictability at long lead times, and can the predictions inform us of the predictability of the real world?

One region of interest is certainly the North Atlantic. The results of *Smith et al.* (2007)

suggest that the assimilation of information is not improving the prediction of surface temperature here. The lack of predictability in the North Atlantic is in disagreement with the potential predictability studies that suggest that the North Atlantic is a region where we should expect the initialisation to improve the predictions, primarily because of the predictability of the AMOC. Therefore, should we be concerned about the apparent lack of skill in the North Atlantic region (i.e. are the predictions not reaching their potential?), or do the results show that the actual predictability of the North Atlantic, or the AMOC, is low?

It is our ambition to also use initialised climate predictions to understand the mechanisms responsible for variability in the real world, and to validate the underlying model (in DePreSys's case, HadCM3) at the process level. However, before initialised climate predictions can be used to understand the important processes in the real world, or to highlight deficiencies in the models, it is important to first assess the prediction system for this role. An important problem of all prediction systems is the initialisation of predictions. DePreSys uses the anomaly assimilation method to remove climate drift, but no independent analysis of the method has yet been performed. Can we trust such an anomaly assimilation system to adequately create initial conditions for predictions? To ensure that informed decisions can be made about the initialisation of decadal forecasts, and for future development of initialisation strategies, it is important that any potential limitations of using such an anomaly assimilation method are adequately discussed.

To improve the understanding of decadal predictions, and in particular those made by DePreSys, this thesis aims to move beyond the mean skill scores presented in *Smith et al.* (2007), and try to understand the performance of DePreSys at the process level. The analysis focuses primarily on the Atlantic sector where the presence of the AMOC should ensure that the potential predictability is high. Validating the evolution of DePreSys hindcasts against the observations, and understanding the success and failures of predictions at the process level, will increase our understanding of the mean skill scores that were shown in *Smith et al.* (2007), and also of DePreSys system itself. A further aim of the thesis is to identify opportunities and priorities for improving initialised decadal climate predictions.

1.5 Thesis structure

To be able to understand the skill of DePreSys, it is first important to understand the observed decadal variability of the ocean. Therefore, an analysis of the characteristics of the recent observed decadal variability in the Atlantic, and the important mechanisms, is presented in chapter 2.

After the analysis of the observed variability, chapter 3 presents a description of the UK Met Office's Decadal Prediction System DePreSys. An initial comparison with *Smith et al.* (2007) highlighted an error in the construction of climatologies, which added a bias to the NoAssim hindcast anomalies. The error was corrected and a new analysis of the mean skill of DePreSys is presented in chapter 4, which tries to expand on the original analysis of skill presented in *Smith et al.* (2007), by beginning to understand the origin of the mean skill and bias of both prediction systems.

Further analysis showed that the DePreSys hindcasts were drifting, especially in the North Atlantic - a surprise given that the anomaly assimilation approach was designed to remove the drift. The analysis of the mean drift of DePreSys and NoAssim is presented in chapter 5.

To analyse the DePreSys hindcasts at the process level, and to highlight specific hindcast errors, chapter 6 presents an analysis of DePreSys's ability to predict a particular case study of observed Atlantic variability.

The analyses in chapters 5 and 6 highlighted two potential problems with the anomaly assimilation method that is used to assimilate observations into DePreSys, and enabled the framing of specific hypotheses about the causes of specific hindcast errors. The two hypotheses are then tested in chapter 7, by making changes to the assimilation stage of DePreSys.

Finally, the main conclusions of the thesis are summarised in chapter 8, with a brief discussion of possible future work.

Chapter 2

Observed decadal variability in the Atlantic Ocean: Characteristics and mechanisms

2.1 Introduction

To be able understand the performance of the UK Met Office's Decadal Prediction System (DePreSys, *Smith et al.* (2007)), and to validate hindcasts against observations (as described in chapter 1), it is first appropriate to seek to understand the observed decadal variability. Therefore, this chapter aims to identify the space-time characteristics of the variability exhibited by the global ocean, with a particular focus on the Atlantic. Characterising the patterns, and evolution, of the variability is a pre-requisite of understanding the mechanisms that drive it.

The chapter is organised as follows. An introduction to ocean observations and ocean analyses is presented in section 2.2. A brief analysis of the variability of the global ocean is presented in section 2.3, which highlights a significant warming of the North Atlantic ocean in the 1990s. Finally, the observed changes in the North Atlantic are compared with an idealised modelling study in section 2.4. The comparison in section 2.4 suggests that the warming in the 1990s was a lagged response to the North Atlantic Oscillation,

and could have been potentially predictable in advance.

2.2 Ocean observations and analyses

Understanding the variability exhibited by the ocean, and the mechanisms that drive such variability, requires knowledge of the subsurface temperature and salinity. The ocean is observed in a variety of manners, using a variety of instruments. Subsurface data has historically been provided by expendable bathythermographs (XBT) and mechanical bathythermographs (MBT), conductivity-temperature-depth (CTD) probes and bottles. More recently, moored buoys (such as the Tropical Atmosphere Ocean project (TAO) array (*Hayes et al.* (1986))) and Argo automated profiling floats (*Wilson* (2000)) have been used. In particular, the deployment of Argo floats has provided an unprecedented density of observations in the top 2000m across the global ocean. Remote sensing is also providing detailed information of the sea surface temperature (SST) and also the ocean's circulation thorough observations of sea surface height.

Historically the observational sampling of subsurface ocean properties is poor, both spatially and temporally (see *Smith and Murphy* (2007), their figure 1). The observations of salinity are also much poorer than temperature. To make a global analysis, the observational information is "maximised" through an analysis that spreads information from data rich regions to data poor regions. The information is usually spread statistically through the use of covariances (*Daley* (1991)), but the detail of each analysis method differs slightly.

In general, there are two types of ocean analyses, which both have strengths and weaknesses. The first are those that are purely based upon a statistical analysis to spread information, and do not use any data assimilation approach (i.e. *Levitus and Boyer* (1994); *Ishii et al.* (2003); *Smith and Murphy* (2007)). The second are those that are based on the assimilation of information into a dynamical model (i.e. *Masina et al.* (2001); *Weaver et al.* (2005); *Balmaseda et al.* (2008)). The purely statistical analyses are computationally cheaper, due to the omission of a dynamical model, but are based upon linear statistics, which are not state dependent. The dynamical ocean analyses use

the model to propagate information into poorly observed regions, but also take account of non-linear dynamics and evolution. Consequently, the dynamical analyses are potentially more accurate even though the assimilation of observations is based upon linear statistics. However, the lack of data, particularly in the deep ocean, makes constraining dynamical analyses difficult, and can lead to significant uncertainties.

2.2.1 Description of ocean analyses used in this chapter

The exploration of the observed decadal variability for this thesis has used four different ocean analyses. The four ocean analyses use practically the same data[1], but the method of analysing the observational data differs. The difference in the analysing methods could create differing results due to the scarcity of observational data, especially in regions with few observations. Knowledge of the uncertainty in the analysed signals is important for understanding the variability and the mechanisms. Examining all four ocean analyses simultaneously provides a crude handle on the observational uncertainty, and provides some indication of whether an observed signal is real. However, caution must be used even when there is good agreement between datasets; signals could be due to artifacts of the observational data. For completeness, a brief review of the important known problems in oceanic data analysis is produced in section 2.2.2, after a brief description of the four ocean analyses (a more in depth description of the ocean analyses is can be found in Appendix A).

Three statistical based ocean analyses are used in this chapter, which are: The World Ocean Atlas (also known as Levitus or WOA-05 (*Levitus et al.* (2005); *Locarnini et al.* (2006)), the Original Met Office ocean analysis (*Smith and Murphy* (2007)), and the Updated Met Office ocean analysis. The three analyses all make use of the large spatial correlations of temperature and salinity anomalies in the ocean. The analysing methods give observed data a region of influence to spread data over. The region of influence, or in other words, the decorrelation length scale, usually extends for 100s of kilometres. Levitus uses a constant sphere of influence at each latitude and longitude, which is of the

[1]The separate ocean analyses will have different quality control procedures before the analysis is performed and so are unlikely to use exactly the same data.

order of 400km. However, the choice of the decorrelation length scale used in Levitus is not based upon physical processes of the ocean. For example, the spatial extent of anomalies in boundary currents is much smaller than anomalies associated with El Niño in the Tropical Pacific.

To construct decorrelation length scales that are based upon the physical processes of the ocean, *Smith and Murphy* (2007) exploit covariances derived from a coupled climate model to create the Original Met Office ocean analysis. Specifically, they use the HadCM3 (*Gordon et al.* (2000)) model to construct covariances. The model is used to create a covariance matrix for each model grid point, which is used to propagate information. So that errors in the calculation of the covariance matrices do not cause erroneous signals, the Original Met Office ocean analysis limits the horizontal influence of information, usually to 20° - 40° in latitude and longitude.

The Updated Met Office ocean analysis is similar to the Original, but instead of deriving model covariances from just one model, 9 versions of HadCM3 that have had perturbations to uncertain parameters in atmospheric parameterisation schemes (*Murphy et al.* (2004); *Collins et al.* (2006b)) are used to create the model covariances. The Updated Met Office ocean analysis does not limit the horizontal influence of the data, but instead uses global covariances. Both the Original and Updated Met Office ocean analysis use vertical covariances, but limit the spread of information with depth. Covariances between temperature and salinity observations are also used to increase the observational density of salinity data.

Finally, the fourth ocean analysis used is the ECMWF Ocean Reanalysis System 3 (ORA-S3; *Balmaseda et al.* (2008)). The ECMWF ocean analysis is a fully dynamical ocean reanalysis that assimilates observations of temperature and salinity into an ocean model, which is forced at the surface by fluxes of heat, momentum and fresh water. The analysis is based upon a modified version of the Hamburg Ocean Primitive Equation (HOPE) model, and provides the oceanic initial conditions for the ECMWF seasonal forecast system. Temperature and salinity are assimilated into the top 2000m of the ocean only, and sea surface height is also assimilated to supply additional information.

2.2.2 A cautionary tale of ocean analyses: The global mean heat content

The sparsity of observations will increase the uncertainty of the ocean analyses, thus, caution is needed when examining the analysed signals. There are also significant changes in the observing system over the past 50 years, not only in instruments used, but also in the number of measurements, and the positions of observations, which can cause difficulties for the creation of ocean analyses. To highlight issues with the ocean analyses, a comparison of the global mean top 700m average temperature for the four different data sets is shown in figure 2.1. A spread between the data sets in the 1950s and early 1960s is most likely to be due to insufficient data. However, in the 1970s, all four ocean analyses agree that there is a large increase in the global mean heat content, before decreasing in the 1980s. The increase in global mean heat content in the 1970s is dominated by an increase in the heat content of the Pacific (not shown). Model integrations of the historical changes in oceanic heat content reproduce the global mean warming trend, but models did not seem able to recreate the observed decadal variability (*Levitus et al.* (2001); *Gregory et al.* (2004)). Are the models incapable of simulating the correct magnitude of observed variability?

The discrepancy between the modelled and observed variability is partly explained by the scarcity of observations, which inflates the observed variability (*Gregory et al.* (2004); *AchutaRao et al.* (2006)). However, the major cause of the observed 1970s warming seems due to a bias in the observations of temperature from XBTs. XBT measurements, which increased rapidly in the early 1970s, especially in the Pacific, were found have a warm bias due to an incorrect calculation of their fall rate (*Gouretski and Koltermann* (2007)). Initial correction for the bias removes much of the multidecadal variability in the global mean top 700m heat content, especially the 1970s increase in heat content, but does not remove the global warming trend (*Wijffels et al.* (2008); *Ishii and Kimoto* (2009); *Levitus et al.* (2009)). Unfortunately, the fall rate correction has not been applied to the observational data sets used in this thesis. Although the fall rate problem is known to affect the global mean result, its influence on the basin scale, and the patterns of variability remain unclear. It is assumed that although in the fall rate problem will affect the

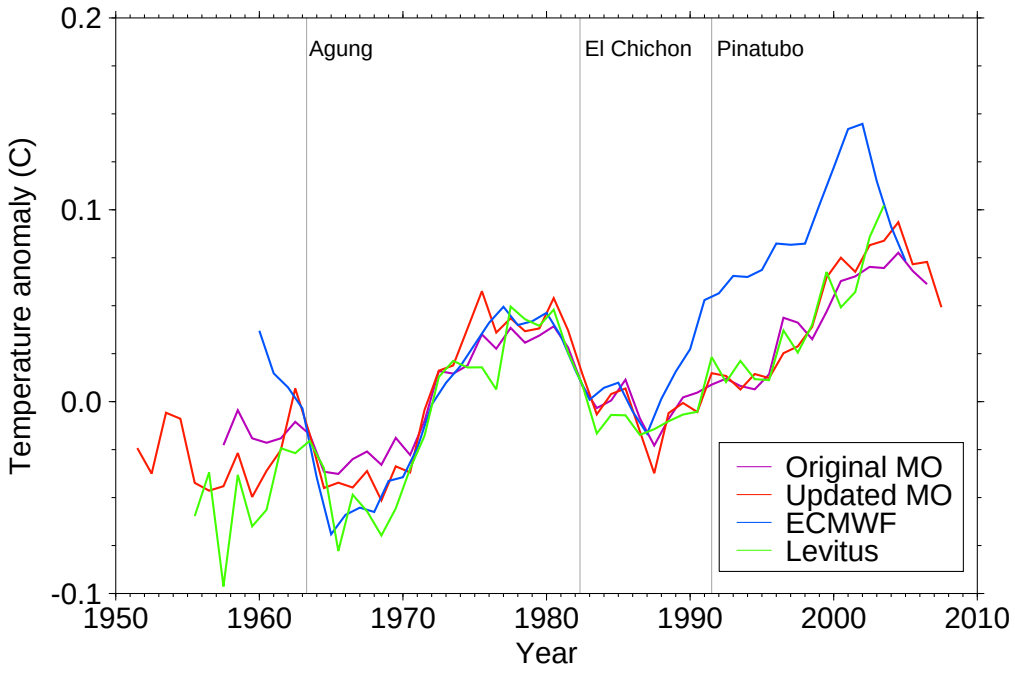


Figure 2.1: The global mean top 700m average temperature anomaly (°C) from the Original Met Office ocean analysis (purple), the Updated Met Office ocean analysis (red), ECMWF ORA-S3 (blue) and Levitus WOA-05 (green).

magnitude, and pattern, of anomalies at the basin scale, particularly in the Pacific where observations of XBTs greatly increased, the effect would not be sufficient to change the general pattern of anomalies, especially near the surface.

From 1990 onwards, the three statistical analyses agree well, but the ECMWF analysis warms at a faster rate. The faster rate of warming in the ECMWF analysis may be due to many factors, including the assimilation procedure and perhaps the model subducting heat to deeper levels. However, a general lack of observational constraint on the model for this depth layer (top 700m) seems important; the four analyses are in better agreement for the top 200m where the observational density is better (not shown).

After 2001, the ECMWF analysis shows a rapid cooling in global mean temperature so that the heat content anomaly is similar to the anomalies seen in the statistical analyses. After 2003, all of the ocean analyses cool. Although it has been argued that the cooling is a real signal of interannual variability (*Lyman et al.* (2006)), the cooling seems likely to be created by greatly increased observational coverage of the ocean due to the Argo

floats, especially in the southern hemisphere (*AchutaRao et al.* (2007)). This is especially true of the ECMWF ocean analysis, which is better constrained by the improved observational coverage. Thus it appears that the cooling signal is not a signal of real climate variability, but is caused by the changes in the observations. *Thompson et al.* (2008) also recently showed that large changes in the observing system can cause variability in the temperature records (in this particular case for Sea Surface Temperature (SST)). Hence, caution must be used when examining oceanic variability when there are large changes in the observing system, such as the deployment of Argo in the 2000s.

2.3 Observed variability of the ocean

2.3.1 Global variability

This section aims to identify the space-time characteristics of the observed ocean variability by examining global anomalies of temperature and salinity, relative to a 1961-1990 climatology. Figure 2.2 shows that, according to the Updated Met Office ocean analysis, the global oceans have not warmed uniformly over the past 50 years or so. The 1996-2005 decade is generally warmer than the 1956-1965 decade, but there is considerable basin and regional scale variability of temperature. There is also considerable basin and regional scale variability in the salinity content, which appears to be correlated with the temperature changes. Significant variability is present in both the Pacific and Atlantic basins and is briefly described here.

In the 1956-1965 decade, a warm salty anomaly is present in the subtropical North West Pacific. This anomaly strengthens in the decade 1966-1975 (see figure 2.2 b and g), and is associated with a warm and salty anomaly in the subtropical South West Pacific, and a cool and fresh anomaly in the East Pacific. This Pacific-wide pattern is linked to the Interdecadal Pacific Oscillation (IPO; *Power et al.* (1999)). Just considering the anomalies present in the North Pacific, the pattern of variability is the cold phase of the Pacific Decadal Oscillation (*Mantua et al.* (1997)). Contrasting panels b and c from figure 2.2, the Pacific regime shift in 1976 (*Trenberth and Hurrell* (1994); *Deser et al.* (1996, 2004);

Mantua et al. (1997)) is clearly seen. The anomalies associated with the IPO/PDO pattern in the previous decade reverse in the 1976-1985 and 1986-1995 decades, as the subtropical North West and South East Pacific are cool and fresh, and the Eastern Pacific is warm and salty. There is evidence that the PDO has shifted phase again in the late 1990s in multiple variables (*Cummins et al.* (2005); *Burgman et al.* (2008)), but the change is not as clear in the decadal mean (see figure 2.2 e).

Now turning to the variability of the Atlantic, in the 1956-1965 decade the subpolar gyre region (SPG) of the North Atlantic is warm and salty, with a cold and fresh tongue in the Gulf Stream extension (GSE). The SPG becomes cooler and fresher in the 1970s and 1980s (see figure 2.2 b) and a) for temperature and g) and h) for salinity). As the SPG cools, the cold, fresh anomalies in the GSE also reverse to warm and salty. Contrasting the decade 1996-2005 with the previous decade, 1986-95, a dramatic change is seen in the region of the SPG, where large negative temperature and salinity anomalies in the earlier decade are replaced by large positive anomalies in the more recent decade.

Figure 2.2 e) clearly highlights the large size of the warming in the North Atlantic, compared to the variability of the global ocean. The warming is seen in both SST (not shown) and upper ocean heat content, and is linked to a slowdown of the SPG circulation (*Hakkinen and Rhines* (2004)). These increases in SST have been implicated in decreased carbon uptake by the North Atlantic (*Schuster and Watson* (2007)), significant changes in marine ecosystems (*Hátún et al.* (2009)) and increased melting of the Greenland ice sheet (*Holland et al.* (2008)). The spatial pattern of the evolution of anomalies in the North Atlantic is also particularly interesting, but how certain can we be that the pattern of anomalies is real? Figure 2.3 shows that although there are global differences between the four oceanic data sets, the large scale features in the North Atlantic agree remarkably well around the time of the warming. All the ocean analyses show an extremely similar evolution of temperature anomalies in the North Atlantic, and the warming of the SPG seen in figure 2.3 e-h occurs before the deployment of Argo floats. Thus, we have confidence that the signal is real, and not caused by an artifact of the ocean analyses or observations.

Because the North Atlantic possesses significant potential predictability on decadal timescale

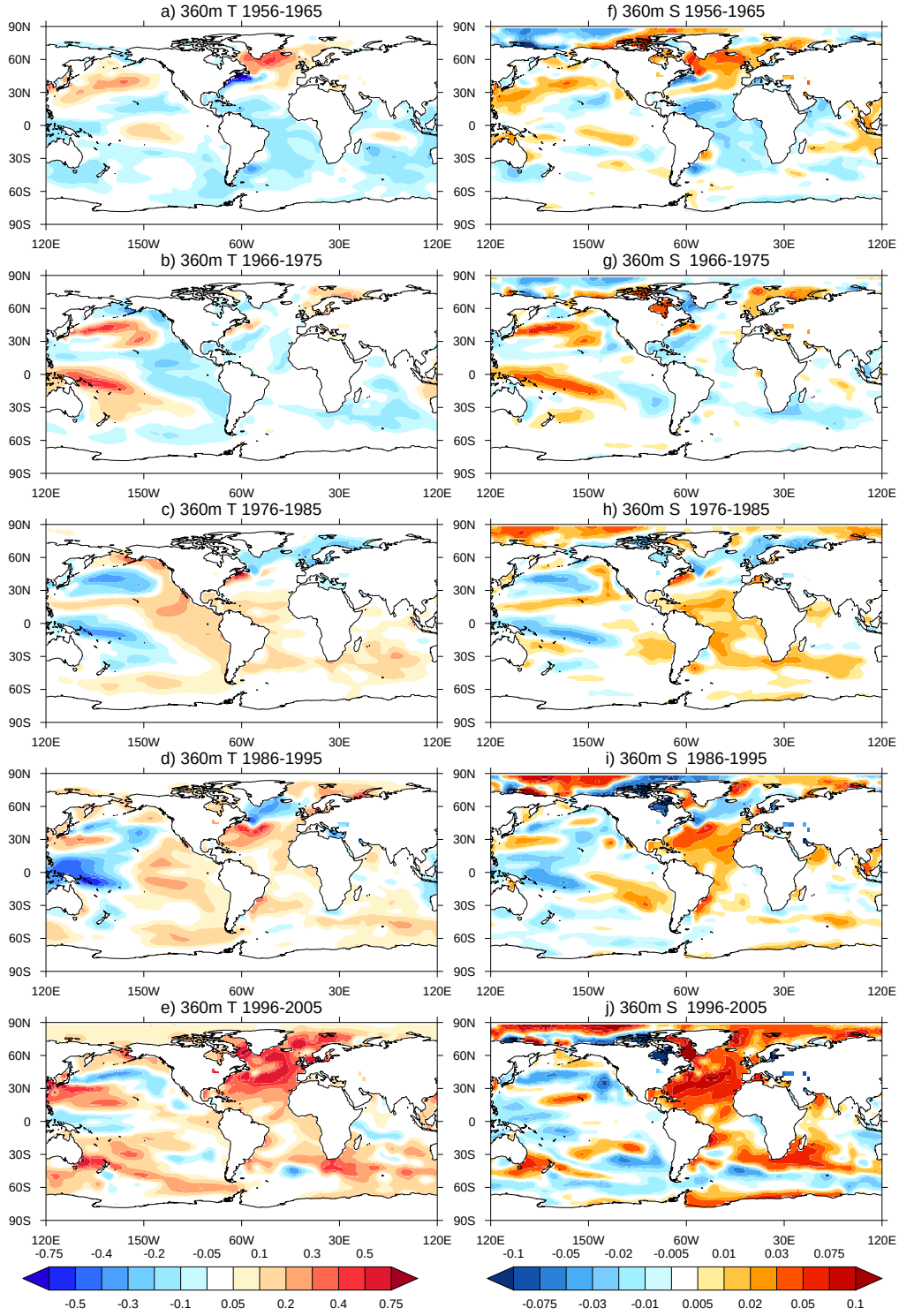


Figure 2.2: a)-e) shows the decadal mean of the top 360m average temperature anomalies (°C), relative to 1961-1990, for the Updated Met Office ocean analysis. f)-j) shows the same, but now for the average salinity anomalies (PSU).

(*Boer* (2000, 2004)), the warming of the North Atlantic could pose an interesting case study for decadal prediction. However, it is first important to understand what caused the warming. Is the warming in the North Atlantic due to anthropogenic forcings, is it a manifestation of natural variability, or is it influenced by both factors? A key question for this thesis is to understand whether the warming was predictable in advance? Before considering whether the warming can be captured by a state of the art decadal prediction system (DePreSys, *Smith et al.* (2007)), it is important to understand the mechanism of the observed warming. Understanding the mechanism will allow us to validate the model at the process level. Therefore, the remainder of this chapter will be focussed on understanding the mechanism of warming.

2.3.2 North Atlantic variability

The dominant mode of atmospheric variability in the Atlantic sector is the North Atlantic Oscillation (NAO) (*Hurrell* (1995)), which plays a major role in driving variability in the North Atlantic Ocean on interannual-to-decadal time scales. The anomalous sensible and latent heat fluxes that are associated with the NAO can drive significant variability of the Atlantic's SST (*Deser and Blackmon* (1993); *Marshall et al.* (2001a)). Such forcing can project onto decadal variability due to the low frequency variability exhibited by the NAO. For example, *Lozier et al.* (2008) showed that the major changes in heat content in the North Atlantic Ocean between the periods 1950-1970 and 1980-2000 could be understood as a response to decadal variations in the NAO. Between these two periods, the SPG cooled (as can be seen in panels a-d of figure 2.2) in response to the positive phase of the NAO that prevailed over most of the later period. The cooling was caused by anomalous upwelling in the SPG due to Ekman divergence, and increased surface heat loss associated with the strengthened westerly winds.

In the winter of 1995/96 the NAO index fell to much lower values than was typical of the 1980s and early 1990s, and relatively neutral values have prevailed since. Thus, the simplest explanation for the warming of the SPG is that the SPG remained slave to the local surface forcing by the NAO. But was this really the case? Such a hypothesis assumes no significant role for non-local dynamical changes in the ocean. However, there

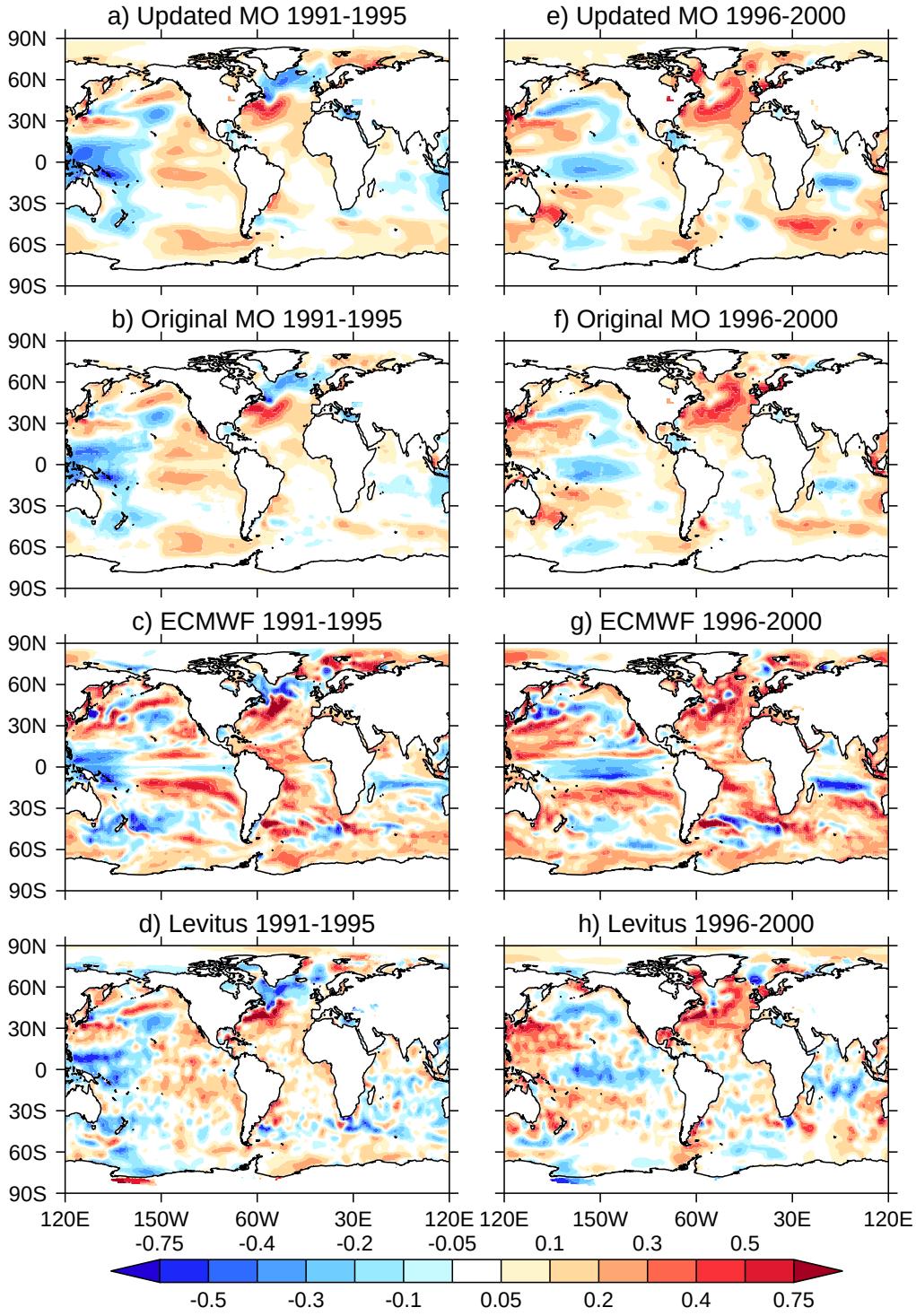


Figure 2.3: a) shows the 1991-1995 mean of the top 500m average temperature anomalies, relative to 1961-1990, for the Updated Met Office Ocean analysis. b)-d) shows the same but for the Original Met Office ocean analysis, ECMWF ORA-S3 and Levitus respectively. e)-f) shows the same as a)-d) but now for the 1996-2000 mean.

is considerable evidence, primarily from model studies, that dynamical changes should be expected on a variety of time scales.

On the shorter time scales, wind stress curl changes associated with the NAO can bring about significant changes in the Ekman transport, with positive (negative) NAO associated with anomalous southward (northward) flow at the latitudes of the storm track (i.e. $\sim$40-60°N; the latitude of the SPG and gyre boundaries) (*Eden and Willebrand* (2001); *Marshall et al.* (2001b)). The anomalous wind stress curl can also set up a fast Sverdrup-like circulation anomaly at the boundary of the subtropical and subpolar gyre (*Marshall et al.* (2001b)). Such fast changes in the circulation can have a significant impact on the overall heat transport of the Atlantic ocean, especially at the latitudes of the subpolar gyre (*Eden and Willebrand* (2001)).

The larger gyre circulation will also be forced by the NAO through Sverdrup balance (*Eden and Willebrand* (2001); *Brauch and Gerdes* (2005)). The strength of the SPG is also partially forced by buoyancy fluxes (*Saenko et al.* (2002); *Hakkinen and Rhines* (2004); *Lohmann et al.* (2009)), which maintain domed isopycnals that contribute to the baroclinic component of the subpolar gyre circulation. Finally, the shape and extent of the gyre can also be affected by the forcing of the wind stress, which can modulate the advection of subtropical waters into the subpolar gyre region (*Rhines and Schopp* (1991); *Hátún et al.* (2005); *Hakkinen and Rhines* (2009)). The complete adjustment of the gyre circulation can take a number of years due to the propagation of baroclinic Rossby waves (*Eden and Willebrand* (2001); *Brauch and Gerdes* (2005)), especially in the presence of topography (*Bellucci and Richards* (2006); *D'Orgeville and Peltier* (2009)), and can, therefore, project on lower frequency variability (i.e. inter-decadal to decadal variability).

The variability in the circulation caused by the NAO can lead to significant heat content anomalies, especially in the Gulf Stream extension region, largely due to the convergence of heat transport in the ocean (*Visbeck et al.* (1998); *Krahmann et al.* (2001); *Bellucci and Richards* (2006)). The temperature anomalies that are created are able to then propagate along the North Atlantic Current, and into the SPG in $\sim$5-10 years (*Krahmann et al.* (2001)), similarly to that seen in the observations (*Sutton and Allen* (1997)). Decadal variability of heat content in the Gulf Stream region can also be forced separately to

changes in the circulation. The thermocline depth can be forced by baroclinic Rossby waves that form due to anomalous wind stress curl in the central Atlantic and propagate westward (*Münnich et al.* (1998); *Schneider et al.* (2002); *Leadbetter et al.* (2007)). The slow propagation of the Rossby waves at high latitudes, and the decadal variability of the NAO, can create oscillations of heat content anomalies that project on to decadal time scale (*Münnich et al.* (1998); *Leadbetter et al.* (2007)).

Models also suggest that longer frequency (i.e. decadal and mutlidecadal) variability can also be excited by the NAO. In particular, several studies have shown that a positive phase of the NAO leads, by several years, a strengthening of the Atlantic Meridional Overturning Circulation (AMOC). (*Eden and Willebrand* (2001), *Dong and Sutton* (2005)). Acceleration of the AMOC typically leads to a warming of the North Atlantic Ocean (e.g. *Knight et al.* (2006)).

The contrast between the instantaneous local response to the NAO, and the lagged non-local response was illustrated very clearly in a recent modelling study by *Lohmann et al.* (2009) (hereafter Loh09). They showed how, in response to persistent positive NAO forcing, the SPG initially cools, but subsequently warms as a result of a spin up in the AMOC, leading to enhanced northward heat transport. Was such a mechanism responsible for the warming of the SPG in the mid 1990s? The relevance of this two-stage response to the observed warming of the SPG will be investigated in the next section (i.e. section 2.4) by comparing the observed evolution of the SPG to that simulated in the experiments of Loh09.

2.4 A comparison of the evolution of the subpolar gyre in ocean analyses and idealised model experiment

2.4.1 Description of model experiments

The model experiment examined is that of Loh09, in which an ocean general circulation model (the Nansen Centre version of the Miami isopycnic coordinate ocean model

(MICOM), at a resolution of approximately 2.4 degrees) was forced with idealised NAO forcing. Specifically, composites of surface fields were constructed from ECMWF ERA-40 reanalysis data to represent positive NAO (NAO^+) and NAO neutral (NAO^n) conditions. NAO^+ years where defined as the 7 years when the winter (December-March) NAO index (also calculated from ERA-40 sea level pressure) exceeded plus one standard deviation. NAO^n years where defined as the 7 years where the index was closest to zero. Following a spin-up procedure, the ocean model was forced with persistent NAO^+ or NAO^n conditions for 40 years. Further details of the model and experimental design can be found in Loh09.

To prevent repetition, the Updated Met Office ocean analyses is used for the observed changes in the comparison. The three other ocean analyses are also briefly examined for the purpose of comparison. However, key differences exist between the observations and the model experiment that first need consideration.

A key difference between the model experiments and the real world is that the model does not include any changes in radiative forcing, which are expected to have contributed to an overall positive trend in sea surface temperatures and upper ocean heat content. Therefore, for the purposes of comparing the spatial evolution of observed anomalies with the results from the model experiments (Section 2.4.2), a linear trend is removed from the analyses. Specifically, a linear trend is fitted to the global mean upper ocean (500m) heat content over the period 1951-2007, and is removed from every individual grid point. Because there was a positive trend in the NAO from 1960s to the 1990s, fitting a separate trend at every grid point removes a significant proportion of the variability that we wish to examine (not shown). However, it must be recognised that removing only the global mean trend is a simple approach, and may leave some differences in overall warming between the analyses and the model experiments.

A further difference between the model experiments and the real world is that the model experiments were designed to be representative of exceptional, persistently positive, NAO conditions, whereas in reality there is considerable interannual variability. Therefore, larger anomalies are anticipated in the model experiments than in observations. For the purposes of comparison, the model results were scaled down. The scaling factor is de-

fined as the ratio of the mean magnitude over the North Atlantic domain (60°W-10°W, 40°N-65°N) of the anomalies from the analyses and the anomalies from the model.

Also during this section, a control run of HadCM3 will be used in understanding the changes in the AMOC. A control run is an integration of the model that is representative of a pre-industrial climate, in which there is no external radiative forcing. Therefore, the variability exhibited by such a model integration is only internally produced. The strength of the AMOC in the control run, otherwise known as the Meridional Overturning Index (MOI), is defined as the mean of the meridional stream function at a depth of 996 m, in a 5 latitude band around the desired latitude (i.e. for 50°N, it is the mean between 47.5°N - 52.5°N), and is similar to that used in *Dong and Sutton* (2005) and *Hawkins and Sutton* (2007). To remove non-linear drift, all the data from the HadCM3 control run has been detrended with a 3rd order polynomial applied at each grid point.

2.4.2 Comparison between observations and model

Figure 2.4 shows the evolution of the SPG upper ocean heat content in the Updated and Original Met Office, ECMWF and Levitus ocean analyses. Also shown is the inverted wintertime (December-March) NAO index. All indices are normalised to unit variance to aid comparison. Up until 1995, temperatures in the upper SPG are anticorrelated with the NAO index, with maximum correlations (of -0.69, -0.70, -0.66 and -0.74 for the different analyses) found when the NAO index leads by one year. As the NAO became more positive through the 70s, 80s and 90s, the SPG cooled due to the local response of the ocean to the atmospheric forcing, as discussed by *Lozier et al.* (2008).

In the winter (not shown) of 1995/96 the SPG warmed rapidly after a period of intense persistent NAO+ forcing from 1988/89-1994/95. The warming coincides with an unusually low NAO index. Following this rapid warming event, there is a persistent offset between the SPG heat content and the inverted NAO index. While the NAO index returns to near neutral levels, the positive heat content anomalies persist. This persistent offset suggests that some other mechanism influenced the relationship between the NAO and heat content. The heat content is not simply responding to the concurrent or previous

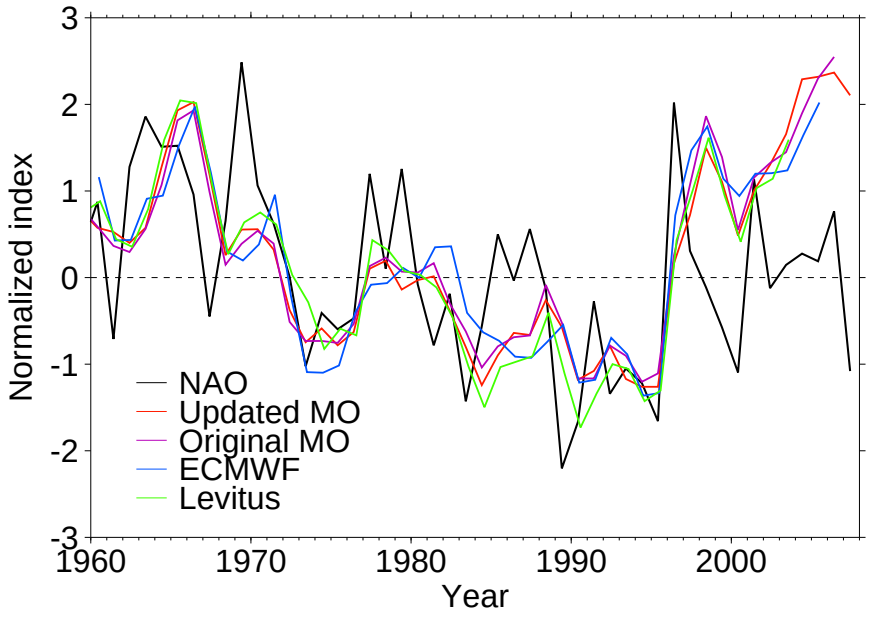


Figure 2.4: Flipped (i.e. multiplied by -1) normalised DJFM NAO index based upon the difference of normalised SLP station data between Lisbon and Reykjavik (black ; NAO Index Data provided by the Climate Analysis Section, NCAR, Boulder, USA, Hurrel (1995)). Normalised top 500m annual average temperature anomalies (°C), relative to the mean for 1961-1990, for the subpolar gyre (60°W-10°W, 50°N-66°N) from the Updated Met Office (red), Original Met Office (purple), ECMWF ORA-S3 (blue) and Levitus WOA-05 (green) ocean analyses. One standard deviation corresponds to 2.1 for the NAO index, and 0.256, 0.233, 0.253 and 0.251 for the Updated Met Office, Original Met Office, ECMWF and Levitus subpolar gyre indices respectively. Note that the temperature of the subpolar gyre and the NAO are anti-correlated to 1995. After 1995 the correlation breaks down, with the warming seen to be more persistent than a simple correlation with the NAO would suggest.

year's NAO conditions.

To better understand the nature of the North Atlantic warming in the 1990s we examine upper 500m heat content anomalies (figure 2.5). The left column shows the evolution in four pentads as seen in the Updated Met Office ocean analysis. From 1985-1989, the SPG is anomalously cold (relative to the 1961-90 reference period), and there is a slight warm anomaly in the northwest subtropical gyre. By 1991-95, corresponding to the highest NAO values, both of these anomalies have intensified greatly and the warm anomaly in the northwest subtropical gyre has begun to spread eastward. In 1996-2000, the rapid warming of the SPG is evidenced by the disappearance of the cold anomalies seen in the previous pentad, and the warm anomaly is reaching northward into the eastern SPG. Finally in 2000-2004, as the overall warming of the SPG continues, this warm anomaly

spreads round the eastern flank of SPG, reaching into the Irminger Sea. This evolution is related to changes in the extent the SPG as discussed by (*Hátún et al.* (2005)).

The right column of figure 2.5 shows the evolution of upper 500m heat content anomalies in the idealised model experiments of Loh09. Because the experiments are idealised, the absolute timing of events cannot be compared with the analyses[2], and nor should we expect detailed agreement of small scale features. However, what is striking is that the major features of the evolution are very consistent between the model results and the analyses. In years 2-6 and 7-11 of the model results, the SPG cools, as expected in response to the positive NAO forcing, and as seen in the analyses. At the same time a warm anomaly appears, and then intensifies, on the northern flank of the subtropical gyre. This anomaly is formed in response to wind stress curl anomalies associated with the positive phase of the NAO that drive the thermocline deeper (*Marshall et al.* (2001b); *Leadbetter et al.* (2007); *Lozier et al.* (2008)) and corresponds to the positive anomaly seen in the analyses between 1991 and 2000. In model years 12-16 and 17-21 this warm anomaly propagates northward, reaching around the eastern flank of the SPG and into the Irminger Sea, in a pattern that is very similar to that seen in the analyses.

The agreement between the evolution of upper ocean heat content seen in the analyses and that seen in the model experiments strongly suggests that the observed warming of the SPG was a lagged response to the persistent positive NAO conditions found in the 1980s, and especially the early 1990s. Furthermore, it suggests that the mechanism described by Loh09 played an important role in the evolution of the real world during the 1990s. A key element of this mechanism involves a strengthening of the AMOC, associated with the spreading of anomalously cold and dense North Atlantic Deep Water from the SPG southward along the western boundary of the Atlantic basin. Figure 2.6 shows a comparison of this deep signal between the idealised experiments and the analyses. In both cases, negative anomalies initially form in the Labrador Sea and wider SPG. These anomalies are seen to intensify and then spread southward, particularly along the western boundary of the Atlantic basin.

[2] In particular it could be argued that years 2-6 of the model experiments should correspond most closely to the real years 1991-95, since these were the years of the highest NAO index, but the initial conditions in the real world and experiment were different

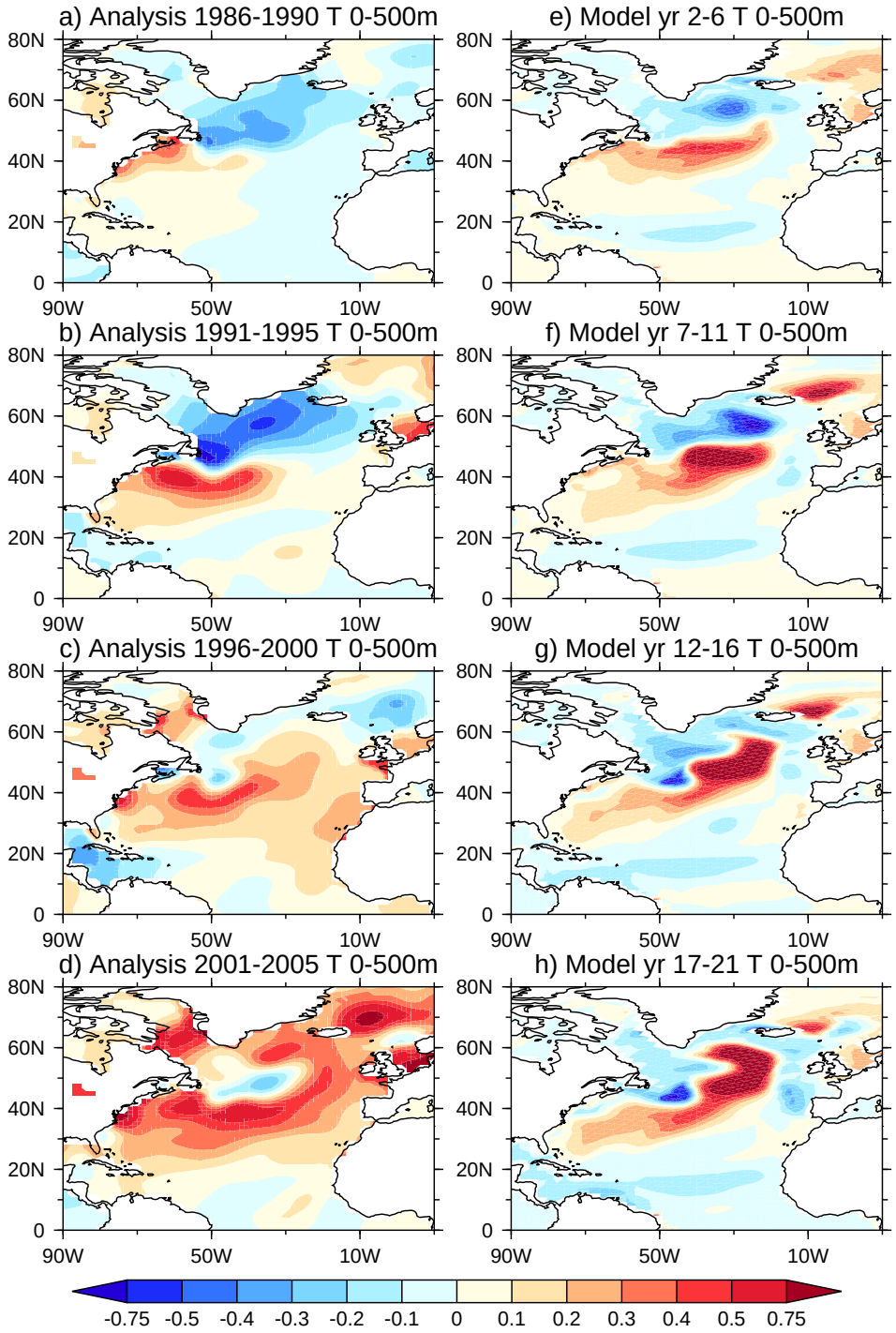


Figure 2.5: Comparison of observed and modelled upper 500m ocean average temperature anomalies (°C) relative to the mean for 1961-1990. a) to d) show pentadal mean anomalies from 1986-2005 from the Updated Met Office ocean analysis. e) to h) shows pentadal mean anomalies from the NAO$^+$-NAO^n experiment. A global mean linear warming trend was removed from the observations and the model anomalies were also scaled to be similar to the observed. See text in section 2.4.1 for details.

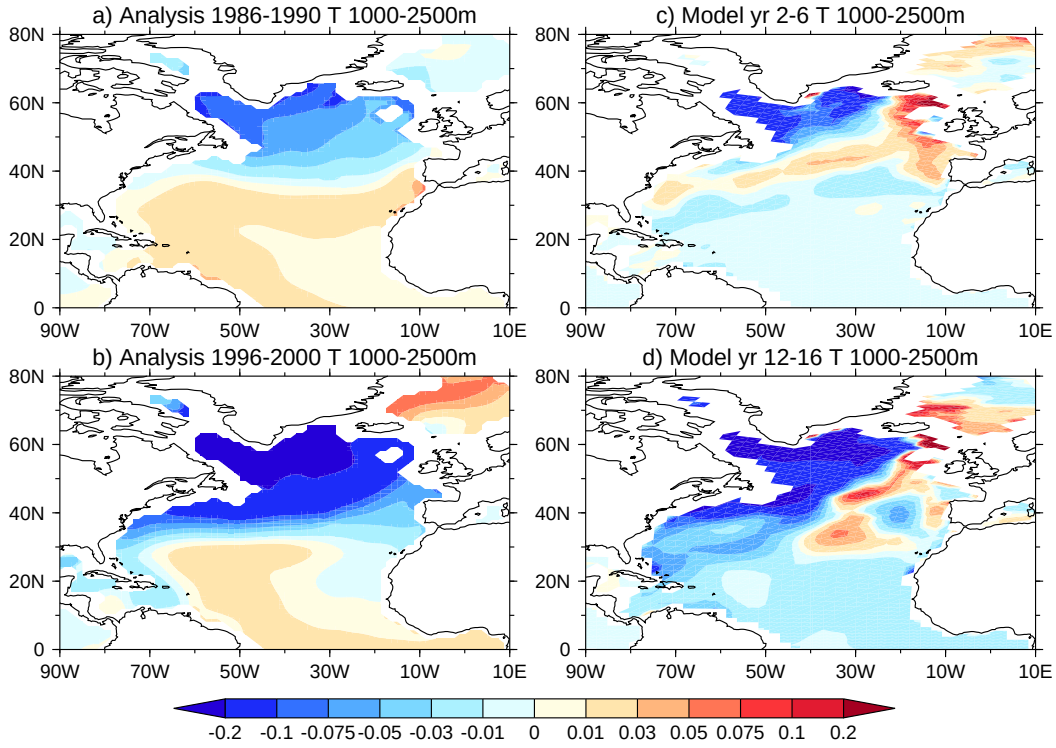


Figure 2.6: Comparison of observed and modelled 1000-2500m average temperature anomalies relative to the mean for 1961-1990. a) and b) show the observed anomalies for the 1986-1990 and 1996-2000 pentads respectively from the Met Office ocean analysis. c) and d) show pentadal mean anomalies from NAO^+-NAO^n experiment for model years 2-6 and 12-16 respectively. As in figure 2.5 the model anomalies were scaled to be similar to the observed, but no trend is removed. See text in section 2.4.1 for details.

The spreading of anomalously cold North Atlantic Deep Water down the western boundary at depth is also seen in AMOC changes in HadCM3 (not shown), but figure 2.7 shows that it is the density change associated with those temperature anomalies that affects the strength of the AMOC at 50°N[3]. In HadCM3, a strengthening of the AMOC at 50°N follows a build up of positive density anomalies in the surface layers of the SPG region (figure 2.7 a and b). A low density signal is also seen to form in the subtropical region to the south of the SPG. Following the peak in AMOC strength, the density anomalies in the SPG reduce (figure 2.7 c), as the temperature increases (not shown).

[3]The spreading of density anomalies also affects the strength of the AMOC at other latitudes. The lead lag correlation of density with the AMOC strength at 30°N is in fact similar to that seen in figure 2.7, but dense anomalies near the surface in the SPG region lead the AMOC changes by up to 3 years more (not shown).

At depth (1000-2400m) a dense signal is seen to build with dense anomalies present downstream of the Denmark strait, and around the continental boundary of the SPG (figure 2.7 d). The density anomalies intensify as the AMOC strengthens and propagate southward in a narrow band against the western boundary (figure 2.7 e). After a peak in the AMOC, density anomalies are still present in the deep western boundary, but there is increasing correlation at lower latitudes and along the eastern boundary (figure 2.7 f), which looks consistent with the emission of Rossby waves from the eastern boundary due to the propagation of a Kelvin wave (*Zhang* (2007)).

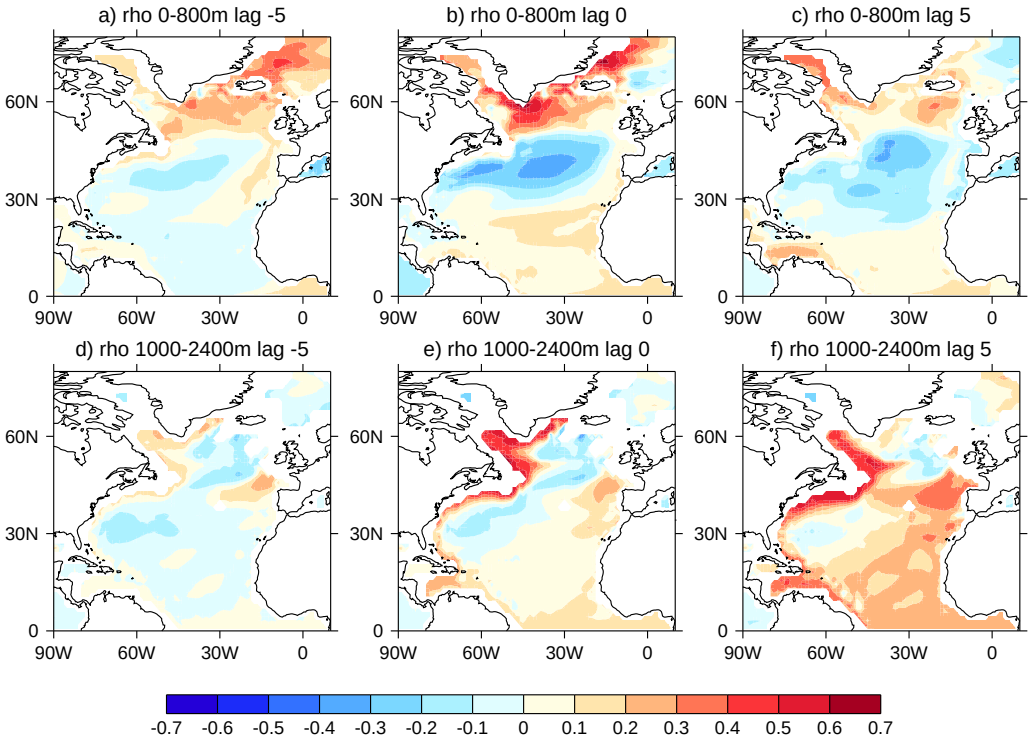


Figure 2.7: a) - c) show the correlation of annual average top 800m density anomalies and the Meridional Overturning Index at 50°N, from 1100 years of the HadCM3 control integration, where the MOI lags the density anomalies by -5, 0 and 5 years respectively. d) - f) as a) - c) but show the correlation of 1000-2400m integrated density anomalies with the MOI at 50°N. A 3rd order polynomial trend was fitted and removed from the MOI and the density anomalies at each grid point to remove non-linear drift.

Can we therefore deduce changes in the real AMOC by examining the observed density anomalies? Figure 2.8 shows the density anomalies calculated from the Updated Met Office ocean analysis. The stratification of the subpolar North Atlantic is weaker than in the subtropics, and so the density anomalies in the subpolar region are usually smaller. Therefore, the density anomalies have been normalised by the observed interannual stan-

dard deviation of density anomalies to highlight where density anomalies have relative importance. In the early 1990s, when the positive NAO was greatest (see figure 2.4), positive density anomalies intensify across the SPG region, and are especially large in the Labrador and Irminger Seas. The magnitude and spatial extent of the positive density anomaly increases and peaks in 1995, with positive anomalies greater than 2σ present in the Labrador and Irminger Seas (figure 2.8 e). In 1996, the density anomaly is much smaller in spatial extent, but a large anomaly is still present in the Labrador sea region (figure 2.8 f). To the south of the SPG, low density anomalies prevail throughout the same period.

As the dense anomalies build in the surface waters of the Labrador sea there is increased deep convection in the Labrador seas area, particularly in the years 1990-1993, when the mixed layer descended to 2000m in depth in the winter (*Lazier et al.* (2002)). A large cooling of the subsurface Irminger Sea is seen in the analysis in 1994 (not shown), which is suggestive of increased deep convection, possibly due to increased Greenland tip jet forcing that is associated with high NAO values seen in the early 1990s (*Pickart et al.* (2003)). However, it must be noted that the cooling signal in the deep Irminger Sea could be due to the spreading of Labrador Sea water into the Irminger Sea region as suggested by *Yashayaev et al.* (2007).

At depth (1200-3000m), the Updated Met Office ocean analysis does resolve positive density anomalies in the SPG region (figure 2.8 g) - l)). The positive density anomalies intensify through the early 1990s as deep convection intensifies in the Labrador Sea, due to the persistently positive NAO. They are also seen to be present down the deep western boundary. The positive density anomalies intensify to a peak in 1995, with positive anomalies in the Labrador Sea region $>3\sigma$. The evidence of positive density anomalies in the regions where the strength of the AMOC might be sensitive to further suggests that an increase in the real world AMOC, particularly at higher latitudes, took place.

However, it must be noted that caution is required when examining the density anomalies, especially at depth due to the poor observations, particularly of salinity. In fact, *only* the Updated Met Office ocean analysis shows such coherent positive density anomalies

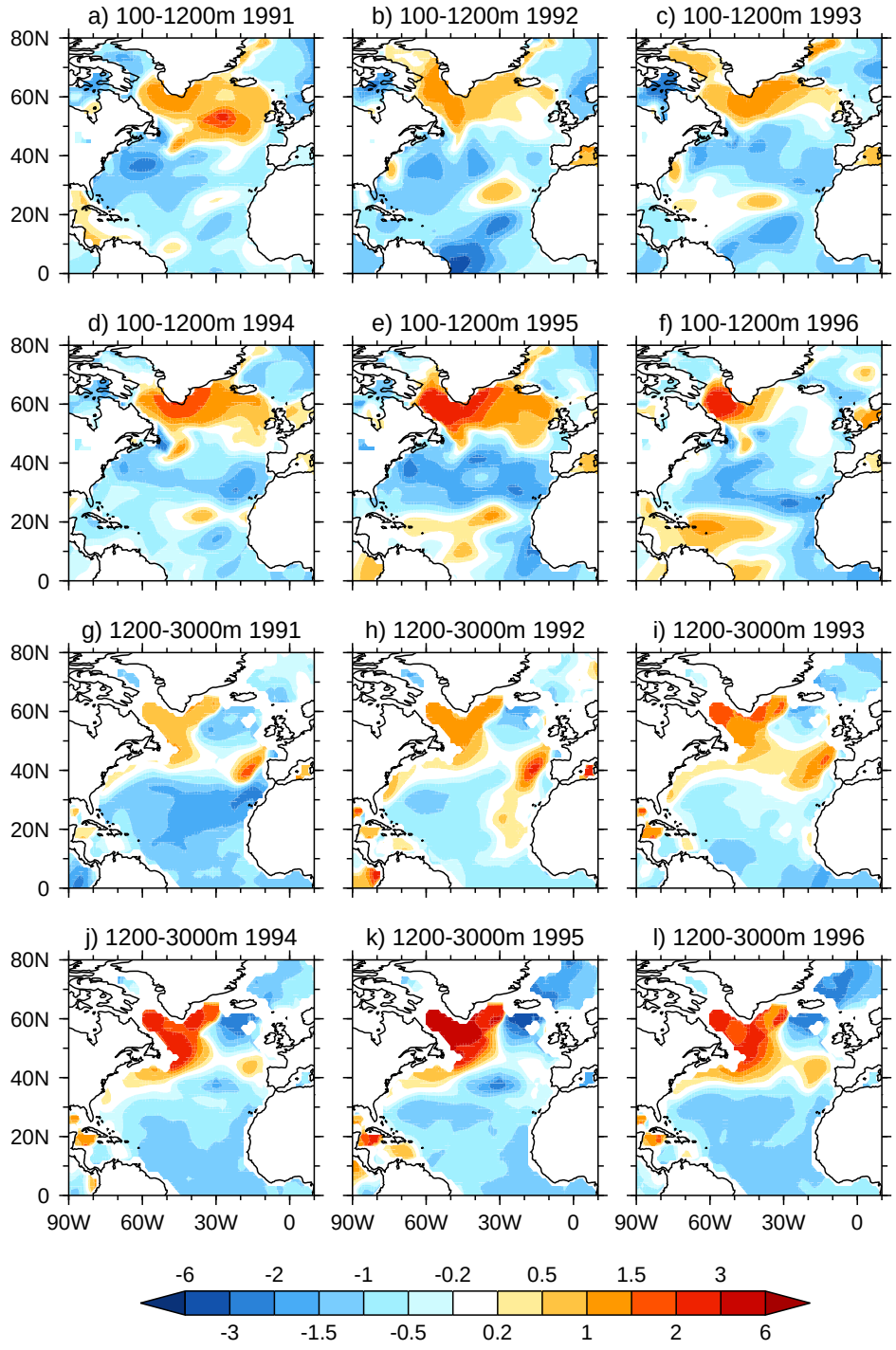


Figure 2.8: a) - f) show the normalised average 100-1200m in situ density anomalies, relative to the 1951-2006 climatology, calculated from the Updated Met Office ocean analysis. g) - l) show the same, but now for the 1200m-3000m layer. To highlight when density anomalies are important, the density anomalies are normalised by the observed interannual standard deviation calculated from the Updated Met Office ocean analysis and applied to each layer separately.

at depth in the North Atlantic (not shown[4]). Although there is much uncertainty in the density anomalies at depth, it remains clear that a significant build up of density anomalies took place in the upper ocean and that significant deep convection was occuring prior to the rapid warming of the SPG, particularly in the Labrador Sea, which is similar to the evolution of the experiments in Loh09.

Changes in the real world AMOC itself can be estimated by assimilating the observations of temperature and salinity into an ocean model. The estimated real world AMOC changes, and the changes from the Loh09 experiments are shown in figure 2.9. The estimated AMOC time series is computed from an assimilation integration using the UK Met Office's Decadal Prediction System (DePreSys), based on the HadCM3 climate model (DePreSys is described in detail in chapter 3). The assimilation involves strongly relaxing the model to a climatology plus "observed" anomalies of temperature and salinity derived from the Met Office ocean analyses[5]. The AMOC index at 50°N shows an increase in the overturning from the 70s into the 90s, and has a well defined peak in 1995, of approximately 3.5 Sverdrups, or approximately 20% of the climatological value. A similar peak in the mid 1990s has been seen in a number of other recent ocean analyses, although not all analyses are consistent (*Anderson et al.* (2009)). A peak in the AMOC in the mid 1990s is also seen in ocean integrations forced by surface fields derived from reanalyses (*Böning et al.* (2006), *Deshayes and Frankignoul* (2008)).

Also shown in figure 2.9 is the AMOC index and northward heat transport derived from the model experiment. In this experiment the AMOC increases because the NAO+ forcing leads to cooling of the SPG and enhanced deep water formation (Loh09). Subsequently, the AMOC anomaly decays, even though the NAO+ forcing is maintained. This decay occurs as the upper SPG starts to warm in response to increasing northward heat transport (Loh09). When put alongside the agreement in the spatial evolution of anomalies that was seen in figure 2.5 and 2.6, the good agreement between the AMOC changes seen in the experiment and those derived from the analyses strongly suggests a common cause: i.e. that persistent NAO+ forcing was the primary factor in both cases.

[4]The density anomalies at depth in the Atlantic from the Original Met Office ocean analysis are shown later in chapter 5.

[5]For this particular assimilation run, the Updated Met Office ocean analysis was used.

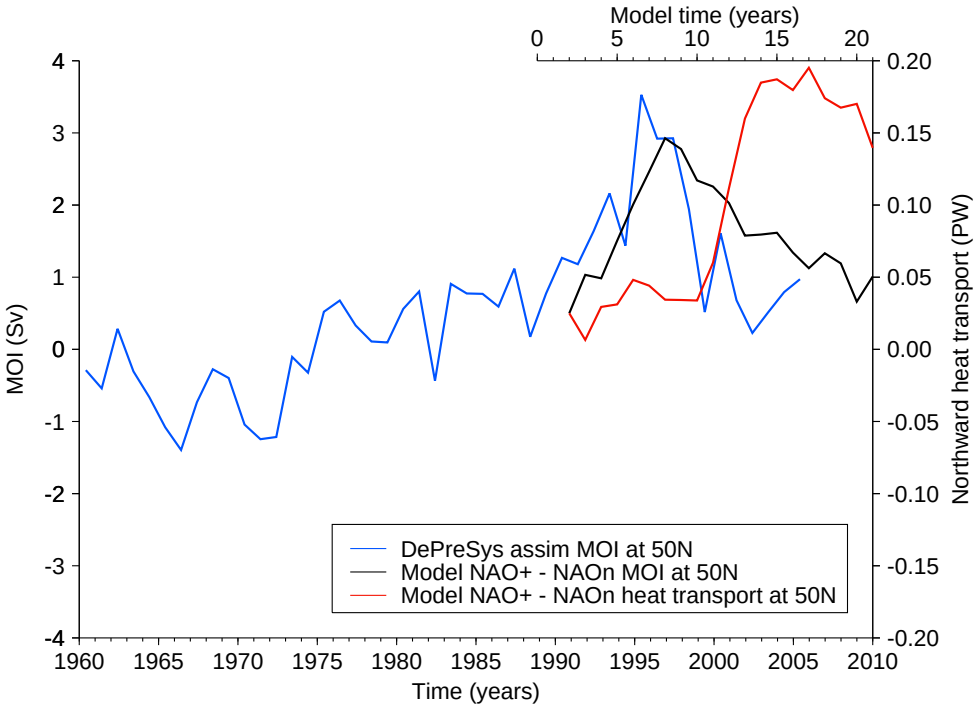


Figure 2.9: Annual average Meridional Overturning Index (MOI) anomaly (relative to 1961-1990) at 50°N from an assimilation run of DePreSys (see text for details; blue), compared with the MOI anomaly from the NAO$^+$-NAO^n experiment (black). The northward heat transport at 50°N from the NAO$^+$-NAO^n experiment, in PW's, is also shown (red). Because the model experiment is idealised the absolute timing of the MOI time series cannot be established. Therefore, the relative timing has been adjusted to highlight the similarity of the MOC variations. See text for further discussion.

The northward heat transport at 50°N in the model experiment increases gradually at first, but then shows a substantial surge of ∼0.15PW (corresponding to ∼25% of the mean value at this latitude) in model years 11-13. Interestingly, this surge in heat transport lags the AMOC increase at the same latitude by ∼5 years. Comparison with figure 2.5 shows that the surge in heat transport occurs at the time when the positive heat content anomaly that forms on the northern side of the subtropical gyre starts to move northward across 50°N. Thus it appears that anomalies in ocean temperature, as well as anomalies in circulation (both gyre and AMOC, Loh09), play an important role in accounting for the increase in northward heat transport. It is appropriate to emphasise the need for caution in interpreting the *timing* of changes in heat transport in the real world. Figure 2.9 suggests that it is very likely that a surge in northward heat transport occured in the real world, but the exact timing of this surge cannot be established with great confidence owing to the

idealised nature of the model experiments, and possible weaknesses in the model itself.

2.4.3 Discussion

The discussion thus far has highlighted the significant similarities between the evolution of the North Atlantic SPG seen in ocean analyses for the period 1986-2005 and the evolution simulated in a model in response to persistent NAO+ forcing. As already noted, in reality, but not in the model experiment, there was an unusually negative NAO index in the winter of 95/96 and, subsequent to this winter, NAO conditions have been near neutral (i.e. near the average for the period 1960-2008). The comparison suggest that the decline in the NAO index was not *essential* for the warming of the SPG. However, this decline almost certainly contributed to the magnitude of the warming, since a lower NAO index is associated with reduced cooling of the SPG through surface heat fluxes. Furthermore, it seems very likely that the unusually negative NAO winter of 95/96 contributed to the *rapidity* of the warming. In addition to the reduced heat flux away from the ocean, there may have been a short lived positive contribution from the Ekman component of the northward heat transport (*Brauch and Gerdes* (2005)). Lastly, it should be noted that the ocean may not have only responded to the NAO, but could have exerted some influence on it. Whilst the sensitivity of the atmosphere to changes in extratropical sea surface temperature appears to be modest (*Kushnir et al.* (2002)), the large warming of the North Atlantic that occured in the 1990s may have been sufficient to have an effect.

In conclusion, this chapter has presented persuasive evidence that the prominent warming of the North Atlantic Subpolar Gyre that occurred in the mid 1990s was a consequence of the prolonged forcing of the ocean by positive NAO conditions over the previous decade. Whilst the exact magnitude and rapidity of warming were influenced by the strongly negative NAO winter of 95/96, and subsequent more neutral NAO conditions, the occurrence of a significant warming was already inevitable before these events took place. The mechanism of warming involves enhanced deep water formation, a temporary spin up of the AMOC and subpolar gyre circulations (Loh09), and a surge in northward heat transport. The analysis of a assimilation run from DePreSys suggest the increase in the AMOC at 50°N may have been $\sim$3Sv. An important implication of these findings

is that they imply that the North Atlantic Ocean has significant memory on interannual time scales: the ocean does not simply respond immediately to atmospheric fluctuations, but integrates the effect of these fluctuations to generate coherent changes in large-scale circulation and other properties. Therefore, the evidence from this study supports other evidence, primarily from models (*Griffies and Bryan* (1997b); *Collins and Sinha* (2003)), that important changes in the North Atlantic Ocean should be predictable on multi-annual time scales.

2.5 Chapter Conclusions

This chapter has described an exploration of the observed variability of the Atlantic Ocean in the recent past, in order to better understand the mechanisms that drive variability, and to understand if such variability is potentially predictable in advance. The main findings are listed as follows:

- The global ocean has warmed over the past 50 or so years, but it has not warmed uniformly. Significant variability is present in both the Atlantic and the Pacific. In particular, the North Atlantic has warmed much faster than the global ocean in the past 15 years, especially in the North Atlantic subpolar gyre (SPG) region.

- Much of the variability in the North Atlantic can be attributed to the changes in the North Atlantic Oscillation (NAO). The variability of the North Atlantic SPG region is highly correlated with the NAO, albeit with a lag of one year. As the NAO became more positive through the 1970s, 1980s and the early 1990s, the SPG region cooled as a consequence of increased cooling through surface heat fluxes, and Ekman suction. Also, warm anomalies developed along the southern edge of the SPG in response to the wind stress curl anomalies associated with a positive NAO.

- In the winter of 1995/96, the North Atlantic SPG underwent a large and sudden warming in which the warm anomalies at the southern edge of the SPG broke out into the SPG. The warming coincided with a large negative NAO event, but after

the warming the correlation between the SPG temperatures and the NAO broke down. The NAO exhibited neutral values, but the warming of the SPG persisted suggesting that another mechanism was involved.

- Comparison of the observed temperature anomalies at the surface and at depth with an idealised model experiment, in which an ocean model was forced with persistent NAO positive forcing, showed that the warming was likely to be due to a strengthening of the Atlantic Meridional Overturning Circulation (AMOC), which lagged the persistent positive NAO. The cooling of the surface waters in SPG region led to greater volumes and deep convection, and hence, increased deep water formation. A cool dense anomaly is seen to spread down the western boundary of the North Atlantic at depth, strengthening the AMOC, and the northward heat transport. Hence, the warming of the North Atlantic was caused, at least in part, by the increase in the northward heat transport. Analysis of an assimilation run from the UK Met Office's Decadal Prediction System (DePreSys) suggests that the strengthening of the AMOC at 50°N could have been $\sim$3Sv.

- The warming of the SPG was not caused by the negative NAO spike of winter 1995/96, but was inevitable before the negative spike occured. The negative NAO of winter 1995/96 was not *essential* for the warming to occur. However, it is likely that the negative NAO in the winter of 1995/95 influenced the magnitude of the warming and also the rapidity.

The lagged warming of the SPG, following the persistent positive NAO forcing highlights the significant memory in the North Atlantic and suggests that such a change should have been predictable in advance. Investigating the predictability of such a case study in an operational decadal prediction system could be informative of the performance of the underlying model, such as the presence, or lack, of important processes. The investigations may also be informative of the limitations of such a decadal forecast system, and could guide investments and research for delivering future improvements to decadal climate forecasts.

The ability of the UK Met Office's Decadal Prediction System (DePreSys) to predict the warming of the North Atlantic in the 1990s will be investigated further in this thesis in

chapter 6. However, it is first appropriate to introduce the DePreSys in detail in the next chapter.

Chapter 3

Description of the DePreSys decadal prediction system and introduction to mean skill scores

To understand the effect of initialising coupled climate models on prediction skill, it is important to understand how the predictions are constructed and validated. This chapter aims to first outline the construction of The UK Met Office Decadal Prediction System (DePreSys), and the hindcast experiments that have been performed at the UK Met Office in section 3.1. For clarity a notation system that will be used throughout this thesis will be introduced in section 3.2, alongside the skill scores used to understand the mean performance of DePreSys. The results of *Smith et al.* (2007) (hereafter Sm07) will be recreated in section 3.3, which highlights an incorrect assumption made in their analysis of skill. It will be shown that the error will likely affect the improvement in the skill of DePreSys relative to un-initialised hindcasts.

3.1 Introduction to DePreSys

Key elements of DePreSys, as will become apparent in this introduction, are climatologies. The construction of DePreSys begins with transient runs of HadCM3 (*Gordon et al.*

(2000)) to create a model climatology that is used in the assimilation process (described below in section 3.1.3). Climatologies are also very important in the evaluation of skill (described in section 3.2.2). Therefore, it is appropriate to first introduce the transient runs, and the model climatology, before proceeding.

3.1.1 Transient integrations

The past (and future) evolution of the Earth's climate system can be understood by simulating the response to changes in radiative forcings (see figure 3.1), which can be broadly split into two groups; anthropogenic and natural. Anthropogenic forcings, such as the emission of radiatively active gases and aerosols into the atmosphere, have been the main driver of the warming seen over the past century (*Solomon et al.* (2007)). However, natural forcings, such as volcanic aerosol and solar variability, also exert a radiative forcing and, hence, can change the climate. In particular, the eruption of volcanic sulphate aerosol into the stratosphere generally acts to cool the global climate (*Robock* (2000)). The cooling associated with volcanic aerosol is clearly seen in the transient integrations in figure 3.1 following large eruptions such as Agung, El Chichón and Pinatubo.

An accurate simulation of the observed global mean warming trend in the twentieth century thus requires the representation of both the major anthropogenic and natural forcings (*Stott et al.* (2000) (for HadCM3) and *Hegerl et al.* (2007)). The integrations that include all the major anthropogenic and natural forcings are, therefore, referred to as "all forcings" integrations. It is also worth stating that the radiative forcing for the real world, and also for realistic simulations of historical and future climate, is never constant and the climate is not in radiative equilibrium. Thus, realistic simulations of time dependent past and future climate are often called "transient" integrations.

Even with a model that could simulate the climate perfectly, an "all forcings" simulation of the global mean temperature through the twentieth century will not be exactly equal to the observed trend. A single "all forcings" transient integration and the real world evolution differ due to natural variability (see figure 3.1), which makes isolating the changes in climate due *only* to the radiative forcing difficult. One way to address the

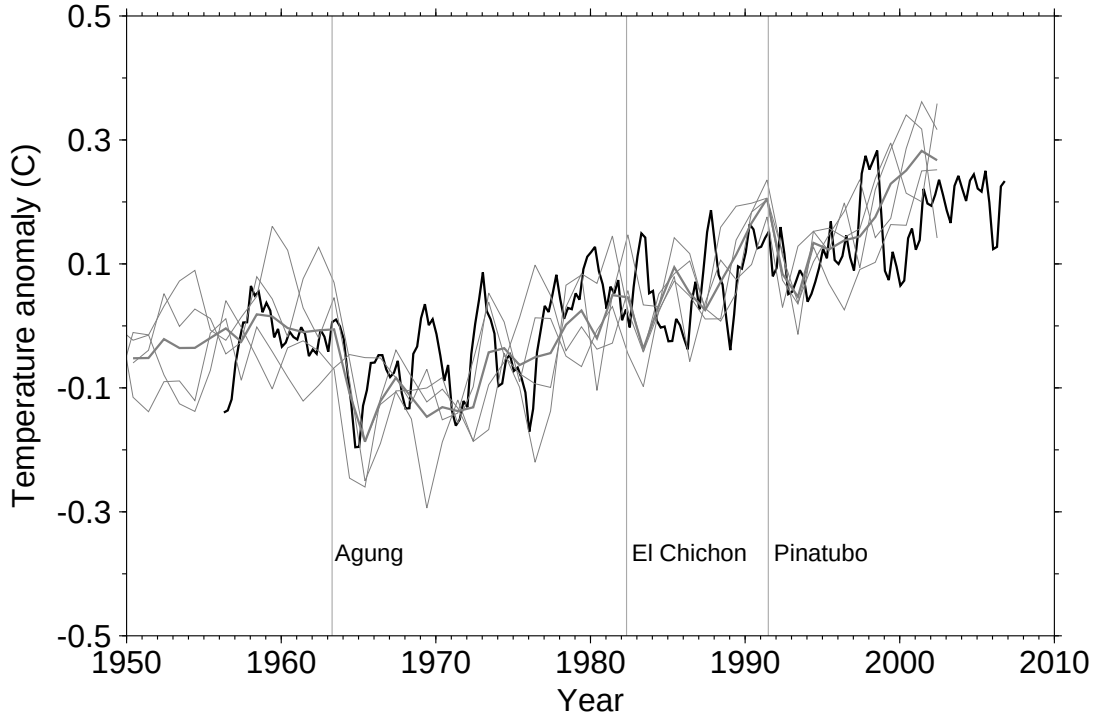


Figure 3.1: The observed global mean top 113m average ocean temperature anomaly from the Original Met Office ocean analysis (black) and the four "all forcings" integrations of HadCM3 from *Stott et al.* (2000) (gray; the thick line shows the ensemble mean and the thin lines the separate members). All anomalies are relative to 1941-1996 climatology. The vertical gray lines show when the eruptions of Agung, El Chichón and Pinatubo took place.

problem of natural variability, in the model at least, is to construct an ensemble of "all forcings" transient runs, which are initialised from a wide range of different initial states taken from a control run (see figure 3.2 a). The control run is a model integration that is usually representative of a pre-industrial climate, with no external radiative forcing applied. Thus, the variability expressed by the control run is only internally produced. The ensemble mean of a sufficiently large ensemble, started across a wide range of initial states, should represent the mean response to radiative forcing by averaging out the natural variability. However, most complex climate models show large amplitude, long time scale natural variability; HadCM3, for example, has centennial modes of variability in the AMOC (*Vellinga and Wu* (2004) and *Hawkins and Sutton* (2007)), which could contaminate the ensemble mean. In order to prevent long time-scale variability contaminating the ensemble mean response to global warming, *Stott et al.* (2000) initialised the transient integrations from states 100 years apart in the HadCM3 control run to ensure

that the long time-scale variability is averaged out.

The transient runs used in the construction of DePreSys are those that were used by *Stott et al.* (2000). More specifically, four transient runs are used that were initialised 100 years apart in a 1300 year control run of HadCM3 (see figure 3.2 a). Time series of the radiative forcings that are applied to all forcings transient runs of HadCM3 to simulate the past are shown in *Gregory and Forster* (2008) (their figure 2). For ease, the transient runs will be called transient run 1 - 4, where transient 1 was initialised earliest in the HadCM3 control run. The following transient members were initialised 100 years apart (i.e. transient 4 was initialised exactly 300 years after transient 1). Each transient run (Y_{ij}) will have its own climatology C_j:

$$C_j = \frac{1}{l} \sum_{i=i_1}^{i_2} Y_{ij} \tag{3.1}$$

where subscript j = 1 - 4 denotes the transient run, the subscript i = time (an index for season and year), and l = number of time points in the interval i_1, i_2. The ensemble mean of the transient integrations as a function of time is therefore given as

$$\langle Y_i \rangle = \frac{1}{m} \sum_{j=1}^{4} Y_{ij} \tag{3.2}$$

where m = the number of ensemble members (4 for the transient runs). The ensemble mean provides an estimate of the mean response of HadCM3 to historical radiative forcings. Finally, the model climatology of HadCM3 will be defined as the ensemble mean of the transient climatologies.

$$C_{mod} = \langle C_j \rangle = \frac{1}{lm} \sum_{i=i_1}^{i_2} \sum_{j=1}^{m} Y_{ij} \tag{3.3}$$

3.1.2 DePreSys

As discussed in chapter 1, DePreSys is constructed to predict the natural climate variability, and, therefore, constrain uncertainty in predictions of the climate systems evolution over the next decade. DePreSys attempts to predict the natural variability by initialising an ensemble forecast from the observed climate state, as shown in figure 3.2 c). Initially the DePreSys ensemble has a small spread, but as the forecast progresses, the ensemble

spread increases until it is equal to the climatological variance of the transient runs. To be able to initialise the observed variability in DePreSys, observational data needs to be assimilated into the model. HadCM3 is not a perfect model, and even though it produces a reasonable representation of the real world climate, considerable biases are present. If the full fields (i.e. the actual observed values) of oceanic and atmospheric variables were assimilated directly into the model, the initial state would lie away from the models preferred (or climatological) state. Once initialised from the observed state the model would drift towards its own climatology. Operational seasonal forecast centres, such as the ECMWF, often remove a mean drift from their forecasts, which is calculated as the mean evolution of an extensive set of hindcasts[1] initialised over a range of climate states, to return "true" forecast anomalies (*Stockdale* (1997)).

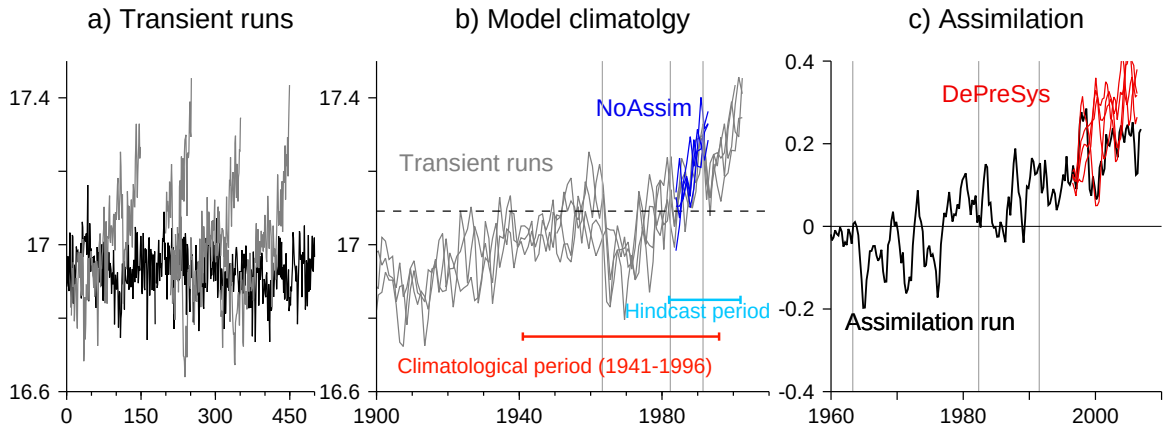


Figure 3.2: a) shows the global mean top 113m annual average ocean temperature from the HadCM3 control run (black) and the four "all forcings" integrations of HadCM3 from *Stott et al.* (2000) (gray). Note that the transient runs are initialised 100 years apart in the control. b) also shows the four "all forcings" integrations of HadCM3 on their real time axis. The climatological period and the hindcast period of *Smith et al.* (2007) are also shown. The annual mean climatology is shown by the dashed black line. *Monthly* mean climatologies are calculated by meaning over all the relevant months in the climatological period. The transient runs also supply the initial conditions for NoAssim (blue). c) shows the global mean top 113m seasonal ocean temperature anomaly from the assimilation run of DePreSys. The assimilation is relaxed to the monthly climatologies *plus* the observed monthly mean anomalies. The assimilation supplies the initial conditions for DePreSys (red). The vertical gray lines in b) and c) show when the eruptions of Agung, El Chichón and Pinatubo took place.

[1]Hindcasts are forecasts that are initialised in the past using the observational data available, and so can be validated against available observational data.

The mean drift over a decade for a climate prediction system like DePreSys could be much larger than the signals that may be predictable on these timescale. Therefore, initialising DePreSys with full fields, and removing the mean drift, was not seen as an attractive option due to potential inaccuracies in the calculation of the mean drift (Sm07). As an alternative approach to full field assimilation, DePreSys assimilates anomalies in order to initialise predictions from the models own climate, hence, removing the problem of drift altogether. The anomaly assimilation method used for DePreSys is explained in depth in the next section.

In addition to attempting to predict the variability of the climate system, DePreSys also tries to predict the response to radiative forcing (which can also be clearly seen in figure 3.2 c), where DePreSys warms at the same rate as the transient runs). In order to predict the changes in the climate due to the response to radiative forcing caused by greenhouse gases and anthropogenic aerosols, the DePreSys hindcasts are forced with the same historical forcings as the transient integrations before the year 1990. After the year 1990 DePreSys hindcasts are then forced with Special Report on Emissions Scenarios (SRES; *Nakicenovic et al.* (2000)) scenario B2 as in *Johns et al.* (2003). Variability in the solar forcing due to the 11 year cycle is represented by repeating the previous solar cycle, and volcanic aerosol is also assimilated into the model following volcanic eruptions such as Pinatubo in 1991. The volcanic aerosol is then removed from the atmosphere with an e-folding time of 1 year. It is not possible to predict the eruptions of volcanoes on decadal time scales. Hence, volcanic eruptions that occur after a hindcast has been initialised are not seen by that hindcast. For example, a DePreSys hindcast initialised in 1988 will not contain volcanic aerosol from the eruption of Pinatubo.

To assess the impact of assimilating the observed information into DePreSys, a hindcast experiment was undertaken at the UK Met Office. 4 member DePreSys ensemble hindcasts were initialised once a season for a 20 year period from 1982 - 2001 (see figure 3.2). These hindcasts were the focus of Sm07, and will be the focus of this thesis. The DePreSys ensemble members were initialised on the days leading up to, and including, the 1st of March, June, September and December. The initialisation of the members on consecutive days gives a crude representation of initial condition uncertainty, and ensures that the ensemble will diverge.

3.1.3 Anomaly Assimilation

The assimilation method used for DePreSys is based upon the simple nudging method (also known as dynamic assimilation or Newtonian relaxation) used in *Anthes* (1974), *Hoke and Anthes* (1976) and *Stauffer and Seaman* (1990). The nudging method adds a term to the models prognostic equations so that the model is relaxed, or nudged, toward a predetermined observational analysis. The dynamical assimilation method effectively makes use of the prognostic model equations to make the observational analysis dynamically consistent with the model; in other words the model has some freedom to adjust to the observations so that its dynamical circulation can come into hydrostatic and geostrophic balance. Therefore, in the assimilation of DePreSys the prognostic equation for a model variable α that is being relaxed to the observed value is now given as:

$$\frac{\partial \alpha}{\partial t} = F(\alpha) + \frac{(A^{\alpha} - \alpha)}{\tau}$$

where $F(\alpha)$ is a complex non-linear function representing the prognostic part of the model equation (i.e. the dynamics and physics that is represented in the model), A^{α} is the state that the variable α is nudged towards (for DePreSys, A^{α} is the model climatology *plus* the observed anomalies), and τ is the relaxation time, or nudging coefficient.

DePreSys assimilates the observed anomalies by relaxing (nudging) the assimilation run to the model climatology (C_{mod}) *plus* the observed anomalies. More specifically the assimilation run is relaxed to the linearly interpolated (in time) monthly mean climatologies *plus* the linearly interpolated observational anomalies (see fig. 3.3). The relaxation is extremely strong, with a relaxation timescale of 3 hours in the atmosphere and 6 hours in the ocean. Monthly mean anomalies of temperature (T) and salinity (S) are assimilated into the ocean, and are calculated from the Original Met Office ocean analysis (*Smith and Murphy* (2007)). 6 hourly anomalies of atmospheric variables, specifically 3D winds (u,v), potential temperature and surface pressure, are also assimilated in this fashion. The assimilation run is relaxed to atmospheric anomalies calculated from ERA-15 (*Gibson et al.* (1997)) for the period 1979-1993 and ERA-40 (*Uppala et al.* (2005)) for 1994-2001. Operational ECMWF atmospheric analyses are used for 2002 onwards.

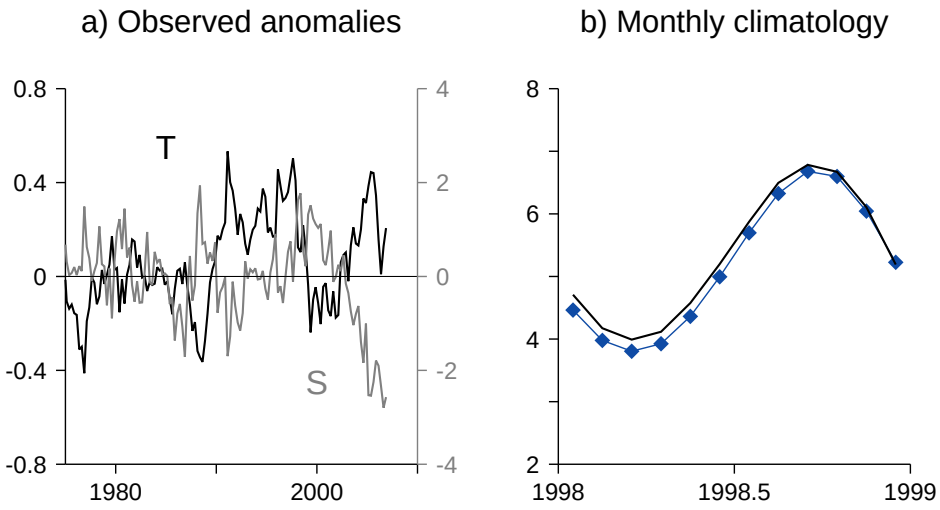


Figure 3.3: a) shows the observed anomalies of *ocean* 0-113m average temperature (T, °C) and salinity (S) that are used in the assimilation of DePreSys, averaged over the North Pacific (140°E-270°E, 40°N-60°N). Note that the salinity anomalies have been normalised for the comparison. b) shows the 12 monthly mean ocean 0-113m average temperature climatologies that are used in the assimilation run of DePreSys (blue dots, °C), calculated over the same region in the North Pacific. Also shown is the monthly mean temperatures from the assimilation run of DePreSys for the year 1998 (black). The assimilation is relaxed to the linear interpolation of the climatologies *plus* the observed anomalies; the trajectory of the assimilations climatology over one year is represented as the straight lines between the monthly mean values. Although only for ocean temperature here, the same method is used for the other variables that are assimilated.Note the difference in the size of the anomalies, and the climatology.

Monthly mean model climatologies for the oceanic and atmospheric variables mentioned above are calculated from "all forcings" transient integrations of HadCM3. The observed climatologies for the oceanic and atmospheric variables are constructed by averaging the observations over the same period. Because of long timescale variability of the ocean, a long climatological period is required, so that the observed climatology is not dominated by natural variability. The climatological period used in Sm07 for the ocean was 1941-1996. Therefore, the climatological period and the period from which the hindcasts are started from (the hindcast period), are not the same. When the hindcasts where originally performed, only ERA-15 was available for the atmospheric observations, which restricted the hindcasts and the climatology to a 1979-1993 period. When ERA-40 was released, the assimilation run and hindcast set were extended to cover the period 1994-2001. However, the climatological period of the atmosphere was moved to 1987-2001

for the assimilation run for 1994 onwards.

Unfortunately, the climatological period and source of the atmospheric anomalies is not the only difference between the hindcasts initialised after 1993 and those initialised before. The original assimilation, from 1979-1993, was only relaxed strongly in the ocean at grid boxes that received information from an observation (either directly or through the spread of information by the covariances used in the ocean analysis). Only relaxing the model where there was observed data enabled the model to, in theory, spread the observed data into data sparse regions. However, when the assimilation run was extended from 1994 onwards, the strong relaxation was applied at every grid point regardless of observational information. Therefore, where there is no observational information the model is relaxed to climatology. Despite the change to the assimilation procedure, it was assumed that the surface is sufficiently well observed that there will be little difference to the skill of the hindcasts, and so Sm07 analysed the skill of the two hindcast sets as one. For the evaluation of skill in this thesis we will also take the DePreSys hindcasts as one hindcast set, but the change in assimilation will be discussed briefly in chapter 5. The differences between the DePreSys hindcasts are summarised in table 3.1

Table 3.1: Summary of the DePreSys hindcasts initialised between 1979 and 1993, and those initialised between 1994 and 2001

Hindcasts	Oceanic Clim	Oceanic Anom	Atmos Clim	Atmos Anom	Ocean relaxation
1979-1993	1941-1996	Met Office Ocean Analysis	1979-1993	ERA-15	only where there are observed anomalies
1994-2001	1941-1996	Met Office Ocean Analysis	1987-2001	ERA-40	All grid boxes relaxing to climatology where there are insufficient observations

3.1.4 NoAssim

To assess whether DePreSys improves predictions of climate variables for the upcoming decade, relative to conventional climate projections, DePreSys is compared to a pre-

diction system that does not assimilate observational information. The hindcasts that do not assimilate observed information are called NoAssim. NoAssim is also based on HadCM3, and is forced in the same way as DePreSys, but its initial conditions are taken from the corresponding time from an ensemble of four transient runs. Thus, the NoAssim ensemble is initialised from a spread of states that are representative of the current climate (see figure 3.2 b) so that the natural variability can be averaged out in the ensemble mean. NoAssim only tries to predict the mean response to radiative forcing, not the natural variability. The NoAssim ensemble, therefore, represents current centennial climate projections.

The initial conditions for the transient runs used to generate the initial conditions for NoAssim were taken from the same ensemble of four "all forcings" transient integrations used to create the DePreSys climatology, so share the same climatological state. However, it appears that the evolution of the transient runs used to generate the NoAssim initial conditions was different to the original transient integrations used to make the climatology (not shown). The climatology of NoAssim may, therefore, be slightly different than the ones calculated from the original transient runs. However, the difference is assumed to be small.

3.1.5 DePreSys2 and DePreSys Perturbed Physics Ensemble (PPE)

Finally, since the completion of the original DePreSys hindcast set used for Sm07 and this thesis, two further hindcast experiments have been undertaken by the UK Met Office. A detailed analysis of the newer DePreSys and NoAssim hindcasts was beyond the scope of this thesis, but a brief analysis of DePreSys2 will be presented in chapter 5. Both the newer systems, **DePreSys2** and **DePreSys PPE** are based upon the methodology of the original **DePreSys**. Both prediction systems assimilate anomalies using the anomaly assimilation method described in section 3.1.3 and assimilate the same variables. Any conclusions made about the methodology or construction of DePreSys should, therefore, also be applicable to these systems. However, there are also some key differences. Hence, it is appropriate to briefly introduce the other prediction systems.

DePreSys2 is extremely similar in construction to the original **DePreSys**, but has a few key differences:-

1. The climatological period used to make the observed and model climatology in the ocean and atmosphere is different. The climatology is constructed from the same sources as was used in DePreSys but the climatological period is now 1951-2006 and 1958-2001 for the ocean and atmosphere respectively.

2. DePreSys2 uses the Updated Met Office ocean analysis for the observed temperature and salinity anomalies in the assimilation run. ERA-40, and then the operational ECMWF analyses are used for the atmospheric variables.

3. The assimilation run for DePreSys2 used a 6 hour relaxation time in the ocean at all grid points for the entire length of the run. Thus, the hindcasts are initialised consistently for all start times, unlike the original DePreSys, where the assimilation technique changed slightly in 1994.

4. The hindcast period for DePreSys2, which covers the years 1960-2005, is much longer than for the original DePreSys (1982-2001) and so covers a wider range of climate states. 10 member ensembles were initialised every 5 years from the 1st November 1960 until 1st November 2005, giving a total of 10 ensembles, and 100 individual members.

The difference between **DePreSys** and **DePreSys PPE** is larger. Sm07 show, in the supporting online materials, that the DePreSys hindcasts are overconfident in their predictions of global mean temperature changes. One source of uncertainty that is not included in the original DePreSys (and DePreSys2) is an estimation of the model uncertainty. Therefore, DePreSys PPE tries to sample the model uncertainty using an ensemble of 9 models with perturbations to the physical parameterisations (taken from *Collins et al.* (2006b)). The 9 models were chosen so that the ensemble would have a spread of climate sensitivities and ENSO variability (*Doblas-Reyes et al.* (2009)). The changes in the model parameters means that the mean global state will change, and so a flux adjustment is applied to the models Sea Surface Temperature (SST) and Sea Surface Salinity (SSS) fields so that they are more similar to the observed (*Collins et al.* (2006b)).

Again the construction methodology for **DePreSys PPE** is very similar to the original DePreSys. Each individual model member has its own assimilation run that is relaxed strongly to each models own climatology. The separate model climatologies are calculated from a single transient run of each perturbed model. The climatological period, hindcast period, and source for the observational anomalies are the same as for DePreSys2. DePreSys PPE has 9 member hindcasts initialised annually in November.

3.2 Notation and skill

3.2.1 Notation

To simplify the large amount of data required in the construction and validation of DePreSys, a clear notation for the different data sets is introduced for the observations, transient runs, assimilation run, NoAssim, and DePreSys hindcasts. Table 3.2 shows the notation and a brief summary of each of the separate data sets, where i = start date (index for season and year), j = ensemble members, and t = hindcast lead time.

The definition of climatologies and anomalies will be important for the understanding of the analysis, and is described here. The model climatology is calculated from the transient integrations and is given by equation 3.3. The analysis climatology for the atmosphere is calculated from ERA-15 or ERA-40 and is given by:

$$C_A = \frac{1}{l} \sum_{i=i_1}^{i_2} X_i \tag{3.4}$$

where l = number of seasons in the time interval i_1, i_2. The Met Office Ocean analysis climatology is actually calculated from the raw observational data before the analysis procedure (*Smith and Murphy* (2007)). The equation 3.4, therefore, allows the recalculation of the analysis climatology for a different period by changing i_1 and i_2.

For the purpose of assessing skill, hindcasts are expressed as anomalies with respect to a suitable climatology. Hence, we define the following anomalies:

Analysis anomaly =

$$X'_i = X_i - C_A$$

Assim anomaly =

$$Z'_i = Z_i - C_{mod,A}$$

NoAssim anomaly =

$$N'_{ijt} = N_{ijt} - C_{NoA}$$

DePreSys anomaly =

$$D'_{ijt} = D_{ijt} - C_{DeP}$$

In principle all of the model climatologies ($C_{mod,A}$, C_{NoA} and C_{DeP}) would be equal to C_{mod}. However, different choices for the climatologies could be made in theory, and the impact of using different climatologies will be examined in this thesis. The time period used to calculate climatologies is defined by the interval i_1, i_2. For oceanic quantities i_1 = DJF 1941, i_2 = SON 1996. For the atmosphere the hindcast anomalies (N'_{ijt} and D'_{ijt}) are made relative to a climatological period of 1979-2001 (i.e.. i_1 = DJF 1979, i_2 = SON 2001), which is different to both the climatologies that were used in the assimilation run for DePreSys. It is unclear whether the use of a different climatological period for the atmospheric anomalies, rather than the same climatologies used in the assimilation, will affect the calculation of skill. The issue of the atmospheric climatology will not be explored in this thesis, but ideally in the future the atmospheric climatologies should be the same for all hindcasts in a particular hindcast set.

Table 3.2: Summary of the data and notation used

data	symbol	Info
Analysis	X_i	Independent, monthly mean analysis made using quality controlled observations of T and S and a correlation matrix derived from a transient run of HadCM3 covering years 1957 - 2006
Transient runs	Y_{ij} j=1,4	Time dependent "all forcings" runs of HadCM3, 1850 - 2001 Also used as the initial states for NoAssim
Assimilation run	Z_i	Initial condition in 1979 taken from a transient all forcings integration. T and S relaxed to model climate (calculated from the ensemble mean of the transient runs 1941-1996) plus analysis anomalies relative to 1941 - 1996 climatology. Atmospheric state relaxed to 1979-1993 (1987-2001) model climatology plus anomalies from ERA-15 (ERA-40) anomalies before (after) 1993
NoAssim	N_{ijt} j=1,4	Time dependent 10 year integrations forced with historical forcings until 1990. After 1990 integrations are forced with IPCC SRES B2. Solar flux is specified by repeating the previous 11-year cycle and volcanic aerosol is reduced with a e-folding time of 1 year. Initialised from transient run Y_{ij} 4 member ensemble hindcasts are initialised each season from 1982 - 2001 (eg 1st Dec, 1st Mar, 1st June and 1st Sept) giving a total of 320 individual hindcasts
DePreSys	D_{ijt} j=1,4	Similar to NoAssim above, but initial state comes from the assim run (Z_i). Initial conditions are taken from consecutive days preceding and including each hindcast start date

3.2.2 Introduction to skill

Skill is a measure of how accurate a particular prediction system is. A prediction system that predicts the future state of the climate system with smaller errors than another prediction system is a more skillful system. There are many ways to calculate the mean skill of a prediction system. The Root Mean Squared Error (RMSE) is widely used in the literature, but it can be sensitive to outliers in the predictions or observations, which leads to large errors. Anomaly correlation, which is not as sensitive to outliers, is also widely used, and was used in the studies of *Keenlyside et al.* (2008) and *Pohlmann et al.* (2009). However, anomaly correlation does not measure quantitative predictions of a variable, only whether the hindcast returns the right sign. The study of Sm07 mainly used the RMSE as a measure of skill, where a lower RMSE means a more accurate hindcast. Therefore, the evaluation of skill in this thesis will also mainly focus upon the RMSE measure of skill. Different measures of skill can show slightly different levels of skill (*Collins* (2002)) and so focusing on RMSE could be seen as a limitation of the study. Hence, another measure of skill that is used in this thesis is the standard deviation of the error, first used in *Haines et al.* (2009).

To calculate the RMSE of a particular hindcast we first start with the hindcast error. The equations for both DePreSys and NoAssim will be the same but for the examples below DePreSys will be used:

$$D_{ijt}^{error} = D'_{ijt} - X'_i \tag{3.5}$$

The use of model and observed anomalies removes the problem of systematic mean model error, which is defined as the difference between the model and the observed climate

$$C_{mod}^{error} = C_{DeP} - C_A \tag{3.6}$$

This is because

$$D_{ijt}^{error} = D'_{ijt} - X'_i$$

$$D_{ijt}^{error} = (D_{ijt} - C_{DeP}) - (X_i - C_A)$$

$$D_{ijt}^{error} = (D_{ijt} - X_i) - C_{mod}^{error}$$

Thus, we are comparing the model anomalies with the observed anomalies. Therefore, DePreSys assumes that the model is capable of predicting the anomalous observed variability, even though the model has systematic mean biases. *Collins et al.* (2001) showed that HadCM3 produces a reasonable reproduction of the variability seen in the observations, suggesting that DePreSys should be able to capture some of the variability correctly. *Schneider et al.* (1999), although using a different assimilation and initialisation technique, showed that assimilating anomalies can produce skillful predictions of ENSO, giving confidence that assimilating anomalies could yield skillful predictions in non-linear systems.

The RMSE of the prediction system as a function of hindcast lead time can be calculated by meaning together the squared hindcast errors for each lead time. The RMSE as a function of lead time is then defined as:

$$D_t^{RMSE} = \sqrt{\frac{1}{lm}\sum_{i=1}^{l}\sum_{j=1}^{m}(\langle D_{ijt}^{error}\rangle)^2} \tag{3.7}$$

where i is the number of hindcast start dates. Even though the analysis concerns only the hindcast anomalies, the model is not perfect and so could have a bias that is dependent on lead time. The comparison between the NoAssim and DePreSys biases could highlight important differences between the prediction systems. The mean bias of DePreSys is defined as:

$$B_t = \frac{1}{lm}\sum_{i=1}^{l}\sum_{j=1}^{m}(D_{ijt}^{error}) \tag{3.8}$$

The RMSE as a function of lead time is calculated for rolling annual means as in Sm07, by averaging seasonal means into a series of annual means. In Sm07, a particular ensemble was combined with hindcasts made in the previous seasons in order to increase each ensembles size. In the current analysis, each ensemble hindcast is kept separate so that it is possible to understand the difference between the evolution of each individual hindcast.

Finally as an alternative to RMSE, *Haines et al.* (2009) used the standard deviation of hindcast errors to also highlight areas where DePreSys has improved skill over NoAssim.

This measure should highlight areas where DePreSys consistently returns lower errors than NoAssim, like the RMSE, but removes the errors due to the mean bias (*Haines et al.* (2009)). The standard deviation of the error is defined as:

$$D_t^{\sigma} = \sqrt{\frac{1}{lm}\sum_{i=1}^{l}\sum_{j=1}^{m}(D_{ijt}^{error} - B_t)^2} \tag{3.9}$$

3.3 Mean Skill Scores of DePreSys

3.3.1 Replication of Smith et al, 2007

In Sm07 oceanic hindcast errors of DePreSys (D^{error}) and NoAssim (N^{error}) were made using the Original Met office ocean analysis, and Surface Air Temperature (SAT) errors were made using the HadCRUT2vOA data set. The calculation of hindcast errors in this thesis involves some differences to those of Sm07. They are as follows:

1. The ERA-40 reanalysis was chosen as the atmospheric observations for the analysis in this thesis. As in the DePreSys assimilation, operational atmospheric analyses where used to extend the ERA-40 data beyond 2002.

2. Sm07 applied a $35^{\circ} \times 35^{\circ}$ spatial smoothing to the hindcast and observed anomalies for the calculation of spatial RMSE. The smoothing of the anomalies could hide small scale features or anomalies, especially in the hindcasts, which could make the understanding of the mechanisms involved more difficult. Therefore, no spatial smoothing will be applied to the hindcasts or observed data in this thesis.

3. Sm07 combined hindcasts from similar start times to increase the size of the hindcast ensembles. Combining of the hindcasts was not performed in this thesis so that the evolution, and, therefore, the successes and failures of individual hindcasts could be identified, and analysed.

These differences in the calculation of skill from Sm07 could affect the apparent skill of DePreSys relative to NoAssim. To expand the analysis of skill beyond Sm07, and

to make useful inferences, the effect of the above changes upon the skill presented in Sm07 should be first understood. Therefore, the RMSE as a function of lead time and of latitude and longitude, for DePreSys and NoAssim, is presented below for comparison with Sm07.

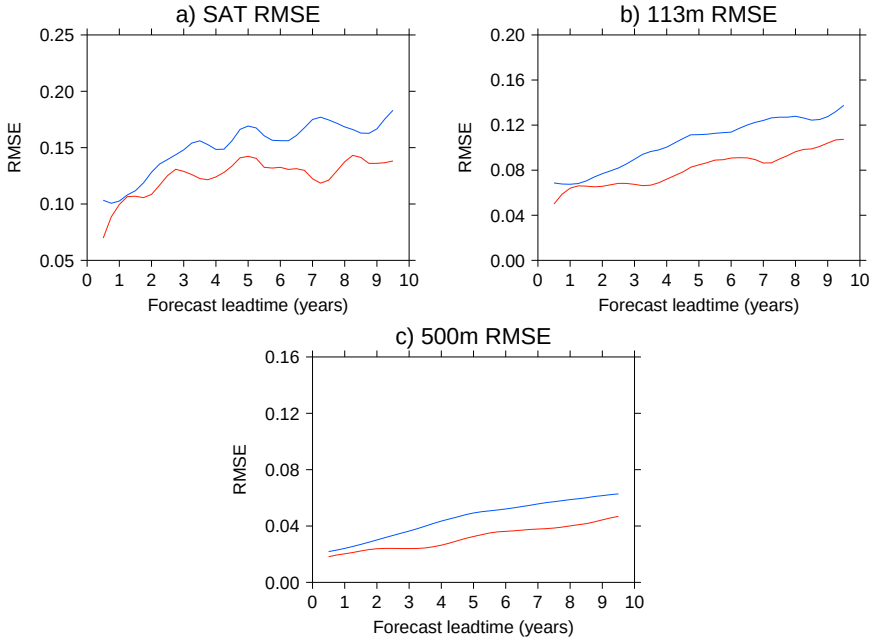


Figure 3.4: Comparison of RMSE as a function of lead time calculated for ensemble mean DePreSys hindcast anomalies (red; $\langle D'_{ijt} \rangle$) compared to the hindcast anomalies calculated from NoAssim (blue; $\langle N'_{ijt} \rangle$) for global mean SAT (a), 113m heat content (b) and 500m heat content (c). SAT anomalies are made relative to the 1979-2001 climatology. Heat content anomalies are made relative to the 1941-1996 climatology.

Figure 3.4 shows the RMSE as a function of hindcast lead time is smaller in DePreSys than NoAssim for a number of global mean variables for the entire length of the 10 year hindcasts, which is consistent with figure 1 in Sm07 (see chapter 1). The differences between the DePreSys and NoAssim RMSE for both the global mean SAT and 113m heat content (fig. 3.4 a and b) do not appear as large as those seen in Sm07, especially in the first couple of years of the hindcasts. The change in the atmospheric observations is a likely cause for the change in the SAT RMSE difference, but the lack of combining hindcasts initialised close in time is likely to also increase the RMSE by increasing the sampling error.

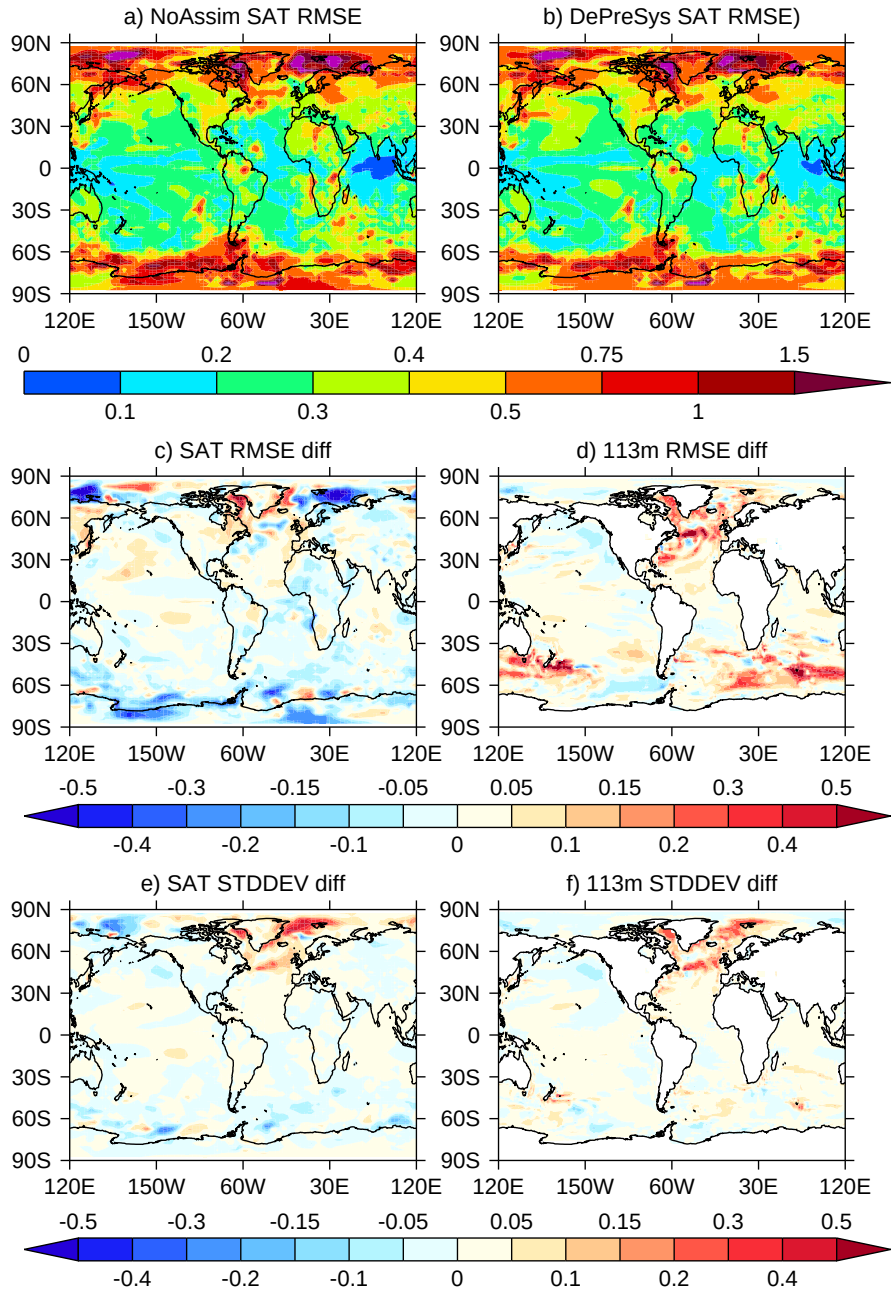


Figure 3.5: The spatial RMSE of 9 year mean SAT for a) NoAssim (N_{xy}^{RMSE}) and b) DePreSys (D_{xy}^{RMSE}). c) shows the difference between the 9 year mean NoAssim and DePreSys RMSE ($N_{xy}^{RMSE} - D_{xy}^{RMSE}$) for SAT. d) as c) but for 113m heat content. e) and f) show the difference in 9 year mean NoAssim and DePreSys error standard deviation ($N_{xy}^{\sigma} - D_{xy}^{\sigma}$) for SAT and 113m heat content respectively

The spatial RMSE and RMSE difference between DePreSys and NoAssim is repeated here for the 9 year mean SAT and 113m heat content anomalies in figure 3.5. Generally figure 3.a and b in Sm07 (shown in chapter 1, figure 1.7) and figure 3.5 a)-b) are comparable with increased RMSE over the North America and African continents, and the

high latitudes of the Eurasian continents, although the magnitude of the RMSE is much greater in figure 3.5 a)-b) than Sm07. It is likely that the lack of regional smoothing is the cause of the larger values of RMSE seen in figure 3.5.

The difference in the RMSE for SAT between NoAssim and DePreSys is shown in figure 3.5 c) and seems quite different to the figure 3.c in Sm07 (see figure 1.7). There is little sign of an improvement of DePreSys over the South American continent and in some areas of the Southern Ocean in figure 3.5 c) and the difference is much more patchy than in Sm07. In Sm07 NoAssim apparently performs better than DePreSys for SAT over a large part of the North Atlantic between 30°N-60°N (their figure 3.c), but NoAssim is only outperforming DePreSys in small areas of the North Atlantic. Figure 3.5 c) does highlight the GIN Sea as an area where NoAssim is performing significantly better than DePreSys. The Improvement of DePreSys over NoAssim in North America seen in Sm07 is also not as clear in figure 3.5 c).

The analysis for 113m heat content is broadly similar between the two studies. Both figure 3.5 d) and figure 3.d in Sm07 show an improvement of DePreSys in the North Atlantic and large areas of the Southern Ocean, especially between the east of the African continent ($\sim$30°E) and $\sim$150°W. A major difference between the studies is in the region immediately south east of the African continent, where in Sm07 NoAssim apparently performs better than DePreSys. In figure 3.5 b) the negative values are not present. The improvement of the performance of DePreSys relative to NoAssim in the North Atlantic and Southern Ocean is consistent with the area of high potential predictability (shown in *Boer* (2000, 2004)), which led Sm07 to suggest that DePreSys is able to capture the observed long timescale variability in these regions better than NoAssim.

The change in the RMSE caused by changing the atmospheric observational analysis highlights the sensitivity of the mean skill scores to the observations, and suggests that hindcasts need to be compared to as many different analyses as possible for a clear picture of skill to emerge. The lack of regional smoothing here is also likely to be a key cause of the difference, not just in the magnitude of the RMSE differences; the nonlinear nature of the RMSE score could also cause differences in sign of the RMSE difference between NoAssim and DePreSys.

Finally, figure 3.5 e) and f) show the difference in the standard deviation of the 9 year mean hindcast errors (following equation 3.9) between NoAssim and DePreSys. The standard deviation of the errors highlights the North Atlantic (North of 40°N) as a region where DePreSys is performing better than NoAssim for both SAT (fig. 3.5 e) and 113m heat content (fig. 3.5 f). The standard deviation of the error also removes most areas where NoAssim is performing better for 9 year mean SAT anomalies. A major disagreement between the standard deviation of the errors and the RMSE is in the Southern Ocean. Both Sm07 and this analysis highlight the Southern Ocean as an area where DePreSys performs better than NoAssim for the 113m heat content when using RMSE as the measure of skill. However, the standard deviation of the error does not show that DePreSys gives an improvement over NoAssim here. Thus, the reduction in RMSE over the Southern Ocean is due to a reduction of bias in DePreSys, *not* an improved ability to predict variability. The reduction in improvement of NoAssim SAT predictions when using the standard deviation of the error (compare figure 3.5 c and e) also shows that DePreSys has a larger SAT bias than NoAssim in some regions of the North Atlantic and Southern Ocean.

The large difference between the spatial RMSE and the standard deviation of the error highlights that there is more to understand about the calculation of the mean skill of DePreSys than was presented in Sm07, and for understanding how DePreSys and NoAssim are actually performing.

3.3.2 Climatological Difference

When comparing hindcast anomalies from DePreSys and NoAssim, it became apparent that there were some extremely large differences in the deep ocean. Figure 3.6 shows the ensemble mean difference between DePreSys and NoAssim for a deep layer of the ocean (3000-5000m) in the first season of hindcasts started in June 1995. Figure 3.6 shows the difference between only one DePreSys and NoAssim ensemble, but the result is representative of all the hindcasts (not shown). Investigations into the cause of these differences revealed that the climatology used for the DePreSys assimilation run was not made from an ensemble of 4 transient runs, but was in fact only constructed from one

transient run[2].

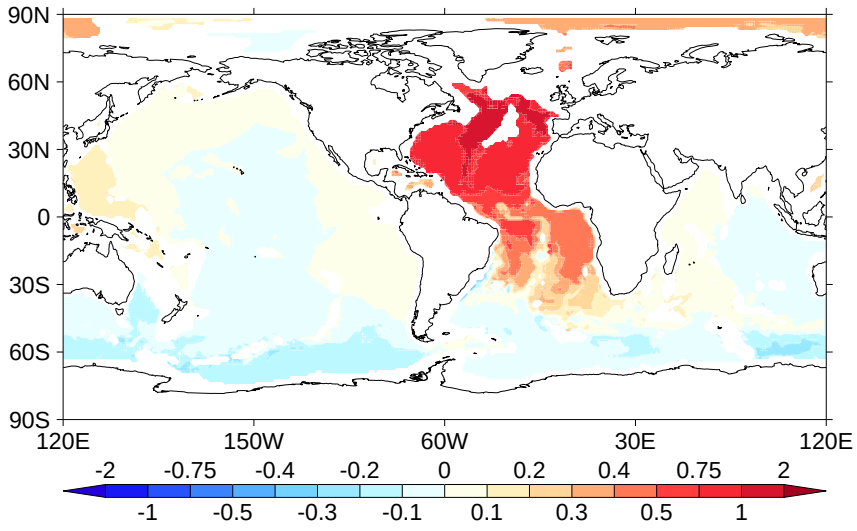


Figure 3.6: Difference in first season ensemble mean 3000-5000m average temperature (°C) between DePreSys and NoAssim hindcasts ($\langle D'_{t=1} \rangle$ - $\langle N'_{t=1} \rangle$) started in June 1995.

The ocean climatology used for the original DePreSys assimilation run, where i_1 = DJF 1941 and i_2 = SON 1996, is

$$C_{mod,A} = \frac{1}{l} \sum_{i=i_1}^{i_2} Y_{i1} = C_1$$

which is not the same as the model climatology C_{mod} calculated from the ensemble mean of the transient ensemble. The difference between the climatologies (C_{diff}) is therefore.

$$C_{diff} = C_{mod} - C_{mod,A} = C_{mod} - C_1 \tag{3.10}$$

Figure 3.7 shows the annual average difference between the two climatologies (C_{diff}). There are large differences in SAT, with C_1 over 1°C warmer than C_{mod} in the European region and Greenland-Iceland-Norwegian (GIN) sea region. Differences greater than $\pm$0.5°C are seen in large areas of the Southern Ocean. In the ocean, C_{diff} is largest in the Southern ocean, near the surface, and throughout the whole depth of the Atlantic.

[2]Note that because the source of the climatology is the same, the climatological error also applies to DePreSys2

C_{diff} in the deep Atlantic (figure 3.7, panel f) is over 1°C and sometimes over 2°C which are extremely large differences at this depth relative to the interannual variability of the HadCM3 control run (not shown).

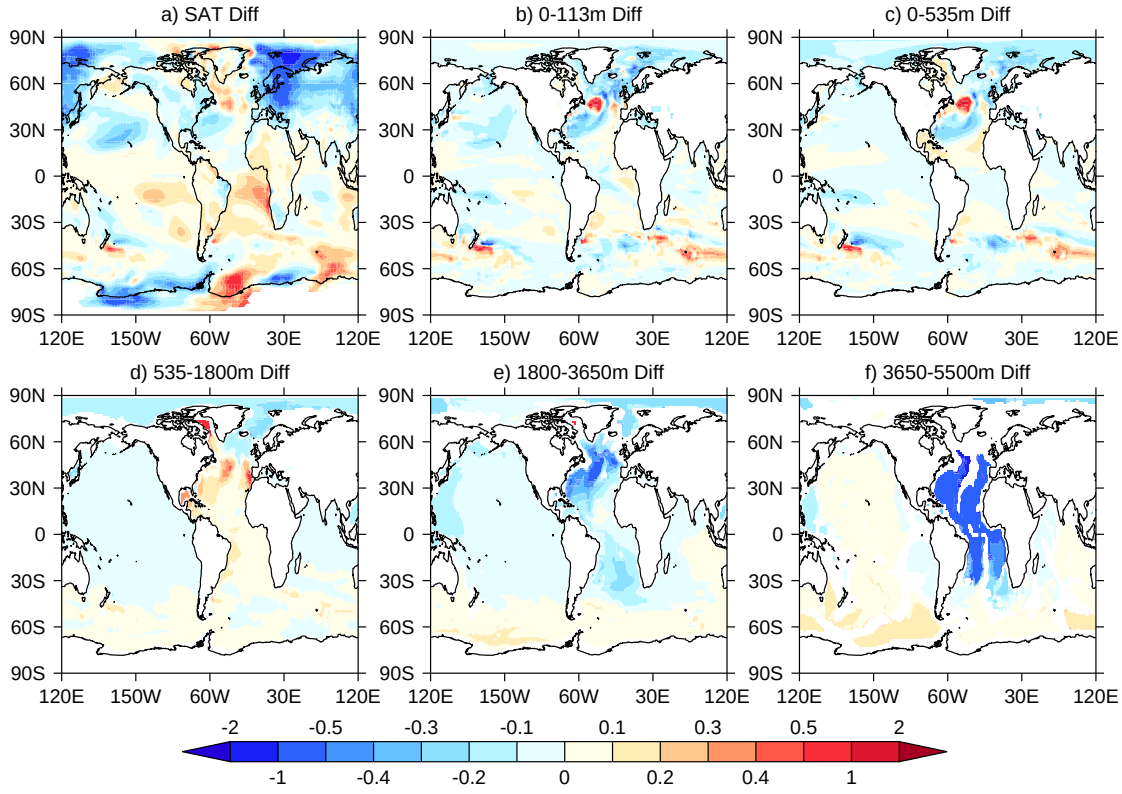


Figure 3.7: Difference between the transient ensemble climatology and the climatology constructed just using the transient 1 ($C_{diff} = C_{mod} - C_1$) for SAT (a; (°C)) for a 1979-2001 climatology and for various average temperatures in layers in the ocean (b-f; (°C)) for a 1941-1996 climatology.

The differences in the climatologies seen in figure 3.7 are caused by model drift. The HadCM3 ocean was originally initialised from rest with observed climatologies of temperature and salinity, and subsequently spun up for 360 model years. The HadCM3 control run, from which the initial conditions for the transient integrations were taken, was started after the 360 year spin up period. However, because of systematic errors in the model, the model has a different preferred state or climatology which it will drift towards. The processes in the ocean, especially the deep ocean, are very slow. Thus, it can take the model many centuries, or even many millennia, of model time to reach the model's preferred state. Figure 3.8 clearly shows the slow drift in the deep ocean. HadCM3 cools in the deep Atlantic for almost 2000 model years from initialisation (see

figure 3.8). Transient 1 was initialised in year 9 of the HadCM3 control run (369 years from the initialisation of HadCM3 from the observed climatology) and is clearly still undergoing strong drifts in the deep and intermediate waters of the Atlantic (see fig. 3.8). The 55 year periods over which the transient climatologies were computed are also shown in figure 3.8, which demonstrates clearly that C_{diff} is caused by the drift of the control run between the initial states of the transient runs, and of continued drift of the transient runs themselves.

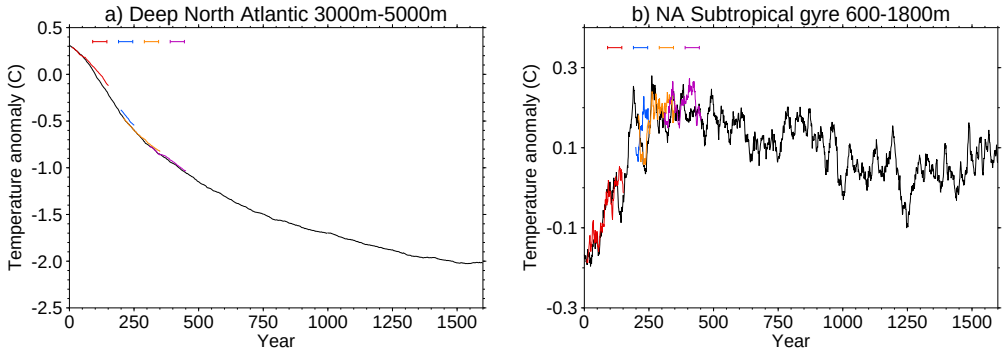


Figure 3.8: a) shows the evolution of North Atlantic (70°W-10°E,0°N-70°N) 3000-5000m average temperature anomalies (°C), relative to the climatology of the first transient run C_1, from the HadCM3 control run (black) and transient 1-4 (red, blue, yellow and purple respectively). Note only the data from 1950 was available for transient 2. b) shows the same but for North Atlantic subtropical gyre (70°W-10°W,20°N-50°N) 600-1800m average temperature anomalies. The time in years shows the model years since the control run was started after a 360 year spin-up. The coloured bars show the time period from where each transient runs 1941-1996 climatology is constructed from

In the analysis of Sm07 the error in the construction of the ocean climatology (i.e. that the ocean climatology used for the assimilation was C_1 not C_{mod}) was not known. Therefore, in Sm07 DePreSys and NoAssim hindcast ocean anomalies were both made using the climatology calculated from the first transient run, C_1.

$$C_{DeP} = C_{NoA} = C_{mod,A} = C_1$$

However, as mentioned earlier, the NoAssim ensemble is initialised from the four transient integrations. Thus, the average of the states found in NoAssim should be close to C_{mod}. Removing C_1 from both DePreSys and NoAssim hindcasts would not return the hindcast anomalies as intended. The NoAssim hindcast anomalies, relative to their own climate C_{mod}, should be:

$$\langle N'_{correct} \rangle = \langle N \rangle - C_{mod}$$

but in the analysis of skill presented in Sm07 the NoAssim hindcast anomalies are:

$$\langle N' \rangle = \langle N \rangle - C_1$$

substituting $\langle N \rangle = \langle N'_{correct} \rangle + C_{mod}$ gives

$$\langle N' \rangle = \langle N'_{correct} \rangle + C_{mod} - C_1$$

but from 3.10

$$\langle N' \rangle = \langle N'_{correct} \rangle + C_{diff}$$

The ensemble mean NoAssim hindcasts anomalies, relative to C_1, will have a bias $\sim C_{diff}$. The extra bias added onto NoAssim will not be exactly C_{diff}, because of the model drift seen in figure 3.8. The NoAssim hindcasts are initialised between 1982 and 2001, not from the years 1941-1996 from which the climatology is constructed. The transient runs are free integrations and so would continue to drift past the 1941-1996 climatological state. In contrast, DePreSys is relaxed strongly to the model climate and the observed anomalies. However, in the deep ocean the observed anomalies are small, and so the DePreSys hindcasts will be essentially initialised from the climatology (not shown).

To confuse the analysis of skill further, the atmospheric climatology for the DePreSys assimilation was constructed from the ensemble of all four transient runs (C_{mod}, not shown), and the DePreSys and NoAssim hindcast atmospheric anomalies where made also using the ensemble mean climatology. After the DePreSys hindcasts have been started from the assimilated state, the atmosphere will move to be in equilibrium with the ocean. Therefore, the atmosphere will move away from C_{mod} towards C_1. Hence, it is not entirely clear what climatology the DePreSys hindcast atmospheric anomalies should be made relative to. Because the atmospheric climatology was C_{mod} in the assimilation run the atmospheric climatology is kept as C_{mod} for the analysis of skill. A bias correction could be applied to remove the mean drift, but as will be highlighted in chapter 4, the calculation and application of a bias correction is not necessarily straightforward.

Because of the above error in the construction of hindcast anomalies, the analysis of skill that was performed in Sm07 was not a fair comparison of DePreSys and NoAssim. The initialisation method was not the only difference between the two prediction systems;

the systems entire climatological state was also different. Also, the presence of a bias equal to C_{diff} (approximately equal to that are seen in fig. 3.7) in the NoAssim hindcast anomalies could change the bias and RMSE of NoAssim, and hence the relative skill of DePreSys to NoAssim. Sm07 claimed that the improvement in skill of DePreSys relative to NoAssim is due to the initialisation from the observed climate state. The error in climatology might mean that the actual relative skill of DePreSys could be different to that presented in Sm07. Hence, an analysis of the effect of the climatology error is important for understanding initialised decadal hindcasts from DePreSys.

It was not possible to directly examine the effect of relaxing the DePreSys assimilation run to C_1 not C_{mod} because it would have involved rerunning the assimilation and all the hindcasts (although this issue will be examined for a selected sub set of the hindcasts as a case study in a chapter 7). However, it is possible to understand the sensitivity of NoAssim's bias and RMSE, and, therefore, the apparent skill of DePreSys, to the climatology used to make the hindcast anomalies. The analysis of Sm07 will now be repeated here but now calculating the NoAssim anomalies relative to their "correct" climatology.

Different choices are possible for the "correct" NoAssim climatology. Ideally, the NoAssim climatology would be the model climatology (i.e. the mean of many transient integrations initialised from non-drifting states), as the climatology of one transient integration could be affected by poorly sampled low frequency variability. Using a single transient climatology, or indeed a separate climatology for each transient integration, could bias the NoAssim hindcast anomalies. However, as already seen in figure 3.8, each separate transient integration is in a different climatological state, due to the drift of the control run. Using the mean model climate for these NoAssim hindcasts may not cause an ensemble mean bias, but each individual NoAssim member would still have its own bias. For the purpose of choosing the best climatology for these NoAssim hindcasts, the sensitivity of NoAssim's RMSE is initially investigated using hindcasts anomalies calculated for 3 different climatologies:

1. The climatology from transient 1 ($C_{NoA} = C_1$). So that

$$N'_{ijt} = N_{ijt} - C_1$$

the same as that used for the DePreSys assimilation run, and for creating the NoAssim hindcast anomalies in Sm07. Call this **Trans1 Clim**.

2. The ensemble mean of all the transient members ($C_{NoA} = C_{mod}$).

$$N'_{ijt} = N_{ijt} - C_{mod}$$

Call this **Mean Clim**.

3. Climatologies calculated for each transient integration separately. Each member of NoAssim has hindcast anomalies which are calculated using the climatology from the transient member which NoAssim was initialised from. E.g.

$$N'_{ijt} = N_{ijt} - C_j$$

Call this **Sep Clim**.

Unfortunately, data prior to 1950 for transient member 2-4 was not available, so it was impossible to construct the 1941-1996 model climatology for the ocean for each transient member separately. The climatological period used for making the hindcast and observed anomalies for the ocean will be swapped to a 1951-2006 climatology for the evaluation of skill to accommodate for the lack of data. The change of the climatological period could also have an effect upon the bias of the system, and so will be investigated first in section 3.3.3 on only hindcast anomalies calculated relative to C_1 (**Trans1 Clim**). After the effect of changing the climatological period has been investigated, the sensitivity of NoAssim's RMSE to the climatology will be investigated in section 3.3.4. For clarity the atmospheric climatological period remains as 1979-2001. Although the NoAssim atmospheric anomalies were made relative to C_{mod} in the original analysis of Sm07, the effect of using **Trans1 Clim** upon the skill of SAT will also be examined for completeness.

3.3.3 Sensitivity to the Climatological Period

It is possible that the change in the climatological period for the calculation of ocean anomalies could cause differences in the bias, and hence skill, of DePreSys and NoAssim.

This is because hindcast error is the difference between hindcast anomaly and observed anomaly.

$$D_{ijt}^{error} = D'_{ijt} - X'_{ijt}$$

$$D_{ijt}^{error} = (D_{ijt} - C_{mod}) - (X_{ijt} - C_A)$$

So changing the climatological period will cause a difference in hindcast error of

$$\Delta D_{ijt}^{error} = D_{ijt}^{error41-96} - D_{ijt}^{error51-06}$$

$$\Delta D_{ijt}^{error} = ((D_{ijt} - C_{mod}^{41-96}) - (X_{ijt} - C_A^{41-96})) - ((D_{ijt} - C_{mod}^{51-06}) - (X_{ijt} - C_A^{51-06}))$$

which becomes

$$\Delta D_{ijt}^{error} = -C_{mod}^{41-96} + C_A^{41-96} + C_{mod}^{51-06} - C_A^{51-06}$$

or

$$\Delta D_{ijt}^{error} = (C_{mod}^{51-06} - C_{mod}^{41-96}) - (C_A^{51-06} - C_A^{41-96})$$

$$\Delta D_{ijt}^{error} = C_{mod}^{change} - C_A^{change}$$

The change in the observed and model climatologies does not have to be the same. Even for a perfect model the difference in the modelled and observed climatologies would be different due to natural variability

$$C_A^{change} \neq C_{mod}^{change}$$

the change in the mean bias of the prediction system caused by changing the climatological period is:

$$B_{clim} = \frac{1}{lm} \sum_{i=1}^{l} \sum_{j=1}^{m} (\Delta D_{ijt}^{error}) \tag{3.11}$$

Therefore, if the climatologies are changed in this way, the hindcast anomalies will gain an extra bias (B_{clim}) equal to the change in hindcast error seen in equation 3.11. The ΔD_{ijt}^{error} is the same for each hindcast and lead time and so

$$B_{clim} = C_{mod}^{change} - C_A^{change}$$

Hence, it should be possible to calculate the differences between the two climatological periods for both the observations and the model climatologies as above and then remove

the bias. However, a further complication in this analysis is that a clean comparison between the two observed climatologies is not possible, as the observed data from the first Met Office ocean analysis only covers 1956-2006. Thus, it is also necessary to change the observational data set (X_i) used for the analysis here. The data set used for the analysis will be the Updated Met Office ocean analysis (described in Appendix A) which covers the period 1950-2007, and has climatological period of 1951-2006[3]. However, the variability and placement of the anomalies in the Updated Met Office ocean analysis will not be the same as the Original Met Office ocean analysis. The climatology of the new analysis is, therefore, not the only difference; the difference in variability, and possibly trends, could also have an effect on the bias. As a result, calculating B_{clim} may not be as simple as only calculating the difference between the change in climatologies.

The use of this new ocean analysis is not ideal; however, an investigation on the effect of the climatology error on the skill of DePreSys relative to NoAssim seems important enough to warrant the change. To be able to investigate whether B_{clim} is large enough to affect the skill of the hindcasts, the mean bias (equation 3.8) will be calculated for DePreSys and NoAssim hindcast anomalies using both the 1941-1996 climatology, and the 1951-2006 climatology calculated from transient 1. The hindcast errors will be calculated using the old analysis anomalies when they are relative to the 1941-1996 climatology, and the new analysis when they are relative to the 1951-2006 climatology. Therefore, B_{clim} is defined as:

$$B_{clim} = B_t^{41-96} - B_t^{51-06}$$

Figure 3.9 shows that changing the climatological period does have a large effect upon the mean bias of the DePreSys and NoAssim hindcasts. A large offset in the mean bias is seen for both global mean 113m and 500m heat content respectively (figure 3.9 panels a and b). If the change in the mean bias is brought about *only* by the change in the climatological period, then the change in mean bias for both NoAssim and DePreSys would be equal for both prediction systems because the same change was applied. The change in the mean bias is extremely similar for both NoAssim and DePreSys for both the 113m and 500m heat content (see figure 3.9 c). The difference between the biases should

[3]note that the error in the Updated Met Office ocean analysis was corrected for this analysis. See Appendix A for details.

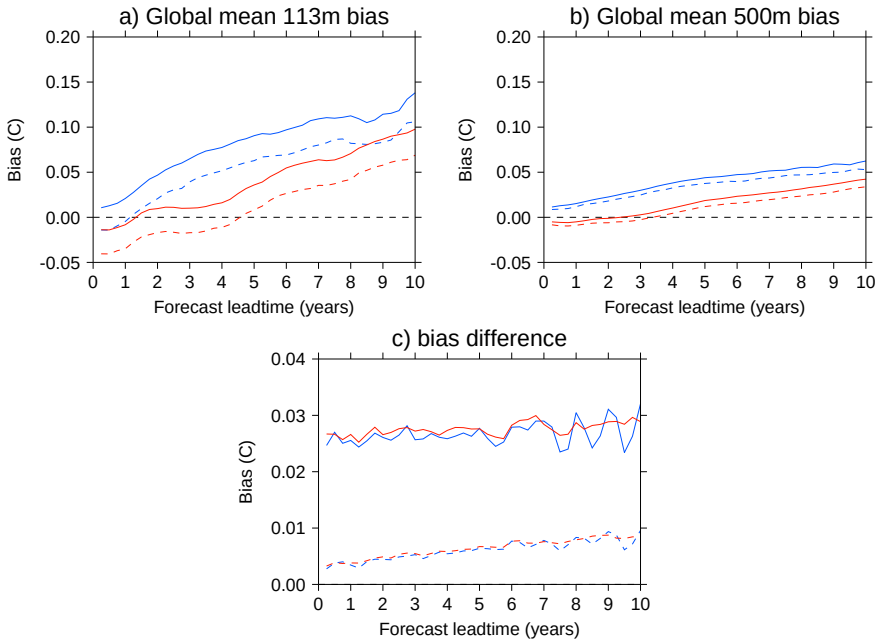


Figure 3.9: Global mean hindcast bias (B_t) for a) 113m heat content and b) 500m heat content for DePreSys (red) and NoAssim (blue). Solid and dashed lines show bias calculated using the 1941-1996 and 1951-2006 climatology respectively. c) shows the difference between the biases for DePreSys and NoAssim (e.g. 1941-1996 minus 1951-2006 bias). In c) solid and dashed lines show the bias difference for global mean 113m heat content and 500m heat content respectively

be constant with hindcast lead time, whereas the difference shown in figure 3.9 c) is not constant with hindcast lead time, especially for the 500m heat content. This increase in the bias is likely to be brought about by the change in the observational anomalies variability and trends.

The change in the mean bias of the prediction systems, seen in figure 3.9 c), does affect the RMSE (not shown), and would make understanding the effect of the climatology error on the results of Sm07 impossible if left uncorrected. Hence, a correction is applied to the hindcast anomalies. The correction for B_{clim} is be defined as the mean of the change in bias for both NoAssim and DePreSys, averaged over the first year of the hindcasts. The same 3D correction will then be applied to the entire length of both the DePreSys and NoAssim hindcasts. The bias correction is calculated as a mean bias averaged over all seasons. It is possible that the offset could be different for each season, but it is assumed that the seasonal difference is small and so the annual mean correction is applied.

3.3.4 NoAssim's sensitivity to the choice of climatology (C_{NoA})

It was shown in section 3.3.2 that an error in the calculation of the C_{DeP} for the ocean, and the use of the C_{DeP} to create the NoAssim forecast anomalies, meant that the NoAssim hindcasts would be left with a residual anomaly C_{diff}. It is important to establish whether the climatology difference will affect the skill of NoAssim in Sm07. The analysis of NoAssim is repeated using the 1951-2006 climatology and the updated Met Office ocean analysis as the observations. The climatological period for the atmosphere will remain as 1979-2001.

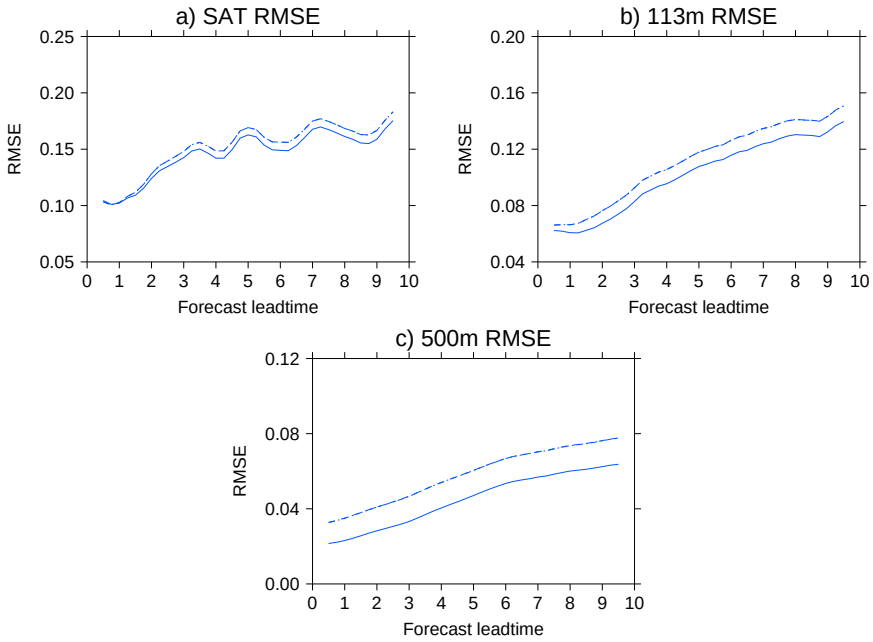


Figure 3.10: RMSE as a function of lead time calculated for NoAssim ensemble mean hindcast anomalies (N_t^{RMSE}) calculated using the three different model climatologies for global mean anomalies of SAT (top left), 113m heat content (top right) and 500m heat content (bottom centre). The solid, dotted and dashed line shows the RMSE for NoAssim when using the transient 1 (**Trans1**) climatology, the mean transient climatology (**Mean**) and the separate transient (**Sep**) climatologies respectively. Using the DePreSys climatology (C_1) to calculate the NoAssim hindcast errors gives a false impression of the skill of NoAssim

Figure 3.10 shows NoAssim's RMSE as a function lead time is sensitive to the climatology difference. The **Trans1** RMSE of the SAT (fig 3.10 a) starts similar to the original but increases to a slightly lower value over the course of the hindcasts. The 113m and the 500m heat content (3.10 b) and d) respectively) show an increase in the RMSE over the whole length of the hindcasts when using **Mean** and **Sep** climatologies. The re-

sults presented in figure 3.10 indicate that the global mean RMSE as a function of time for NoAssim was underestimated in Sm07, especially for the ocean heat content. The RMSE calculated using the **Mean** and **Sep** climatologies are the same here because the analysis was performed upon the ensemble mean anomalies, which should be the same for the **Mean** and the **Sep** climatology by construction.

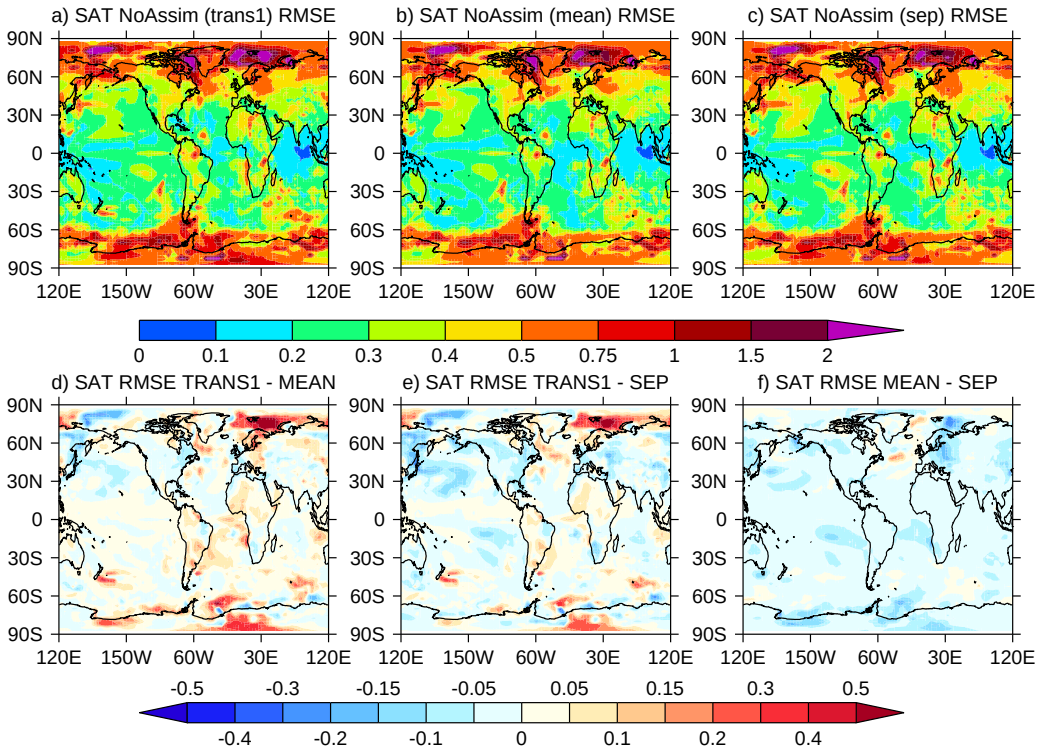


Figure 3.11: panel a-c show 9 year mean surface air temperature (SAT) RMSE for NoAssim hindcasts (N^{RMSE}) where the hindcast anomalies are calculated using the **Trans1 Clim**, **Mean Clim** and **Sep Clim** climatologies respectively. Panel d) and e) shows the impact using the **Mean Clim** and the**Sep Clim** compared to the **Trans1 Clim** respectively. Panel f) shows the difference in RMSE when using the **Mean Clim** and the **Sep Clim**.

The sensitivity of the SAT spatial RMSE is examined in figure 3.11 for 9 year mean anomalies in a similar way to Figure 3 of Sm07. Qualitatively, the difference between the RMSE calculated using the three different hindcast anomalies is small (see figure 3.11 panels a) to c)). However, there are subtle differences in the SAT RMSE calculated using different climatologies. When the climatology is changed between **Trans1** and the **Mean** or **Sep** climatologies the largest differences seem to be located in the Antarctic and GIN sea regions (figure 3.11 d) and e) respectively), but there are subtle changes

across much of the globe. The areas where RMSE differs is geographically consistent with the difference in the SAT climatology (See figure 3.7 a)). The largest difference to the RMSE is caused by swapping away from **Trans1 Clim**, but there is still a large difference between the **Mean** and **Sep** climatologies (figure 3.11 f).

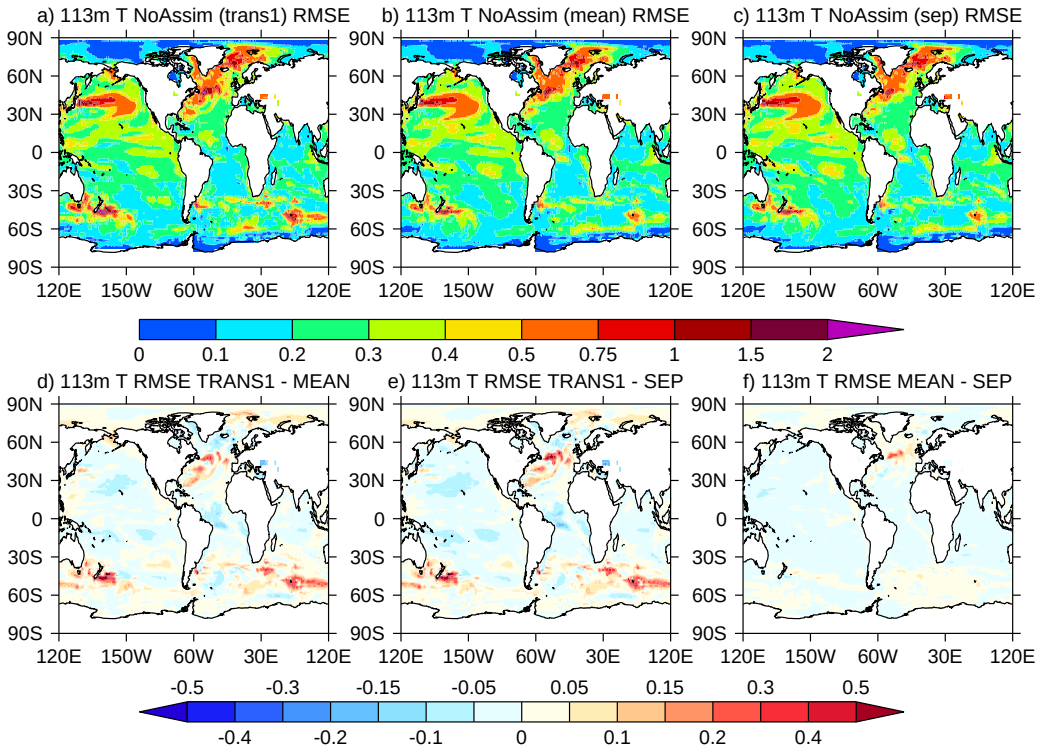


Figure 3.12: As for figure 3.11 but now for 113m average temperature.

In addition to the changes in SAT spatial RMSE, there subtle changes in the top 113m heat content RMSE shown in figure 3.12. Again, qualitatively speaking the difference in spatial RMSE (shown in figure 3.12 a)-c)) is small, for the three different climatologies. Again, changing the climatology does change the RMSE. Using the **Mean** or **Sep** climatologies shows that the original analysis overestimates the RMSE in the areas of the Southern Ocean as well as the subtropical North Atlantic (20° - 50°N). There is also a slight underestimation of the RMSE in the Pacific (slightly clearer when using the **Sep Clim**) and in the subpolar North Atlantic (50° - 70°N). The areas where RMSE differs is also geographically consistent with the difference in the 113m climatology (See figure 3.7 b))

3.3.5 Summary and discussion

In the analysis of Sm07, an error was made in the construction of the model climatology for the ocean. The ocean climatology was made from just one transient run, and not an ensemble of four. The DePreSys and NoAssim hindcast anomalies were made relative to the climatology derived from transient 1. Unfortunately, the transient runs were undergoing significant drift. The NoAssim hindcasts initialised from the later transient runs were in a different climatological state than transient 1, and so significant bias is present in the NoAssim hindcast anomalies.

The RMSE for the NoAssim hindcasts is sensitive to the climatology used to construct the hindcast anomalies. Because of the drifting transient runs, and because there are relatively large changes in RMSE between hindcast anomalies made relative to the **Mean** and **Sep** climatologies, it was decided to construct the NoAssim hindcast anomalies using **Sep** climatology for the re-evaluation of skill in chapter 4 . Unfortunately, to be able to use separate climatologies, the climatological period also needs to be changed due to a lack of transient run data. The change in the climatological period also causes a change in the mean bias of the hindcasts, which will affect the evaluation of skill. Thus, to enable a direct comparison of skill with the results of Sm07, a bias correction is applied to correct for the change in climatological period.

For future NoAssim prediction systems, it seems important to use transient runs with minimal drift. In that case the **Mean Clim** should be still used, so that poor sampling of low frequency variability in the calculation of individual transient climatologies do not affect the skill scores.

3.4 Chapter Conclusions

This chapter has introduced the construction of the UK Met Office's Decadal Prediction System (DePreSys), and the hindcast experiment that was undertaken for Sm07 has been described. A notation was introduced for clarity, as well as the basic measures of skill that will be used throughout this thesis. The results of Sm07 were also reproduced. The

main results are summarised below.

- The two measures of skill used here, the difference in Root Mean Squared Error (RMSE), and the difference between the standard deviation of the errors, agree that DePreSys is performing better than NoAssim at predicting 9 year mean 113m heat content in the North Atlantic subpolar gyre region. However, there are significant differences between the measures of skill for the prediction of 9 year mean 113m heat content in the Southern Ocean. The results suggest that the improvement in DePreSys over NoAssim seen in the Southern Ocean in the analysis of Sm07 was due to a reduction in bias, not improved predictions of variability.

- The RMSE and standard deviation difference measures of skill also disagree for SAT in the North Atlantic region. For RMSE, NoAssim is performing better than DePreSys, but the standard deviation of the error shows that the difference is again due to a mean bias.

- Further investigation showed that an error was made in the construction of ocean climatologies (i.e. that $C_{DeP} = C_{NoA} = C_{mod,A} = C_1$ not C_{mod}). Drifts in the HadCM3 control run caused a significant difference between C_1 and C_{mod}, which led to the creation of a bias in the NoAssim hindcast ocean anomalies. Thus, the analysis presented in Sm07 was not a correct comparison of skill between DePreSys and NoAssim. The improvement of DePreSys over NoAssim in the Southern Ocean was found to be largely an artifact of their analysis, and not due to the initialisation of low frequency variability as suggested in that study.

- The relative skill of DePreSys, when compared to NoAssim, is very sensitive to a wide range of choices that can be made in the analysis process. Changing the climatological period for the analysis can add a bias to the hindcasts, which will affect the RMSE of both systems. If the climatological period needs to be changed, then a bias correction can be calculated by computing the difference in the change of both the model and observed climatologies.

- In future it is important to ensure that new versions of both DePreSys and NoAssim use transient runs that are initialised from well spun-up states. Also, the DePreSys

climatology should be computed from the ensemble mean of the transient runs, to avoid a difference in climatological state between DePreSys and NoAssim.

To assess the impact of the error in the construction of the ocean climatology on the perceived skill of DePreSys, the analysis of Sm07 is repeated in chapter 4 but now using the NoAssim hindcast anomalies calculated using the **Sep Clim**. An extension of the analysis of Sm07 will also be presented that will examine the bias of DePreSys and NoAssim, and will also seek to further understand the origin of the improved skill.

Chapter 4

Analysis of mean skill and bias in DePreSys and NoAssim

4.1 Introduction

Chapter 3 showed that *Smith et al.* (2007) made an error in the construction of the model climatology that was used to create the hindcast anomalies for NoAssim. More specifically, the model climatology for the ocean was constructed from *only* the first transient run, and not the mean of all four, as originally thought. The remaining transient runs, and, therefore, the NoAssim hindcasts, were in a different climatological state due to the drift of the HadCM3 control run. The error led the NoAssim hindcast anomalies to be biased and will likely affect the mean skill scores that were presented in *Smith et al.* (2007). Thus, it is important to revisit the analysis of mean skill to understand more fully the improvement gained in the initialisation of DePreSys.

This chapter is organised as follows. The mean skill of DePreSys relative to NoAssim is presented first in section 4.3 after correcting for the climatology error. Biases in both the DePreSys and NoAssim hindcasts could degrade the skill of both systems. Therefore, the mean bias of DePreSys and NoAssim is examined in section 4.4, and the origin of the bias is discussed. Finally, the mean measures of skill, such as those presented in section 4.3 of this thesis, still do not explain why a given prediction system may be performing better or

worse than another. The question of whether the improved skill is just due to DePreSys bias correcting the initial temperature of the ocean is examined in section 4.5.

4.2 Significance testing

Now that the climatology problem has been solved, the significance of the differences in RMSE between DePreSys and NoAssim will now be tested. Testing the significance of the differences in RMSE will require the evaluation of individual independent significance tests at every grid point (i.e. multiple independent tests). However, because of spatial correlation in the fields of errors, the individual significance tests are not independent. The presence of spatial correlations could give a false impression of the actual significance of the whole field. Therefore, the field significance of the RMSE difference maps will also be calculated using the method of *Livezey and Chen* (1983).

At each grid point, a non-parametric Mann-Whitney test (also known as a Mann-Whitney Wilcoxon test, (*Mann and Whitney* (1947); *Wilcoxon* (1945))) is used to evaluate whether the difference in the mean squared errors of DePreSys and NoAssim are significant. Monte Carlo simulation is then used to calculate the area of the error field that is erroneously correlated with a random series, which represents the null hypothesis (i.e. there should be no correlation with the random series (*Wilks* (1995))).

For each iteration of the Monte Carlo simulation, a random series of errors is constructed, and is correlated with the series of errors at each grid point. If the errors at different grid points were independent, then the area of the field that is significantly correlated with the random series should be, on average, approximately equal to the significance level (i.e. for significance at the 5% level, the null hypothesis would be erroneously rejected over 5% of the area). However, if there is spatial correlation in the errors, the fraction of the total area that erroneously rejects the null hypothesis (i.e. that is significantly correlated) will likely be larger (*Livezey and Chen* (1983); *Wilks* (1995)).

To calculate whether the field of interest (i.e. the spatial field of the difference in RMSE) is its self significant (i.e. it has field significance), the area of the globe that has signifi-

cantly different errors (calculated using the independent tests) is compared to the distribution of the area that is correlated with the random time series. More specifically, if the area with significantly different errors is larger than the 95th percentile of the distribution found from the Monte Carlo simulation, then the field is said to have field significance. To ensure that the distribution of area that is significantly correlated with a random series is robust, 3000 Monte Carlo iterations are performed (3000 iterations is the same as that used in *Elmore et al.* (2006)). Finally, it must be noted that the field significance shown in this thesis is intended to show when there is significant differences between DePreSys and NoAssim, *not* where the improvement in DePreSys over NoAssim is significant.

4.3 Analysis of DePreSys Skill

To examine the effect of initialising the DePreSys hindcasts from the observed state of the climate, the RMSE of NoAssim (now calculated using a separate climatology for each NoAssim member (i.e. **Sep Clim**)) is compared to the RMSE calculated for the DePreSys hindcasts. For clarity, it is worth restating that the oceanic hindcast anomalies are calculated relative to 1951-2006 climatological period. Also, note that the Updated Met Office ocean analysis is used for the observational analysis[1], unlike figure 3.4 which uses the 1941-1996 climatology and the Original Met Office ocean analysis as the observations. So that the change in the climatological period does not affect the skill scores, a bias correction is applied, which was calculated as stated in section 3.3.3.

Figure 4.1 shows the RMSE as a function of lead time for SAT, 113m heat content and 500m heat content (panels a), b) and c) respectively). The difference in the RMSE is greatest for the 113m and 500m heat content. The figure is very similar to figure 3.4, suggesting that the bias caused by the change in the climatological period has been corrected. However, figure 4.1 highlights the increase in the NoAssim RMSE for the global mean values of 113m and 500m heat content, which suggests that DePreSys is actually performing better than NoAssim, in a global mean sense, than the analysis of *Smith*

[1]Note that the observed anomalies are made relative to a modified 1951-2006 climatology to remove the climatology error described in Appendix A.

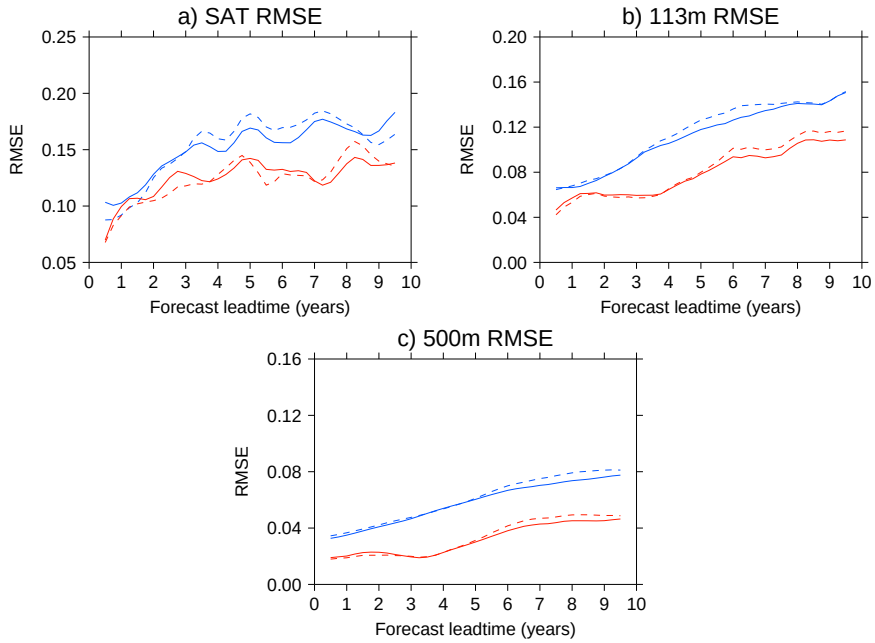


Figure 4.1: a) shows the comparison of RMSE as a function of lead time for SAT, calculated for ensemble mean DePreSys (D_t^{RMSE}; red) hindcast anomalies compared to ensemble mean NoAssim (N_t^{RMSE}; blue) hindcasts anomalies. The NoAssim hindcast anomalies are calculated using **Sep Clim**. b) and c) shows the same for the 113m and 500m heat content respectively. The solid lines show the RMSE calculated using all of the hindcasts in the 1982-2001 period, the dashed lines show the RMSE calculated when the hindcasts that are initialised with volcanic aerosol are omitted.

et al. (2007) suggested. The greater skill for the heat content compared to SAT is due to the higher predictability of the ocean, compared to the noisier atmosphere (*Griffies and Bryan* (1997b)).

The hindcasts that include the volcanic aerosol when initialised could perform differently to the rest of the hindcasts, especially if HadCM3 has an unrealistic response to the volcanic forcing. The dashed lines in figure 4.1 show hindcasts initialised with volcanic aerosol have a small effect on the mean skill. There is no clear indication as to whether including the volcanic aerosol improves or degrades the mean performance of the system as there are reductions and increases in the RMSE for different variables and lead times.

Figure 4.2 a) - c) shows the difference between DePreSys and NoAssim RMSE for 9-year mean anomalies of SAT, 113m heat content and 500m heat content, when the climatology error is *not* corrected for, as in the analysis of *Smith et al.* (2007) (i.e. NoAssim anomalies are calculated using **Trans1 clim** for the ocean and **Mean clim** for SAT). Figure 4.2 a)

and b) compare very well with figure 3.5, which gives us further confidence that the bias created by changing the climatological period has been corrected. The difference in RMSE for 9-year mean anomalies of 500m heat content (fig. 4.2 c)) is very similar to that of the 113m heat content (fig. 4.2 b)), and shows improved skill of DePreSys in the Southern Ocean, and North Atlantic.

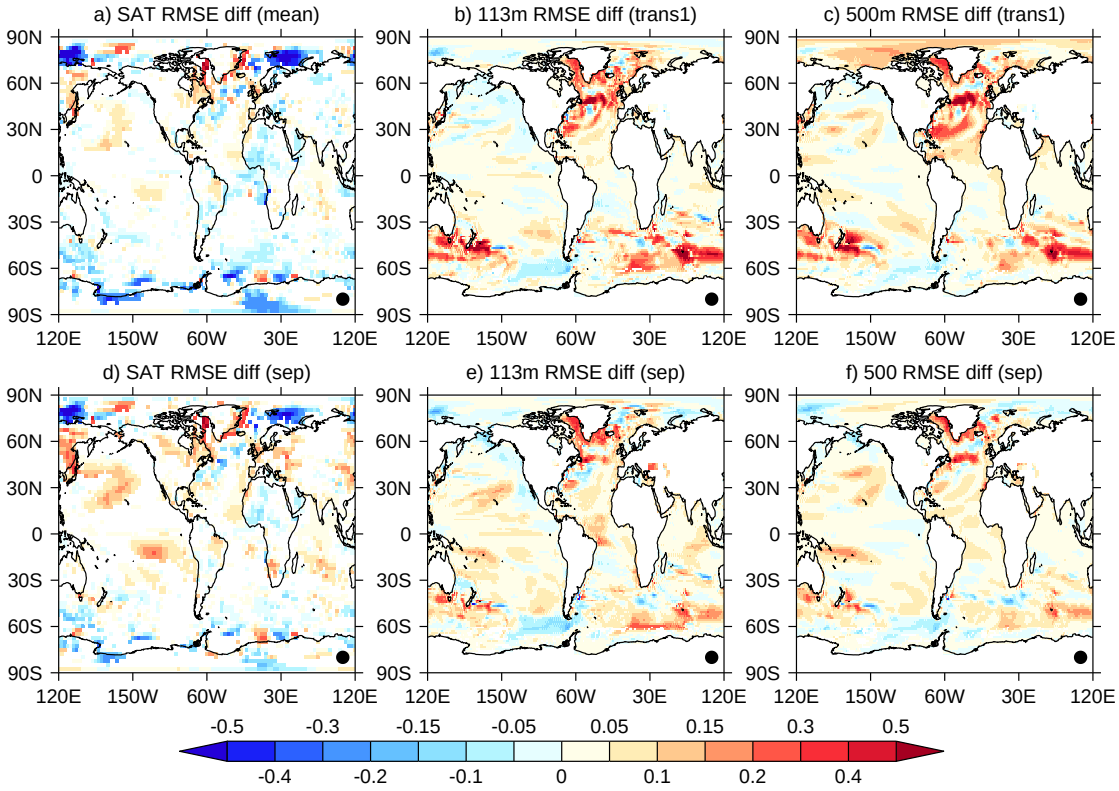


Figure 4.2: a) shows 9-year mean NoAssim - DePreSys RMSE ($N_{xy}^{RMSE} - D_{xy}^{RMSE}$) for SAT using the **Mean Clim** to calculate the NoAssim hindcast anomalies. b) and c) shows the same for 113m and 500m heat content when using the **Trans1 clim** to calculate the NoAssim anomalies. d-f shows the same as a-c but now using the **Sep clim** for the NoAssim hindcast anomalies. Only the grid boxes with a significant difference at the 5% level using independent tests are coloured. A black dot in the bottom right corner of a panel shows that that panel has field significance (see section 4.2 for details). The apparent improvement of DePreSys in the Southern ocean is mostly an artifact of the climatology error.

When correcting the NoAssim hindcast anomalies for the climatology error, the patterns of RMSE do change significantly (figure 4.2 d) to f)). Interestingly, when using **Sep Clim**, an improvement in SAT over the land in the Northern Hemisphere becomes more apparent for DePreSys. In figure 4.2 d) it can be seen that DePreSys gives improved predictions of SAT over Western Europe and much of North America when **Sep Clim** is

used. The regions where DePreSys improves SAT predictions is now in better agreement with the potential predictability studies of *Boer* (2004) and *Boer and Lambert* (2008), and could be due to improved predictions of North Atlantic SST, which may give predictive skill for SAT in Europe and North America (*Sutton and Hodson* (2005)). There are also improvements in 9-year mean anomalies of SAT in the North and tropical East Pacific.

For the ocean variables, the largest change in the RMSE between NoAssim and DePreSys is in the Southern Ocean. The relatively large improvement of DePreSys compared to NoAssim in the Southern Ocean south of India and Australia, seen in Figure 4.2 b) and c), is much smaller when using **Sep Clim** (see fig. 4.2 e) and f)). The result suggests that much of the improvement of DePreSys in the Southern Ocean presented in *Smith et al.* (2007) (see figure 1.7) was not due to improved prediction here, but was mostly an artifact of the analysis. More specifically, it was due to the error in the climatology used to create the NoAssim anomalies. There is also a reduction in the RMSE improvement in the North Atlantic, another area that was affected by the climatology difference seen in figure 3.7. Using the **Sep Clim** has highlighted other areas of the ocean where DePreSys seems to be performing better than NoAssim, especially in the Tropical Atlantic and Pacific.

Changing the climatology used to create the NoAssim hindcast anomalies has little effect on the standard deviation of hindcast errors. This is because standard deviation measure already removes the mean bias caused by the climatology difference (C_{diff}). Therefore, only the effect on RMSE score will be shown for the remainder of the thesis.

The improved predictability is not removed entirely from the Southern Ocean, suggesting that DePreSys may actually be performing better than NoAssim here. A note of caution is encouraged however. As already noted, the standard deviation of the errors (figure 3.5 f) does not show an improvement of DePreSys relative to NoAssim in the Southern Ocean. As mentioned before, the mean difference between DePreSys and NoAssim hindcasts will be $\sim C_{diff}$ because the climatological period and the hindcast period are not the same. DePreSys is initialised close to the background state C_{dep} because of the assimilation. However, the transient runs are free to evolve, and will continue to drift. The mean difference between NoAssim and DePreSys could, therefore, be different to C_{diff}.

The improved predictability of DePreSys, when compared to NoAssim in the Southern Ocean, could still be due to the drift of the transient runs which gives a mean bias to NoAssim, but not DePreSys. Hence, it is vital that for future iterations of DePreSys, the transient runs used in the creation of the system, and the NoAssim hindcasts, should be initialised from a well spun up and consistent state.

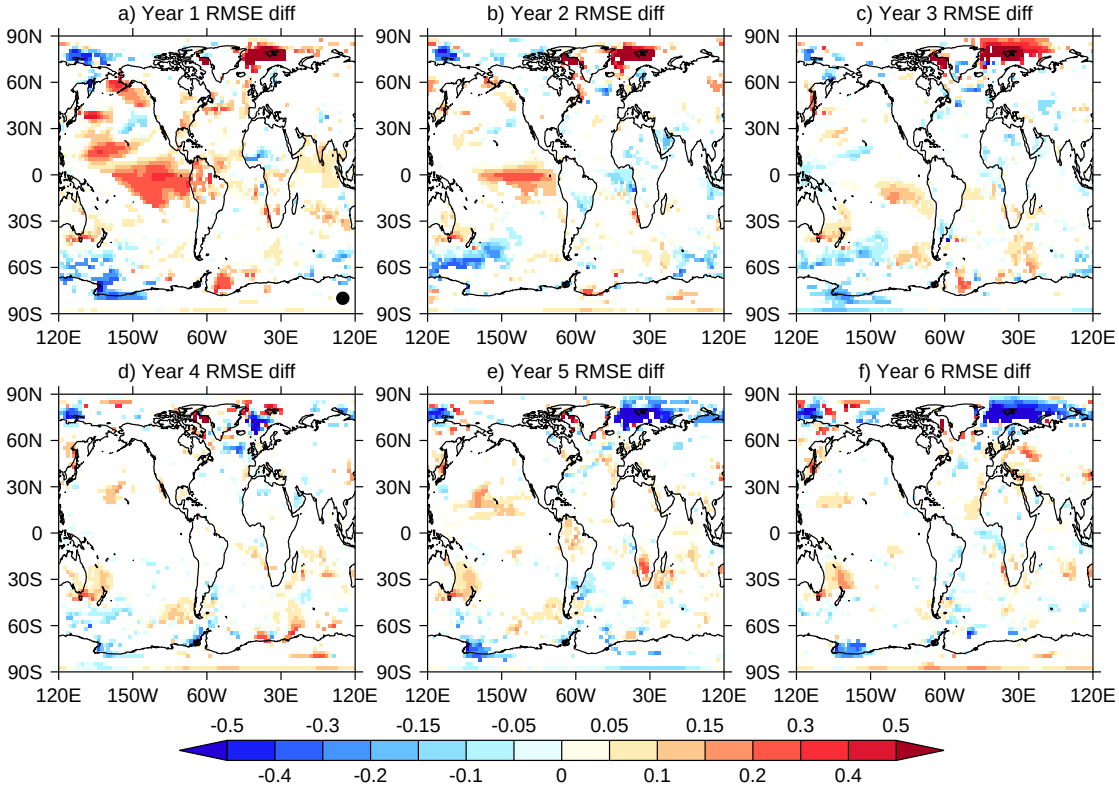


Figure 4.3: NoAssim - DePreSys RMSE ($N_{xy}^{RMSE} - D_{xy}^{RMSE}$) for SAT calculated separately for the first 6 years of the hindcasts. Each year of the hindcast is shown from year one (a) to year 6 (f). Only the grid boxes with a significant difference at the 5% level using independent tests are coloured. A black dot in the lower right hand corner shows that the field has significance at the 5% level. Note that the NoAssim hindcast anomalies are made using the **Sep Clim**.

The 9-year mean RMSE plots are useful for beginning to see some the differences between DePreSys and NoAssim, but they only show long timescale differences. The largest difference in spatial RMSE between DePreSys and NoAssim should be in the first few years where DePreSys may be able to predict some of the changes dynamically. Figure 4.3 a) shows that DePreSys does indeed improve the prediction of SAT over a large part of the globe for the first year of the hindcasts. The largest improvement is seen in the Tropical Central and Eastern Pacific and is likely to be associated with the

prediction of ENSO, which is predictable at this lead time (*Latif et al.* (1998)). However, it is only the first year of the SAT predictions where the difference in RMSE has field significance. The second year (see fig. 4.3 b)) shows a reduction in this improvement of skill, but some skill remains in the Tropical Pacific, which may result from being able to predict the transitions between El Niño and La Nina. Although this analysis uses a different metric of skill to *Haines et al.* (2009), who show the difference between standard deviations of the error, the results for the first two years agree well. After the first two years the improvement in skill for DePreSys is patchy, which is to be expected for SAT, as the atmosphere has low potential predictability (*Griffies and Bryan* (1997b); *Boer and Lambert* (2008)). There is an intriguing signal in the GIN sea region, which shows that, initially, DePreSys has a much lower RMSE compared to NoAssim. However, after year 4, the difference is reversed and DePreSys has greater RMSE than NoAssim. DePreSys also seems to perform worse than NoAssim at predicting SAT over much of the Southern Ocean and areas of the Arctic.

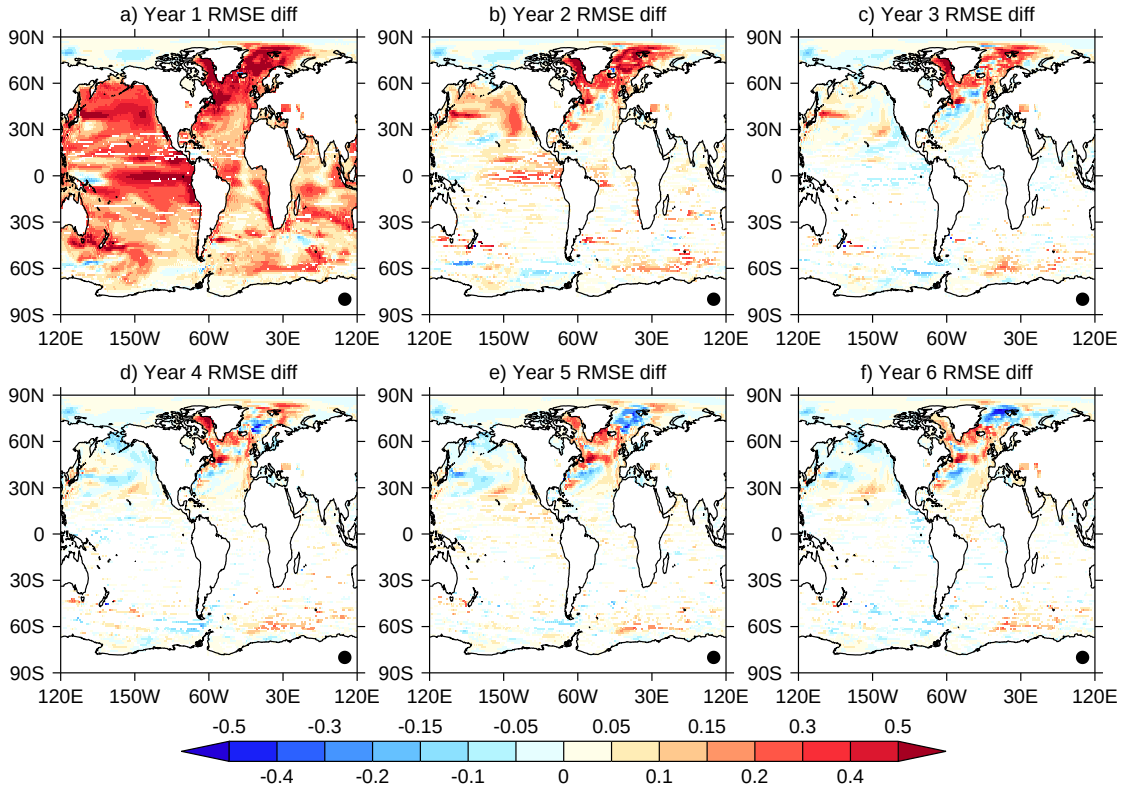


Figure 4.4: As for figure 4.3 but now for 113m heat content

Much of the improvement in the predictions of SAT seen in figure 4.3 a) and b) are co-

incident with the largest improvements in the predictability of the top 113m heat content of the ocean (see figure 4.4 a and b). Although most of the world's oceans show an improvement in the predictions of heat content anomalies for the first year, the largest improvements for DePreSys are seen in the Tropical and North Pacific and the North Atlantic, especially the subpolar Atlantic and the GIN seas. In the second year the improvement reduces quickly, with the largest improvement present in the subpolar North Atlantic and hints of improvement in the Tropical and North Pacific. After the third year there is little evidence of a systematic improvement in skill of DePreSys relative to NoAssim, except perhaps in the North Atlantic, but it is important to note that all 6 years have field significance.

Haines et al. (2009) argue that their results show that the improvement in the predictions of 9-year mean anomalies is dominated by the improvement in skill of the first year. Although it is clear that the improvement in skill for DePreSys hindcasts is larger for the first few years, the sum of the skill improvement for all the years calculated separately is not comparable with skill of predicting 9-year mean anomalies. Imagine an observed 9-year sinusoidal oscillation of temperature in the North Atlantic. For a series of annual mean RMSE, a prediction would have to also follow the sinusoidal oscillation for a skillful forecast. However, a skillful prediction of the 9-year mean anomalies would just need the anomalies to sum to zero, regardless of their actual trajectory. Thus, care must be used when validating 9-year mean anomalies. It seems important to validate the evolution of the hindcasts in more detail, beyond the use of 9-year means, to confirm that the initialised hindcasts are delivering improved predictions.

4.4 Mean bias of DePreSys and NoAssim

DePreSys and NoAssim are not perfect prediction systems. The model used to make the predictions is not perfect. *Smith et al.* (2007) showed that both DePreSys and NoAssim develop a bias in the global mean SAT, but what is the bias for other variables, such as ocean heat content? Examining the patterns of bias in initialised climate predictions, and their development, may be useful in highlighting problems with the model that lead to

the mean errors, and for understanding the mean skill. Therefore, this section examines the time varying mean bias of DePreSys and NoAssim, and briefly discusses its origins.

4.4.1 Mean hindcast bias

The evolution of the global mean biases of DePreSys and NoAssim (calculated using equation 3.8) is shown in figure 4.5. The NoAssim SAT bias starts below zero, indicating that the hindcast anomalies are on average smaller than the observed anomalies, but increases over the course of the hindcast period (see figure 4.5 a)). The increase is more rapid in the first few years, increasing to 0.1°C by the third year, and then appears to increase more slowly. The DePreSys bias also increases over the course of the hindcasts. Interestingly however, the DePreSys bias has a relatively large and rapid cooling of $\sim$0.05°C over the first couple of seasons. Following the rapid cooling the increase in DePreSys SAT bias is slower than NoAssim over the first few years, but the rate of increase increases towards the end of the hindcasts.

A similar increase also appears in the 113m and 500m heat content bias for both DePreSys and NoAssim (see figure 4.5 b) and c) respectively). Both DePreSys and NoAssim heat content biases increase over the course of the hindcasts, but now the NoAssim hindcasts start with a much larger initial bias for the ocean. The 113m heat content for NoAssim again increases at a slightly faster rate in the first two years compared to the later increases. DePreSys 113m heat content bias again seems to increase at a slightly slower rate than the NoAssim bias initially, but the rate of increase accelerates towards the end of the hindcasts. The initial slow increase in the DePreSys bias compared to NoAssim is best seen for the 500m heat content bias (fig 4.5 c)).

The omission of future volcanoes would lead us to expect that NoAssim and DePreSys hindcasts would warm up too rapidly. However, Figure S6 in *Smith et al.* (2007)'s supporting online materials shows that the mean bias of SAT for hindcasts that do not include future volcanic eruptions, also increases from $\sim$-0.01°C to $\sim$ 0.15°C. The similarity between *Smith et al.* (2007)'s figure S6 and figure 4.5 a) here suggests that the lack of future volcanic eruptions is not the sole cause of the enhanced warming of the model compared

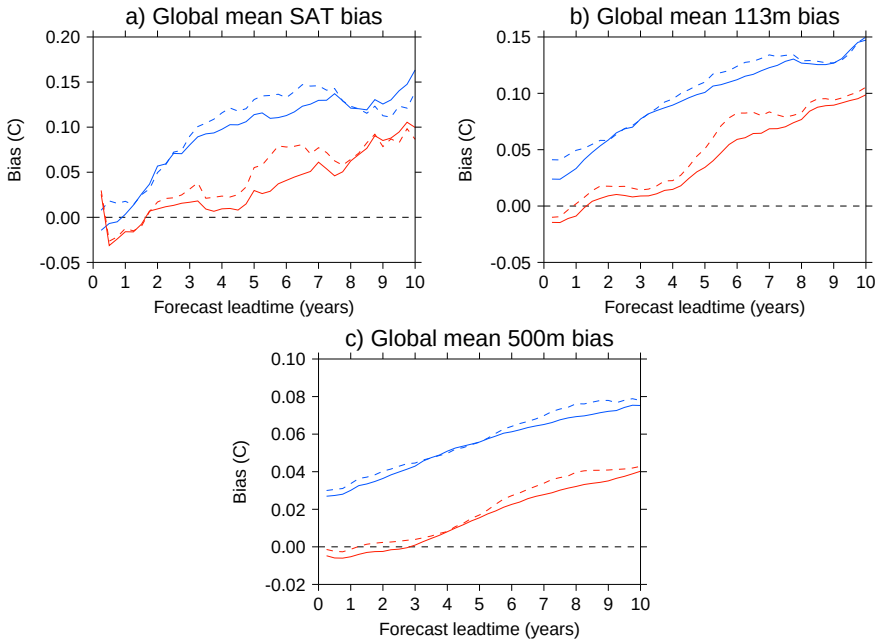


Figure 4.5: a) shows the mean bias calculated for NoAssim (blue) and DePreSys (red) for global mean SAT. b) and c) show the same but now for global mean 113m heat content and 500m global mean heat content. The solid lines show the bias calculated using all the hindcasts, dashed lines show the bias when the hindcasts containing volcanic aerosol were omitted. Note that all NoAssim hindcast anomalies were made relative to **Sep Clim**.

to the observations.

In theory the initial bias of DePreSys should be zero, and so the bias of the first season should, in theory, be extremely close to zero. What is the cause of the initial cold heat content offset for the DePreSys hindcasts? A possible cause could be the inclusion of volcanic aerosol in some of the hindcasts, which will act to cool the climate. If the response of HadCM3 to El Chichón and Pinatubo eruptions was too strong it could create a negative first season bias. Removal of the hindcasts that include volcanic aerosol from El Chichón and Pinatubo does show that the initial cold offset for DePreSys heat content is partially caused by these hindcasts (see figure 4.5 b) and c)). The NoAssim bias calculated when the hindcasts including volcanic aerosol are omitted is also warmer, as expected.

Interestingly, the large decrease in the DePreSys SAT bias (fig. 4.5 a)) does not change when the volcanoes are omitted. The large decrease must have an explanation other than volcanic aerosol. Examining the spatial structure of the first season bias in DePreSys SAT

shows that there is a large warming over the South American and African continents, as well as a warming over the South Tropical Atlantic (not shown). It is possible that this warming and cooling of global mean SAT in DePreSys is due to an imbalance (or initial shock) related to the mismatch of Oceanic and Atmospheric climatologies (i.e. ocean climatology is C_1 and the atmospheric climatology is C_{mod}). However, further analysis would be needed to understand the origin of the change.

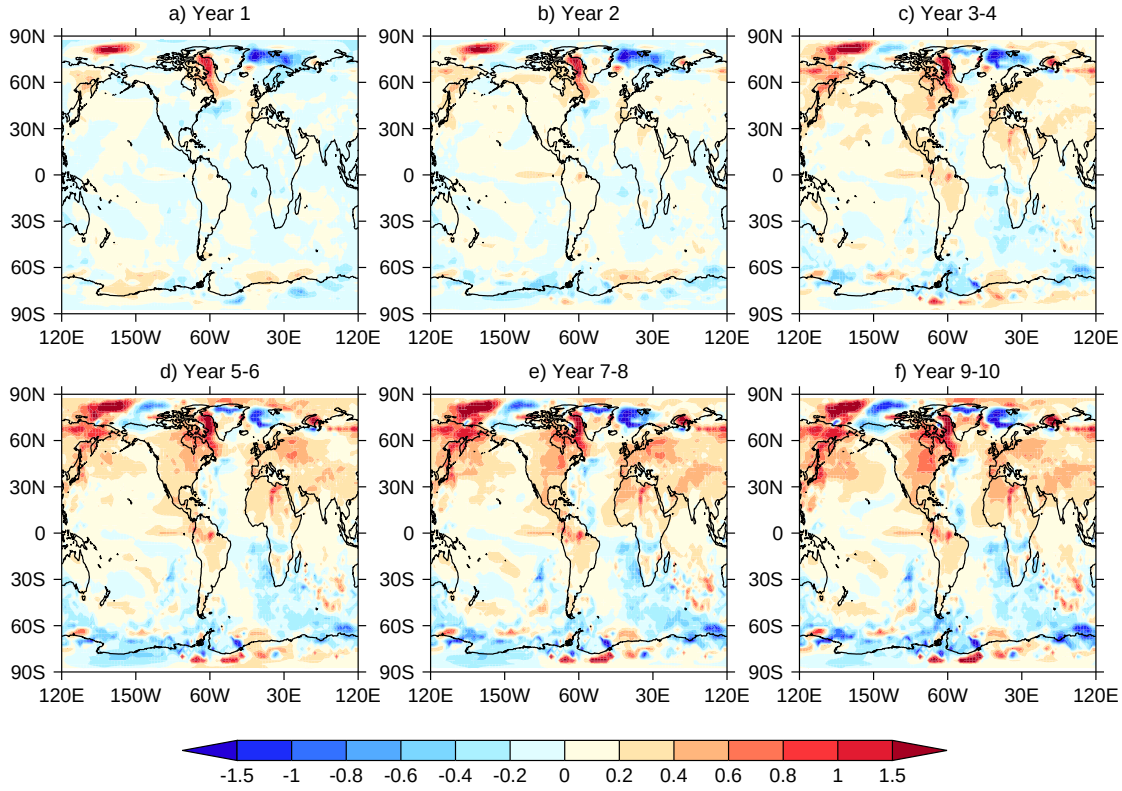


Figure 4.6: The mean NoAssim SAT bias calculated using all NoAssim hindcasts (anomalies calculated using **Sep Clim**) as it develops with hindcast lead time

Figure 4.6 shows that the increase in NoAssim SAT bias is not spatially uniform, and varies with lead time. Initially the bias is relatively small over most of the globe, but there are large biases in the polar regions of the Northern Hemisphere (fig. 4.6 a)). The biases in the polar regions of the North Atlantic are somewhat stationary with lead time. Throughout the course of the hindcasts it is clear that a warm bias develops over much of the continental land masses, especially in the Northern Hemisphere. In years 9 and 10 of the hindcasts there is a mean bias of over 0.4°C over most of Europe and eastern North America. The entire Asian continent also has a warm bias in excess of 0.2°C. There is

also an increase in negative bias in the Southern Hemisphere south of 60°S. The Southern ocean is likely to be a region which is poorly represented in the ERA-40 reanalysis due to a lack of observations. Hence, signals in bias here are unlikely to be as reliable as changes in the Northern Hemisphere where the observational density is higher.

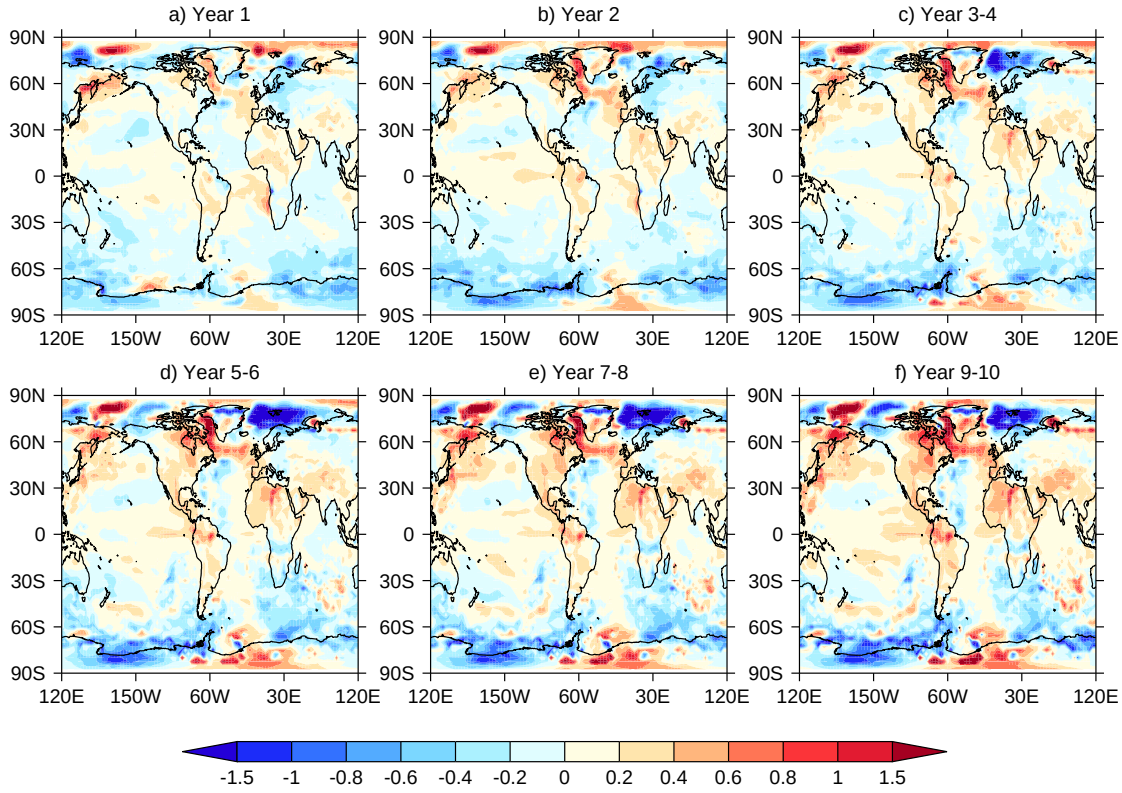


Figure 4.7: As for figure 4.6 but for DePreSys SAT

Figure 4.7 shows the spatial pattern of the mean SAT bias for DePreSys. The first year SAT bias from DePreSys is cooler than that for NoAssim in the Southern Ocean and Eastern Europe. As the hindcasts progress many of the biases develop in a similar way to the NoAssim bias, with an increase in a positive bias over the North and South American continents and Northern Africa. The increase in positive bias over Europe and Asia in NoAssim seen in figure 4.6, is not seen in the DePreSys mean bias, although a positive bias does develop over the South Asia. A large difference between the SAT biases in NoAssim and DePreSys is the negative bias in the GIN sea region which, is seen to develop over the course of the DePreSys hindcasts, and peaks at over -1.5°C. The cold bias over the GIN Sea was later found to be model drift. The reduced increase in DePreSys's positive bias over Western Europe was, therefore, not only due to improved predictability

here, but instead was likely to be caused largely by the drift.

Figure 4.8 shows the evolution of 113m heat content bias for the NoAssim hindcasts as a function of lead time. Initially there are relatively large positive biases in the North Atlantic subpolar gyre region and the North Pacific of approximately $\sim$0.5°C. There are also smaller warmer anomalies in the Western Tropical Pacific and in the Southern and Indian Oceans. There are also cool biases in the Eastern Pacific and in the subtropical North and South Atlantic. As the hindcasts progress, the warm bias generally increases over much of the globe, especially in the North Pacific and the Southern Ocean. The bias decreases in the North Atlantic, especially in the subtropical gyre region.

Figure 4.9 shows the evolution of 113m heat content bias for the DePreSys hindcasts. The first year mean bias of DePreSys (Fig. 4.9 a)) is smaller than that seen in NoAssim (Fig. 4.8 a)) as expected, but shares a similar structure. For example DePreSys has a first year warm bias in the North and west tropical Pacific, and a slight warm anomaly in the North Atlantic subpolar gyre. As the DePreSys hindcasts progress, the positive bias increases in the North Pacific and Southern ocean, and by year 9-10 (Fig. 4.9 f)) is extremely similar to the 9-10 year NoAssim bias (Fig. 4.8 f)). However, there are notable differences between the two systems' biases at later lead times. A large cold bias, greater than -0.7°C develops in the GIN seas in DePreSys, which peaks in magnitude in year 5-6 (Fig. 4.9 d)).

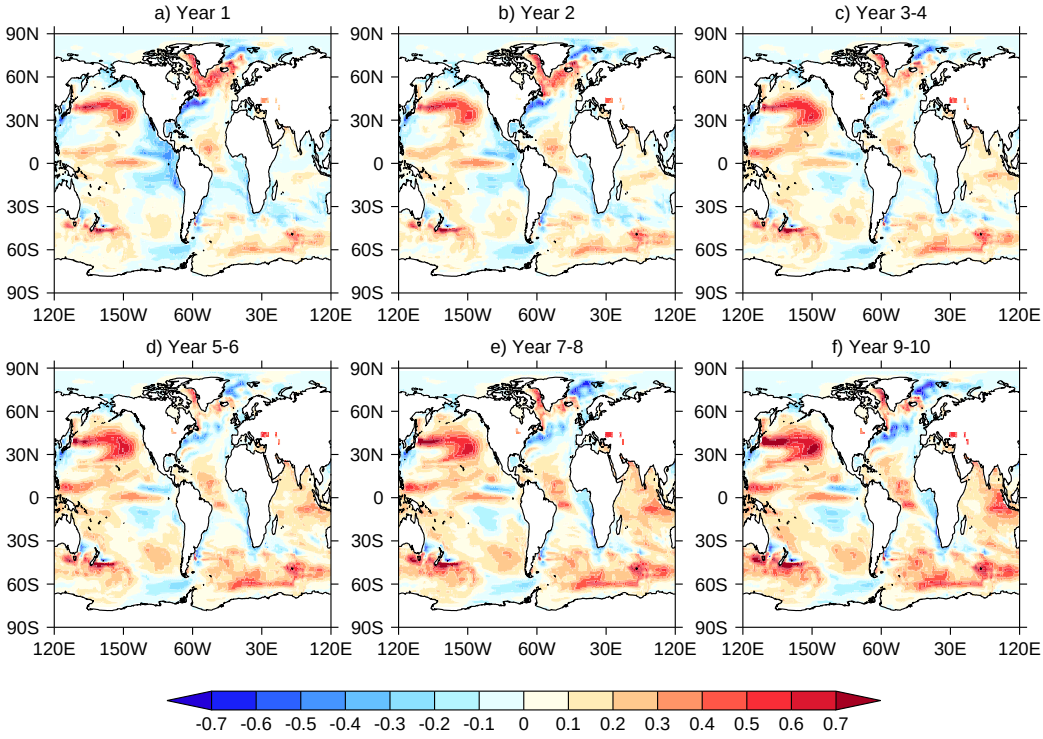


Figure 4.8: As for figure 4.6 but for NoAssim 113m average temperature anomaly

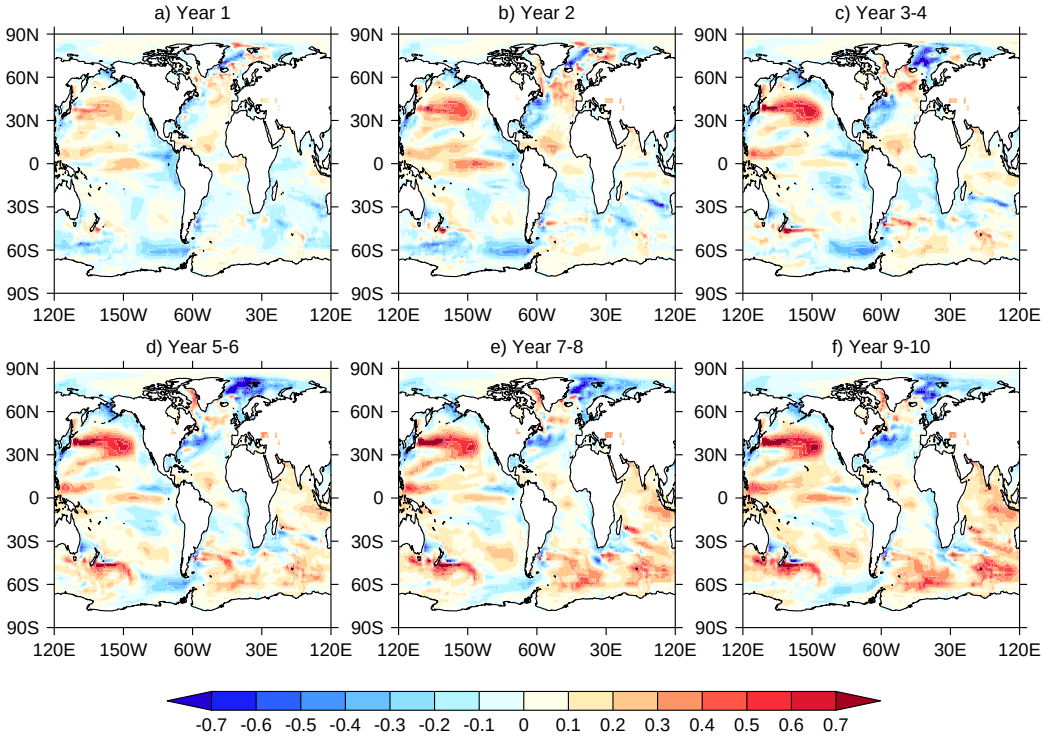


Figure 4.9: As for figure 4.6 but for DePreSys 113m average temperature anomaly

4.4.2 Understanding bias and drift

The bias that is seen in the DePreSys and NoAssim systems is likely to be attributable to a range of factors. Trying to understand the contribution of every cause is likely to prove difficult and time consuming, and so no attempt has been made to quantify the contributions in this thesis. However, it is intended that the evolution of the mean bias will be investigated further in the future. Thus, a short discussion of the likely causes will be undertaken in this section.

Firstly, HadCM3 is not a perfect model. Although the bias attributed to the mean model error has been removed by using anomalies, if the model is unable to sufficiently recreate the spatial pattern and magnitude of the observed variability, then biases may arise. For example, *Collins et al.* (2001) showed that the spatial pattern in the Tropical Pacific associated with HadCM3's El Niño, spreads too far west into the West Pacific warm pool, and hence the positive anomaly seen in the warm pool region in figures 4.8 and 4.9, may be caused by model error. In a non-linear system it is, therefore, possible that a bias may result from a model's inability to represent the variability of the climate system.

If the model error were the only source of mean bias for NoAssim then, with good sampling (e.g. a long hindcast period covering a large variety of model and observed states), then the bias caused by model error should be a stationary pattern (i.e. it should not vary with lead time). For DePreSys, the mean bias should be zero when it is initialised. As the DePreSys ensemble diverges to the climatological variance, then the mean bias of DePreSys should increase so that it is equal to the mean bias of NoAssim. In this case, the amplitude of DePreSys would increase to the NoAssim bias, but would then remain stationary. It is clear that from section 4.4.1 that the pattern of bias is not stationary in DePreSys and NoAssim suggesting, that there must be another source of bias.

Bias is also likely to arise due to the presence of low frequency variability in the observations, and a relatively short hindcast period. Imagine that NoAssim was a perfect prediction system, with a perfect representation of the observed variability and infinite ensemble members. In a stationary climate the ensemble mean hindcast anomalies of "perfect" NoAssim would be zero at all lead times (see figure 4.10). However, there is

only one realisation of the real world that exhibits low frequency variability. The observed anomalies will not be zero at all hindcast lead times. Therefore, the hindcast error for a "perfect" NoAssim ensemble at each lead time would be equal to the negative of the observed anomaly because:

$$\langle N_{it}^{error} \rangle = \langle N_{it}' \rangle - X_i'$$

but $\langle N_{it}^{anom} \rangle = 0$ in a perfect NoAssim prediction system, so

$$\langle N_{it}^{error} \rangle = 0 - X_i'$$

$$\langle N_{it}^{error} \rangle = -X_i'$$

Thus, it is possible to isolate the bias caused by the variability in the observations over the hindcast period if we assume that NoAssim is a perfect prediction system. This contribution to the first season bias from the observed variability (B_t^{obsvar}) for the hindcast period used for NoAssim, which is 1982-2001, is, therefore, equal to

$$B_1^{obsvar} = \frac{1}{80} \sum_{i=JJA1982}^{MAM2002} (-X_i')$$

and the nth season bias would be equal to

$$B_t^{obsvar} = \frac{1}{80} \sum_{i=i_1+n}^{i_2+n} (-X_{i+n}') \tag{4.1}$$

However, NoAssim is not trying to predict a stationary climate. Perfect NoAssim will be predicting the global warming trend of the real climate system. The ensemble mean anomaly for the first season in "perfect" NoAssim will be zero, relative to the warming trend. Therefore, the warming trend in the observations will also contribute to the calculation of B_t^{obsvar}. To ensure that the calculated B_t^{obsvar} represents the variability of the ocean, and not the global warming trend, a trend is removed from the observed anomalies. More specifically a linear global mean warming trend, calculated over the period from 1951-2007, was removed from each grid point in the same way that it was removed for the analysis of the rapid warming of the North Atlantic in chapter 2.

Figure 4.11 shows the B_t^{obsvar} contribution for 113m heat content calculated according to equation 4.1. The spatial pattern of the bias contribution from the observational variability in year one (fig. 4.11 a)) is remarkably similar to the first year bias of NoAssim

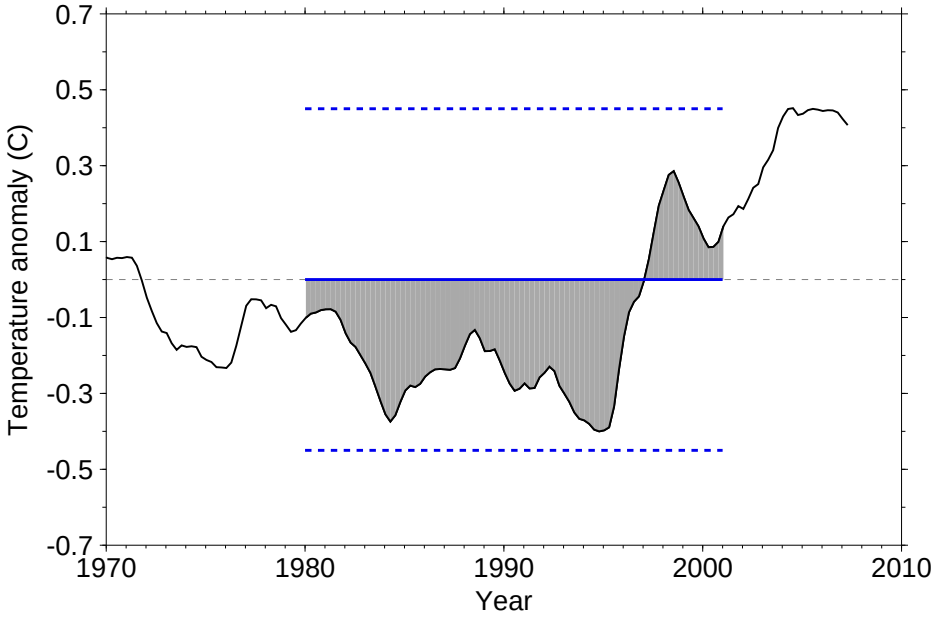


Figure 4.10: Schematic highlighting the origin of bias created by the low frequency natural variability of the real world (B^{obsvar}). The black line shows the linearly detrended rolling annual average observed top 500m heat content for the North Atlantic subpolar gyre. The blue lines represent the mean, and min and max bounds, of a perfect NoAssim system, which simulates the mean response to global warming perfectly and has infinite ensemble members and, therefore, has a zero ensemble mean hindcast anomaly for all lead times. The solid blue line is the ensemble mean first season prediction of 500m heat content from perfect NoAssim hindcasts over the hindcast period 1982-2001, the dashed lines represent the ensemble spread. The shaded grey area highlights the first season hindcast error for all perfect NoAssim hindcasts, which would be averaged to calculate the mean bias. The first season hindcast error clearly does not add up to zero if summed over the whole hindcast period. The first season bias of perfect NoAssim is therefore dependent upon the inverse sum of the observed variability ($-X_i'$).

(fig. 4.8 a)), but the magnitude of the bias is not the same. The reduction in the magnitude was not due to the removal of the global mean trend. When the global mean trend is not removed the magnitude of the warm bias in the Pacific is reduced (not shown). NoAssim is not a perfect prediction system, so it should not be surprising that not all of the bias originates from the observed variability. The bias in fig. 4.11 does change with hindcast lead time, but generally does not evolve in the same way as the mean bias of NoAssim develops with hindcast lead time. In the North Atlantic subpolar gyre, B_t^{obsvar} does evolve similarly to NoAssim's mean bias. B_t^{obsvar} is positive initially, but the bias reduces over the first 6 years of the hindcasts. The reduction in bias in the North Atlantic subpolar gyre suggests that an important contribution to the reduction in NoAssim mean

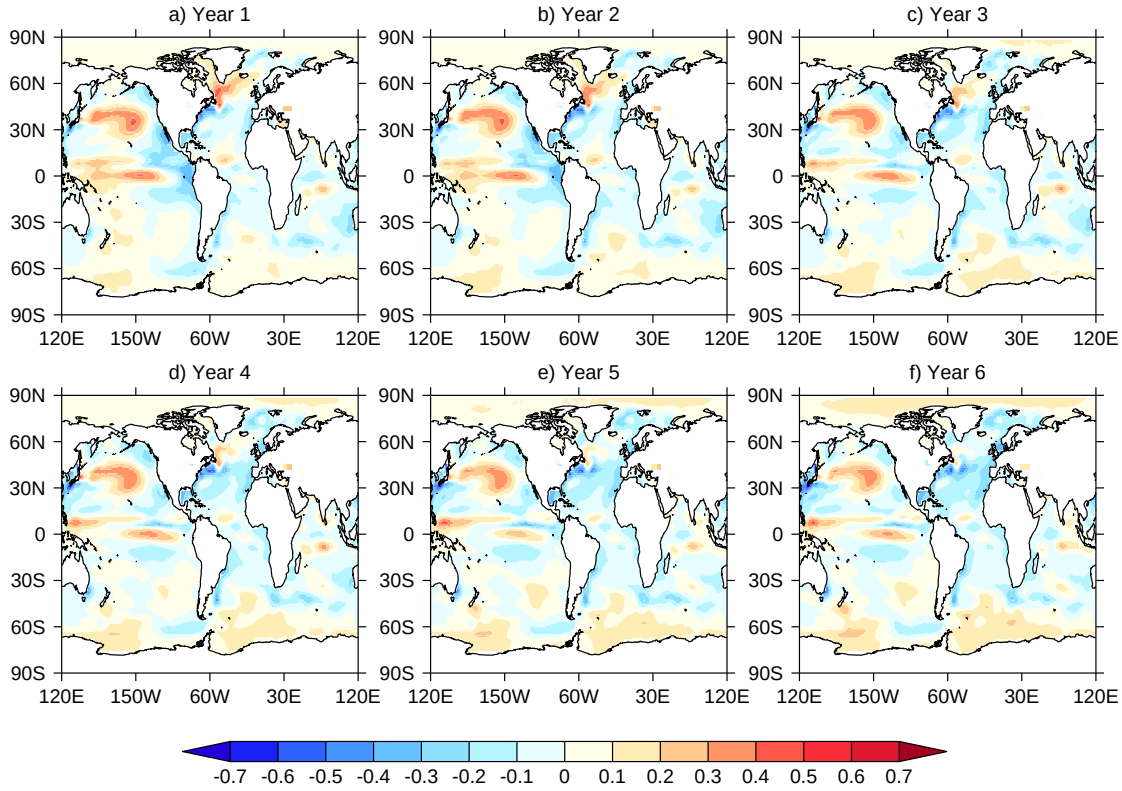


Figure 4.11: The bias caused by the observed variability (B_t^{obsvar}) for the 113m average temperature anomalies calculated from the updated Met Office ocean analysis. A linear global mean warming trend has been removed from every grid point to remove the contribution of the bias that may be caused by errors in the rate at which the NoAssim hindcasts warm due to radiative forcing.

bias here is due to observed variability. However, the warm bias in the North Pacific and Southern Ocean's seen in figure 4.11 do not warm like in figure 4.8, suggesting there is another cause for the increase in bias here.

The contribution to the bias caused by the natural variability of the observations (i.e. B_t^{obsvar}) will cause the mean NoAssim bias to be different for different hindcast periods. B_t^{obsvar} *should therefore be considered when analysing the mean bias and skill of a decadal prediction system.* One possible way to minimise B_t^{obsvar} could be to make the climatological period and the hindcast period equal if using an anomaly assimilation approach. B_1^{obsvar} (i.e. B_t^{obsvar} for the first season) should equal zero if the hindcast period is the same as the climatological period, as the sum of the observed temperature anomalies will go to zero.

Errors in the radiative forcing, or the model's response to the radiative forcing, e.g. the model's climate sensitivity and ocean heat uptake, are also likely to contribute to biases seen in both DePreSys and NoAssim. For example, the upper ocean is warming faster than the observed ocean throughout much of the Pacific, especially the North Pacific, and the Southern Ocean. If the model is unable to subduct as much heat to the deep ocean as the real world, then the upper ocean would warm faster than the observed ocean, contributing to a warm bias.

Finally, the drift of the transient runs, and in particular transient 1, would cause the bias to change with hindcast lead time for NoAssim. The DePreSys assimilation run used a climatology from a drifting transient run and so that drift is also likely to be included in the DePreSys hindcasts. Therefore, the drift of the transient runs will also contribute to the bias for DePreSys. The evolution of the drift in DePreSys will be examined further in chapter 5.

The mean bias of the DePreSys and NoAssim hindcasts could be a rich source of information that remains relatively unused. The contribution here is only meant to be a start, and more work in this area is encouraged. The validation of the hindcasts against the observations allows the exploration of model deficiencies at the process level. Model errors remain the largest uncertainty in projections of the future climate (*Hawkins and Sutton* (2009a)) and significant investment is required in reducing the model error. Understanding the evolution of hindcast errors or bias, in DePreSys and NoAssim, could allow the exploration of the deficiencies in both the construction of DePreSys and of the underlying model, potentially leading to a reduction in the uncertainty due to natural variability and model error.

4.5 Understanding the source of improved skill

Section 4.3 showed that DePreSys is improving predictions of SAT and upper ocean heat content, especially in the in the first couple of years. The assimilation of observations is the main difference between DePreSys and NoAssim, which led *Smith et al.* (2007) to reason that the initialisation of low frequency variability is the cause of the improve-

ments. However, Section 4.4 showed that both NoAssim and DePreSys posses a time dependent mean bias. Crucially however, DePreSys has a smaller mean bias in comparison to NoAssim for this particular hindcast period. In general the assimilation of information into DePreSys could improve predictions for two general reasons:

1. **Bias correction** - The assimilation of information into DePreSys effectively removes the initial bias of NoAssim and there is no improved predictability of the variability.

2. **Prediction** - Assimilating the observed information allows DePreSys to predict the variability.

Thus, is the improved skill of DePreSys largely due to the correction of initial bias, rather than the prediction of variability? Knowledge of the relative importance of both sources of improved skill will be important for understanding the performance, and the benefit, of initialised decadal predictions. For example, it is conceivable that bias of NoAssim and DePreSys will be different for a different hindcast period. The analysis of skill for hindcast initialised from a different hindcast periods, or in fact forecasts, could show very different results for the mean skill *only* because of effect of bias on the skill scores.

Examining the NoAssim hindcasts, figure 4.12 a) shows a clear initial bias in global mean 500m heat content. The ensemble mean of the NoAssim hindcasts starts warmer than the observed anomalies because of a cooling of the observed heat content in the early 80s. The NoAssim hindcasts then continue to warm, and stay warmer than the observations for the entire length of the hindcasts. In contrast DePreSys hindcasts are initialised close to the observations due to the assimilation, and although the DePreSys hindcasts seems to be warming at a similar rate to NoAssim, the hindcasts lie much closer to the observations. The lower RMSE of DePreSys compared to NoAssim is, at least in part, because of the smaller *initial* bias (see figure 4.12 b).

Separating the contributions to the improved skill will likely prove to be extremely challenging; there is no obvious way to separate the sources. Therefore, this section presents a brief analysis of two methods that attempt to separate the contributions. The first method attempts to remove the initial bias of each individual NoAssim hindcast, and is presented

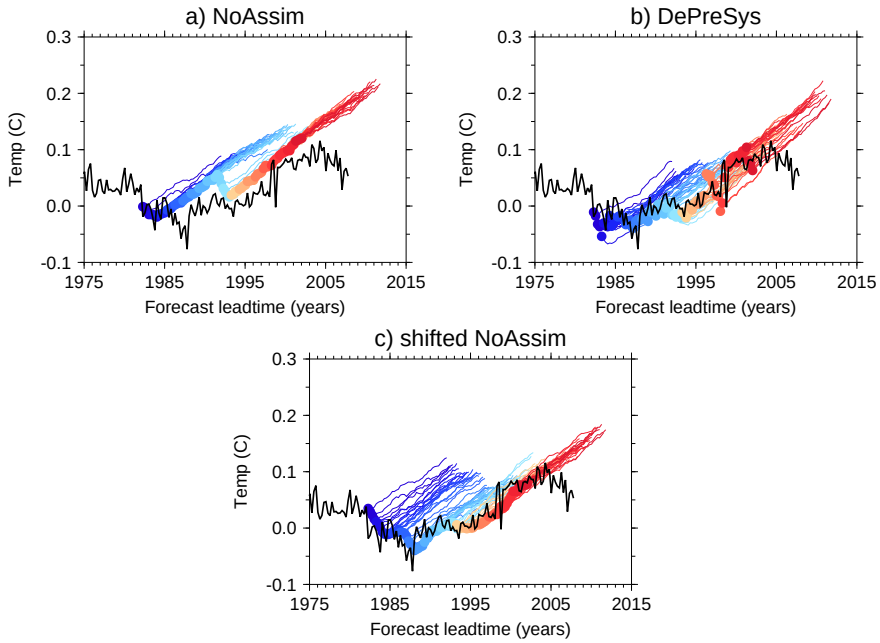


Figure 4.12: a) shows the observed global mean top 500m average temperature anomaly in black and overlaid in colour are the ensemble mean NoAssim hindcasts. The dot shows the first seasonal mean and the change in colour is to aid the eye. b) and c) are as a) but shows the DePreSys and shifted NoAssim respectively. Note that the NoAssim hindcast anomalies were made relative to **Sep Clim**.

in section 4.5.1. The second method removes the mean bias from both NoAssim and DePreSys, and is presented in section 4.5.2.

4.5.1 Shifted NoAssim

A simple way to correct the bias present in the NoAssim hindcasts is to offset the initial bias using a similar method to *Laepple et al.* (2008). In that study, the CMIP-3 all forcing transient integrations were used as a prediction system by offsetting the initial bias of the ensemble relative to the observations. The offsetting is achieved by removing the difference between the previous 7 years of the transient ensemble, and the observations, from the future years of the projection. *Laepple et al.* (2008) claim that their method gives a more skillful prediction of global mean SAT than NoAssim as presented in *Smith et al.* (2007), and is only out-performed by DePreSys for the first two years.

For the early NoAssim members we cannot calculate the previous 7 years of transient

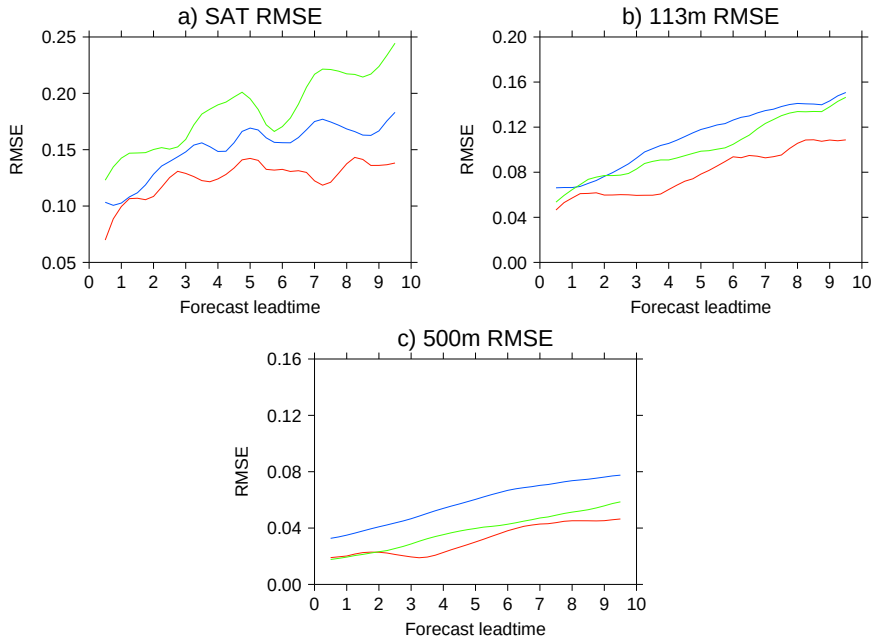


Figure 4.13: a) shows a comparison of RMSE as a function of lead time calculated for DePreSys (red) hindcast anomalies compared to the hindcasts anomalies calculated from NoAssim (blue) and shifted NoAssim (green) for SAT. b) and c) shows the same but now for the 113m heat contentand 500m heat content. Note that all NoAssim hindcast anomalies were made relative to **Sep Clim**.

integration due to the lack of the transient runs. We can only calculate the difference between the actual hindcast, and the previous observational anomalies. Calculating the offset using a 7 year observational anomaly would not be ideal, as the offset will cause the NoAssim hindcasts to lag behind the observed variability. An initial offset, equal to the difference between the initial season of the NoAssim hindcast and the previous 2 years of observations, was applied to the whole length of the hindcast. A 2 year averaging period was chosen as a compromise between a long period that would add lagged variability to the "initial conditions", and a short one, which would add high frequency variability. Figure 4.12 c) shows the Shifted NoAssim hindcasts of global mean 500m heat content anomalies. The initial temperature of the Shifted NoAssim hindcasts is much closer to the observations than the original NoAssim hindcasts. The initial bias correction effectively makes NoAssim take the observed long timescale variability into account, rather than the original NoAssim that exhibits almost no variability in its initial conditions.

Figure 4.13 shows that shifting NoAssim hindcasts affects the skill. The shifting of NoAssim did not improve the RMSE of global mean SAT (see fig 4.13 a)), and in fact

seems to make the RMSE worse. The higher RMSE of shifted NoAssim compared to NoAssim for global mean SAT seems to be because the observed global mean SAT is highly variable, especially on interannual time scales. Shifting the NoAssim global mean SAT to the high or low peaks in the observed SAT introduces an ensemble mean bias to the whole length of the Shifted NoAssim hindcast, increasing the RMSE (not shown). Shifting NoAssim has a more positive impact on the oceanic heat content. For 113m heat content, Shifted NoAssim initially performs better than NoAssim for the first year but does not outperform DePreSys. The improvement of DePreSys over shifted NoAssim for the first year is again likely to be due to DePreSys' ability to predict ENSO. The shifted NoAssim then generally performs slightly better than NoAssim for the remainder of the hindcast length suggesting that the initial bias contributes to the RMSE of NoAssim.

The largest difference for the RMSE is for the 500m heat content in figure 4.13 c). The shifted NoAssim's RMSE is reduced when compared to regular NoAssim, which was expected from figure 4.12 that shows that the NoAssim hindcasts have a warm. The RMSE of Shifted NoAssim and DePreSys are equal for the first year and a half of the hindcasts, which suggests that much of the skill seen in the top 500m arises because the assimilation effectively bias corrects the oceanic heat content. The shifted NoAssim RMSE is greater than the DePreSys RMSE after the second year of the hindcasts, which could suggest that the initialisation of variability in the DePreSys hindcasts may be important.

One way to check whether DePreSys is predicting the variability is to calculate the RMSE for each individual hindcast. So far, the mean RMSE as a function of lead time has been examined, but not all the hindcasts are likely to be equally skillful. The skill of each hindcast can be found by calculating the mean of the squared hindcast error of all lead times. The RMSE as a function of hindcast start time for a particular lead time t_1 to t_2 is:

$$D_i^{RMSE} = \sqrt{\frac{1}{lm}\sum_{t=t_1}^{t=t_2}\sum_{j=1}^{m}(D_{ijt}^{error})^2} \tag{4.2}$$

The RMSE as a function of start time for NoAssim, DePreSys and Shifted NoAssim is shown in figure 4.14. The RMSE as a function of hindcast start time highlights important sources of hindcast errors for both DePreSys and NoAssim. The eruption of Pinatubo (in 1991) clearly causes increased error for the NoAssim members initialised prior to the eruption (see figure 4.14 c) DePreSys also has increased error from the 1997/1998

El Niño (see figure 4.14 a). Although there is a significant difference between the mean RMSE for DePreSys and NoAssim throughout the hindcasts, there is no significant difference between Shifted NoAssim and DePreSys for the first three years of the hindcasts (fig 4.14a) and b)). Therefore, figure 4.14 suggests that, for the 500m heat content at least, DePreSys is not predicting the global mean changes better on average than a prediction system that does not assimilate information. The improved skill here is due primarily to the *initial bias correction* and not improved predictability.

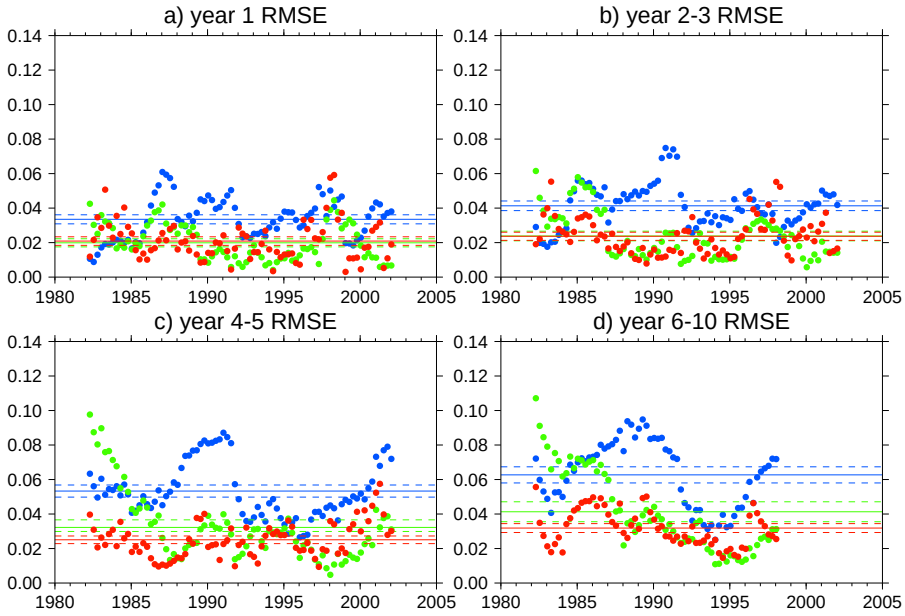


Figure 4.14: a) comparison of RMSE as a function of start time (D_i^{RMSE}) for the 1st year global mean top 500m average temperature anomaly of the NoAssim (blue), DePreSys (red) and shifted NoAssim (green) forecasts. The dots show the ensemble mean RMSE as a function of start time. The solid line shows the mean of the RMSE as a function of start time, and the dashed lines show the 95% confidence interval for the mean. b), c) and d) the same as a) but for the 2-3 year, 4-5 year, and 6-10 year RMSE respectively. Note that the NoAssim hindcast anomalies were made relative to **Sep Clim**.

The improvement of DePreSys relative to Shifted NoAssim at long lead times for 500m heat content (fig. 4.13 c)) seems to be explained by the shifted NoAssim hindcasts from 1982-1987. The Shifted NoAssim hindcast between 1982 and 1987 have a larger RMSE than the majority of shifted hindcasts after 1987, the later of which are similar to the DePreSys errors. The shifted NoAssim hindcasts from 1982-1987 hindcasts are "initialised" when the observed heat content was cooling, which is not captured by NoAssim and hence shifted NoAssim (see figure 4.12). Interestingly, figure 4.12 b) does seem to

show that DePreSys is capturing the cooling in the early 1980s, however, more work will be needed to identify the mechanism. Nevertheless, the large increase in RMSE for NoAssim in the late 80s early 90s will make a significant contribution to the difference in the mean RMSE as a function of lead time between NoAssim and DePreSys in figure 4.1 leading to increased RMSE (see figure 4.13 c)).

An obvious a limitation to this analysis is that the 500m heat content is poorly observed. Noise associated with poor observational sampling will make a meaningful verification of global mean heat content as a function of start time challenging. For example, the cooling after 2005 is poorly predicted by both NoAssim and DePreSys, which leads to increased RMSE (see figure 4.13 c)). However, as discussed in chapter 2, the cooling after 2005 seems mostly due to an increased number of observations in poorly observed areas of the globe (*AchutaRao et al.* (2007)), particularly the deployment of Argo floats to the Southern Hemisphere, and is, therefore, not a predictable signal. Subsampling the hindcasts by only validating the ocean heat content at grid boxes that contain an observation (such as *AchutaRao et al.* (2006)) could be considered in the future to remove false signals associated with observational sampling. Regardless of the global mean issues, the RMSE as a function of start time may be an important tool for understand the mean skill scores, and may prove more useful in understanding how hindcasts perform in smaller regions, where mechanisms exist to improve predictions, such as for El Niño.

4.5.2 Mean bias correction

As discussed briefly, the Shifted NoAssim hindcasts are subject to some problems due to the construction of the initial bias correction. Perhaps a more robust correction would be the removal of the mean bias, which is shown in section 4.4.1 and is less sensitive to sampling errors. The mean bias for the first season could be applied, but it was decided that the mean bias as a function of lead time would be removed to investigate the total biases effect on the skill. More specifically, the mean bias, calculated by averaging all of the hindcast errors together, was removed from every hindcast. The removal of the mean bias that is directly calculated from the same hindcasts could artificially inflate the skill by including the information from that hindcast. This is especially true for decadal

climate predictions, since a proportion of the bias is contributed from the low frequency natural variability of the observations over the hindcast period (i.e. B^{obsvar}). Correcting the bias using the hindcast being corrected will also tune the hindcast to the variability seen in the observations, artificially inflating the skill of that hindcast. A cross validation approach (*Michaelsen* (1987)) could be employed, where a separate bias correction for each hindcast is calculated by removing that hindcast from the calculation of bias. However, it is assumed that the difference would be negligible because of the large data set (80 hindcasts) and because the bias caused by the low frequency variability will affect all the hindcasts. Therefore, the same bias correction is applied to each hindcast for this analysis, but the effect of cross-validating the bias correction should be explored in the future.

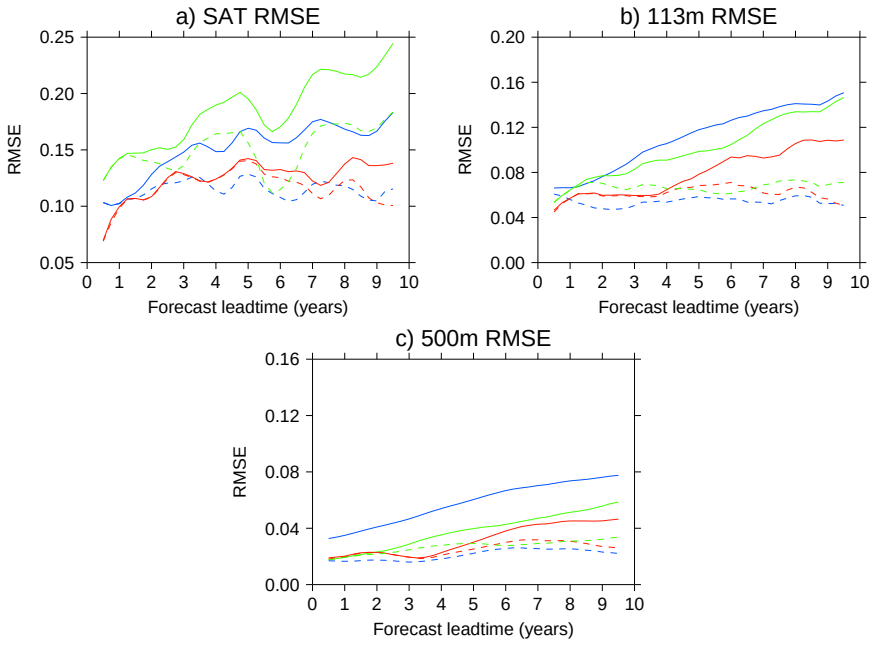


Figure 4.15: a) shows the comparison of RMSE for predictions of SAT as a function of lead time calculated for DePreSys (red), NoAssim (blue), and shifted NoAssim (green) hindcast anomalies. B) and c) show the same but now for the 113m heat content and 500m heat content respectively. The solid lines and the dashed lines show the RMSE as a function of lead for the for the normal and bias corrected hindcasts respectively. Note that all NoAssim hindcast anomalies were made relative to **Sep Clim**.

Figure 4.15 shows the effect of the bias correction upon the global mean RMSE as a function of hindcast lead time for NoAssim, DePreSys and Shifted NoAssim. All of the prediction systems show a reduced bias at long hindcast lead times with the RMSE also remaining relatively flat for entire length of the hindcast. The reduced RMSE at long

lead times demonstrates that much of the error here originates from the mean model bias, brought about by deficiencies in the model, especially the fact that it warms too quickly. *DePreSys now only out performs NoAssim at predicting the global mean temperature in the first year for both SAT and 113m heat content, and does not out perform NoAssim for 500m heat content at any lead time.* The worst performer after the bias corrections have been applied is Shifted NoAssim, which again is most likely due to its construction (i.e. the averaging period was not large enough to average out the high frequency variability).

The largest change after the bias correction has been applied is in NoAssim's 500m heat content RMSE. As already explored in section 4.5.1, there is a large initial bias between the observations and NoAssim because the observations cool in the early 1980s, and then remain cooler than NoAssim for the remainder of the hindcast period. The mean bias for NoAssim here is dependent on the observed variability, and will not be the same for all periods. Removing the bias calculated from this hindcast set could degrade the skill if it were removed from other hindcasts, and especially forecasts. *Extreme care should be taken when applying bias corrections to decadal forecasts*. The mean bias should be calculated from hindcasts initialised over a long period, and should cover a large spread of initial oceanic states to ensure that the bias is not just hindcast evolution. *The bias contribution from the observed variability should be minimised by initialising hindcasts across the climatological period, or should be removed from the bias before a bias correction is applied*.

Although the bias correction suggests that DePreSys is only able to predict the *global mean* temperature better than un-initialised hindcasts for just over a year it is at the regional-scale where the largest potential predictability resides (*Collins* (2002); *Boer and Lambert* (2008)). Figure 4.16 shows the regional impact of bias corrected 9-year mean anomalies. The bias corrected SAT does not show that DePreSys is able to improve hindcasts of 9-year mean anomalies over much of the continental land masses (fig. 4.16 d), although the positive impact in eastern Europe remains. Also, DePreSys does not appear to be able to offer much improved prediction of heat content over much of the ocean. This is especially true in the Southern Ocean, where the improved skill is removed when removing the mean bias, which further suggests that the apparent improvement seen here was essentially due to the drift of the transient integrations, and not improved prediction

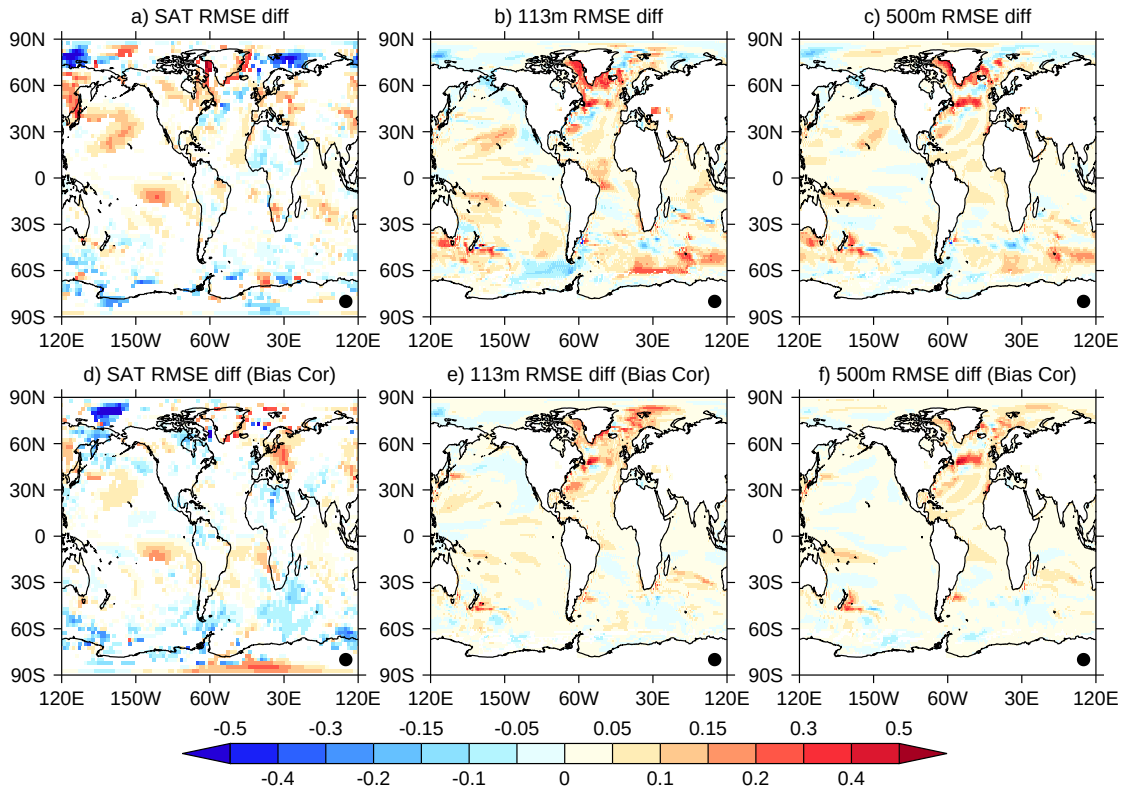


Figure 4.16: Shows 9-year mean NoAssim - DePreSys RMSE for SAT (a), 113m (b) and 500m (c) heat content when using the separate climatologies to calculate the NoAssim anomalies. d-f shows the same but with using a bias correction for both the NoAssim and DePreSys forecast anomalies. Only the grid boxes with a significant difference at the 5% level using independent tests are coloured. A black dot shows that the field is significant at the 5% level. Note that all NoAssim hindcast anomalies were made relative to **Sep Clim**.

of the variability. However, DePreSys does show a mean improvement over NoAssim is the North Atlantic.

Figure 4.17 d) shows the spatial improvement of RMSE as a function of time for 113m heat content. DePreSys still shows an improvement over NoAssim for the first few years of the hindcasts, especially in the North Atlantic and North Pacific. Therefore, DePreSys does seem able to capture some of the *regional* variability in areas that we expect to be predictable (*Boer and Lambert* (2008)). The improvement in skill at the regional-scale is not just from the correction of the initial bias, but the prediction of variability does seem play a role too, although only for a couple of years. More work is needed in understanding the specific mechanisms that improve the skill of DePreSys[2]. After

[2]An in depth analysis of DePreSys's ability to predict changes in the North Atlantic subpolar gyre is

the first 2 years the RMSE for NoAssim and DePreSys is very similar outside the North Atlantic region. Also note that the bias corrected RMSE difference for years 5-6 does not have field significance, further suggesting that the main difference between DePreSys and NoAssim is in the first few years.

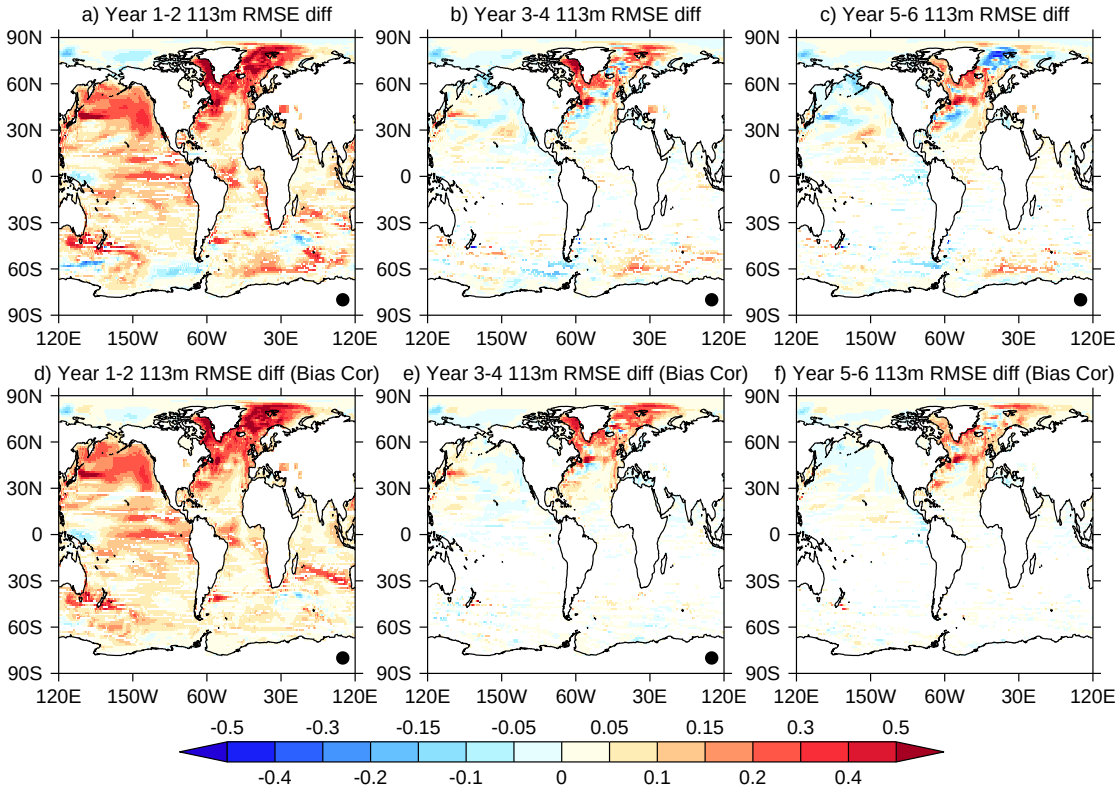


Figure 4.17: a), b) and c) show the NoAssim - DePreSys RMSE for means of hindcast years 1-2, 3-4 and 5-6 113m heat content respectively. d-f shows the same but for the bias corrected 113m heat content. Only the grid boxes with a significant difference at the 5% level using independent tests are coloured. A black dot shows that the field is significant at the 5% level. Note that NoAssim hindcast anomalies were made relative to **Sep Clim**.

4.5.3 Summary and discussion

The analyses presented here were performed to try to understand the causes for the apparent improvement in skill that the assimilation of observational information provides. For example, was the improved skill due to predicting low frequency variability, or was

presented in chapter 6.

the improvement due to the initial bias correction that the assimilation provides. Understanding the contributions to the improvement from both seems key to understanding the performance of decadal climate predictions, and to understand what benefit they provide.

Understanding the relative contribution from each will be difficult, and the analyses presented here should not be considered the definitive analysis, but instead as the first steps. However, progress has been made. Both the shifting of NoAssim (i.e. to remove the initial bias), and removing the mean bias of both prediction systems, showed that the correction of the initial bias does seem to provide a significant source of improved skill. Much of the improvement at long lead times in predictions of the *global mean* temperature seems to be due to the correction of the *initial bias*, not improved prediction of the variability. At the *regional-scale*, DePreSys does seem able to predict temperature for a number of years ahead, especially in the North Atlantic and the North Pacific, suggesting that DePreSys may be predicting variability in these regions.

The analysis here does have some limitations, particularly in the construction of Shifted NoAssim, which could be significantly improved in the future. Also the bias correction used here will artificially improve both systems skill because it contains the bias created by the observed variability. Nevertheless, the results have clearly shown that a significant proportion of the improvement offered by DePreSys is due to simply removing the initial bias. The results suggest that, to truly understand the improvement that initialised climate prediction provides, much more work is needed in understanding how, and why, the initialised hindcasts improve the predictions. Care must be taken when analysing mean skill of decadal predictions as the skill of both NoAssim and DePreSys is sensitive to the mean bias. The construction of hindcast experiments, and the validation of skill should try to minimise the effect of the bias on the skill.

Finally, it is also worth restating that the bias correction analysis presented here is not an argument for bias corrections to be applied to decadal climate predictions, rather, the bias correction is used here understand the origin of skill. As discussed in the section 4.5.2 the application of bias corrections should be performed with great care, especially as a significant proportion for the bias is due to the observed low frequency variability.

4.6 Chapter Conclusions

This chapter has presented the skill of DePreSys relative to NoAssim after correcting the error in the construction of NoAssim hindcast anomalies. The mean analysis of DePreSys was then extended beyond what was presented in *Smith et al.* (2007). The evolution of mean SAT and upper ocean heat content bias of the DePreSys and NoAssim hindcasts with lead time has also been investigated, and a brief discussion of the origins was presented. The causes of improved skill in DePreSys were also investigated. The main results are as follows:

- Taking account of the error in the calculation of the NoAssim hindcast anomalies shows that DePreSys still appears more skillful at predicting heat content anomalies in many regions of the worlds oceans, such as the North Atlantic and Southern Ocean, but the relative improvement was reduced somewhat.

- The correction of the climatology also showed that DePreSys is performing better than NoAssim at predicting 9-year mean SAT anomalies in Western Europe and North America, agreeing better with previous potential predictability studies, and implying that improved predictions of the Atlantic SSTs may be the cause.

- Although DePreSys shows some improved skill of 9-year mean anomalies, the major improvement is seen in the first year of the hindcasts. The improvement in SAT resides mostly in the Tropical Pacific, and suggests that predictions of El Niño are improving predictions in this region for at least a year. The first year's 113m heat content prediction from DePreSys compared to NoAssim is more skillful over most of the globe, but the improvement quickly reduces over the first couple of years.

- DePreSys and NoAssim both show a mean bias with respect to observations. The biases have a number of causes, but we think that they are primarily caused by: the model's inability to correctly simulate the patterns of variability; errors in the radiative forcing; errors in the model's response to radiative forcing (climate sensitivity errors, climate feedback errors and process errors (e.g. the model can not subduct

enough heat into the deep ocean)); the drift of the transient runs, especially transient 1. Finally multidecadal and interannual variability of the observations also contribute to the mean bias, but this contribution would change for each hindcast period used.

- NoAssim has a larger mean bias than DePreSys at all lead times. However, crucially NoAssim has a significant initial bias at zero hindcast lead time, but DePreSys does not by construction. Attempting to remove the initial bias, or the mean bias, suggests that the correction of the initial bias is a significant contributor to the improved skill. Much of the improvement at long lead times in hindcasts of *global mean* temperature seems to be due to the correction of the *initial bias*, not improved prediction of the variability. At a *regional* level, DePreSys does improve the predictions of temperature for a number of years ahead, especially in the North Atlantic and North Pacific, suggesting that DePreSys may be predicting variability in these regions. However, more work is needed to understand the mechanisms that improve the skill.

- Correcting the bias reduces the error of the hindcast at long lead times, showing that the mean model bias is a large contributor to the hindcast error. However, caution is encouraged for the application of bias corrections for decadal climate forecasts. Because of the slow evolution of the ocean, a long hindcast period is required to ensure that the hindcasts are initialised across a wide range of climate states. DePreSys needs to be initialised across a wide range of states so that predicted variability is not considered as a mean bias. Ideally NoAssim hindcasts should be initialised across a wide range of initial states, so that the contribution from the observational variability can be averaged out.

Although the mean skill scores show the general performance of the prediction systems, they are not informative of why a particular prediction system is better or worse than another. It could be easy to suggest that an improvement of DePreSys over NoAssim is caused by the initialisation of low frequency variability, as reasoned by *Smith et al.* (2007). However, beginning to move past these mean skill scores and understanding the origin of the RMSE seen in the analysis has already highlighted the correction of the

initial bias in NoAssim as playing an important role in improving predictions.

The analysis of the observations in chapter 2 suggested that there have been large, potentially predictable, fluctuations in the Atlantic Meridional Overturning Circulation (AMOC). Is it therefore surprising that the improvement in predicting upper ocean heat content in the North Atlantic is not larger than that seen in figure 4.17? So far the analysis of DePreSys has focussed upon the mean skill of the prediction systems, and not on understanding the processes that are behind the improved predictability of DePreSys. Understanding the mean skill scores of a prediction system, and identifying areas of the prediction system could be improved to enable more skillful predictions, is dependent on understanding the processes at work in the hindcasts. To start to understand the processes at work in DePreSys and NoAssim hindcasts, the mean evolution of the AMOC in the DePreSys and NoAssim hindcasts will be presented in the next chapter, which highlights some potential problems with the anomaly assimilation method used to initialise DePreSys.

Chapter 5

Analysis of mean drift in DePreSys and NoAssim

5.1 Introduction

The transient runs, which are key to the construction of NoAssim and DePreSys, were shown to be drifting in chapter 3, especially in the North Atlantic. The NoAssim hindcasts are initialised from the transient runs, and so will also contain drift (see figure 5.1), which will contribute to NoAssim's time dependent bias (as discussed briefly in chapter 4). The drift of NoAssim will, therefore, contaminate the NoAssim hindcasts, and could potentially diminish the skill of NoAssim. In principle, DePreSys should show no drift, since it is relaxed to a model climatology. However, the climatology used as the background state of the assimilation for DePreSys was computed from a drifting transient run. Due to the strong relaxation to climatology, and poor observations at depth, DePreSys is initialised from climatology at depth, i.e. with near zero anomalies. Therefore, it is likely that DePreSys hindcasts would also drift similarly to transient 1, which will contribute to the time dependent bias of DePreSys (see figure 5.1).

However, the climate system is non-linear; it is not clear whether the drift of DePreSys would be at the same rate as the transient run and NoAssim. The drift of DePreSys could be different; for example it could occur faster than that seen in NoAssim. In this case,

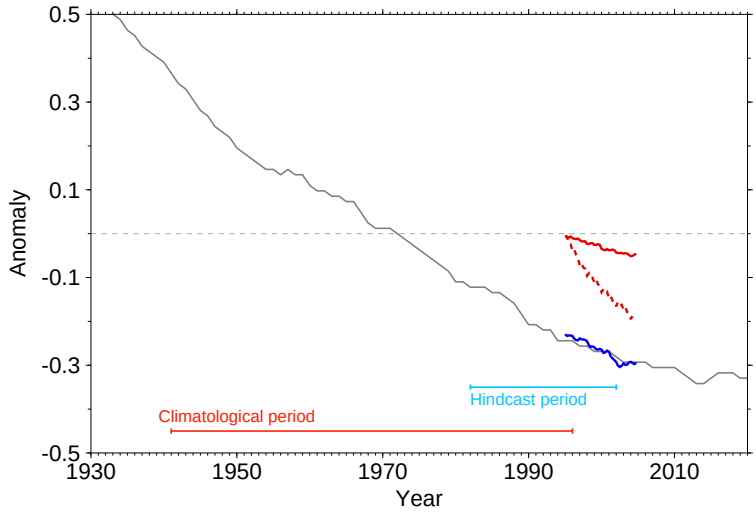


Figure 5.1: Schematic highlighting the problem of drift in the NoAssim and DePreSys systems for a model variable. The drift of transient 1, over the climatological and hindcast period is shown in grey. NoAssim hindcasts (blue) initialised from such a transient run would confine to drift as the transient run. We would also expect some sort of drift in a DePreSys hindcast (red) that is initialised from the climatology, since the climatology is constructed from a drifting transient run. The solid red line shows a DePreSys hindcast that exhibits the same drift as that of NoAssim. The drift could however be different, and could occur more quickly (red dashed line), or could even have the opposite sign.

the skill of DePreSys would be degraded relative to NoAssim. Therefore, understanding the mean evolution of the DePreSys and NoAssim hindcasts, in the presence of drift, is important for a complete understanding of the skill scores that were shown in the previous chapter. This chapter will present an investigation into the mean drift of DePreSys and NoAssim, with a particular focus on the North Atlantic. An analysis of the mean drift in the Atlantic Meridional Overturning Circulation (AMOC) for both DePreSys and NoAssim is presented in section 5.2, followed by an analysis of mean density drift in section 5.3. A brief analysis of the AMOC and density drifts in DePreSys2 will be presented in section 5.4, which shows that the drift is not just a feature of the original DePreSys. Finally, section 5.5 contains a brief disussion of the possible causes of the more rapid drifts found in DePreSys.

5.2 The AMOC in the DePreSys and NoAssim hindcasts

Figure 5.2 shows the strength of the AMOC for both NoAssim and DePreSys. The initial condition data is not available, and so the first annual mean AMOC strength for each hindcast, represented by the dots, is the state that is closest to the initial state. The initial conditions of NoAssim are taken from four all forcings transient integrations, and so the first annual mean of the NoAssim hindcasts should be representative of the evolution of the transient runs. Therefore, figure 5.2 shows the NoAssim ensemble members separately according to which transient run the initial conditions are taken from (e.g. NoAssim 1 was initialised from transient 1). The forcing of the original transient run and the NoAssim hindcasts are somewhat different, in that the volcanic aerosol forcing reduced back to zero as the NoAssim hindcasts progress. Hence, the NoAssim hindcasts and the transient run (here represented by the dots) should diverge. Nevertheless, figure 5.2 shows that NoAssim is capable of predicting many of the large changes in the MOI of the parent transient runs, up to a decade in advance, suggesting that the MOI at 50°N, in HadCM3 at least, could be potentially predictable up to a decade in advance in agreement with "perfect model" studies (*Collins and Sinha* (2003); *Collins et al.* (2006a)).

The transient runs, and the NoAssim hindcasts, show that HadCM3 possesses large amplitude, decadal timescale variability in the MOI at 50°N, with variations of up to 4 Sverdrups in many of the NoAssim hindcasts over a 10 year lead time. However, similar long timescale variability is *not* seen in the original DePreSys assimilation run (see black line in figure 5.2 e)[1]. The AMOC strength in the assimilation run is dominated by high frequency (multi-seasonal) variability around a mean overturning at 50°N of $\sim$17Sv. The observational analysis presented in chapter 2 suggests that there may have been large changes in the AMOC strength at 50°N in the hindcast period, so is a lack of large amplitude decadal variability in the original DePreSys assimilation run is a concern?

Investigating the difference between the assimilation runs of the original DePreSys (figure 5.2 e), and DePreSys2 (figure 2.9), highlights a potential problem with the anomaly assimilation method. Figure 5.3 shows that the lack of large amplitude, multidecadal

[1]Note that this is a different assimilation run to that shown in figure 2.9. The assimilation run here uses the Original Met Office ocean analysis as the observed anomalies.

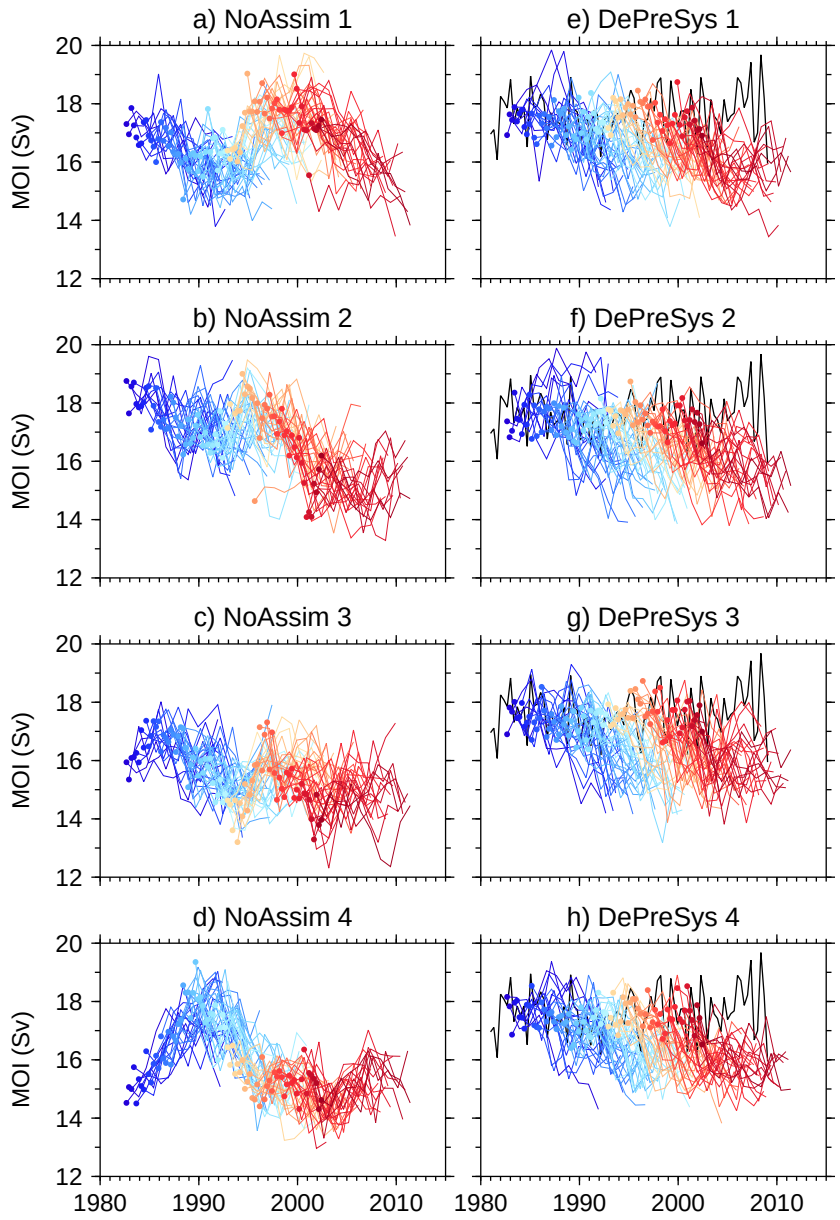


Figure 5.2: The evolution of annual mean Meridional Overturning Index (MOI) at 50°N from the NoAssim (a)-d)) and DePreSys (e)-h)) hindcasts. The dots show the first annual mean of the MOI, the closest hindcast state to the initial condition that is possible, and the coloured plumes show the evolution of the MOI in each hindcast. The colours are purely intended as an aid to the eye. The NoAssim members are separated according to the transient run from which their initial conditions are taken. a)-d) shows NoAssim hindcasts which are taken from the first, second, third and fourth transient run respectively. The individual DePreSys hindcasts members are also shown separately to highlight the similarity between the DePreSys hindcasts. The black line in e)-f) shows the seasonal evolution of the MOI at 50°N from the DePreSys assimilation run.

variability in the assimilation run of the original DePreSys is due to a lack of large magnitude, positive density anomalies assimilated into the model. Although positive density anomalies are resolved in regions where the AMOC in HadCM3 is sensitive to density anomalies in the Updated Met Office ocean analysis (see figure 2.8), the original ocean analysis does not resolve large or spatially coherent density anomalies. The strong relaxation in the anomaly assimilation method to the model climate *plus* the observed anomalies, at all depths, means that the assimilation relies heavily on the ocean analysis for an adequate representation of density anomalies. Even though positive density anomalies are assimilated at the surface in the late 1980s and early 1990s, the assimilation run is heavily constrained. Therefore, if positive density anomalies are not observed at depth, the AMOC will not strengthen in the assimilation, as seen in the original DePreSys (figure 5.2 e). Therefore, the strong relaxation removes the models ability to remember what forcing has already been applied.

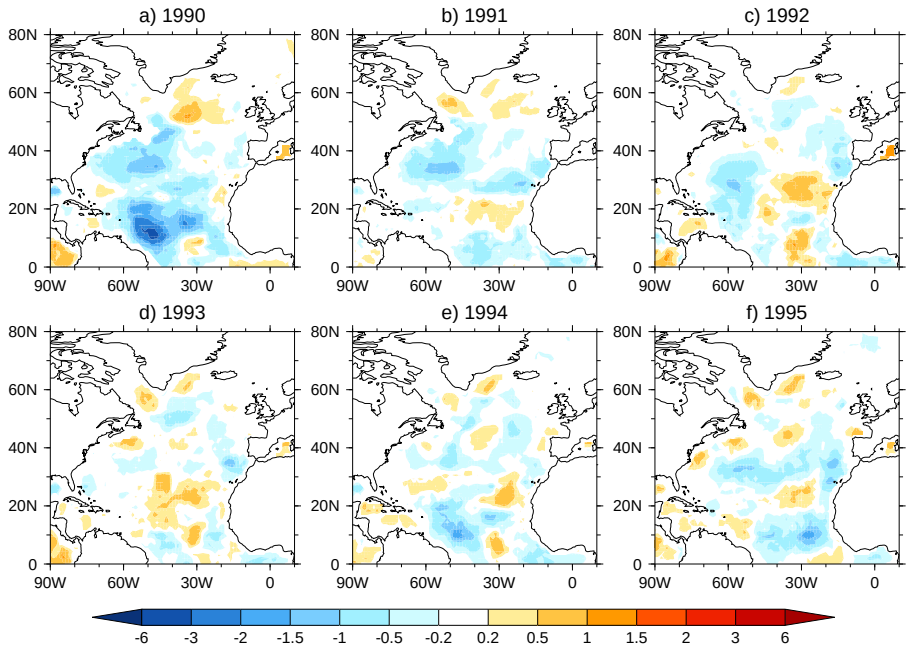


Figure 5.3: Annual mean 800-3000m in situ density anomaly for the original DePreSys assimilation. The density anomalies are calculated by placing the monthly mean observed anomalies of temperature and salinity from the Original Met Office ocean analysis on to the model climatology (C_1). Therefore, the density anomalies are representative of the original assimilation run. The density anomalies have been normalised by the standard deviation of annual average density anomalies in the HadCM3 control run.

In contrast to NoAssim, DePreSys is a poor predictor of the future assimilation MOI. If relatively small density anomalies were assimilated at depth in the hindcasts then no

large changes in the hindcasts predicted AMOC would be expected. However, many of the DePreSys hindcasts show rapid decreases in the MOI at 50°N, of order 2-3 Sverdrups, over just a couple of years. The rapid decrease in the AMOC strength is most obvious for the DePreSys hindcasts which are initialised after 1994, where the decrease seems to occur within 2 years of DePreSys being initialised. The DePreSys hindcasts initialised in the 1980s do not have an immediate reduction in the MOI at 50°N like those initialised from 1994 onwards, but most show a rapid reduction in the MOI at a later hindcast lead time. Only a few hindcasts early in the hindcast period (1982-1983) do not show a rapid decrease in the MOI at 50°N over the course of the hindcast period (figure 5.2 f is the clearest example). However, it is unclear whether the MOI would decrease in these hindcasts if they were continued beyond 10 years.

Although a reduction in the AMOC strength may well be expected for the hindcasts initialised near to 1995 (i.e. near the observed warming of the North Atlantic when we expect AMOC to be near its peak), it is a surprise that most of the hindcasts show a similar large and rapid decrease in the AMOC strength. To understand the structure and extent of the changes in AMOC, the mean anomalous evolution of the Atlantic meridional stream function is calculated taking the mean of all of the hindcasts:

$$\overline{D_t^{\psi}} = \frac{1}{lm} \sum_{i=1}^{l} \sum_{j=1}^{m} D_{ijt}^{\psi} - C_{DeP}^{\psi} \tag{5.1}$$

where D_{ijt}^{ψ} is the Atlantic stream function from a DePreSys hindcast and C_{DeP}^{ψ} is the climatology of the Atlantic stream function. The ocean velocities for the transient integrations, or assimilation run, were not available for the construction of the Atlantic Meridional stream function climatology. Therefore, a climatology is defined as the mean of the first season from all the DePreSys hindcasts. The first season is close to the initial condition of the hindcasts, and is therefore taken to be representative of the assimilation run. Because DePreSys hindcasts are initialised seasonally 4 times a year, the climatology is, therefore, an annual average stream function climatology:

$$C_{DeP}^{\psi} = \frac{1}{lm} \sum_{i=1}^{l} \sum_{j=1}^{m} D_{ij1}^{\psi} \tag{5.2}$$

Figure 5.4 shows the mean evolution of the Atlantic meridional stream function anomalies (in Sverdrups, (Sv)) in the DePreSys hindcasts. A steady weakening of the AMOC of

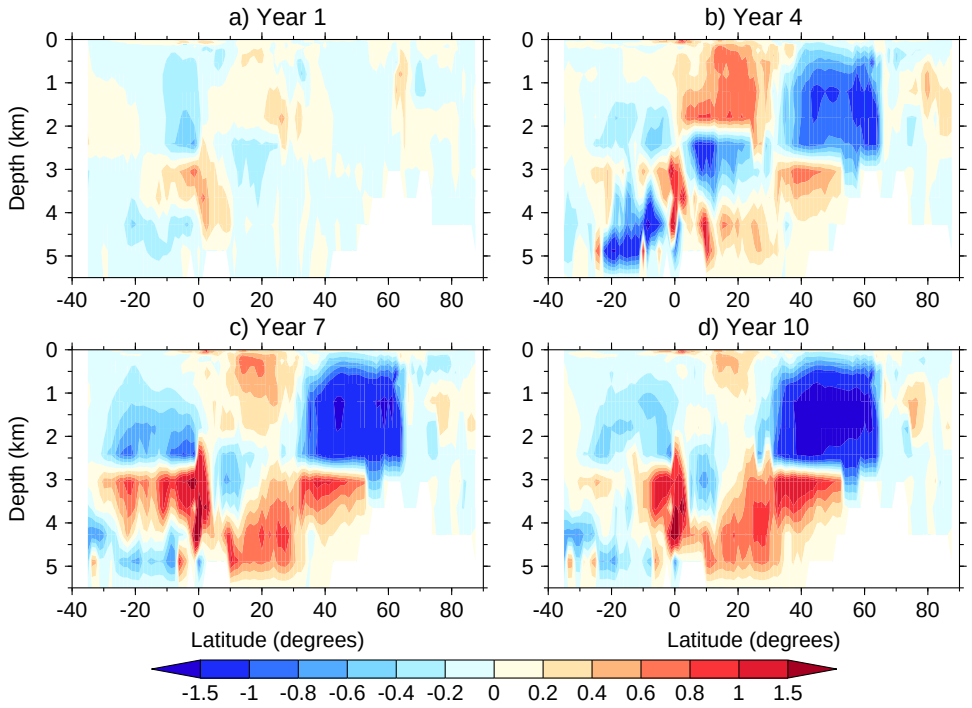


Figure 5.4: The mean evolution of the Atlantic Stream function in the DePreSys hindcasts ($\overline{D_t^{\psi}}$) relative to a stream function climatology calculated from the mean of the first seasons from the same DePreSys hindcasts (Sv). The anomalies have been constructed so that a positive anomaly indicates a strengthening of the circulation disregarding the sign of the original stream function.

$>$1.5 Sv is seen in the North Atlantic between 30°N-60°N (from 0-2500m in depth) over the 10 years of the hindcasts. A non-linear evolution is seen in the tropical North Atlantic between 0°N-30°N with an initial strengthening of $\sim$0.6 Sv at 2000m, with a local weakening of the Antarctic Bottom Water (AABW) cell immediately below of $\sim$-1 Sv after 4 years. As the hindcasts progress past year 4, the mean anomalies in the tropical North Atlantic from 0-3000m begin to reduce toward zero. Finally, beneath the AMOC cell there is a strengthening of the AABW cell over much of the North Atlantic, especially between 10°N-50°N, with mean increase of $\sim$1 Sv.

The anomaly assimilation method was constructed so that it would minimise large adjustments or drifts. Therefore, the presence of such large adjustments in the DePreSys hindcasts is a concern. Thus, it is important to understand the origin of such adjustments; for example is the adjustment just a manifestation of the drift seen in transient 1 from which the climatology is constructed? To begin to understand the cause of the mean evo-

lution of the DePreSys Atlantic stream function (figure 5.4), the mean evolution of the NoAssim 1 hindcasts, which have the most similar climatology to DePreSys, is shown in figure 5.5 for comparison. The mean NoAssim Atlantic Meridional stream function anomalies are calculated in a similar fashion to the DePreSys anomalies (equation 5.1), by calculating a climatology by averaging all the first seasonal means from the NoAssim hindcasts as in equation 5.2. However, because the transient runs, and hence the NoAssim members, each have their own climatology, the stream function anomalies for NoAssim are calculated relative to separate climatologies given by:

$$\overline{N_t^\psi} = \frac{1}{lm}\sum_{i=1}^{l}\sum_{j=1}^{m} N_{ijt}^\psi - C_{NoA,j}^\psi$$

where $C_{NoA,j}^\psi$ is the climatology of NoAssim member j and is given by:

$$C_{NoA,j}^\psi = \frac{1}{l}\sum_{i=1}^{l} N_{ij1}^\psi$$

Figure 5.5 shows that NoAssim 1 hindcasts also have a reduction in the AMOC strength north of 40°N over the course of the 10 year hindcasts. The evolution of the NoAssim 1 stream function is similar to that seen in DePreSys, although the reduction in AMOC strength is smaller. The slow down of the AMOC above 40°N continues through the whole length of the NoAssim 1 hindcasts (fig. 5.5) to a maximum reduction in AMOC strength of 1 Sv, which is accompanied by a increase in the AABW cell north of 10°N. However, the evolution is not similar to DePreSys everywhere. The complex change of overturning strength seen in the tropical North Atlantic of DePreSys (fig. 5.4) is not present in the NoAssim hindcasts. Also, in contrast to the strengthening response of the AABW cell in the DePreSys hindcasts, the AABW cell decreases significantly in strength in the South Atlantic.

The weakening of the AMOC in the North Atlantic in NoAssim is likely to attributable to a number of factors. Firstly, the weakening signal seen in figure 5.5 could be the manifestation of the drift seen in the HadCM3 control run (see figure 5.1). The transient runs, from which the NoAssim hindcasts are initialised, are also moving towards HadCM3's preferred state. Hence, the NoAssim hindcasts will also contain the drift.

Secondly, in HadCM3, like in many coupled climate models, the AMOC strength de-

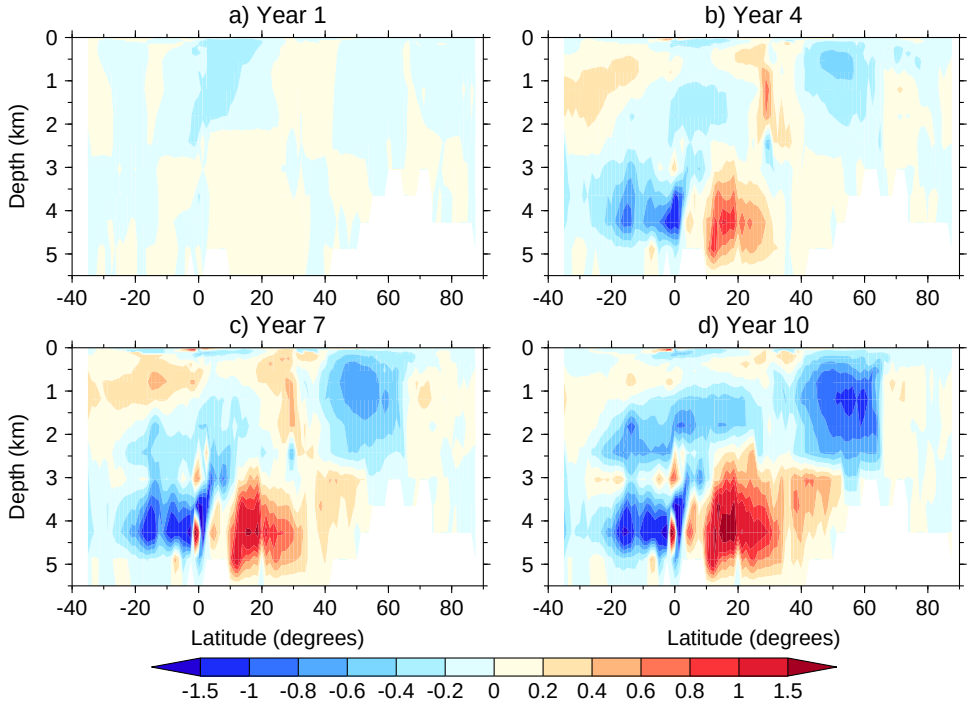


Figure 5.5: As figure 5.4 but for NoAssim 1.

creases in response to increased radiative forcing through a reduction in oceanic heat loss and an increase in freshwater inputs at high latitude (*Thorpe et al.* (2001); *Gregory et al.* (2005)). However, the AMOC response to radiative forcing is unlikely to be important for the end of the 20th Century (*Meehl et al.* (2007)).

Thirdly, as already seen in figure 5.2, the AMOC at high latitudes exhibits large amplitude multi-decadal variability. Thus, is it possible that the signal seen in figure 5.5 is due to sampling error? The NoAssim hindcasts are all initialised over a relatively small time frame (only 20 years). Therefore, it is impossible to quantify the contribution of long timescale drift or natural variability separately. Looking again at figure 5.2, it is clear that transient 1, and the other transient runs, are exhibiting significant multidecadal variability through out the hindcast periods. Is the similarity, in the North Atlantic AMOC evolution, between NoAssim 1 and DePreSys just due to chance? Figure 5.6 shows the mean evolution of the other 3 NoAssim members (i.e. NoAssim 2-4). Although the magnitude is smaller, there is still a reduction of the AMOC north of 40°N, suggesting that a significant part of the AMOC changes in figure 5.5 are due to the transient run drift.

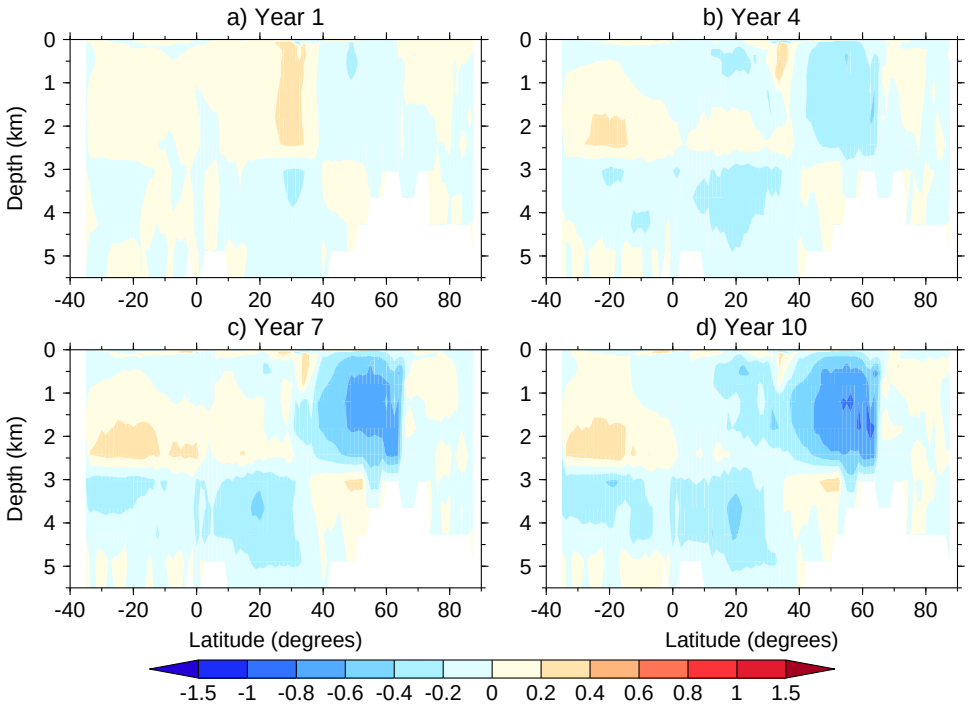


Figure 5.6: As figure 5.4 but now for the average of NoAssim 2-4.

The similarity between DePreSys's and NoAssim's mean meridional stream function evolution in the North Atlantic suggests that the mean AMOC change in the DePreSys hindcasts is, at least in part, a manifestation of the drift in the transient runs, and in particular transient 1. Although it appears that the drift of the transient runs are affecting both NoAssim and DePreSys, it is important to note that the changes in the AMOC for DePreSys are larger in magnitude and spatial coverage, and appear more rapid than is explained by the transient runs alone. The increased drift and the similarity in the evolution of the AMOC in almost all of the DePreSys hindcasts suggests that the evolution of the Atlantic stream function in DePreSys is also a consequence of the initialisation. In particular, it suggests a lack of balance in the initial state of the hindcasts.

5.3 Density

As highlighted in the previous section (5.2), most of the DePreSys hindcasts exhibit rapid and large changes in the AMOC, particularly in the North Atlantic, which is a major

concern. Large adjustments or drifts are also seen in the density fields of the DePreSys hindcasts, and will be discussed in this section. For clarity it is important to first state a number of points about the density analysis.

As mentioned in chapter 3, the seasonal data from the transient runs is unavailable before 1950, and so it is impossible to calculate separate T and S climatologies for the 1941 - 1996 period for transient members 2-4. Because of the non-linear equation of state for sea water, the initial density anomalies for the DePreSys hindcasts may be sensitive to the climatological period. Therefore, it was decided to keep the climatological period as 1941-1996 when calculating the DePreSys density anomalies, rather than swap to the 1951-2006 climatology used for the analysis of skill in chapter 4. Therefore, for the comparison of DePreSys and NoAssim, the DePreSys density anomalies will only be compared with anomalies from NoAssim 1, which share a similar climatology.

Another complication for the comparison of mean density anomalies arises because of the change in the anomaly assimilation method from 1994 onwards, as described in chapter 3. The start dump for the assimilation run was taken from a 1979 state in a later transient run (e.g. transient 3 or 4), not transient 1, from which the climatology used in the assimilation was calculated (not shown). Over much of the ocean, especially near surface, which is best observed, the assimilation run was quickly pulled towards the climatological state of transient 1 (C_1), but in areas of the ocean which are poorly observed, the deep ocean and the Arctic for example, the assimilation moves toward C_1 much more slowly. *Hence, there are large density anomalies relative to C_1 in the deep ocean and the Arctic for the DePreSys hindcasts initialised prior to 1994*. From 1994, the strong relaxation pulls DePreSys to C_1 and the hindcasts were initialised with near zero density anomalies at depth. For clarity, only the density anomalies for the hindcasts initialised from 1994 onwards will be analysed from the original DePreSys.

The transient runs, from which the NoAssim initial conditions are taken are free to evolve, and continue to drift relative to their climatology; thus, the initial density anomalies for NoAssim are also non-zero (see figure 5.1). In order that the evolution of the mean density anomalies, or drift, from NoAssim 1 can be compared with DePreSys, the initial density anomaly of the NoAssim hindcasts will be classed as a bias, and is removed from

every lead time. The mean NoAssim 1 anomaly with the first season removed as a bias is given as:

$$No^{\rho\prime}_{i1t} = N^{\rho\prime}_{i1t} - N^{\rho\prime}_{i11}$$

Thus, the mean evolution of NoAssim 1 density anomalies is given as:

$$\overline{No^{\rho\prime}_{t}} = \sum_{i=1}^{l} No^{\rho\prime}_{ijt} \tag{5.3}$$

As for the Atlantic meridional stream function, taking the mean of all NoAssim hindcasts allows the internal variability to be removed, and should leave just the mean evolution.

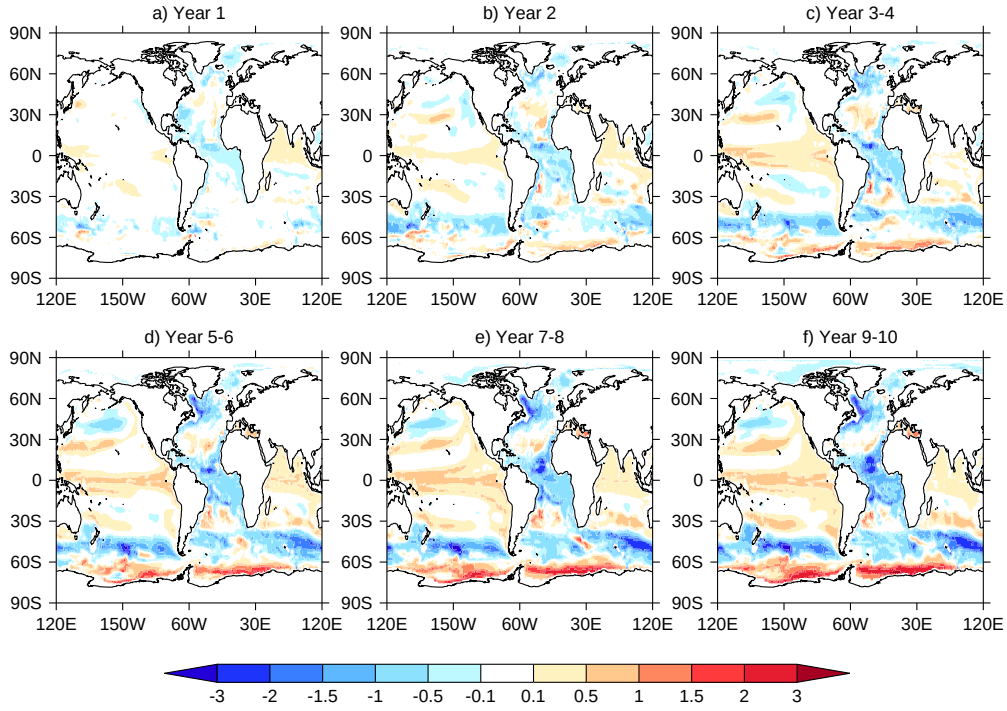


Figure 5.7: The 800-3000m mean integrated in situ density anomalies, relative to the 1941-1996 density climatology calculated from the first transient run (C_1), averaged over the DePreSys hindcasts ($\overline{D^{\rho\prime}_t}$) initialised from 1994 - 2001. The density anomalies are normalised by the interannual standard deviation from the HadCM3 control run.

Figure 5.7 shows the mean evolution of density anomalies in the DePreSys hindcasts. A low density anomaly is seen to originate down stream of the Greenland Scotland ridge in the first year, and spreads westward in year 2 (fig. 5.7 a and b). By year 3-4 (fig. 5.7 c) a low density signal is present in the Labrador sea and is beginning to spread down the western boundary of the North Atlantic. At the end of the 10 year hindcasts the

low density signal has continued to strengthen to a magnitude more than 3 (2) standard deviations (when all DePreSys hindcasts are averaged together; not shown) away from the original climatology of the assimilation run. The propagation of the low density signal seen in figure 5.7 is consistent with the weakening of the AMOC in HadCM3 (see figure 2.7).

Away from the North Atlantic, interesting changes in the mean density of DePreSys are also seen in the Southern Ocean (fig. 5.7). In the first season (not shown), strong dipoles of density anomalies are present in the Drake passage region of the Southern Ocean and south of Australia ($\sim$150°W). These dipoles, initially small in spatial extent, expand as the DePreSys hindcasts continue, until a large dipole of density is present over much of the Southern Ocean, with high density to the South at the edge of Antarctica, and low density to the North particularly in the Southern Indian and Pacific Oceans. The change in density in the Southern ocean for the 800-3000m layer is also present in the top 500m of the ocean in the DePreSys hindcasts with mean anomalies greater than 2σ (not shown).

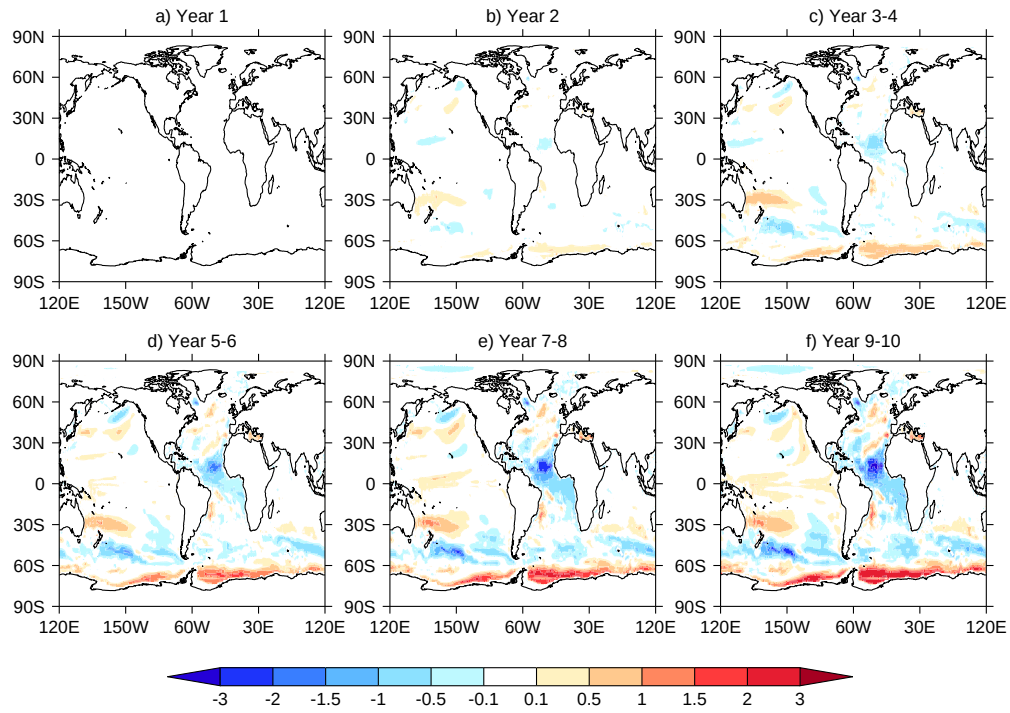


Figure 5.8: As figure 5.7 but for 800-3000m mean NoAssim 1 density anomalies ($\overline{No^{\rho\prime}_{t,1}}$).

Looking at global evolution of the mean density anomalies in NoAssim 1, a large low density signal propagating down the deep western boundary of the North Atlantic is *not*

seen (see figure 5.8). There is small reduction in density in the subpolar North Atlantic, especially in the Labrador Sea, with some propagation down the deep western boundary, but the drift is much smaller than seen in DePreSys. Elsewhere, there are similarities between the mean evolution of DePreSys and NoAssim in the Southern Ocean. The NoAssim hindcasts also have an increase in the integrated 800-3000m density anomalies at the coast of Antarctica, similar in magnitude to DePreSys. The increase in 800-3000m integrated density anomalies along the coast of Antarctica, in both NoAssim and DePreSys, is therefore, likely to be a manifestation of the drift seen in the transient integrations: transient 1 in particular. However there are also differences which can not be explained by the drift in transient 1. For example, the decrease in density over much over the Southern Indian and Pacific Ocean's between 40°S-60°S is not as rapid in NoAssim 1 as it is in DePreSys (see fig. 5.7).

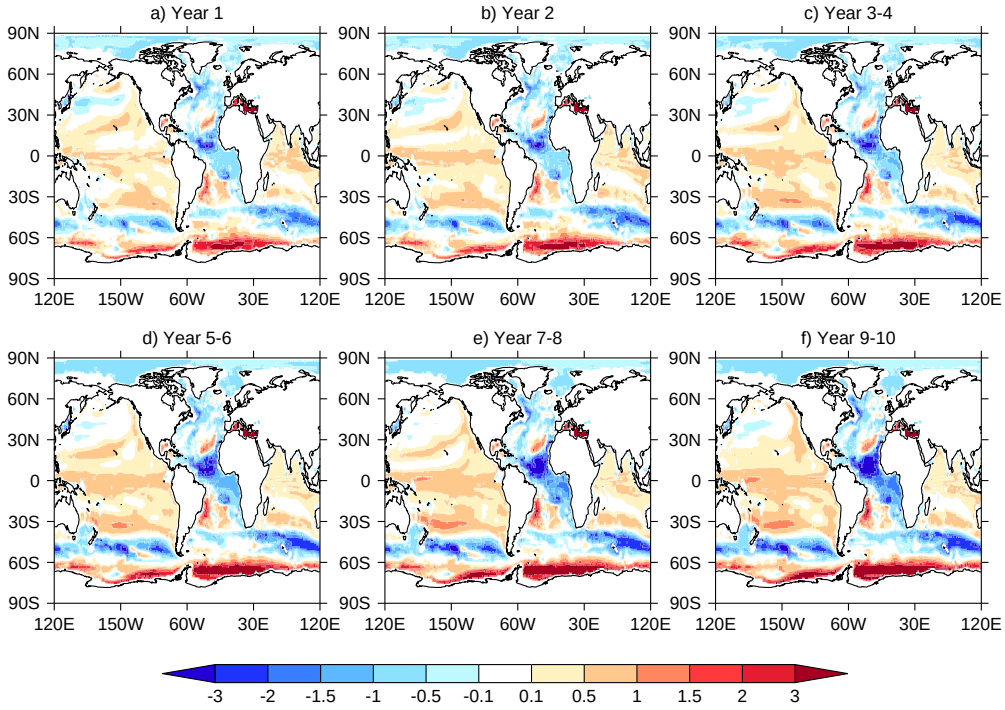


Figure 5.9: As figure 5.8 but without the first season density anomaly removed as a bias ($\overline{N^{\rho\prime}_{t,1}}$).

Interestingly, when examining NoAssim 1 mean density anomalies relative to C^{ρ}_{1}, without the first season's anomaly removed (see figure 5.9), the mean density anomalies of NoAssim are found to be of similar sign and magnitude to the mean evolution of DePreSys after 10 years. Assuming that the drift is roughly linear, then the drift seen in

the first season of NoAssim 1 is the drift of transient 1 over 25 years (by taking the central points of the climatological period (1968 for 1941-1996) and the NoAssim hindcasts period (1992 for 1982-2001)). The similarity of DePreSys's mean density drift, and the density anomalies of the NoAssim hindcasts suggests that the DePreSys hindcasts are adjusting towards the preferred state of HadCM3, at a rate that is, at least initially, faster than the drift of the transient runs.

Figure 5.10 clearly shows the long-timescale drift of the transient integrations used to create the climatology for DePreSys. In the Southern Ocean (60°S-80°S) HadCM3 has an increasing 800-3000m density anomaly, relative to C_1, for each successive transient run. In the Deep Western North Atlantic (DWNA; 75°W-40°W,35°N-65°N) each successive transient run has a lower 800-3000m density relative to C_1. Note that it is clear from figure 5.10 that NoAssim's initial conditions are taken from transient integrations that have a slightly different evolution to the transient run used to supply the climatologies for the DePreSys assimilation run. Although the NoAssim members are initialised from different transient runs they generally follow the evolution of the original transient runs and drift at the same rate as the transient runs.

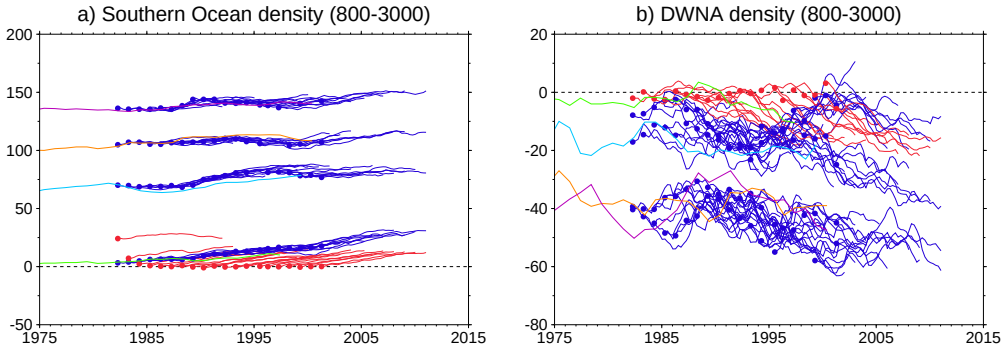


Figure 5.10: Integrated density anomalies, relative to the 1941-1996 DePreSys climatology (C_1), for a) the Southern Ocean (60°S-80°S) and b) the deep western Atlantic (DWNA) region of the North Atlantic (75°W-40°W,35°N-65°N) for DePreSys (red) and NoAssim hindcasts (blue). For clarity only DePreSys and NoAssim hindcasts initialised from the 1st of March are shown. Also shown is the density anomalies for the four transient integrations, which are also relative to the 1941-1996 DePreSys climatology (C_1). Transient 1, 2, 3 and 4 are shown by the green, light blue, orange and purple lines respectively.

For these two deep regions of the ocean the DePreSys hindcasts are initialised extremely close to a zero initial anomaly; the observed temperature (T) and salinity (S) anomalies

are small here, or non existent, so the assimilation is constrained to the climatological T and S. The first few initial years of DePreSys hindcasts are the only hindcasts to be initialised with a non-zero density anomaly in the Southern Ocean as the assimilation is relaxed towards the C_1 climatology. The March 1982 hindcast is initialised with a lower density in the Southern Ocean than the NoAssim 1 members, and interestingly the increase in density is slower than the members of DePreSys initialised from the climatological state. In the North Atlantic, the observational coverage is much better than in the Southern Ocean. The DWNA density in the assimilation is pulled towards C_1 much faster than in the Southern Ocean, and so all the DePreSys hindcasts are initialised with an initial anomaly close tp zero (figure 5.10 b). The first few hindcasts do not see a large reduction in the DWNA density anomalies, but after the first couple of DePreSys hindcasts the density anomalies begin moving rapidly towards the evolution of transient 1.

In summary, it is clear that in the original DePreSys hindcasts undergo significant drifts, especially in the AMOC, which undergoes large, and rapid, adjustments in almost all 320 hindcasts. The reduction in AMOC strength in DePreSys is associated with a low density signal propagating down the deep western boundary of the North Atlantic. The NoAssim 1 hindcasts also show a drift of the AMOC and density, but the changes are not as large or as rapid. There are also large drifts present in the Southern ocean in both DePreSys and NoAssim, which appears to be generally similar. Therefore, the drifts seen in DePreSys, in AMOC and the density are, at least in part, due to the drift in the original transient runs. The Southern Ocean drift in the transient runs appears to be a major contributor to the bias in both systems. However, additional biases in DePreSys, such as the dipoles of density anomalies in the Drake Passage, and the rapid nature of the DePreSys drift, suggests that there may be a lack of balance in the initial conditions that also leads to adjustments.

5.4 Is the drift still present in a updated version of DePreSys? (DePreSys2)

Is the drift seen in sections 5.2 and 5.3 specific to the construction of the original DePreSys, or specific to the fundamental design of DePreSys? DePreSys2 used a consistent assimilation method, and covers a wider range of climate states than DePreSys, especially in the North Atlantic where the original DePreSys hindcasts are dominated by the warming of the subpolar gyre. Therefore, it would appear wise to make use of this data set to help understand the cause of the drifts seen in the original DePreSys. The comparison presented here is brief and is by no means an exhaustive analysis of DePreSys2.

A key feature of the original DePreSys hindcasts is the rapid decrease in the overturning strength in the North Atlantic (see figures 5.2 and 5.4). The decrease in the AMOC in the original DePreSys could just be the models response to the rapid warming of the subpolar gyre. Therefore, it is important to also understand the AMOC changes in DePreSys2.

Figure 5.11 shows the MOI at 50°N from the assimilation run of DePreSys2 (note it is the same data that is used in figure 2.9). The evolution of the MOI in the DePreSys2 assimilation shows an increase in the AMOC strength from ∼16Sv in the 1965, to ∼18Sv by 1990, before an increase to a peak of ∼21Sv in 1995, followed by a rapid decrease back to ∼18Sv in by 2005. The evolution of the MOI at 50°N in the DePreSys2 assimilation is clearly very different to the MOI at 50°N in the original DePreSys assimilation run which shows no interdecadal variability (fig.5.2 e). The strengthening of the AMOC in the DePreSys2 assimilation run is due to the updated ocean analysis resolving a coherent, positive density anomaly in the deep subpolar North Atlantic, which is spreading down the deep western boundary of the Atlantic (see figure 2.8).

Also shown in figure 5.11 is the MOI at 50°N from the DePreSys2 hindcasts. It is clear that many of the DePreSys2 hindcasts show a large weakening of the AMOC at 50°N, as seen in the original DePreSys hindcasts (figure 5.2). The DePreSys2 hindcasts, like the DePreSys hindcasts, seem to move toward an MOI of ∼16Sv, which is lower than the mean overturning state of the DePreSys and DePreSys2 assimilation runs. The changes

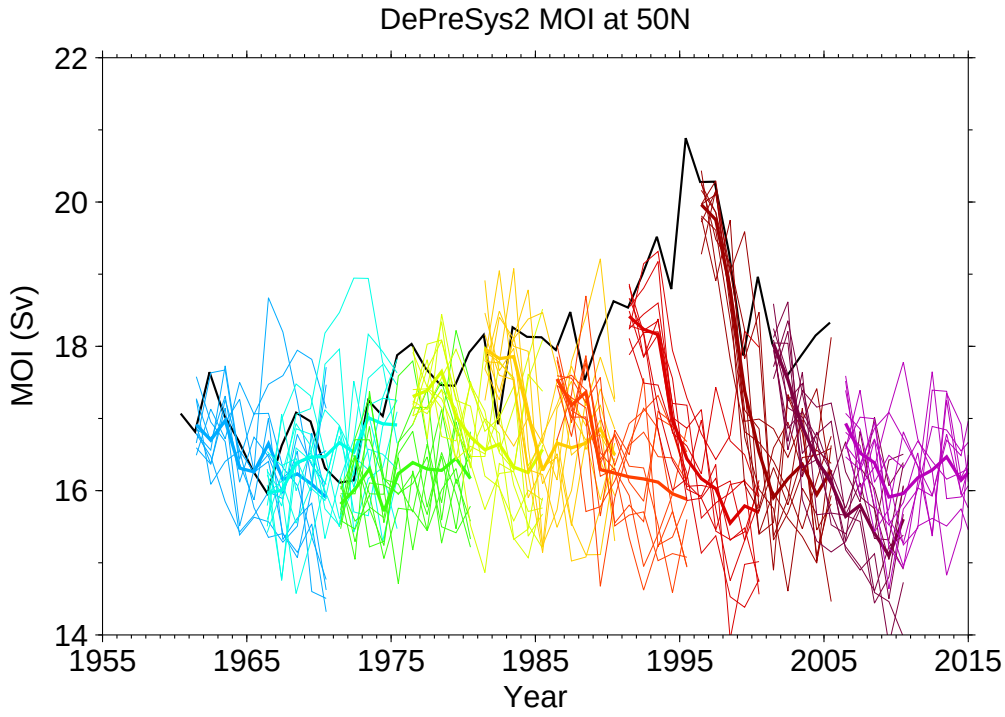


Figure 5.11: The evolution of the MOI at 50°N from the DePreSys2 data set. The MOI from the assimilation run is in black. The coloured lines show the MOI from the 10 DePreSys ensembles. The thin lines show the individual ensemble members and the thick coloured line denotes the ensemble mean. The changing colour for each ensemble member is to aid the eye.

in DePreSys2 MOI can also be extremely rapid, with an annual mean decrease of 4Sv in a couple of years for the November 1995 hindcast.

The mean evolution of the Atlantic stream function from the DePreSys2 hindcasts can be evaluated in a similar way to DePreSys and NoAssim (e.g. using equation 5.1). Because the DePreSys hindcasts are only initialised in November an annual mean climatology is calculated by taking the average of the first annual mean of all the DePreSys2 hindcasts to give a climatology:

$$C^{\psi}_{DeP,2} = \frac{1}{4 \times lm} \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{t=1}^{4} D^{\psi,2}_{ijt}$$

Figure 5.12 shows the mean evolution of the Atlantic stream function in DePreSys2, which should compare with figure 5.4. The first year is omitted from figure 5.12 because the mean anomaly for the first year is zero by construction. In the North Atlantic in particular, the mean Atlantic meridional stream function evolution of DePreSys and

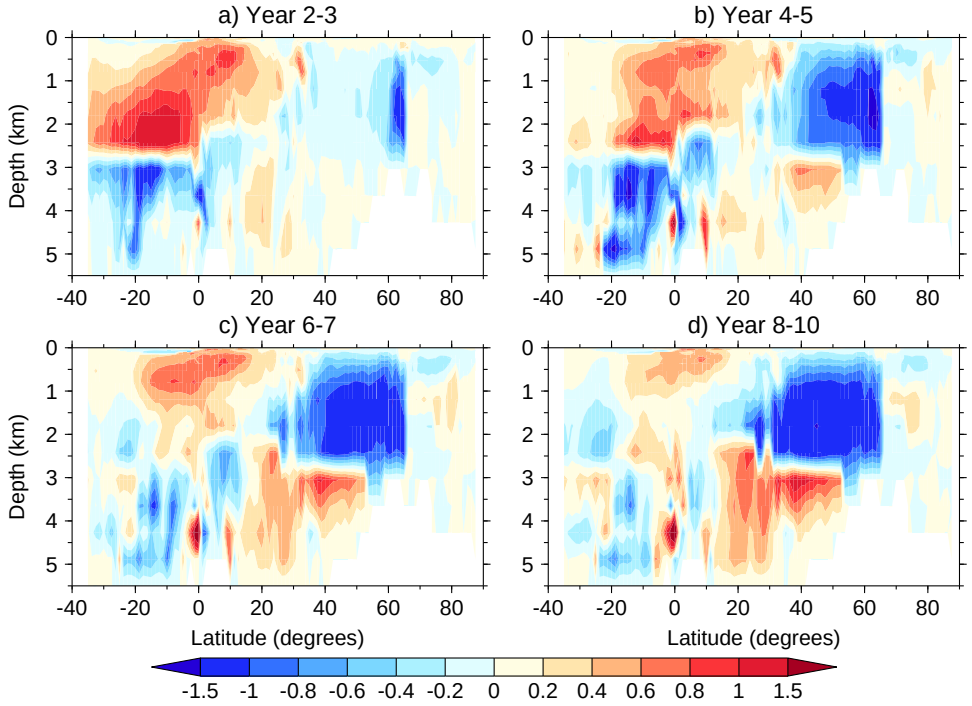


Figure 5.12: The mean evolution of the Atlantic meridional stream function in the DePreSys2 hindcasts, relative to a stream function climatology calculated from the mean of the first year from the same DePreSys2 hindcasts.

DePreSys2 are very similar. The DePreSys2 hindcasts see a large reduction (>1Sv) in North Atlantic AMOC strength, between 30°N and 70°N. An increase in the AABW cell strength below the decrease in the AMOC is also present in DePreSys2, similar to DePreSys. The mean evolution of the stream function is somewhat different at the latitudes south of 20°N: an increase of ∼1Sv in the AMOC strength between 40°S - 20°N, peaking in year 2-3 before reducing to zero. Below the increase in AMOC strength between 40°S - 20°N there is an initial weakening of the AABW cell greater than 1Sv which peaks in years 4-5, before reducing towards zero.

Finally, examining the evolution of the mean density anomalies from DePreSys2, which are constructed in the same way as those for NoAssim 1 (i.e. the first seasons density anomaly is removed as a bias), shows that the same density drift is present in the DePreSys2 hindcasts. A low density signal begins to propagate down the western boundary of the North Atlantic after 3-4 years, which continues to intensify to a mean density anomaly greater than -2σ. A positive density anomaly also builds up along the coast of

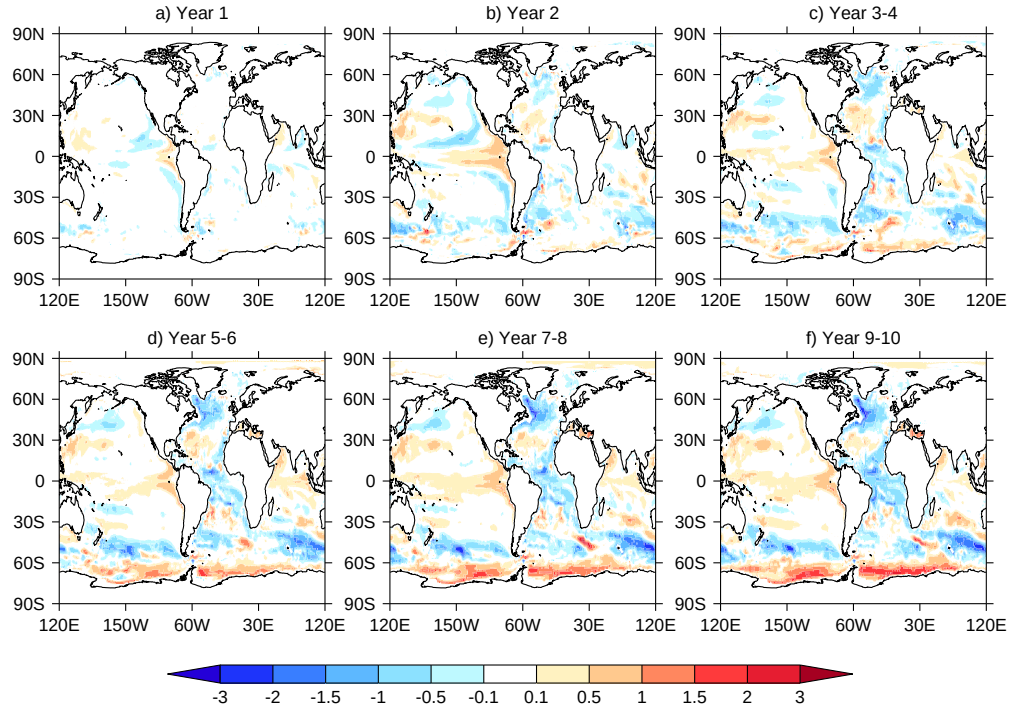


Figure 5.13: The 800-3000m integrated in situ density anomalies averaged over the DePreSys 2 hindcasts (1960-2005). The density anomalies are normalised by the interannual standard deviation of density anomalies from the HadCM3 control run

Antarctica, with low density anomalies seen in the South Pacific and Indian oceans.

The similarity of the AMOC and density drift for DePreSys and DePreSys2 informs us that the drift is not only specific to the original DePreSys. Therefore, the presence of the rapid changes could highlight issues with the fundamental construction of DePreSys. The similarity of the drift also gives us confidence that the drift, seen in the North Atlantic in the original DePreSys hindcasts, was not just a response to the initial conditions; i.e. that the reduction in the AMOC was not just a reduction in overturning strength from a peak in the mid 1990s.

5.5 What is causing the drift?

The potential predictability of climate variables on decadal timescales, in the Atlantic sector, is a direct consequence of long time scale variability in the strength of the AMOC

(*Griffies and Bryan* (1997b); *Collins and Sinha* (2003); *Knight et al.* (2005); *Sutton and Hodson* (2005); *Collins et al.* (2006a)). Therefore, if initialised climate predictions are to improve forecasts of the climate over the upcoming 10 years in the Atlantic sector, relative to un-initialised predictions, correct initialisation of the AMOC is an important priority. The drifts seen in the AMOC in DePreSys will, at best, limit the ability of the system to improve predictions of climate in the North Atlantic relative to un-initialised climate predictions. Trying to understand the cause of the drifts, and to identify improvements, is, therefore, crucial for delivering more skillful forecasts in the future, and for initialised climate predictions to attain their potential. This section of the thesis will focus upon discussing potential causes for the large drifts that are seen in the DePreSys hindcasts.

5.5.1 The construction of the climatology

As already shown in figures 5.4 and 5.5, the drift in the AMOC in the DePreSys hindcasts is, at least in part, explained by the drift in the transient runs: in particular transient 1. However, why do the changes occur faster than that seen in NoAssim? Continued drift towards HadCM3's preferred state would be expected by initialising the hindcasts from a state that is drifting, as is the case with the NoAssim hindcasts (see figure 5.1). However, the DePreSys hindcasts are often initialised from the climatological mean state, especially away from the surface or in areas of the ocean that is poorly observed. The climatology is an average of states from a transient integration undergoing non-linear drift. In a non-linear system there should be no reason to assume that the drift should be the same as that seen in the original transient run. Therefore, the climatology used by DePreSys and DePreSys2 in the assimilation may be a particular state that encourages the evolution seen in the previous sections; in other words, hindcasts that are initialised from that particular climatology could all evolve in the way seen in DePreSys and DePreSys2.

Ignoring the issue of climate drift in the original transient integration for now, the hindcasts could still posses an average adjustment or drift if the climatology does not sample low frequency variability sufficiently. The climatology used to initialise DePreSys and DePreSys2 was only taken from one transient run. As discussed in chapter 3, HadCM3

possesses multi decadal and centennial variability in the North Atlantic (*Vellinga and Wu* (2004); *Hawkins and Sutton* (2007)). If the climatology was calculated over a period where, for example, the AMOC was weakening due to natural variability, then, assuming the climate system is linear, the AMOC in the hindcasts would be likely to be initialised with a tendency for a weakening AMOC. The climate system is not linear; therefore, the evolution of hindcasts initialised from such a climatology is likely to be more complicated.

If the drift is caused by imbalance in the model climatology (C_1) caused purely by the construction of the climatology, then it follows that the anomaly assimilation technique could be sensitive to the climatology used. In other words, hindcasts (or forecasts) are dependent on the climatology used in the assimilation stage, and not just on the observed anomalies that are assimilated. This led to hypothesis 1:

Hypothesis 1:- DePreSys hindcasts can be sensitive to the model climatology that is used in the assimilation.

If such dependence exists, it could cause problems for a prediction system that uses an anomaly assimilation method, especially when the observational coverage is poor and the model is often initialised from climatology. For example, how does one decide on the "best" climatology to use for the assimilation. Hindcast successes may be failures when using a different climatology. The sensitivity of DePreSys hindcasts to the climatology (**Hypothesis 1**) will be tested in chapter 7.

5.5.2 The non-linear equation of state

The anomaly assimilation method relaxes the assimilation run to monthly mean T and S fields. Hence, the background density field for the assimilation run is the density calculated from these climatological means ($\rho(\overline{T}, \overline{S})$), not the mean density ($\overline{\rho}(T, S)$). The equation of state for sea water is non-linear and, therefore, does not guarantee that the density calculated from mean T and S, is the same as the mean density:

$$\overline{\rho}(T, S) \neq \rho(\overline{T}, \overline{S})$$

It follows that relaxing to mean T and S fields does not guarantee that the initial states will be balanced, and that no initial adjustments will take place.

Is the difference between $\overline{\rho}(T, S)$ and $\rho(\overline{T}, \overline{S})$ large enough to cause a problem in the initialisation of DePreSys? Figure 5.14 shows the difference between the 1941-1996 density climatology calculated from the annual mean 1941-1996 T and S climatologies used for the original DePreSys assimilation run, and the density climatology calculated from annual mean density fields over the period 1941-1996 from transient 1. The density climatology used in the DePreSys assimilation is clearly overestimating the density by $>0.5\sigma$ in many areas of the top 800m of ocean (see figure 5.14 a and b), especially in the North Atlantic, such as the Gulf Stream Extension, and also in the Irminger, Labrador and the GIN seas. The Pacific is also a region with a large over estimate of the climatological density. In the top 113m there is an overestimation in the North Pacific, especially in the Kuroshio Extension, and in the South Pacific. Between 113m and 800m there is a large over estimate in the density in the Western Tropical Pacific of approximately $\sim 0.5\sigma$.

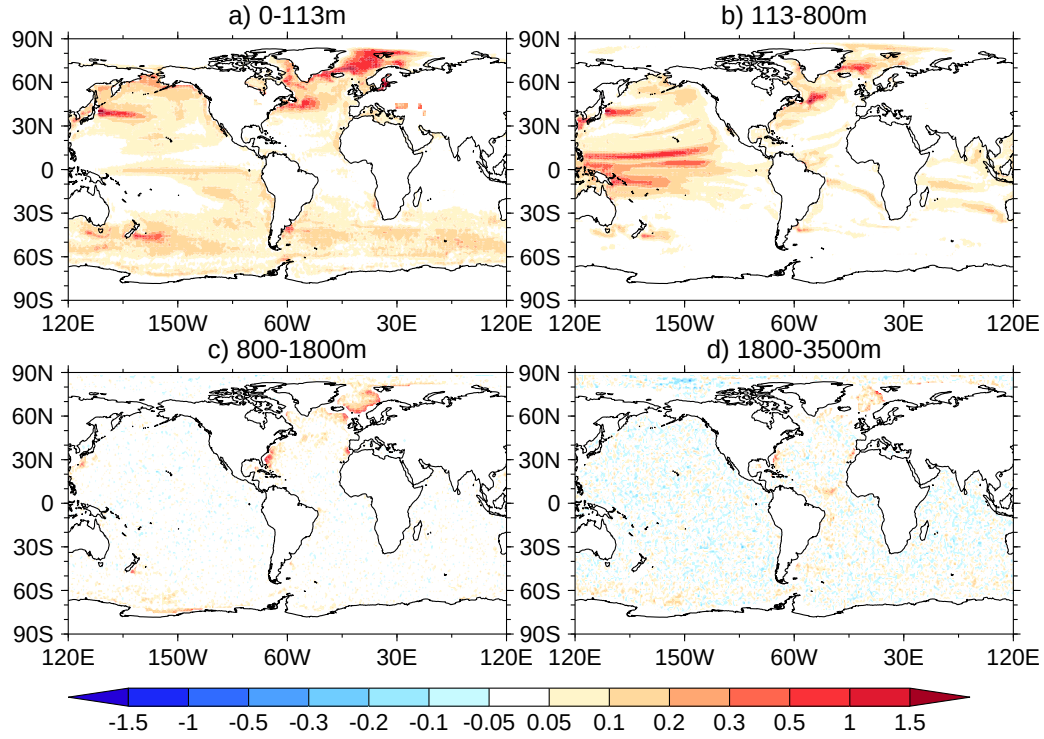


Figure 5.14: The difference between in situ density calculated from annual mean 1941-1996 T and S C_1 climatology, and the mean in situ density calculated over the same period from Transient 1. The difference has been normalised by the interannual standard deviation of densities from the HadCM3 control run.

Below the top 800m there are also differences in the climatologies, especially in the North Atlantic. Interestingly, an over estimation of the density in the deep western boundary between 20°N-40°N $>0.5\sigma$. There is also an overestimation of the density climatology along the continental shelf in the GIN Sea and along the Greenland-Scotland ridge, and a slight overestimation of the density climatology along the boundary of the Labrador Sea. Finally, between 1800-3500m, there is an over estimate of density anomalies in the GIN Sea, and along the deep western boundary of the North Atlantic.

Although figure 5.14 shows that there is a significant difference between the density climatology derived from the mean T and S fields, rather than the mean density itself, it only shows the differences between densities calculated from annual average fields. The difference between the monthly climatologies would be more relevant, because of the use of monthly mean climatologies in the assimilation. However, the monthly mean data was not available from the transient runs. Nevertheless, the difference between monthly mean climatological density and the density calculated from the monthly mean T and S could be important. The difference between $\overline{\rho}(T, S)$ and $\rho(\overline{T}, \overline{S})$ is likely to be different for each monthly climatology. Therefore, imbalance, and hence adjustment, caused by this effect could be different for hindcasts initialised at different times of the year, or for hindcasts initialised from a different climatology (i.e. calculated over a different period, or from a different source). For example, DePreSys PPE, which uses a different climatology for each ensemble member, could have a different common adjustment for each separate ensemble member.

5.5.3 The use of Dynamic Assimilation

The strong relaxation applied in the assimilation run could also be causing problems, especially as the observational coverage of the ocean is poor. Using a strong relaxation in atmospheric models can lead to large adjustments once the relaxation is removed (*Ramamurthy and Carr* (1987); *Stauffer and Seaman* (1990)), suggesting that the model is not in dynamical balance. *Bao and Errico* (1997) highlighted that even with a relatively long relaxation time, the relaxation terms that are added to the prognostic model equations, can be relatively large. While the relaxation is applied, the model comes into balance

with the model equations, of which the nudging term is a significant proportion. Therefore, once the relaxation is removed, the model state created by the dynamic assimilation may not be in balance with *only* the model equations.

The relaxation time in DePreSys is extremely short and it is likely that the prognostic equations of HadCM3 are dominated by the relaxation term. Thus, the strong relaxation applied in the assimilation run could cause the DePreSys initial states to be unbalanced, especially as the observed anomalies may not be consistent with the model itself. Also, there is no effort to ensure that the observations of T and S are in balance before the assimilation stage. Imagine an extreme example of the assimilation of the observed T and S creating a convectively unstable water column. If the model had enough freedom, the instability would be mixed out. However, with the strong relaxation applied in the DePreSys assimilation run, the instability would instead be maintained until the monthly mean anomalies of T and S did not cause such instability, or until the relaxation was removed. In the case of the latter, the model would then quickly convect and adjust to remove the instability. The idea that the strong relaxation causes some of the imbalance fits with the rapid and instant reduction in the MOI of DePreSys hindcasts from 1994 onwards, where the ocean relaxation was applied as 6 hours everywhere (see figure 5.2).

5.6 Chapter Conclusions

Chapter 3 showed that the transient runs used in the construction of DePreSys and NoAssim were undergoing significant drift, especially in the Atlantic. Therefore, similar drifts would be expected in the DePreSys and NoAssim hindcasts. This chapter presents an analysis of the mean evolution of the DePreSys and NoAssim hindcasts that was undertaken to investigate the potential drift. The main findings are as follows:

- DePreSys (and DePreSys2) hindcasts share a common drift - a surprise given that the anomaly assimilation method was designed to remove the problem of drift. Large, and rapid, drifts are present in the Atlantic Meridional Overturning Circulation (AMOC), especially in the North Atlantic at 50°N where the AMOC reduces

$\sim$2Sv in only a few years. There are also large changes in density in other areas of the global ocean: notably in the Southern Ocean, and especially in the Drake Passage.

- The NoAssim hindcasts show a similar drift to DePreSys in the AMOC and density fields. Therefore, the drift in DePreSys is, at least in part, a manifestation of the drift in the original transient run used to construct the climatology. However, the drift in DePreSys is usually larger, and occurred at a faster rate than the drift seen in NoAssim, which suggests that the initial conditions for the DePreSys hindcasts may not be in balance.

- The larger, and more rapid, drift seen in DePreSys could be a non-linear response to the drift in the original transient integration. The sub sampling of low frequency variability in the model, for example, the construction of the climatology from a period where the AMOC was decreasing in the transient integration may also be expected to cause a mean drift.

- By relaxing the model to climatologies of mean temperature (T) and salinity (S), the anomaly assimilation method ignores the non-linear equation of state for sea water. The relaxation of the assimilation run to a density climatology that is a function of mean T and S could cause imbalance. The density derived from mean T and S fields is not the same as the mean density field (i.e. $\overline{\rho}(T, S) \neq \rho(\overline{T}, \overline{S})$): a fundamental limitation of the anomaly assimilation method. The difference in density fields, and hence the drift caused, will be different for every climatology calculated.

- Imbalance could also be caused by the use of an exceptionally short relaxation time in the assimilation stage. When short relaxation terms are used the prognostic equations can become dominated by the relaxation term. In the assimilation run the model comes into balance with the prognostic equations plus the relaxation term. Therefore, once the relaxation is removed the model state is no longer in balance with only the prognostic equations.

Understanding the origin of the drift and rapid adjustments seen in DePreSys is an important goal and should be investigated further. The relaxation to mean T and S fields and

the strong nudging are integral parts of the current anomaly assimilation method. Testing the importance of such effects, especially the relaxation to mean T and S, will require significant changes to the anomaly assimilation technique, and will not be investigated further in this thesis.

If the drift is caused by the use of the particular climatology C_1 that is used in the DePreSys and DePreSys2 hindcasts, then it follows that the hindcasts could be sensitive to the climatology used in the assimilation run; the hindcasts are not just reacting to the observed anomalies. This led to the hypothesis:

Hypothesis 1:- DePreSys hindcasts can be sensitive to the model climatology that is used in the assimilation.

which can be tested relatively easily by changing the climatology in the assimilation. Hypothesis 1 will be tested in chapter 7.

The difference between $\overline{\rho}(T, S)$ and $\rho(\overline{T}, \overline{S})$ highlights that the anomaly assimilation method ignores the non-linear equation of state. Even if the relaxation to mean T and S does not contribute to imbalance, is the implicit assumption of a linear equation of state a good one? The next chapter will investigate this assumption by analysing the effect of the non-linear equation of state upon the initialisation of density anomalies in DePreSys.

Chapter 6

Case study in decadal climate prediction: the rapid warming of the North Atlantic in the mid 1990s

6.1 Introduction

Chapters 4 and 5 both focussed on understanding the **mean** performance of DePreSys and NoAssim. Although mean skill score analyses, like those presented in chapter 4 and by *Smith et al.* (2007), are informative as to whether a particular prediction system is giving better predictions in general, they are not informative to reasons why the initialised predictions are performing better than the un-initialised. The mean skill scores give little insight into the mechanisms that are involved in improving the predictions. Is DePreSys actually able to predict changes seen in the climate system, or is the assimilation of information just correcting the initial bias of the transient runs, as suggested in chapter 4? Perhaps more importantly, the mean skill scores do not give insight into when, where or why DePreSys hindcasts fail. Implementing improvements to the design of the prediction system, or indeed the underlying model, is absolutely dependent upon understanding what causes the hindcasts to fail.

This chapter moves beyond the mean skill scores. The work contained here was origi-

nally completed before the discovery of major drifts in the DePreSys hindcasts, and so the original conclusions have been somewhat altered. Nevertheless, the analysis contained here highlights an interesting limitation of the anomaly assimilation method. Initially, section 6.2 highlights the importance of examining the performance of DePreSys on a hindcast by hindcast basis for a particular case study, the rapid warming of the North Atlantic subpolar gyre (SPG) in the 1990s, which has already been discussed in chapter 2. The rapid warming of the North Atlantic was one of the largest events in the hindcast period for DePreSys and NoAssim (1982-2001), and presents a unique challenge for initialised climate predictions. Analysing the evolution of individual hindcasts for this event allowed the framing of hypotheses for the reasons behind prediction successes and failures, which pointed towards a limitation of the anomaly assimilation method. The assimilation of information into DePreSys creates errors in the assimilated density anomalies. The origin of the density anomaly errors is discussed in section 6.3.1 and the potential impact of the density errors upon the prediction of the rapid warming is discussed briefly in section 6.3.2.

6.2 Rapid warming case study

To gauge the improvement, or indeed the degradation, in prediction skill caused by the assimilation of observed information into DePreSys, the NoAssim hindcasts are examined first for perspective. The ensemble mean NoAssim hindcast average temperature anomalies for the North Atlantic SPG, are compared to the observed heat content anomaly in figure 6.1. The initial conditions of the NoAssim and DePreSys hindcasts were not available for the analysis, and so the first season of the hindcasts is assumed to be representative of the initial conditions, and is highlighted by the dot. The ensemble means of the “initial condition” for the NoAssim hindcasts do not lie near the observations for much of the hindcast period, as expected due to the lack of assimilation. A warming trend is apparent in the NoAssim “initial conditions” after 1995, but the observed warming is not contained by the envelope of NoAssim hindcasts, giving further evidence that the warming of the SPG was not just a response to radiative forcing.

The "initial conditions" of the NoAssim hindcasts show large multidecadal variability over the course of the hindcast period. The cooling in the early 1980s and in 1993-95 follow the two large eruptions that took place in this period, El Chichón and Pinatubo. This suggests that variability in the SPG heat content could be connected with variability in radiative forcing, particularly due to sulphate aerosols, and suggests that the eruption of Pinatubo in 1992 could have contributed to the observed cooling of the SPG between 1992 and 1995. However, caution is strongly encouraged for the hypothesis that the variability is being created by volcanic forcing when the spread of the initial season of the NoAssim hindcasts (and hence the transient runs) is examined. The variability of the ensemble mean does not lie outside of the spread in the transient runs, suggesting that the variability of the ensemble mean "initial conditions" of NoAssim (i.e. the cooling after Pinatubo, before a warming after 1995) is likely caused by sampling error. There does seem to be a reduction in the ensemble spread of the initial conditions in 1994 and 1995, but the difference is unlikely to be significant due to the small size of the sample (4 members).

The DePreSys hindcasts start much closer to the observed anomalies than the NoAssim hindcasts, as expected. The envelope of ensemble mean temperature from the DePreSys hindcasts seems to capture the cooling and warming of the SPG in the early 1980s (fig 6.1). However the envelope of the DePreSys hindcasts does not capture the cooling of the SPG in early 1990s, and instead appears to be generally warmer than the NoAssim hindcasts. More positively, the envelope of ensemble mean DePreSys hindcasts is able to capture the warm SPG from 1996 onwards, unlike the envelope of ensemble mean NoAssim hindcasts. Examining the DePreSys hindcasts in detail (see figure 6.2) reveals that the rapid transition of the SPG from cold to warm in 1995/1996 is captured with remarkable levels of skill by many of the DePreSys hindcasts that were initiated immediately before the transition of the SPG. The June and September 1995 hindcasts especially, seem to capture the warming with remarkable skill, and with very small ensemble spread, including the initial warming and then cooling before warming again.

Examining the other DePreSys hindcasts of SPG heat content more closely (figure 6.2) it quickly becomes apparent that many of the hindcasts initialised between 1990-1994 also warm rapidly and show small ensemble spread (for example March 1990 and 1991).

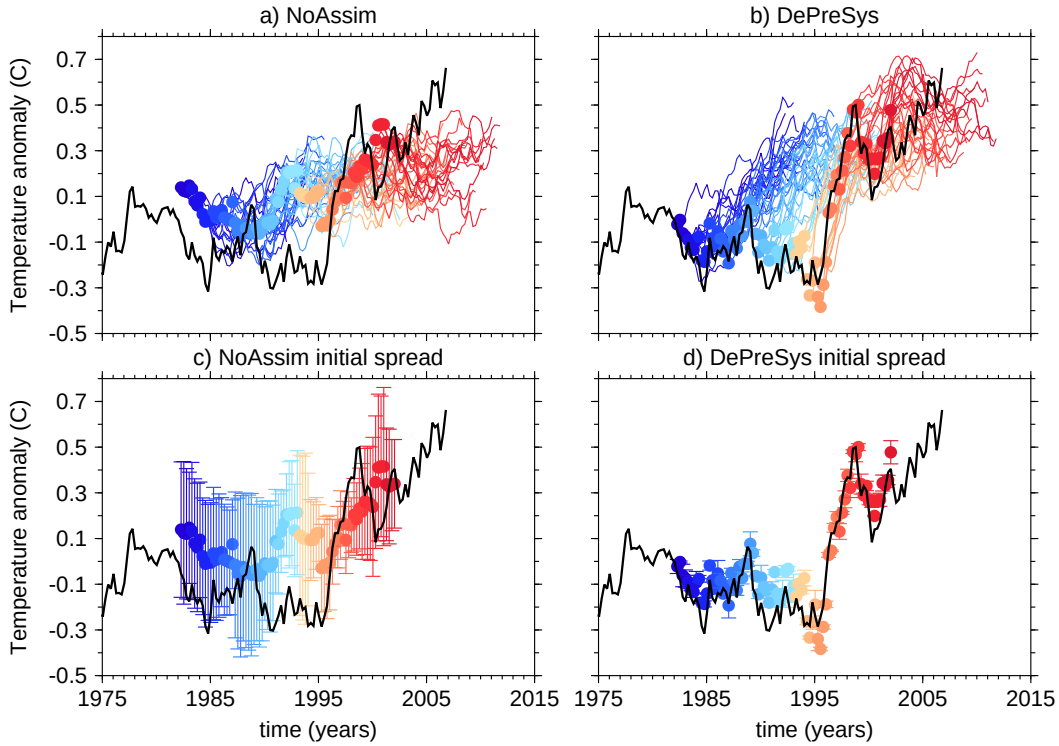


Figure 6.1: Ensemble mean hindcast plumes of North Atlantic SPG (60°W-10°W, 50°N-66°N) top 500m average temperature anomalies for NoAssim (a)) and DePreSys (b)) compared to the observed (black) from the Original Met Office ocean analysis. The dots highlight the first seasons average temperature anomaly from the hindcasts (the closest state to the initial condition) and the changing colours denote the separate hindcasts. c) and d) show the initial conditions compared to the observed average temperature anomalies, as is a) and b), but also shows the 95% confidence interval of the ensemble mean, to highlight the initial spread of the hindcasts.

Although not all of the hindcasts warm instantly after initialisation (September 1991, for example, shows no warming for over a year) the majority of the hindcasts in this period do warm rapidly within a few seasons from initialisation. The small ensemble spread of these DePreSys hindcasts indicates vastly overconfident predictions of the change in SPG heat content. The rapidly warming hindcasts quickly develop forecast errors and these errors are likely to be the cause of the initially rapid increase in positive bias in upper ocean heat content of the North Atlantic SPG in the first few years of the DePreSys hindcasts seen in figure 4.9.

By warming before the observed heat content in the SPG, the early rapid warming hindcasts in DePreSys between 1990 and 1994 degrade the mean skill of DePreSys relative

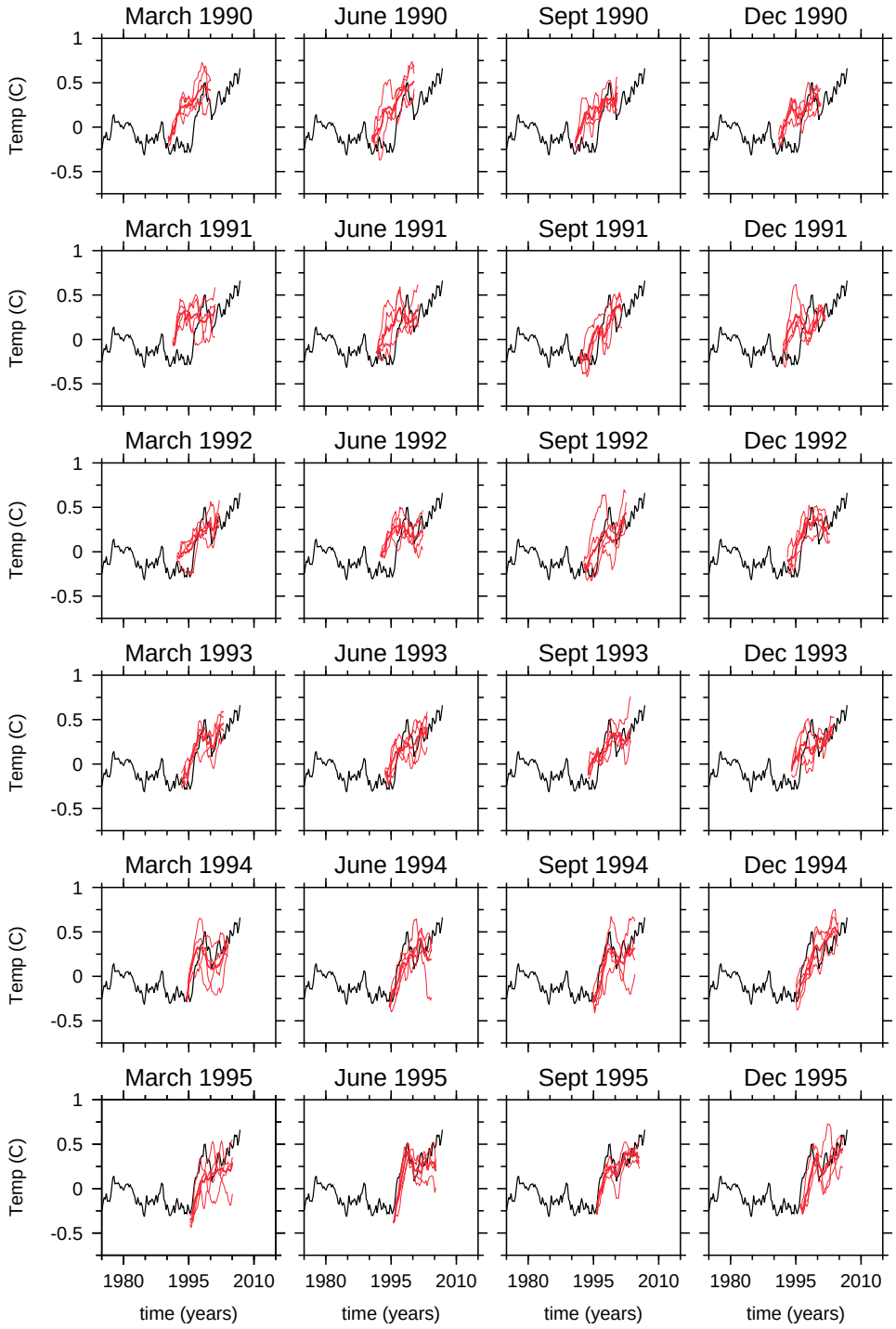


Figure 6.2: Hindcast plumes of North Atlantic SPG (60°W-10°W, 50°N-66°N) 500m temperature for the 1990-1995 DePreSys hindcasts (red), observed North Atlantic SPG 500m heat content calculated from the Original Met Office ocean analysis is shown in black. The thin red lines show each member of the ensemble and the thick red line shows the ensemble mean.

to NoAssim in the North Atlantic, especially in the first couple of years. If the DePreSys hindcasts initialised between 1990-1994 did not warm rapidly, then the skill of DePreSys in the 2nd and 3rd year of figure 4.4 would be higher in the North Atlantic. Thus, it is vitally important to understand why the DePreSys hindcasts warm early in this period, and give such over confident predictions. If the hindcasts warm rapidly due to a deficiency in either the construction of DePreSys or HadCM3 then understanding why the early warming hindcasts occurred could improve future decadal predictions.

Initially to understand why the DePreSys hindcasts warmed rapidly, the understanding of the rapid warming obtained in chapter 2 was used to frame hypotheses for the early warming hindcasts. The proposed mechanism of the warming, outlined in chapter 2, involves a build up of positive density anomalies across the subpolar gyre in the early 1990s, in response to the persistent positive NAO forcing. The increase in density anomalies could have caused a strengthening of the AMOC in the mid 1990s through an increase in deep convection, which possibly led to the rapid warming of the SPG through an increase in the northward heat transport. If the warming in the real world was dependent on changes in the circulation, and especially a strengthening of the AMOC, then the evolution of the DePreSys hindcasts will be dependent on the density anomalies that are assimilated into the initial conditions for the hindcasts. Therefore, understanding why the hindcasts warm rapidly started with an analysis of the assimilated density.

6.3 Density errors

In HadCM3, as shown in chapter 2, the AMOC is highly correlated with density anomalies in the SPG region (see figure 2.7). One possible hypothesis is that the model reacted too strongly to the density anomalies that are assimilated into the North Atlantic in the early 1990s, which caused the AMOC to strengthen too early, or too strongly. In other words, deep convection, and hence deep water formation, in HadCM3 was too sensitive to the density anomalies that were assimilated into the North Atlantic the early 1990s. The overestimation of deep convection and deep water formation may have caused a strengthening of the AMOC and hence caused the SPG to warm too early. However, analysis of

the stratification in the regions of the ocean with positive density anomalies in the early 1990s (i.e. the Labrador and Irminger Seas) showed that HadCM3 is *more* stably stratified in these regions than the ocean analyses, suggesting that deep water formation in HadCM3 was unlikely to be too sensitive to the assimilated density anomalies.

Although the mean stratification of the model cannot explain the rapid early warming hindcasts, it is nevertheless a good example of how systematic errors can affect the evolution and accuracy of hindcasts, even when only anomalies are assimilated. HadCM3 is more stably stratified than the observations and, therefore, DePreSys is unlikely to be able to simulate the same magnitude of AMOC strengthening which would occur in the real world following positive density anomalies in the North Atlantic.

6.3.1 Errors in assimilated density anomalies

The anomaly assimilation technique outlined in chapter 4 adds the observed anomalies of temperature and salinity onto the model climate, so that the correct anomalies of temperature and salinity can be initialised in the model. However, to be able to predict any dynamical changes in the ocean's state, such as a strengthening or weakening of the AMOC, it is the density anomalies that need to be initialised correctly. Therefore, the purpose of the assimilation of temperature and salinity observations is also to initialise the density field, and hence the low frequency variability of the ocean.

As already stated in chapter 3, the models and the observed temperature and salinity climatologies are different, hence why it was chosen to assimilate anomalies. Figure 6.3 highlights the size of the mean model error (C_{mod}^{error}) for both temperature and salinity for a number of layers in the ocean. In the top 100m there are large temperature and salinity errors in the North Atlantic south of Newfoundland and the southern area of the Greenland-Iceland-Norwegian (GIN) seas, with errors of over 5°C and $\sim$0.5PSU (see figure 6.3 a and d). These are both areas of large gradients in temperature and salinity (not shown). In particular the errors south and west of Newfoundland are associated with errors in the Gulf Stream and North Atlantic Current. In HadCM3, the gulf stream separates from the east coast of North America further north than in observations, and the

North Atlantic Current is too zonal (*Gordon et al.* (2000)). Large errors in temperature and salinity are also present at depth in the ocean 100-3000m). Mean model errors are in excess of 2°C and ∼0.5PSU over much of the Atlantic (see fig. 6.3 b, c, e and f). Large temperature errors >3°C are seen in the tropical Atlantic and the GIN sea region in the 500-1200m layer, which is larger than much of the mean model error at the surface. The Arctic Ocean is also warmer and more saline in the model compared to observations.

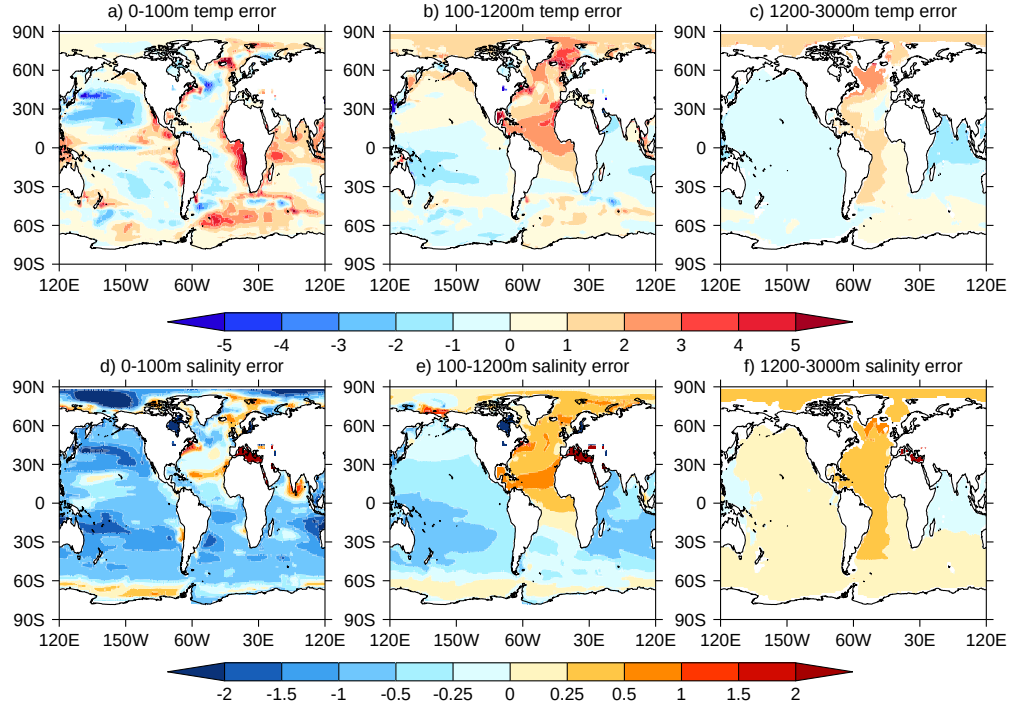


Figure 6.3: a)-c) shows the mean model error (C_{mod}^{error}) for temperature (°C) for 0-100m, 100-1200m and 1200m-3000m layer respectively. d)-f) show the same but now for salinity in practical salinity units (PSU).

The equation of state for sea water is a non-linear function of temperature, salinity and pressure. However, the anomaly assimilation method used by DePreSys adds the observed anomalies of temperature and salinity onto the model climatology, effectively ignoring this non-linearity. The anomaly assimilation approach implicitly assumes that the same density anomaly is created. However, as figure 6.3 highlights, large model errors are present over large volumes of the ocean, including regions which may be sensitive to density anomalies. For example, model errors seen in the North Atlantic for the 1200-3000m layer (see fig. 6.3 c)) have a similar pattern to the spread of North Atlantic Deep Water along the western boundary of the Atlantic (*Bower et al.* (2009)). This is a region that is important for the AMOC strength in HadCM3 (see figure 2.7), which raises the

question: Is the implicit assumption of linearity for the creation of density anomalies a safe assumption to make?

To understand the sensitivity of the assimilated density anomalies to the mean model error (C_{mod}^{error}), the density anomalies are calculated by placing the observed anomalies of temperature and salinity onto both the observed and model climatologies. The observed anomalies of temperature and salinity are taken from the Original Met Office ocean analysis and are defined as:

$$T_i' = X_i^T - C_A^T$$

$$S_i' = X_i^S - C_A^S$$

where i denotes the month (i.e. it is an index of month and year). The climatological period used to define both C_A^T and C_A^S is 1941-1996, as for the original hindcasts. Thus, the **observed** density field is given by:

$$\rho_{obs}(i) = \rho(C_A^T + T_i', C_A^S + S_i')$$

Density anomalies are calculated by removing the density climatology, which is calculated from the climatological mean fields of temperature and salinity.

$$\rho_{obs}'(i) = \rho_{obs}(i) - C_{obs}^{\rho}$$

where

$$C_{obs}^{\rho} = \rho(C_A^T, C_A^S)$$

For the assimilation run, T_i' and S_i' are added to the model climatology for temperature and salinity ($C_{mod,A}^T$ and $C_{mod,A}^S$). The strong relaxation of the assimilation run will ensure that assimilated density anomalies should be very close to the density anomalies calculated off-line here. However, it is possible that some small differences may arise, especially due to the interpolation in time between monthly mean fields. Therefore, for clarity, the density anomalies calculated here that use *only* the monthly mean model climatology and monthly mean observed anomalies will be called **Quasi-Assim**.

$$\rho_{Q-Assim}(i) = \rho(C_{mod,A}^T + T_i', C_{mod,A}^S + S_i')$$

Quasi-Assim density anomalies are calculated by removing the density field calculated from just the assimilation climatology ($C_{mod,A}$).

$$\rho_{Q-Assim}'(i) = \rho_{Q-Assim}(i) - C_{Q-Assim}^{\rho}$$

where

$$C^{\rho}_{Q-Assim} = \rho(C^{T}_{mod,A}, C^{S}_{mod,A})$$

If the anomaly assimilation method was working perfectly then the observed and assimilated density anomalies would be equal ($\rho'_{obs}(i) = \rho'_{Q-Assim}(i)$). If they are not the same, then the difference in the density anomalies will come *only* from the change of climatologies. The error is defined as the Quasi-Assim density anomaly minus the observed.

$$\rho'_{error}(i) = \rho'_{Q-Assim}(i) - \rho'_{obs}(i) \tag{6.1}$$

The calculation of the density anomalies from the temperature and salinity fields is computationally much cheaper than the use of an assimilation run. Therefore, for a clearer understanding of the effect of the mean model error on the initialisation of density anomalies, the density anomalies have been calculated for the entire observed period 1957-2007 and not just for the hindcast period.

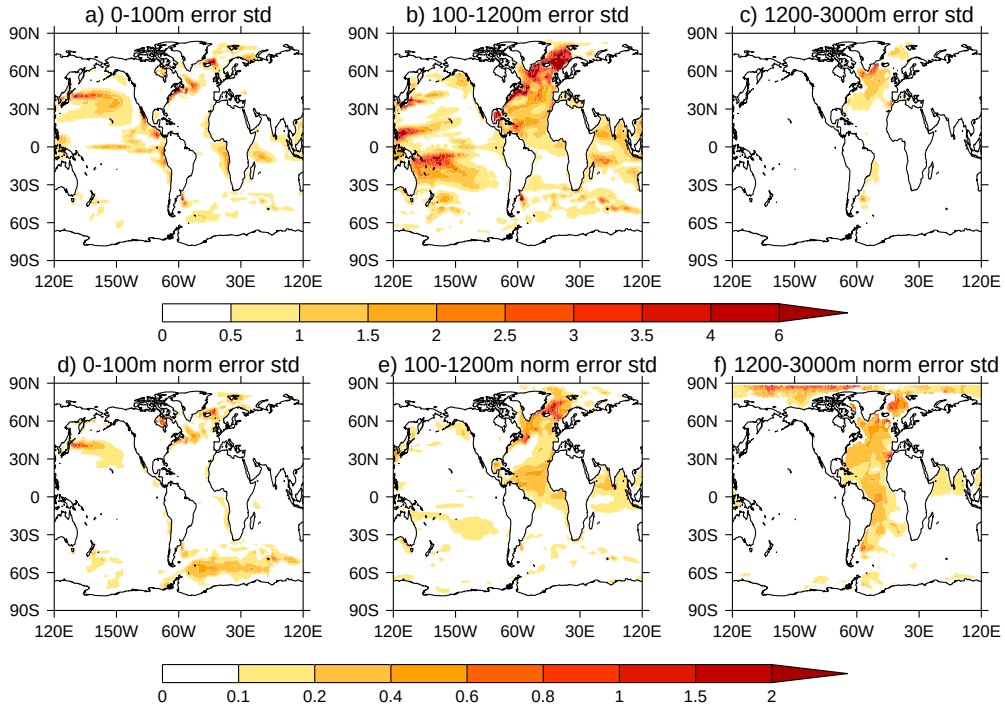


Figure 6.4: a), b) and c) show the standard deviation of the assimilated density anomaly error (ρ'_{error}) for the 0-100m, 100-1200m and 1200-3000m total integrated in situ density anomaly (kg m^{-3}) respectively. d), e) and f) show standard deviation of the assimilated density anomaly error after it is normalised by the standard deviation of the observed density anomalies (ρ'_{obs}). All anomalies are calculated using the Original Met Office ocean analysis.

The errors in the assimilated density anomalies are unlikely to be systematic, because the error will depend on both the temperature and salinity anomalies that are present at each ocean analysis time. Thus, it would be of little use to look for a mean error. Figure 6.4 a)-c) highlights where errors frequently occur by plotting the standard deviation of the density anomaly errors. It is clear that errors in the assimilated density anomalies occur over large areas of the top 3000m of the world's ocean, and, as expected, occur in regions where the mean model error is large (see fig. 6.3).

In the top 100m, errors are present where the errors in the Gulf Stream path are located. There are also errors in the North Pacific, and possibly of greater concern, the tropical East Pacific. The density anomaly errors in this region could cause problems in the initialisation of ENSO variability, which can have a large effect on the global climate system (*Wallace et al.* (1998)). Therefore, the initialisation of density anomalies here is not only important for successful predictions of ENSO variability, but also to avoid spurious variability that could swamp predictable signals across the globe. In the 100-1200m layer (fig. 6.4 b), errors are extensive in the North Atlantic, and especially in regions such as the GIN, Irminger and Labrador seas. Large magnitude density anomaly errors are also present in the Western Pacific for the 100-1200m layer, and in the Indian and Southern Oceans. In the deeper ocean (1200-3000m) (fig. 6.4 c) errors are again occuring across much of the North Atlantic, and especially in Irminger and Labrador Seas.

Figure 6.4 shows that the mean model error causes errors in the assimilated density anomalies, but are these anomalies relevant or are they small and negligible? To highlight the relative importance of the density errors figure 6.4 d) - f) again shows the standard deviation of the density errors in the selected depth ranges, but now normalised by the interannual standard deviation of the observed density anomalies. The normalisation highlights areas where the magnitude of the density errors is large compared to the inter-annual variability; in other words, where the density errors are large enough in magnitude to make a significant impact on the evolution of the assimilation and hindcasts.

The magnitude of the mean density anomaly errors in the Pacific, especially in the 100-1200m layer (figure 6.4 b) are not large when compared to the observed natural variabil-

ity. In contrast, the mean density anomaly errors in the North Atlantic subpolar gyre and GIN seas regions are relatively large compared to the observed variability. The AMOC strength in HadCM3 is correlated with density anomalies in these regions (see figure 2.7). Therefore, the presence of large errors in the assimilated density anomalies could make the correct initialisation of the AMOC in DePreSys difficult. The potential predictability on decadal time scales in the Atlantic sector is strongly associated with the potential predictability of low frequency variability in the AMOC (*Griffies and Bryan* (1997b); *Latif et al.* (2006)). Thus, the assimilated density anomaly errors could, in principle, strongly diminish the capability of DePreSys to make skillful predictions at lead times greater than a few years, especially in the Atlantic sector.

It is important to note that the density anomalies shown here are a mean error over the period 1950-2007. The actual errors in the assimilated density anomalies are non-linearly dependent upon the T' and S', and will therefore be different for each season or month that the density anomalies are calculated. The density errors at any one time could be much larger than the mean anomalies shown in the figure 6.4.

6.3.2 Did the density error affect the early warming hindcasts?

The annual average observed and Quasi-assim density anomalies are compared in figure 6.5 for the period immediately prior to the rapid warming, when the DePreSys hindcasts become keen to warm rapidly (figure 6.2). The stratification of the subpolar North Atlantic is much weaker than the stratification at lower latitudes, and so the size of the density anomalies in this region is usually smaller than anomalies in the tropics. Therefore, to highlight the importance of the magnitude of the density anomalies, all of the data in figure 6.5 has been normalised by the observed interannual density variance.

As the observed positive density anomaly in the subpolar gyre intensifies over the 1990s, it is clear from figure 6.5 that the anomaly assimilation method used to initialise DePreSys is greatly over-estimating the size of the positive density anomaly in the North Atlantic, especially in the Irminger Sea region. The annual mean assimilated density anomaly is over 0.5 standard deviations larger than the observed density anomaly in

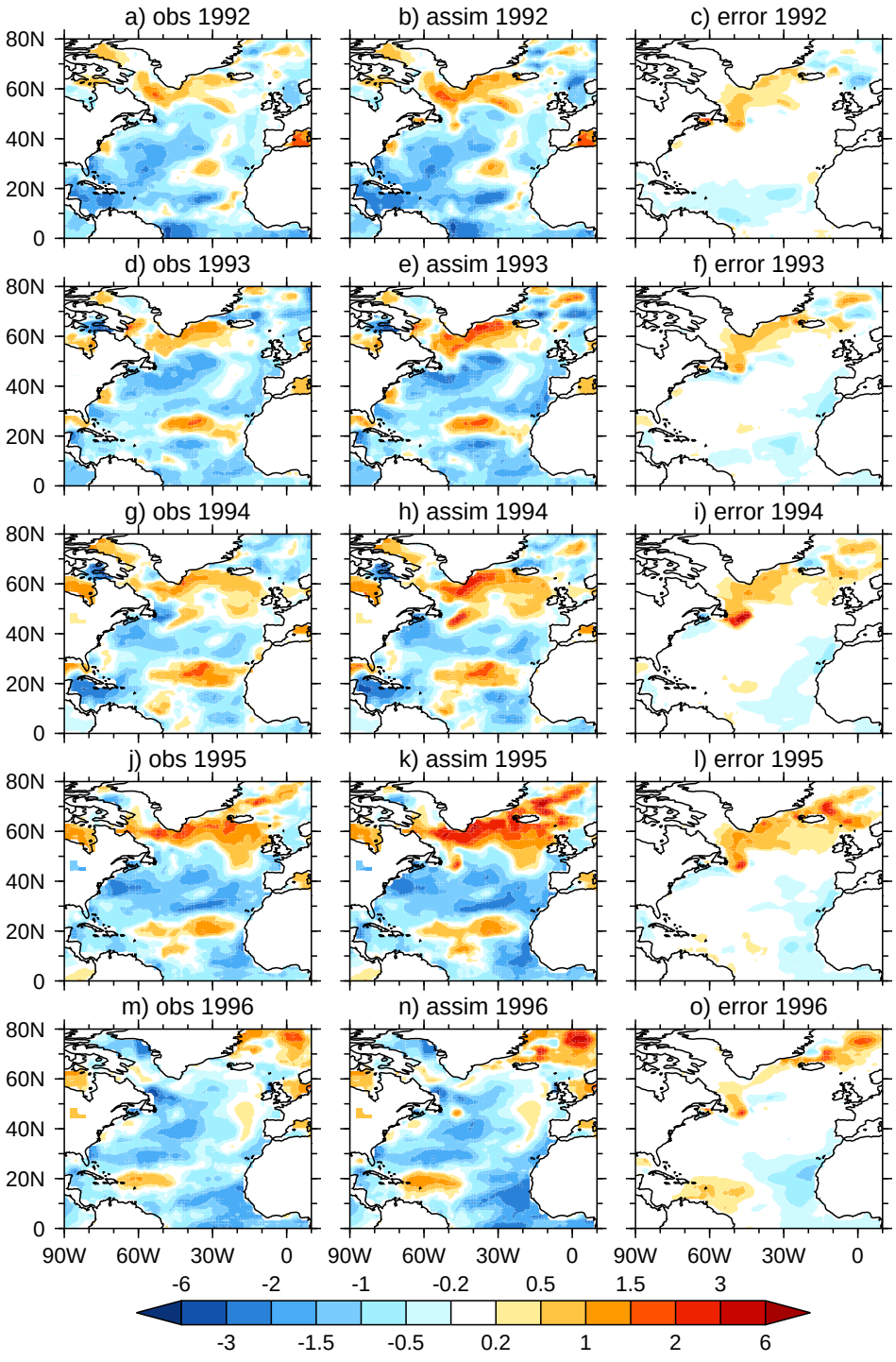


Figure 6.5: a) and b) show the observed and assimilated annual 100-1200m mean in situ density anomaly for 1992 when using the Original Met Office ocean analysis. c) shows the error in the annual mean density anomaly (b minus c). d) - o) show the same for the years 1993-1996. All the plots are normalised by the interannual standard deviation of the observed density anomalies to highlight where the anomalies are large enough to be important.

1992, and increases until it is over 1 standard deviation larger by 1995, highlighting the non-linearity. The largest density errors seem to occur over the Grand Banks in 1994, and to the North East of the Greenland Scotland ridge, two areas with the largest model error seen in figure 6.3.

To place the size of the annual mean density errors in the North Atlantic in 1994/1995 in perspective, figure 6.6 shows the global annual mean density anomalies and errors. The density errors in the Atlantic at this time are much larger than seen over the rest of the global oceans. Again, it is important to note that the density errors presented in figure 6.6 are annual mean errors; at any instant of time, there could be large errors present across the global ocean that could affect the DePreSys hindcasts.

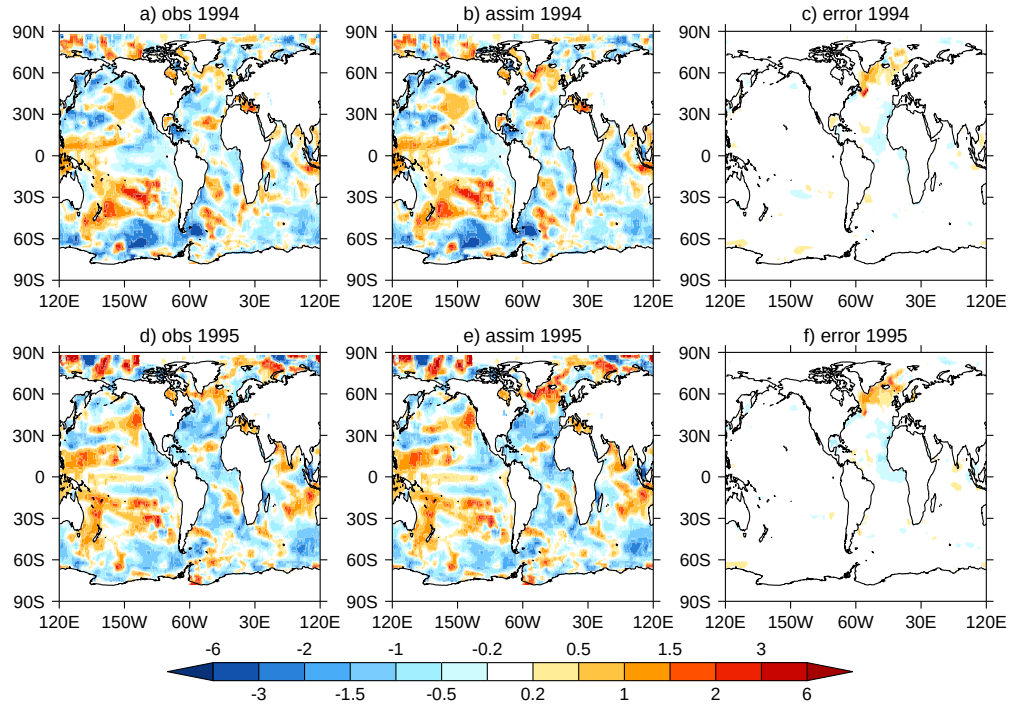


Figure 6.6: a) and b) show the observed and assimilated annual 100-1000m mean in situ density anomaly for 1994 when using the Original Met Office ocean analysis. c) shows the error in the annual mean density anomaly (b minus c). d) - f) show the same for 1995. All the plots are normalised by the interannual standard deviation of the observed density anomalies to highlight where the anomalies are large enough to be important.

The presence of errors in the assimilated density anomalies will lead to errors in the evolution of the DePreSys hindcasts. The positive density anomaly errors in the North Atlantic in the early 90s occur in a region of the North Atlantic that the strength of

the AMOC in HadCM3 is sensitive to, and could also affect the strength of the SPG circulation. This knowledge originally led to the following hypothesis:

Hypothesis 2:- Errors in the assimilated density anomalies caused a spin up of the overturning that was too large, or too early, which caused the rapid warmings that are seen in the DePreSys hindcasts.

However, investigating the AMOC strength in the original DePreSys assimilation and hindcasts, showed that the density anomalies errors had little effect on the strength of the AMOC (see figure 5.2) because there is little change in the AMOC strength over the hindcast period. As already discussed in chapter 5, there is no increase in the AMOC strength in the assimilation run of the original DePreSys because the original analysis did not resolve large, or spatially and temporally coherent density anomalies (see figure 5.3). There are positive anomaly errors found at depth at this time (not shown), but they are not large. However, large density errors are present at the surface just prior to the rapid warming (see figure 6.5), and so could still affect the evolution of the hindcasts, including the rate of warming in the SPG.

Unlike the original DePreSys, the assimilation of information into DePreSys2 does yield an increase in the strength of the AMOC in the 1990s (see figure 5.11) because the Updated Met Office ocean analysis does resolve large positive density anomalies at the deep western boundary of the North Atlantic (see figure 2.8). However, is the size of the density anomalies in DePreSys2 sensitive to the climatology as hypothesised? Figure 6.7 shows that assimilating the Updated Met Office Ocean analysis does cause a significant overestimation of density anomalies in the SPG region and along the deep western boundary. Hence, it is highly likely that the increase in AMOC strength is overestimated in the DePreSys2 assimilation. Hence, the presence of such errors in the assimilated density anomalies could make the correct initialisation of the AMOC difficult. The effect of the density errors on the evolution of the hindcasts, and the strength of the AMOC will be investigated in chapter 7.

Finally, it must be noted that missing processes may also affect the early warming hindcasts. DePreSys hindcasts initialised in the years leading up to the observed rapid warming, may always warm because they are not able to predict the intense positive NAO of

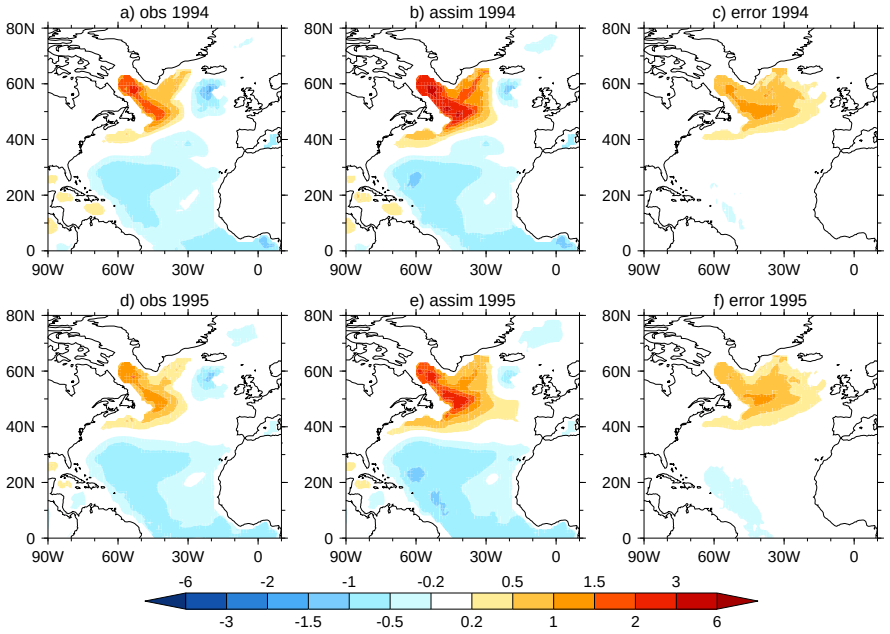


Figure 6.7: a) and b) show the observed and assimilated annual 1200-3000m mean in situ density anomaly for 1994 and c) shows the error in the annual mean density anomaly (b-c) when using the Updated Met Office Ocean analysis. i.e. the assimilated density anomalies are those which are assimilated into **DePreSys2**. d) - f) show the same for 1995. All the plots are normalised by the modelled standard deviation, calculated from the HadCM3 control run, to show where the anomalies are large enough to be important

the early 1990s that cooled the observed SPG (as discussed in chapter 2). However, if the lack of DePreSys's ability to predict the positive NAO was the only reason why the SPG warmed in the DePreSys hindcasts, the ensemble mean hindcast anomalies would not be expected to show such a sudden change in heat content. The ensemble mean of the hindcasts would likely show a more gradual warming, although some members may show a rapid transition if they simulate a large negative NAO event. Thus, the non-linear density errors may not cause the hindcasts to warm in general, but they may be contributing to the rapid nature of the warming in many DePreSys hindcasts, and hence, produce over confident predictions.

6.4 Chapter Conclusions

This chapter has focussed on moving past the mean analysis of DePreSys, and has instead taken a case study approach to understand how DePreSys is performing on a hindcast by

hindcast basis for the rapid warming of the North Atlantic in the mid 1990s. The main findings are as follows:

- Moving past the mean skill scores and trying to understand the specific hindcast successes and failures, at a process level, is an important route for understanding, *and improving*, decadal prediction systems.
- DePreSys hindcasts initialised just prior to the rapid warming of the North Atlantic capture the warming with remarkable skill. NoAssim is not able to capture the warming, giving further evidence that the warming of the SPG is not a response to radiative forcing. However, many DePreSys hindcasts begin warming rapidly too early, with very little ensemble spread, and thus give over confident predictions. The rapidly warming DePreSys hindcasts in the early 1990s will degrade the mean skill of DePreSys, relative to NoAssim, in the North Atlantic.
- The assimilation of observed anomalies of temperature and salinity, onto the model climate, leads to errors in the assimilated density anomalies. This is because the anomaly assimilation method effectively ignores the non-linear equation of state. This is a *fundamental limitation of the anomaly assimilation method*.
- The density errors occur frequently across the world's oceans, but most concerning, they occur regularly in the North Atlantic where the AMOC strength of HadCM3 is sensitive to density anomalies, and in the Tropical Pacific. Due to the non-linearity of the equation of state, the size and sign of the error in the assimilated density error will be different for each anomaly of temperature and salinity. Therefore, it is impossible to construct simple bias correction.
- Just prior to the rapid warming in the North Atlantic, the anomaly assimilation greatly overestimates the size of the assimilated density anomalies in the North Atlantic near the surface, and could be contributing to the early rapid warmings of DePreSys.

The presence of non-linear errors in the assimilated density anomalies could change the evolution of the DePreSys hindcasts, and, therefore, could affect their accuracy and reliability. The North Atlantic could especially be affected by the errors in the assimilated

density anomalies, as the evolution of the AMOC is so dependent on initialisation of density anomalies. Hence, it is important to investigate the effect of the assimilated density anomalies on the evolution of the DePreSys hindcasts. The effect of the presence of the assimilated density errors will be investigated in chapter 7. More specifically:

Hypothesis 2:- Errors in the assimilated density anomalies caused a spin up of the overturning that was too large or too early, which caused the rapid warmings that are seen in the DePreSys hindcasts.

will be tested by altering the density anomalies in the initial conditions of DePreSys hindcasts.

Chapter 7

Improving the UK Met Office Decadal Prediction System: Hypothesis testing experiments

7.1 Introduction

Chapter 6 examined the ability of DePreSys to capture the rapid warming of the North Atlantic subpolar gyre (SPG). It was shown that immediately prior to the rapid warming of the SPG, DePreSys predicts the timing and magnitude of the event with remarkable skill for lead times of up to a decade. However, many of the hindcasts initialised in the early 1990s warm too early, years before the event is observed, with small ensemble spread. Therefore, this chapter aims to explain why there are sometimes large errors in the DePreSys hindcasts of the rapid warming and, in particular, why do many of the hindcasts warm early?

Chapters 5 and 6 highlighted some potential problems with the anomaly assimilation method, and the initialisation of DePreSys. Firstly, the DePreSys hindcasts were shown to be undergoing significant drift a surprise given the anomaly assimilation method was designed to remove forecast drift. The drift in DePreSys seemed to be, at least in part, a manifestation of the drift in the original transient runs (shown in chapter 3). However,

the drift in DePreSys occurred faster in some regions than the drift seen in NoAssim and the transient runs, especially in the North Atlantic. It was postulated that imbalances in the model climatology led to large and rapid adjustments in the hindcasts. If the mean drift seen in DePreSys was due to that particular climatology (i.e. the drift was due to the drift of transient 1, or the sub-sampling of climate states), it follows that the mean evolution of the hindcasts will be different for other climatologies. This led to the following hypothesis:

Hypothesis 1:- DePreSys hindcasts can be sensitive to the model climatology that is used in the assimilation.

Secondly, it was shown that the anomaly assimilation method also ignores the non-linearity of the equation of state for sea water. The assimilation implicitly assumes that placing the same anomalies of temperature and salinity onto the model climatology will create density anomalies equal to the observed density anomalies. Chapter 6 showed that the assumption of linearity can lead to significant errors in the assimilated density anomalies. Most concerning was that errors frequently occur in potentially sensitive areas, such as the North Atlantic. Immediately before the rapid warming of the North Atlantic SPG, the anomaly assimilation significantly overestimates the observed density anomalies in the North Atlantic, which could cause errors in the hindcasts. This knowledge led to a second hypothesis:

Hypothesis 2:- Errors in the assimilated density anomalies caused a spin up of the overturning that was too large or too early, which caused the rapid warmings that are seen in the DePreSys hindcasts.

Testing the hypotheses, and understanding the relative importance of the highlighted problems on the performance of the DePreSys hindcasts, will allow steps to be made to minimise the problems in the future. Therefore, this chapter attempts to test the hypotheses by making changes to DePreSys. More specifically, changes are made to the initial conditions of hindcasts by making changes to the assimilation stage of DePreSys. The changes to the DePreSys system are described in section 7.2. The results of the experiments are presented in section 7.3, which is followed by a brief discussion of the implications in section 7.4.

7.2 Experiment description

The hypotheses are tested using prediction system experiments. That is, ensemble hindcasts made using the standard DePreSys configuration are compared to ensemble hindcasts that have had some perturbation made to them. Due to a number of changes that have occurred since the original DePreSys was configured, especially the use of the Updated Met Office ocean analysis, the experiments are made using the DePreSys2 set up that is described in section 3.1.5.

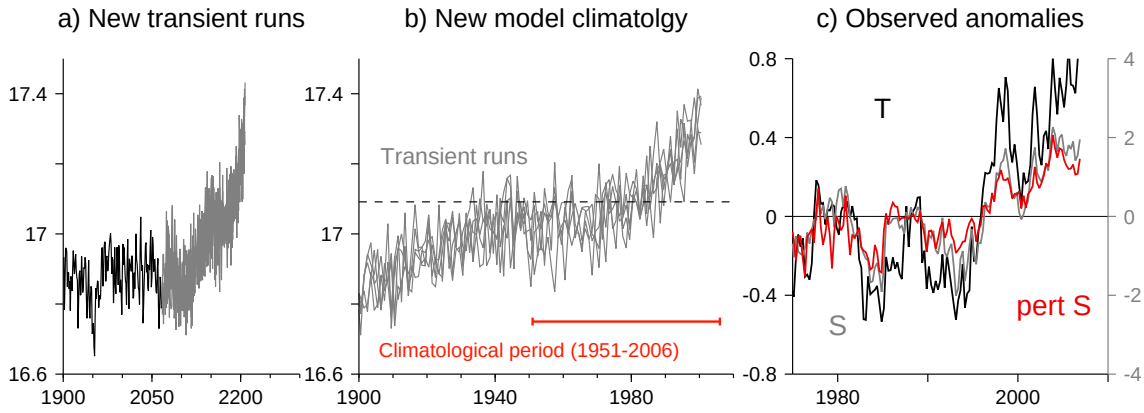


Figure 7.1: a) shows the global mean top 113m annual average ocean temperature from the HadCM3 control run (black) and the six "all forcings" integrations of HadCM3 used for the **new climatology** (gray). Note that the transient runs are all initialised at the same point of the control. b) also shows the six "all forcings" integrations that are used for the new climatology of HadCM3 on their real time axis. The new climatological period, which is used for all experiments, is also shown. c) shows the top 113m seasonal ocean temperature (T) and salinity anomalies (S) that are assimilated into DePreSys, averaged over the North Atlantic Subpolar gyre (60°W-10W°E, 50°N-66°N). Both the original salinity (grey), which is used in the **control** and **new climatology** experiments, and the perturbed salinity (red), which is used in the **perturbed salinity** experiment, are shown for comparison. Note that the salinity anomalies have been normalised for the comparison.

The perturbations are applied to the initial conditions of the hindcasts by modifying the assimilation stage of DePreSys. To understand the sensitivity of DePreSys hindcasts to a different climatology (i.e. hypothesis 1), the climatology used in the assimilation will be changed. The change in the climatology is explained in detail in section 7.2.1. Hypothesis 2, that the presence of non-linear density errors caused the SPG to warm too early in the DePreSys hindcasts, will be tested by modifying the assimilated density

anomalies so that they are the same as the observed anomalies. The density anomalies are modified by perturbing the salinity anomalies that are assimilated into DePreSys; the method is described in detail in section 7.2.2.

The focus of the experiments is on understanding why there are sometimes large errors in the DePreSys hindcasts of the early warming in the North Atlantic SPG. It was decided to test the hypotheses upon a hindcast that warmed too early, and a hindcast that appeared to capture the warming correctly. The December 1991 hindcast was chosen for the initial condition that warmed too early, and the December 1995 hindcast was chosen as the hindcast that captures the warming well. Figure 7.2 shows the plumes from the original DePreSys hindcasts for the two start dates that were chosen. To be able test the impact of the imposed changes, 10 member, unperturbed hindcasts were run for the two start dates.

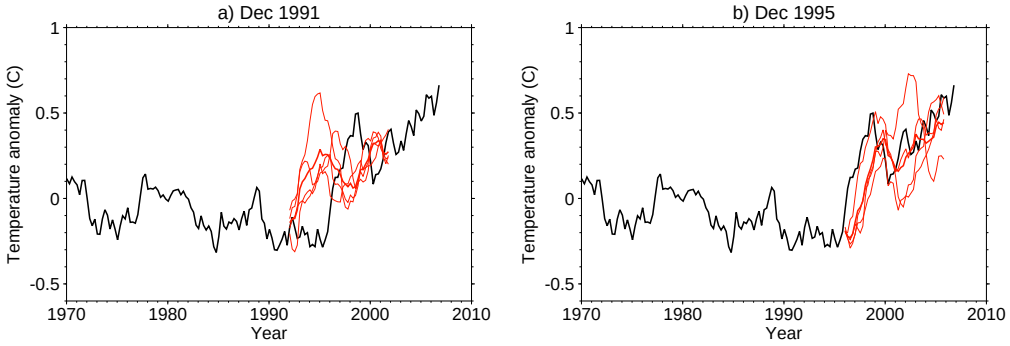


Figure 7.2: a) shows the observed top 500m average temperature anomalies, relative to 1941-1996, in the North Atlantic subpolar gyre (60°W-10°W, 50°N-66°N; black) from the Original Met Office ocean analysis. Overlain is the original December 1991 DePreSys hindcasts (red). The thick and thin red lines denote the ensemble mean and separate ensemble members respectively. b) shows the same as a) but now for the original December 1995 DePreSys hindcast.

7.2.1 New climatology

To test whether the DePreSys hindcasts are sensitive to the climatology used as the background state in the assimilation stage, a different, *well spun up*[1] climatology will be used. Calculating a new climatology using multiple new transient runs would be a computationally expensive exercise. Fortunately, 6 transient integrations had already been performed

[1] A well spun up climatology is one that is not undergoing significant drift in the deep ocean.

as part of the Grid for Coupled Ensemble Prediction project (GCEP, *Haines et al.* (2009)). Therefore, for ease and simplicity, the *new* climatology (C_{new}) was calculated from the 6 transient integrations already performed as part of GCEP. The transient integrations were initialised approximately 2000 model years after the original initialisation of the HadCM3 control run, and show little drift (Alan Iwi, personal communication).

The 6 transient runs were initialised in 1860 from the same state in the control run (see fig. 7.1 a), but were all given random sea surface temperature (SST) perturbations to ensure that the ensemble member diverged. The use of only SST perturbations, rather than initialising transient runs from different deep ocean states (as was intended for the original DePreSys climatology), could mean that the transient runs do not sample a sufficient spread of deep ocean states. However, the use of the new climatology (C_{new}) will still enable the exploration of whether DePreSys hindcasts are sensitive to the climatology that is used. The climatological period for the new climatology is the same as that used in DePreSys2. More specifically, the climatological period for the ocean is 1951-2006 (see fig 7.1 b) and for the atmosphere is 1958-2001, so that the same observational anomalies can be used in the assimilation.

Figure 7.3 shows the difference between the new climatology (C_{new}) and the DePreSys2 climatology (C_1), for ocean temperature. The difference between C_{new} and C_1 is similar to the difference between C_1 and C_{mod} seen in figure 3.7, but the magnitude of the difference is larger due to the continued drift of the control run. The largest differences between C_{new} and C_1 occur in the Atlantic, which may be associated with further drift of the circulation of the Atlantic in the HadCM3 control.

7.2.2 Perturbed salinity

To minimise the errors in the initial density anomalies, a simple perturbation to the assimilated salinity anomalies is performed (i.e. see fig. 7.1 c). The perturbations to the salinity ensure that the assimilated density anomalies are roughly equal to the observed density anomalies, allowing the effect of the initial density errors on the evolution of the hindcasts to be tested. Other choices could be made to minimise the errors in the assim-

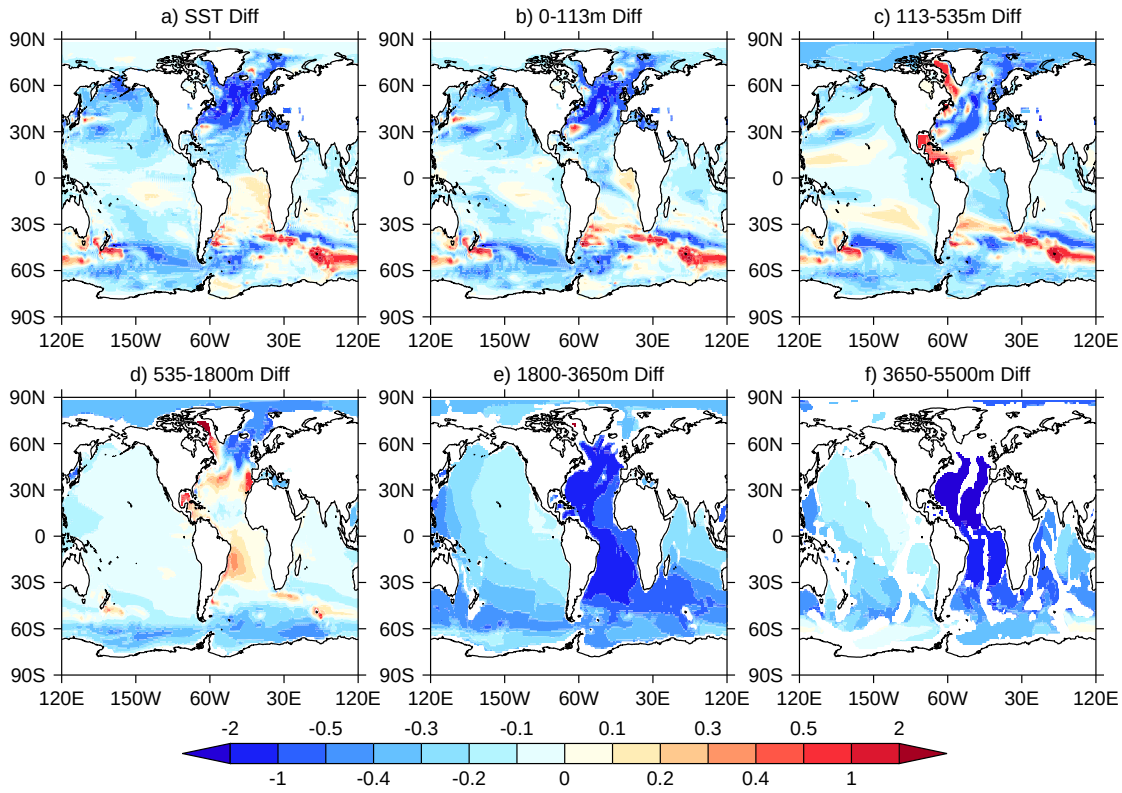


Figure 7.3: Difference in temperature between the transient 1 climatology and the new climatology (C_{new} C_1) for SST (a; (°C)) for a 1951-2006 climatology and for various layers in the ocean (b-f; (°C)).

ilated density anomalies, such as a perturbation to both temperature and salinity, but a perturbation to *only* the salinity was performed for three reasons:-

1. Temperature anomalies affect the atmosphere, and hence climate, more directly than salinity anomalies.

2. Observations of salinity are historically poorer, in space and time, than those of temperature. Therefore salinity is less well constrained.

3. Perturbing the salinity only is simpler compared to a perturbation of both temperature and salinity.

The linear interpolation between monthly mean anomalies in the assimilation run (described in chapter 3) will cause the density error to be slightly different for each day. The perturbation to the salinity could be calculated for each individual day, but it is assumed

that the calculation of the perturbation for monthly means is sufficient to minimise the majority of the density error. Therefore, the salinity perturbations were calculated by using a simple iterative scheme that minimised the assimilated density anomaly error to within 0.05 standard deviations of the observed interannual standard deviation.

The iterative scheme works by first calculating the error in the assimilated density anomalies for each monthly mean. Using the notation set out in section 6.3.1, the density error for each month is defined as the Quasi-Assim density anomaly minus the observed density anomaly:

$$\rho'_{error}(i) = \rho'_{Q-Assim}(i) - \rho'_{obs}(i) \tag{7.1}$$

where i denotes the time (an index which represents the month and year). Quasi-Assim is the density anomaly calculated when placing the observed monthly mean anomalies of temperature (T'_i) and salinity (S'_i) onto the monthly mean model climatologies ($C_{mod,A}$). The observed density anomaly is calculated by placing T'_i and S'_i onto the observed monthly mean climatologies (C_{obs}). For clarity the density anomalies are calculated by removing the density field calculated from the 1951-2006 climatologies for the observed and model. I.e. for the Quasi-Assim density anomalies:

$$\rho'_{Q-Assim}(i) = \rho_{Q-Assim}(i) - C^{\rho}_{Q-Assim}$$

where the density climatology is given as:

$$C^{\rho}_{Q-Assim} = \rho(C^{T}_{mod,A}, C^{S}_{mod,A})$$

and the full density field is given as:

$$\rho_{Q-Assim}(i) = \rho(C^{T}_{mod,A} + T'_i, C^{S}_{mod,A} + S'_i)$$

Grid boxes with an error greater than 0.05 standard deviations (σ) will receive a first guess perturbation to salinity. The $\rho_{Q-Assim}(i)$ becomes:

$$\rho^{pert}_{Q-Assim}(i) = \rho(C^{T}_{mod,A} + T'_i, C^{S}_{mod,A} + S'_i + S^{pert}_n) \tag{7.2}$$

where S^{pert}_n is the salinity perturbation, and n denotes the iteration number. When the density error is positive, the first guess, S^{pert}_1, is equal to -1σ to reduce the density, and when the density error is negative S^{pert}_i is equal to 1σ to increase the density.

The density error is then recalculated using the new $\rho^{pert}_{Q-Assim}(i)$. If a grid boxes density anomaly error was not corrected by the first guess salinity perturbation, then the second guess is calculated. Comparing the original density error with the change in density caused by the previous iterations salinity perturbation ($\Delta\rho^{pert}_{n}(i)$), allows the next salinity perturbation to be calculated. Assuming that the calculation of density is linear, the fraction of the previous salinity perturbation that is needed to minimise the density error is given as:

$$frac = \frac{\rho'_{error}(i)}{\Delta\rho^{pert}_{n}(i)} \tag{7.3}$$

The next guess for the salinity perturbation is calculated by multiplying the previous perturbation by the fractional change required calculated from equation 7.3, and is given as:

$$S^{pert}_{n+1} = frac \times S^{pert}_{n} \tag{7.4}$$

The new salinity perturbation (S^{pert}_{n+1}), is then used to update the perturbed density (equation 7.2) for the next iteration. The scheme iterates the perturbation of salinity until the error in the assimilated density anomalies is reduced to below 0.05σ.

The size of the initial salinity perturbations is analysed to check that the perturbations used are reasonable. The initial condition for the 1st of December of a particular year, in the assimilation run, is the linear interpolation of the November and December monthly mean anomalies, of that particular year. It follows that the salinity anomaly, for the 1st of December initial conditions can be calculated offline by simply taking the average of the November and December monthly mean anomalies. Therefore, the perturbation to the salinity for the initial time is given as the mean observed salinity minus the perturbed salinity.

Figure 7.4 shows that in the 100-1000m layer, salinity perturbations are present in many regions for the December 1991 and December 1995 initial conditions (see figure 7.4 c and f). Although the salinity field is altered, the changes are smaller than many of the observed anomalies, and do not distort the original pattern of anomalies (see figure 7.4 a and b, d and e). For the December 1991 initial condition, the initial salinity is reduced by $> 0.5\ \sigma$ in the North Atlantic subpolar gyre, especially in the Irminger sea, and along the Greenland Scotland ridge. For the December 1995 initial condition, a reduction of up to

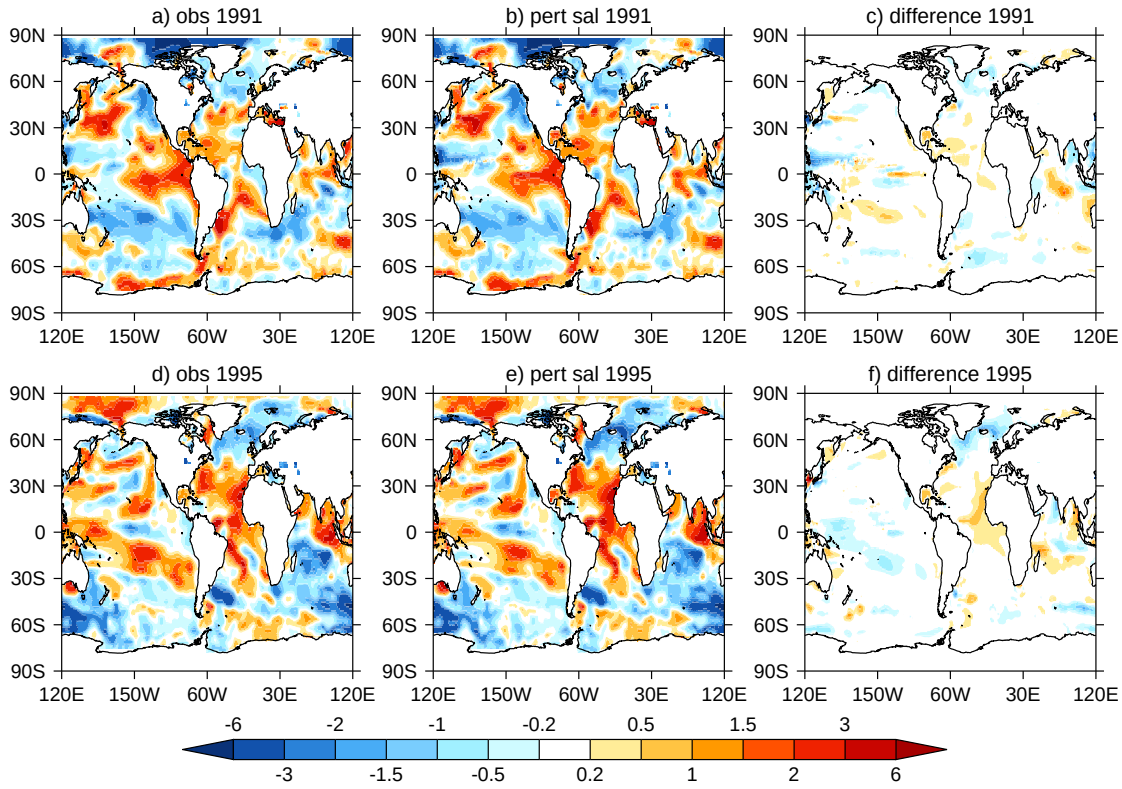


Figure 7.4: a) shows the observed 100-1200m salinity anomaly for average of November and December 1991 monthly means (in order to give the 1st of December) from the Updated Met Office ocean analysis, relative to the 1951-2006 climatology. b) shows the same for the perturbed salinity anomalies and c) shows the difference between the observed and perturbed salinity (perturbed salinity - observed). d)-f) shows the same but for 1995. All anomalies are normalised by the observed interannual standard deviation of the salinity anomalies.

1.5 σ is seen across much of the North Atlantic SPG initial condition.

Substantial perturbations are also seen in ocean basins other than the Atlantic. In the December 1991 initial condition, the largest perturbation occurs in the West Pacific, with a reduction in salinity of up to 2 σ (see figure 7.4 c). Both the 1991 and 1995 initial conditions show large perturbations to the salinity in the Kuroshio current region, with a reduction in salinity along the South coast of Japan, of $\sim$ 1.5 σ. Salinity perturbations are also seen in the Indian and the Southern Ocean for both hindcasts.

In the deeper ocean (1200-3000m), there are also significant perturbations to the assimilated salinity, especially in the North Atlantic as expected (see figure 7.5 c and f). The perturbations are again well within the range of natural variability, and make only a small

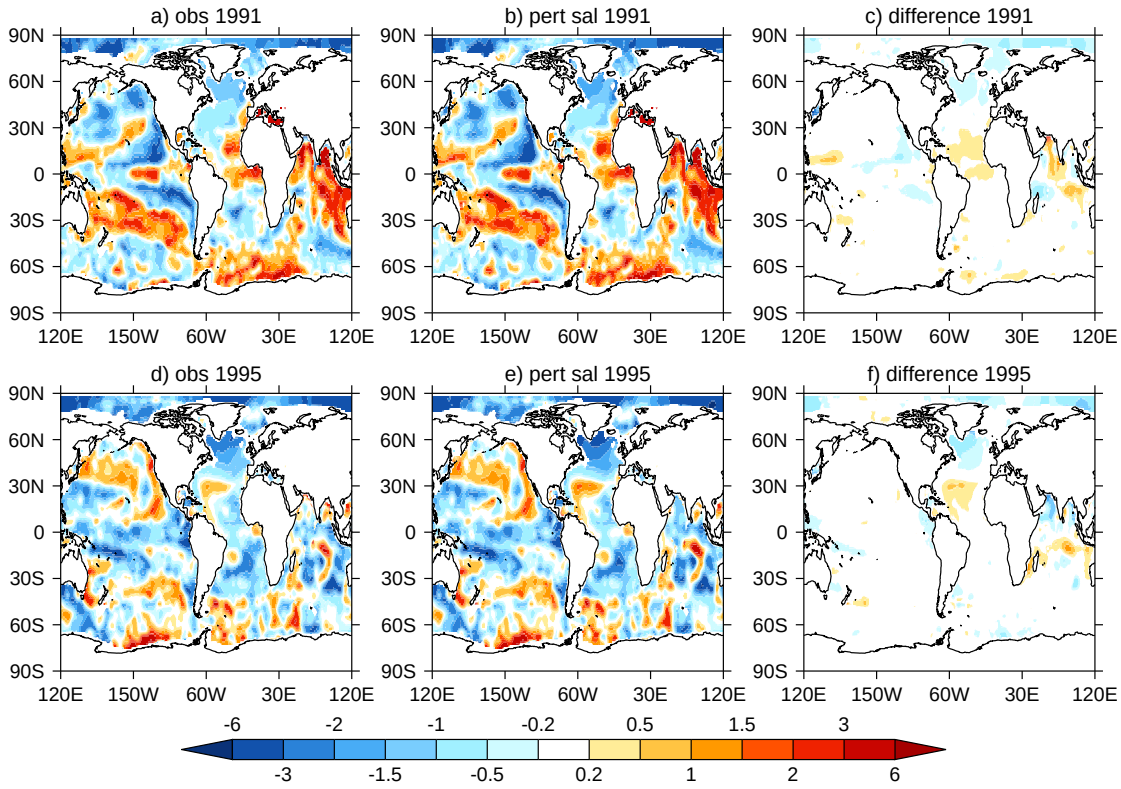


Figure 7.5: As figure 7.4 but for the 1200m-3000m layer.

impact on the overall salinity anomalies (see figure 7.5 a and b, d and e). In the North Atlantic SPG the salinity is reduced by less than 0.5 σ for the December 1991 initial condition, but in 1995, when the overestimate of the density is larger, the salinity is reduced by up to 1 σ. Salinity is also increased in the Tropical Atlantic for both hindcasts. There are also small changes in the salinity in the Arctic Ocean, areas of the Pacific Ocean, and in the Indian Ocean.

7.2.3 Summary of prediction experiments

A summary of the experiments performed is seen in table 7.1. The problem of the non-linear errors in the assimilated density anomalies (described in chapter 6) was originally discovered before the drifts in the AMOC and density (which are detailed in chapter 5). The perturbed salinity hindcasts were under way when the drift in the DePreSys system was discovered. After finding the large adjustments in the original DePreSys and De-

PreSys2, we were concerned that the rapid warming hindcasts could be a manifestation of the adjustment alone. Therefore, it was decided to also investigate the sensitivity of the hindcasts to the climatology. The December 1991 and 1995 hindcasts were repeated using the new climatology to test hypothesis 1. In order to complete the new climatology experiments in a timely manner, it was decided not to run the December 1995 perturbed salinity hindcast. All of the new climatology hindcasts were also run with 4 members (instead of 10 as originally planned) to minimise computational cost. To ensure a fair comparison, the new climatology hindcasts will only be compared with 4 members of the unperturbed hindcasts.

The new climatology hindcasts will also be affected by the presence of the non-linear density errors, but due to the non-linearity of the climate system, the effect of the density errors could be different to the original climatology hindcasts. Therefore, so that the affect of the salinity perturbations upon the new climatology hindcasts can be investigated, two further hindcasts have been performed for the December 1991 and 1995 initial conditions, which include both the change to the climatology, and the assimilated salinity field.

Table 7.1: Summary of prediction experiments performed for this chapter

Experiment name	December 1991	December 1995
Unperturbed (Old Climatology)	10 members	10 members
New Climatology	4 members	4 members
Perturbed Salinity (Old Climatology)	10 members	-
New Climatology + Perturbed Salinity	4 members	4 members

7.3 Results of prediction experiments

7.3.1 New Climatology

The use of the new climatology had significant impact on the evolution of heat content anomalies in the North Atlantic. Figure 7.6 a) shows that the North Atlantic subpolar gyre (SPG) in the unperturbed December 1991 hindcast warms rapidly after initialisation. However, the new climatology 1991 hindcast does not (figure 7.6 b). In contrast to the unperturbed hindcast, the new climatology hindcast cools for the first year, before a slower warming more similar to the observations. The unperturbed December 1995 hindcast does not warm initially, but instead remains cool for the first year (see figure 7.6 b), similarly to the original DePreSys December 1995 hindcast (figure 6.2). After the first year, the unperturbed December 1995 hindcast ensemble mean warms by 0.8°C in just 4 years. The new climatology December 1995 hindcast also warms rapidly, but does not warm to the same extent as the unperturbed December 1995 hindcast with only a $\sim$ 0.4°C warming, which is also more in line with the observed warming of $\sim$ 0.4°C by the early 2000s.

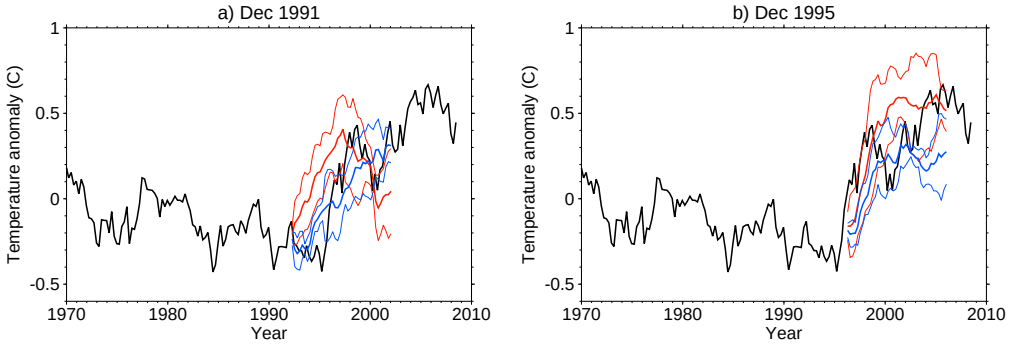


Figure 7.6: Observed top 500m average temperature anomalies, relative to 1951-2006, from the Updated Met Office ocean analysis for the North Atlantic subpolar gyre (60°W-10°W, 50°N-66°N; black). a) shows the same for the December 1991 **unperturbed** and **New Clim** DePreSys hindcasts in red and blue respectively. The thick and thin coloured line shows the ensemble mean and 95% confidence interval respectively. b) shows the same as a) but for the December 1995 hindcasts

Using the new climatology also changes the evolution of global heat content anomalies in both the 1991 and 1995 hindcasts. With only 4 members, it is difficult to prove that all the differences are significant. Nevertheless, there are some significant differences present.

Figure 7.7 shows the difference in the unperturbed and new climatology upper ocean heat content (0-113m) for the 1991 hindcasts only (but significant changes also occur in the 1995 hindcast). The differences in the heat content anomalies are also coincident with similar anomalies of SST (not shown), which could impact upon the atmospheric circulation.

In the first year, the unperturbed hindcast is warmer than the new climatology hindcast across the subpolar gyre (see figure 7.7 a). There are also significant differences over the Southern Ocean, Eastern Pacific and in the Greenland-Iceland-Norwegian (GIN) Sea. By the second year, the GIN Sea in the unperturbed hindcast is much cooler than the new climatology hindcast. Both the unperturbed December 1991 and December 1995 hindcasts develops a cold bias in the GIN sea, relative to observed anomalies, which is similar to the development of bias in the original DePreSys (see figure 4.9). The new climatology hindcasts do not develop a similar cold bias in the GIN sea (not shown), but instead span the observations in the ensemble spread. Therefore, it seems likely that the development of a large bias in the GIN Sea's of the original DePreSys was caused by the drift of the transient run and is removed using the new climatology. Elsewhere, as the hindcasts progress, the unperturbed hindcast cools relative to the new climatology hindcast in the subtropical North Atlantic, and the subtropical North Pacific (see figure 7.7 b-d). Although only the first four years are shown here, there are significant differences present globally throughout the entire length of the hindcasts (not shown).

Changing the climatology has affected the evolution of the hindcasts significantly. Hence, DePreSys hindcasts are sensitive to the climatology that is used in the assimilation run as postulated in **hypothesis 1**. Positively, it appears that changing from the original climatology used for DePreSys (C_1), to the new, improved climatology (C_{new}) that is not taken from a drifting transient run, and better samples the climatology, improved the predictions of the rapid warming of the SPG. What is the cause of the improved predictions of the subpolar gyre?

Using the new climatology has a significant effect upon the ocean circulation of the hindcasts. In particular, the strength of the AMOC is sensitive to the climatology. The new climatology hindcasts have a significantly weaker overturning over most of the Atlantic

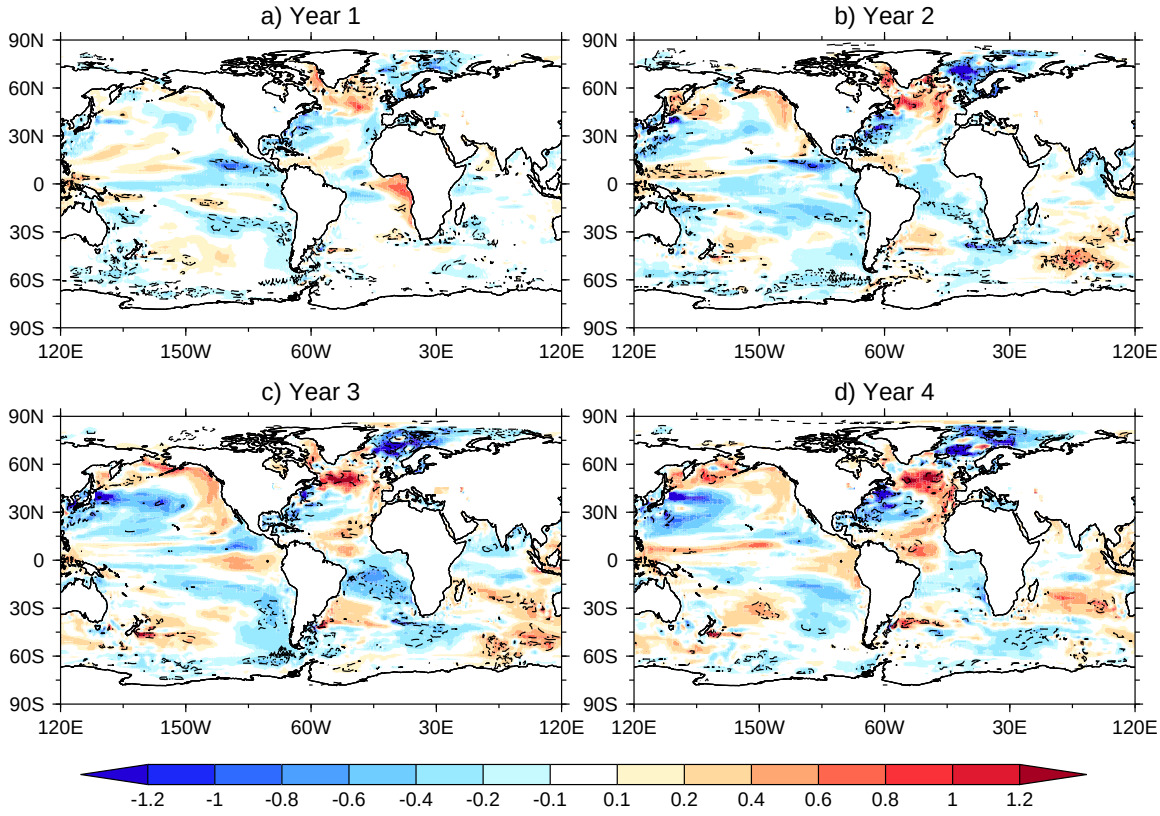


Figure 7.7: a) shows the difference between the first annual mean 0-113m average temperature anomalies between the **unperturbed** and **new climatology** DePreSys hindcasts initialised in December 1991. b)-d) show the same but for the 2nd, 3rd and 4th annual mean respectively. The black dotted and dashed lines show where the difference is significant at the 1% and 5% levels respectively calculated using a T-Test

(40°S-60°N) for the entire length of the hindcast (see figure 7.8 for the MOI at 40°N and 50°N; note that they remain separate for the entire length of the hindcasts). All of the 1991 and 1995 hindcasts show a decline in overturning strength at 40°N and 50°N, but note that the reduction in overturning strength for the new climatology hindcasts is reduced.

Interestingly, the MOI at 50°N in the new climatology December 1995 hindcast increases and peaks after a hindcast lead time of one year (see figure 7.8 d). An increase in the AMOC would be expected from the initialisation of positive density anomalies in the deep North Atlantic. A similar increase is not seen in the unperturbed December hindcasts, but perhaps more interestingly, is seen in the new climatology December 1991 hindcast does show an increase for one year. The difference in the hindcasts predictions suggests that the AMOC is responding to initial condition anomalies more clearly when

using the new climatology, in comparison with the old climatology where both hindcasts are dominated by a common drift.

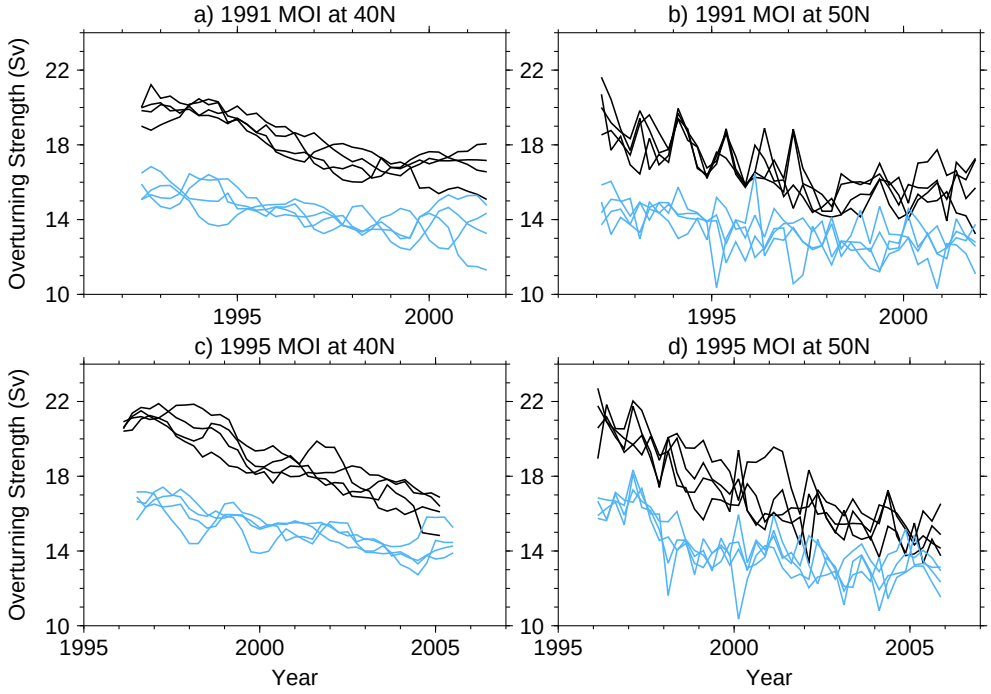


Figure 7.8: a) shows the MOI at 40°N for **unperturbed** (black) and **new climatology** (blue) DePreSys hindcasts initialised in December 1991. b) as a) but show the MOI at 50°N. c) and d) as a) and b) but show the December 1995 hindcasts. Note that the MOI at 40°N are shown as a rolling annual mean to remove the seasonal cycle.

The weakening of the AMOC strength also affects the northward heat transport of the Atlantic. Figure 7.9 shows that the northward heat transport in the North Atlantic is stronger in the unperturbed hindcasts than the new climatology hindcasts, especially in the subtropical and subpolar gyres (between 20°N-60°N). The difference in heat transport is largest for the 1995 hindcasts (see figure 7.9 b), with the northward heat transport reduced by $>$0.1 PW ($>$ 10% of the mean northward heat transport in the unperturbed hindcasts) between 40°N-50°N, throughout the entire length of the hindcast. The reduced warming of the SPG for the new climatology hindcasts (see figure 7.9) therefore seems to be, in part, due to a reduced AMOC strength and hence a reduced northward heat transport. Therefore, the early warming of the SPG in the unperturbed 1991 hindcast is likely to be partly caused by the too-strong AMOC that is established when using the original climatology. The too strong AMOC causes excess northward heat transport, and

hence is likely to contributes to the rapid early warmings of the SPG.

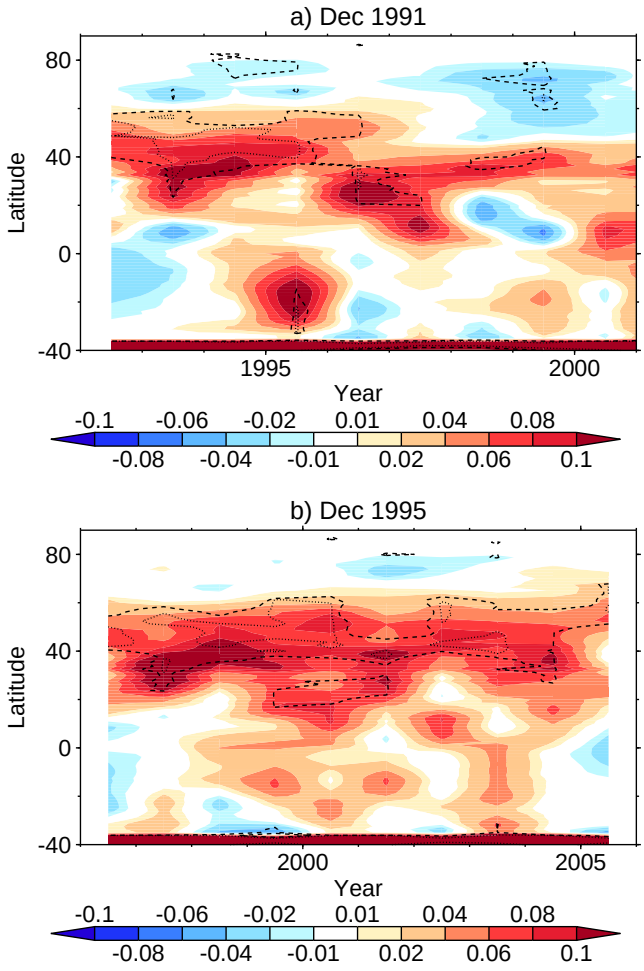


Figure 7.9: a) shows the difference in Northward heat transport of the Atlantic (PW) for the December 1991 **unperturbed** and **new climatology** DePreSys hindcasts (unperturbed - new climatology). b) shows the same for the 1995 hindcasts. The dotted and dashed black contours show where the differences are significant at the 1% and 5% level respectively.

Finally, can the hindcasts inform us of the drift that is seen in the original DePreSys and DePreSys2? With only two initial conditions using the new climatology, it is impossible to prove whether the evolution of the anomalies in the new climatology hindcasts is drift, or a response to the initial conditions. However, figure 7.10 shows that there are some interesting differences in the Atlantic between the evolution of 800-3000m density anomalies in the unperturbed and new climatology hindcasts. The evolution of density anomalies in the unperturbed December 1995 hindcast is very similar to the mean evolution of the drift in DePreSys2 (see figure 5.13). Although the unperturbed December 1995 hindcast starts with a high density anomaly in the SPG, the high density anomaly is

replaced by a large, low density signal that originates in the Labrador Sea, and propagates down the deep western boundary. The low density drift is also present in the Tropical Atlantic. The evolution of 800-3000m density anomalies in the unperturbed December 1991 hindcast is very similar to that seen in the unperturbed December 1995 hindcast (not shown), and highlights the dominance of the mean drift upon the unperturbed hindcasts.

The dominance of the mean drift on the unperturbed hindcasts allows a crude comparison with the new climatology hindcasts. Large differences in the evolution of the new climatology hindcasts compared to the evolution of the unperturbed hindcasts would suggest that the new climatology hindcasts are not subject to the same mean drift seen in the DePreSys and DePreSys2. However, it is important to stress that caution is required. Although the comparison may show whether the new climatology hindcasts are not subject to the mean drift seen in DePreSys and DePreSys2, it of course does not show that a common drift has been removed entirely. A large number of hindcasts would be required to make a comment upon common drift when using the new climatology. In fact, the discussion in chapter 5 suggests that it is very unlikely that a common drift would be removed entirely from a prediction system that uses such an anomaly assimilation methodology as DePreSys.

When using the new climatology (C_{new}), the evolution of the density anomalies is different compared to the unperturbed hindcasts in the North Atlantic (figure 7.10). The positive density anomaly in the North Atlantic reduces over the first 5 years, and is replaced by a low density signal, but the strong ($>2\sigma$) density anomaly propagating down the deep western boundary is not present. A low density anomaly develops across much of the Atlantic, and in the central Atlantic, but the decrease is again smaller than the mean drift seen when using the original climatology (C_1). The difference in the evolution of density anomalies suggests that the drift in the North Atlantic has been reduced by using a well spun up climatology. Interestingly, there is little change in the Southern Ocean (not shown). In chapter 5 it appeared that the Southern Ocean drift in DePreSys was mainly due to the drift in the original transient run. However, the presence of large adjustments in the Southern Ocean in the new climatology hindcasts further suggests that the imbalance was not just caused by the drift in the original transient run, but that there are other sources of imbalance in the initial conditions.

In conclusion, changing the climatology used in the assimilation run to C_{new} has had a significant effect on the DePreSys hindcasts, and positively, seemed to improve the predictions of the rapid warming of the North Atlantic SPG. The hindcasts that used the original climatology seem to be warming too early, or with too large a magnitude, in part because the AMOC is initialised too strongly when using the original climatology (C_1). The use of C_{new} also seemed to reduce the large drifts in the North Atlantic, relative to the hindcasts that use C_1. However, much of the drift, especially in the Southern Ocean, was not changed, suggesting that sources of imbalance remain in the initial conditions of the DePreSys hindcasts that use an improved climatology.

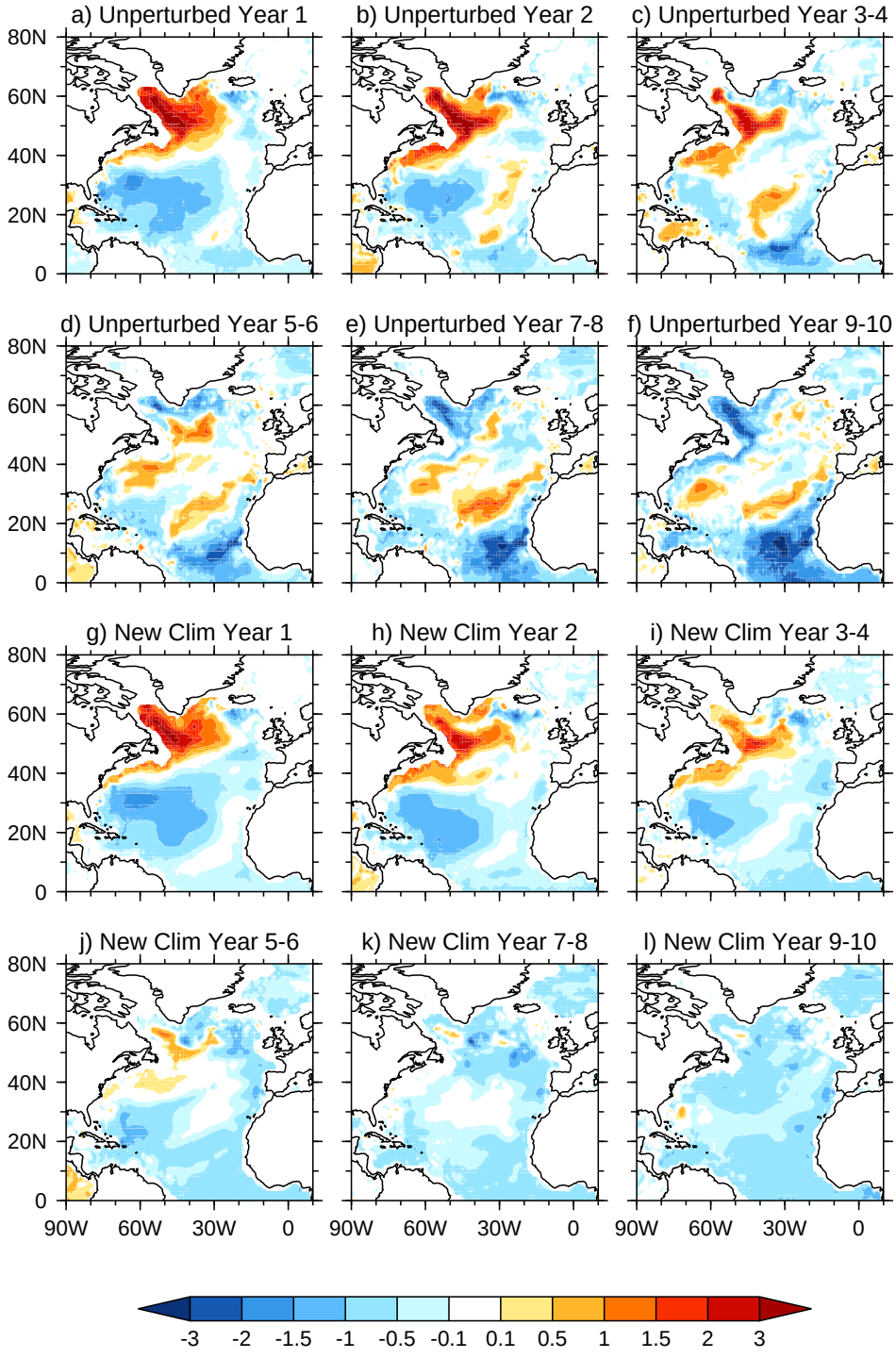


Figure 7.10: a) - f) shows the 800-3000m integrated in situ density anomalies from the **unperturbed** December 1995 hindcast. k) - l) shows the same for the **new climatology** December 1995 hindcast. The density anomalies are relative to $C_1^{1951-2006}$. All the anomalies have been normalised by the interannual standard deviation of density anomalies calculated from the HadCM3 control run.

7.3.2 Perturbed Salinity

The perturbed salinity experiment was designed to test the effect of the overestimated density anomalies in the North Atlantic in the early 1990s. More specifically, did the presence of such errors lead to the early warming hindcasts? However, figure 7.11 suggests that minimising the density errors in the 1991 hindcast had little impact upon the evolution of temperature in the SPG. The ensemble mean SPG temperature warms rapidly in the perturbed salinity hindcast, and is not statistically different to the evolution of the SPG temperature anomaly in the unperturbed hindcast. Therefore, the presence of the density errors in the initial conditions of the December 1991 hindcast did not cause the hindcast to warm rapidly after initialisation. As discussed in the previous section, it appears that the climatology, C_1, was a more important factor.

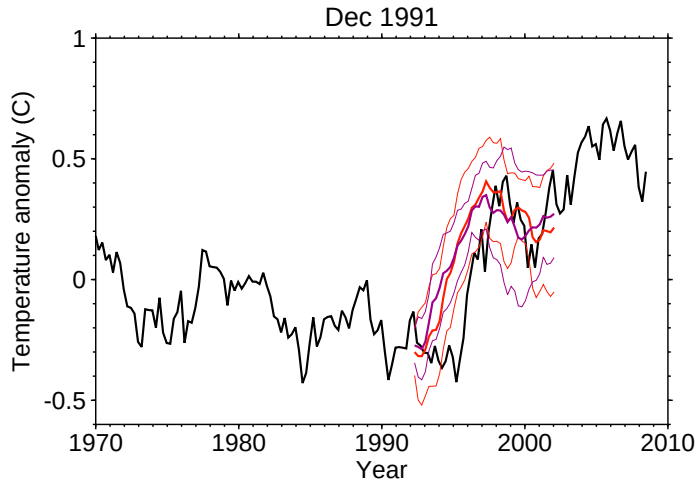


Figure 7.11: Observed top 500m average temperature anomalies, relative to 1951-2006, from the Updated Met Office ocean analysis for the North Atlantic subpolar gyre (60°W-10°W, 50°N-66°N; black). a) shows the same for the December 1991 **unperturbed** and **Perturbed salinity** DePreSys hindcasts in red and blue respectively. The thick and thin coloured line shows the ensemble mean and 95% confidence interval respectively.

Although it seems that the over estimation of the density anomalies in the SPG region did not cause the rapid early warming of the SPG, the density errors do affect the hindcasts and, in particular, the strength of the AMOC (see figure 7.12 a)). The effect on the AMOC strength, due to the corrections of the density errors, is considerably smaller than the effect of changing the climatology ($\sim$ 4 times less at 40°N). Nevertheless, the unperturbed and perturbed salinity 1991 hindcasts are initialised with a statistically significant

difference in AMOC strength (not shown). In the North Atlantic, the AMOC is significantly stronger in the unperturbed hindcast between 20°N-60°N, for the first 3 years of the hindcast, with the difference peaking at >1Sv (∼ 5%) in years 2-3 (figure 7.12). The reduction in the overturning strength in the perturbed salinity hindcasts is consistent with the reduced density anomalies in the initial conditions.

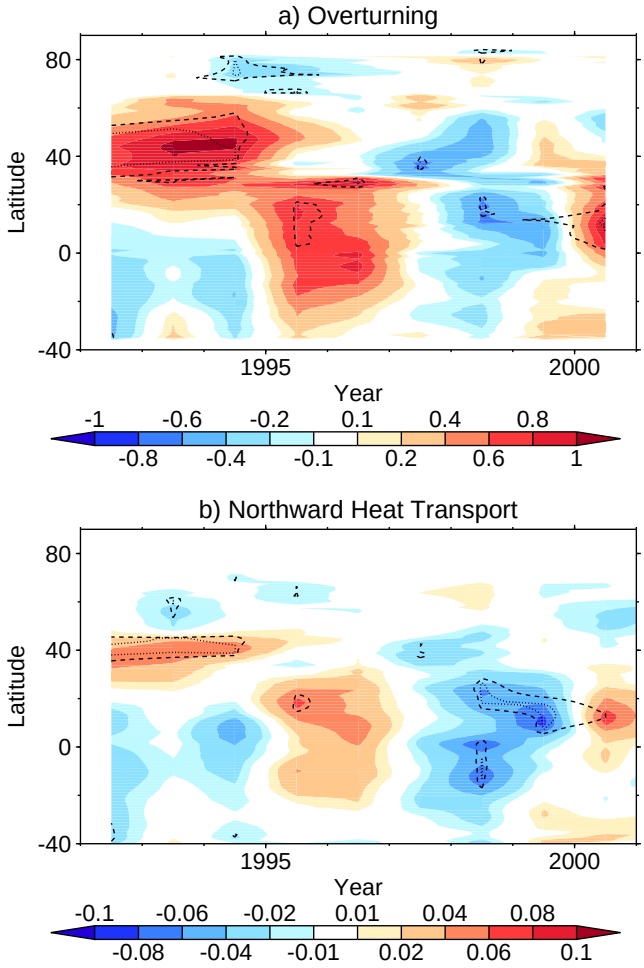


Figure 7.12: a) shows the difference in overturning strength in the Atlantic (Sv) for the December 1991 **control** and **Perturbed Salinity** DePreSys hindcasts (Unperturbed - perturbed salinity). b) shows the same for the 1995 hindcasts. The dotted and dashed black contours show where the differences are significant at the 1% and 5% level respectively.

The changes in the AMOC strength also affect the northward heat transport in the hindcasts (see figure 7.12 b)). The difference in the northward heat transport between the unperturbed and perturbed salinity hindcasts is correlated with the difference in the AMOC. In the first 3 years, the northward heat transport at 40°N in the unperturbed hindcasts is ∼0.05 PW (∼10%) larger than the perturbed salinity hindcasts.

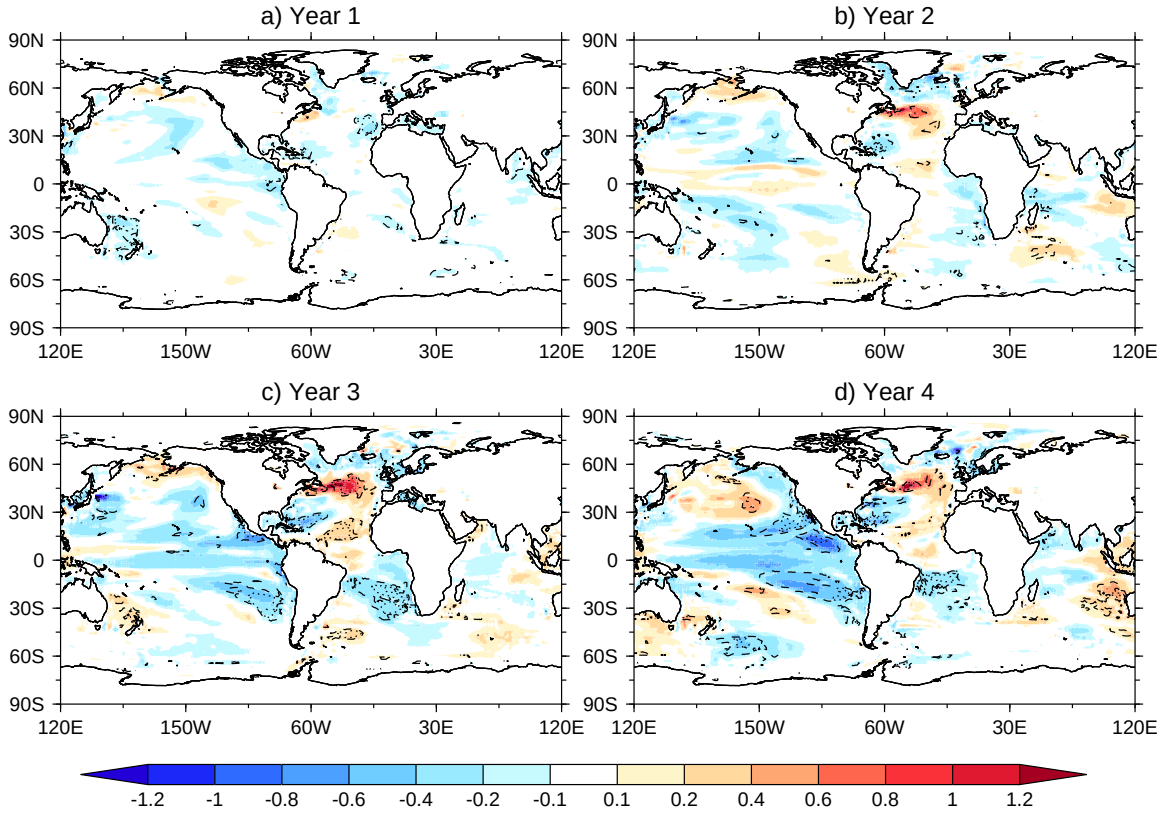


Figure 7.13: a) shows the difference between the first annual mean 113m average temperature anomalies between the **unperturbed** and **Perturbed salinity** DePreSys hindcasts initialised in December 1991. b)-d) show the same but for the 2nd, 3rd and 4th annual mean respectively.

There are also significant differences between the evolution of the heat content anomalies between the unperturbed and perturbed salinity hindcasts (see figure 7.13). Again, differences in SST are coincident with the differences in 113m heat content (not shown). In the first year, there is little difference between the two hindcasts. However, in year 2-4 the unperturbed hindcast is significantly warmer than the perturbed salinity hindcast in the North Atlantic current region to the south of the SPG (and is, therefore, not included in the SPG region used in figure 7.11). This difference in the south of the SPG is most likely due to the larger northward heat transport (see figure 7.12 b). Temperature anomalies in this region could affect the North Atlantic storm track (*Kushnir et al.* (2002)). Thus, predicting anomalies of heat content here is important. The North Atlantic current region remains warmer in the unperturbed hindcast until year 6, but the difference is not significant beyond year 4.

Significant differences in heat content anomalies are also seen in the Pacific. As the un-

perturbed hindcast progresses the Tropical Eastern Pacific is cooler, and the Subtropical gyre in the North Pacific is warmer than the perturbed salinity hindcasts. The anomalies are similar in pattern to the Pacific Decadal Oscillation (PDO; *Mantua et al.* (1997)), and could suggest that the density errors are also affecting the circulation of the Pacific. However, the growth of significant differences in the Pacific (i.e. years 3 and 4, see figure 7.13 c and d) lags the growth of a significant difference in the North Atlantic (year 2, see figure 7.13 b). Therefore, the timing suggests that the change in the Pacific is more likely to be a remote response to the Atlantic change (similar to that seen in *Dong and Sutton* (2002a) or *Zhang and Delworth* (2007)), rather than a direct impact of the Pacific salinity perturbations seen in figure 7.4.

In conclusion, the positive assimilated density anomaly errors present in the initial conditions did not cause the SPG region to warm early in the unperturbed December 1991 hindcast. This is in disagreement with **Hypothesis 2**, which may suggest that the warming of the SPG, in the DePreSys hindcasts at least, was not due to the AMOC. However, the presence of the assimilated density errors *does* have a significant effect on the strength of the AMOC, the northward heat transport in the North Atlantic, and the evolution of heat content and SST (not shown) anomalies. Significant differences are present for 10 years. The sensitivity of the hindcasts to the assimilated density errors shows that steps should be made to minimise this effect in the future.

7.3.3 New climatology and perturbed salinity

Figure's 7.14 and 7.15 show that the errors in the assimilated density anomalies also have an effect upon the AMOC strength and the northward heat transport in the new climatology hindcasts. Significance testing is more difficult for these hindcasts as there are only 4 members for each of the hindcasts. The perturbed salinity December 1991 hindcast initially has a stronger AMOC than the new climatology December 1991 hindcasts, especially between 20°N-40°N (see figure 7.14 a). The difference between 1991 hindcasts is smaller in magnitude than the difference in the 1995 hindcasts (see figure 7.15 a). In the first 2 years, the new climatology December 1995 hindcasts AMOC is $>$1Sv stronger than the perturbed salinity December 1995 hindcast. The difference is significant at the

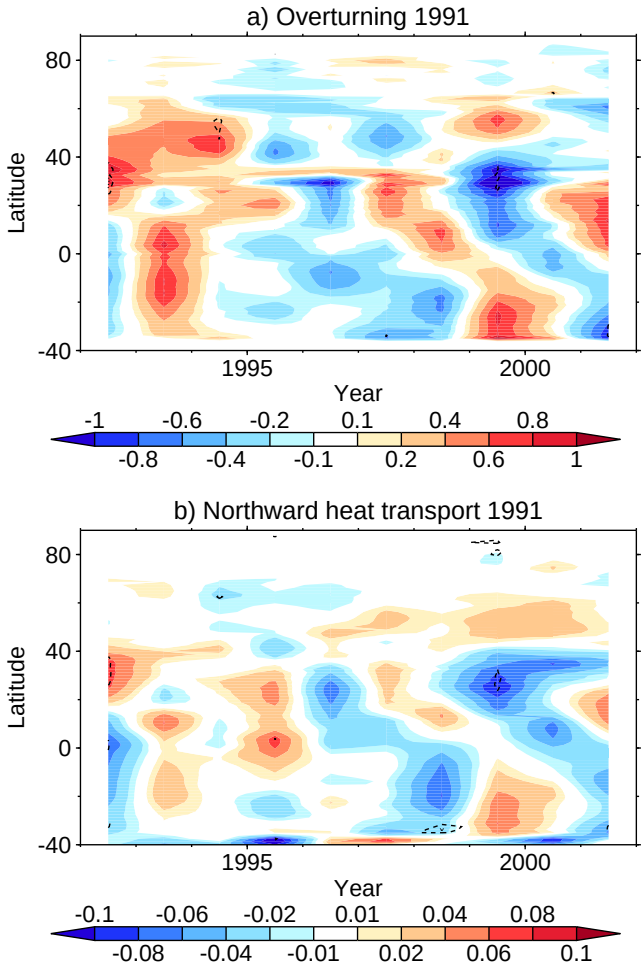


Figure 7.14: a) shows the difference in Overturning strength in the Atlantic (Sv) for the December **1991 New Clim** and **New Clim + Perturbed Salinity** DePreSys hindcasts (New Clim - New Clim + perturbed salinity). b) shows the same for the northward heat transport in the Atlantic. The dotted and dashed black contours show where the differences are significant at the 1% and 5% level respectively.

1% level, which is consistent with the reduction of salinity (figure 7.15 a). After 1997, there is no longer a difference in the AMOC strength at 50°N, but the AMOC between 40°S-40°N is stronger in the new climatology hindcast, especially between 20°N-40°N. The smaller impact upon the 1991 hindcast is expected due to the smaller initial perturbation (not shown).

Again, coincident with the changes in the AMOC strength, are changes in the northward heat transport (see figure's 7.14 b and 7.15 b). The new climatology December 1991 hindcast has a significantly larger northward heat transport of up to ∼0.06 PW (∼ 8%) between 20°N-40°N (see figure 7.14 b) in the first year only. The difference between the 1995 hindcasts is again greater. In the first two years, the new climatology December

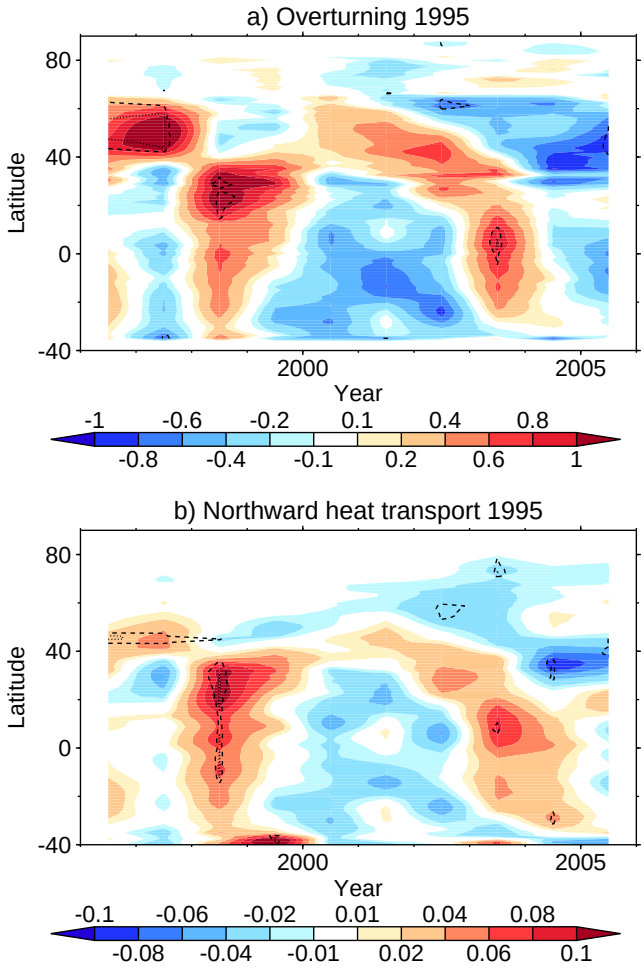


Figure 7.15: a) shows the difference in overturning strength in the Atlantic (Sv) for the December **1995 New Clim** and **New Clim + Perturbed Salinity** DePreSys hindcasts (New Clim - New Clim + perturbed salinity). b) shows the same for the northward heat transport in the Atlantic. The dotted and dashed black contours show where the differences are significant at the 1% and 5% level respectively.

1995 hindcast has a significantly larger northward heat transport of up to $\sim$0.04 PW ($\sim$ 8%) between 40°-50°. After 1997, the new climatology December 1995 hindcast is $>$0.1 PW ($\sim$ 10%) larger than the perturbed salinity December 1995 hindcast.

The presence of errors in the assimilated density anomalies significantly affects the evolution of the new climatology hindcasts. The affect of the density errors on the new climatology hindcasts is similar to their affect on the unperturbed hindcasts. Interestingly, the affect of the density errors upon the new climatology December 1991 hindcast is smaller than their affect on the unperturbed December 1991 hindcast (see figure 7.12). It is perhaps tempting to suggest that, in the North Atlantic at least, the new climatology is less prone, or less sensitive, to the non-linear density errors when compared to

the original climatology. However, more research would be required to confirm this and to understand the sensitivity of the new climatology to the non-linear density errors in other ocean basins. Nevertheless, the sensitivity of the new climatology hindcasts to the salinity perturbations is further evidence that the creation of the density errors due to the non-linear equation of state should be an important consideration, and could contribute significantly to forecast error.

7.3.4 What is the mechanism of the subpolar gyre warming in the DePreSys hindcasts?

Chapter 6 showed that the original DePreSys hindcasts seemed to be able to predict the rapid warming of the North Atlantic SPG in the mid 1990s. Using the mechanism described in chapter 2, it is tempting to assume that the successful prediction of the SPG warming was due to the initialisation of a stronger AMOC. However, with the knowledge of the large adjustments in DePreSys shown in chapter 5, was this really the case in the DePreSys hindcasts? Is the warming due to a dynamical response, is the warming just caused adjustments, or is it simply caused by the assimilation of warm temperature anomalies immediately to the south of the SPG? As the **new climatology** hindcasts seem to improve the predictions of the SPG warming relative to the **control** hindcasts, they offer a chance to further investigate the mechanism responsible for the warming in the DePreSys hindcasts.

To understand the changes in the heat content of the SPG, the northward heat transport of the new climatology hindcasts will be analysed further. The heat transport diagnostics in HadCM3 calculate the flux of energy across a section of ocean at each time step for three different components according to *Bryan* (1969). Firstly, the overturning component of the heat flux, H_O, across a section is defined as the integral over all depths of the zonal mean velocity (V) multiplied by the zonal mean temperature (T).

$$H_O = \rho c_p \int [V][T] dz \tag{7.5}$$

where the zonal mean is over the basin of interest, ρ is the density of the ocean (considered constant), c_p is the specific heat capacity of sea water, and the [ ] brackets show

the zonal mean. The heat transport that is associated with the gyre transport, H_G, is also calculated by the diagnostic, and is given as the integral of the departures of V and T away from the zonal mean [V] or [T].

$$H_G = \rho c_p \int [V^* T^*] dz \tag{7.6}$$

where the * indicates the deviation away from the zonal mean.

The final component is called the diffusive component (H_D), and is associated with the gradient of temperature with latitude, and can be thought of to represent the heat transport associated with eddies. The total heat transport, H, is given as $H = H_O + H_G + H_D$.

Unfortunately, the northward heat transport diagnostic of *Bryan* (1969) does not reveal the relative importance of the advection of temperature anomalies relative to the time mean (T') compared to the importance of dynamical changes in the ocean circulation relative to the time mean circulation (V'). Initially the heat transport calculated by the model will be investigated, but the importance of V' and T' will be investigated briefly for the overturning component later in this section by using the models V and T fields.

To be able to investigate the anomalous northward heat transport in the hindcasts, it is important to define a climatology or reference. With only two start dates, the lack of data makes estimating the northward heat transport climatology extremely difficult. However, the major difference in the ensemble mean northward heat transport between the new climatology hindcasts is in the first few years. After the first few years the ensemble mean northward heat transport in all the hindcasts tends towards a similar value (not shown). Thus, a climatology is estimated as the mean of the last 5 years of every new climatology hindcasts (i.e. 4 experiments $\times$ 4 ensemble members $\times$ 5 years $=$ 80 years of data). Although not ideal, the primary reason for defining a climatology is to compare the variability of the hindcasts.

The estimated annual mean climatology of the northward heat transport for the new climatology hindcasts is shown in figure 7.16. The climatology is largely similar to that found in *Dong and Sutton* (2002b), giving confidence that the features in the climatology are robust. Over most of the Atlantic, the total northward heat transport (H) is dominated by the overturning component (H_O), and the gyre (H_G) and diffusive (H_D) components

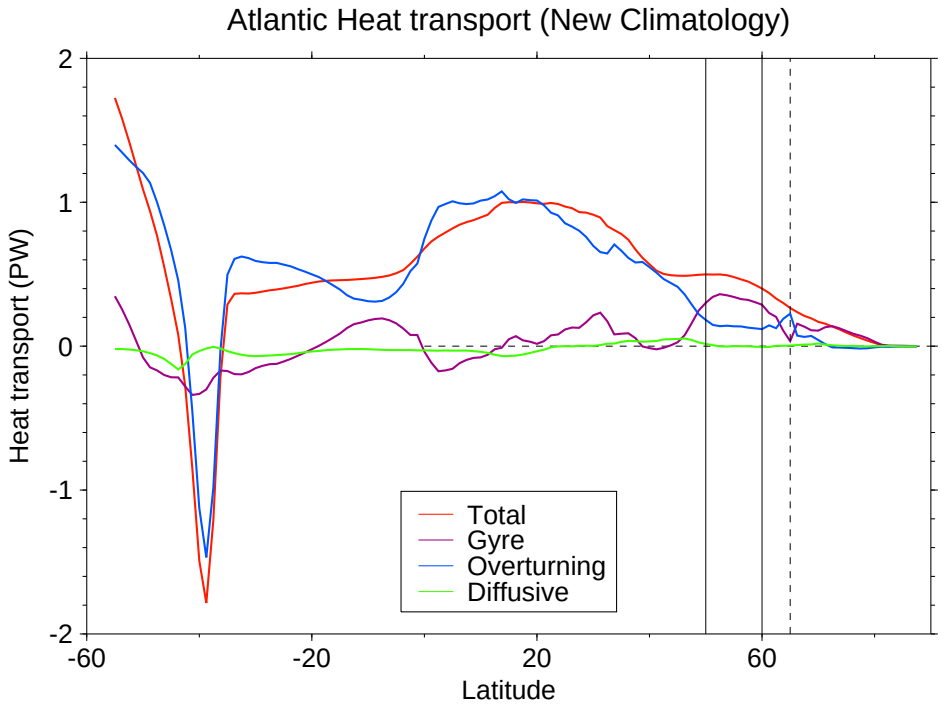


Figure 7.16: Shows the annual mean northward heat transport climatology of the Atlantic calculated from the final 5 years of all four **new climatology** experiments. The total northward heat transport (H, red) and the separate components for the gyre (H_G, purple), overturning (H_O, blue) and diffusive (H_D, green) transports calculated by the model are also shown. The latitudes of 50°N and 60°N are highlighted by the vertical grey lines. 65°N is also highlighted with the vertical dashed line.

are relatively small. At the latitudes of the North Atlantic subpolar gyre (i.e. 50°N - 60°N) H_G becomes more important than H_O. However, using the decomposition of H put forward by *Bryan* (1969) it is difficult to know the relative importance of both H_G and H_O at the latitudes of the SPG because the overturning and gyre circulation are not independent of each other (*Saenko et al.* (2002)). For example, the strong gradient in temperature across the SPG is, in part, due to the presence of the northward branch of the AMOC (*Hátún et al.* (2005)). The strong gradient in temperature will contribute towards H_G not H_O, thus, increasing the importance of H_G here. Although, the heat transport diagnostics have some problems in the SPG, examining the components will give a first order estimate of the important mechanisms for the variability in H in the hindcasts, especially form changes in the circulation its self.

As the hypothesised mechanism described in chapter 2 involves a surge in H, first it is important to establish whether there was indeed a change in H in the DePreSys hindcasts,

and what components were important. Figure 7.17 a) shows clearly that the new climatology December 1995 hindcast is initialised with a significantly larger H in the North Atlantic between the latitudes of 30°N - 60°N, with H between 40°N ∼0.1 PW (∼20%) larger than the climatological value. The vast majority of the increase in H is caused by and increase in H_O (see fig. 7.17 c), and with only a small contribution associated with H_G (see fig. 7.17 b). However, the anomalies associated with the gyre are not completely insignificant and do affect the total heat flux (note the similarity between fig. 7.17 b and d). For example, in the first year of the new climatology December 1995 hindcast, the gyre transport between 40°N - 50°N is anomalously low, which reduces the total H at these latitudes. The new climatology December 1991 hindcast also shows a similar increase in H, between 30°N - 60°N for the first 3 years, which is dominated by H_O, but the increase is weaker, and does not extend to lower latitudes as in the 1995 hindcast (not shown).

To investigate the impact of the changes in the northward heat transport on the warming of the SPG the change in the heat budget of the SPG has been calculated for the new climatology hindcasts. The heat budget of a volume of the Atlantic that is integrated zonally across the entire basin can be written as.

$$\Delta T = O_C - A_f \tag{7.7}$$

where ΔT is the change in annual mean temperature of the volume of ocean integrated form the surface to the bottom. O_C is the convergence of the northward heat transport in the volume of ocean, and is given as the northward heat transport into the box at the south minus the northward heat transport out of the box at the north (i.e. $O_C = H^{in} - H^{out}$). A_f is the atmospheric flux out of the top of the ocean. The atmospheric flux data is not available for the analysis, so A_f will be represented as the residual from equation 7.7 (i.e. $A_f = O_C - \Delta T$).

As shown in figure 7.16, at the latitudes of the SPG, H becomes dominated by the gyre transport. However, when using the heat transport diagnostic from *Bryan* (1969) in HadCM3, the overturning component becomes much more important at ∼65°N (see fig. 7.16 dashed vertical line). The change in the relative importance of the two components of the heat flux at 65°N significantly effects the calculation of the heat budget of the

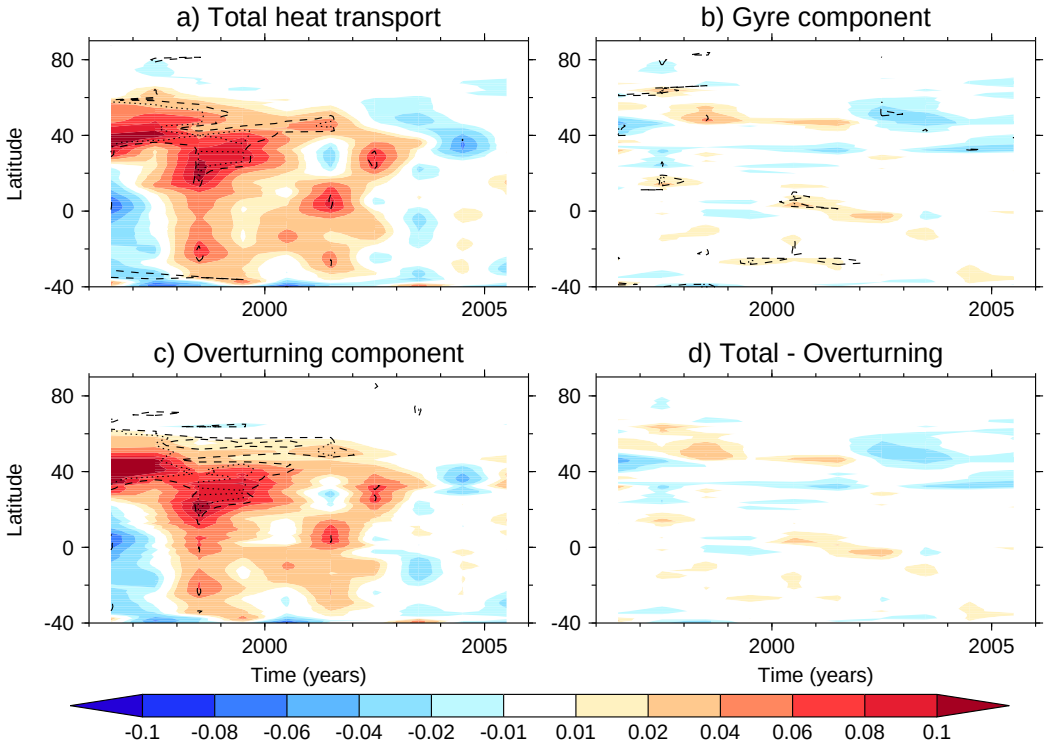


Figure 7.17: a) shows the total annual mean oceanic northward heat transport (H) anomaly from the **new climatology** December 1995 hindcast (PW). b) shows the contribution of the total heat transport that is due to the gyre transport (H_G), and c) shows the contribution that is due to the overturning (H_O). d) shows the difference between H and H_O. All anomalies are expressed relative to a climatology made from all of the **new climatology** hindcasts, see text for details. The black contours show where the anomaly is found to be significantly different from the climatology at the 5% (dashed) and 1% (dotted) levels by using a T-test.

SPG as defined previously (i.e. 50°N to 66°N); the overturning acts to cool the SPG on average (not shown). Therefore, for the purpose of exploring the mechanism proposed in chapter 2, the budget will be calculated on a section of the Atlantic between 50°N and 60°N. The new region warms similarly to the whole SPG defined as 50°N to 66°N (not shown).

Figure 7.18 a) and b) shows the energy budget for the SPG region in the new climatology DePreSys hindcasts. It is clear from figure 7.18 that for the heat budget of the SPG, in HadCM3 at least, the convergence of the overturning (i.e. H_O) component is the most important part of the ocean heat transport. In the climatology (represented by the stars) H_O supplies twice the amount of heat as H_G and H_D combined (although this is likely to be sensitive to the boundaries of the volume of ocean that is of interest). Hence, changes

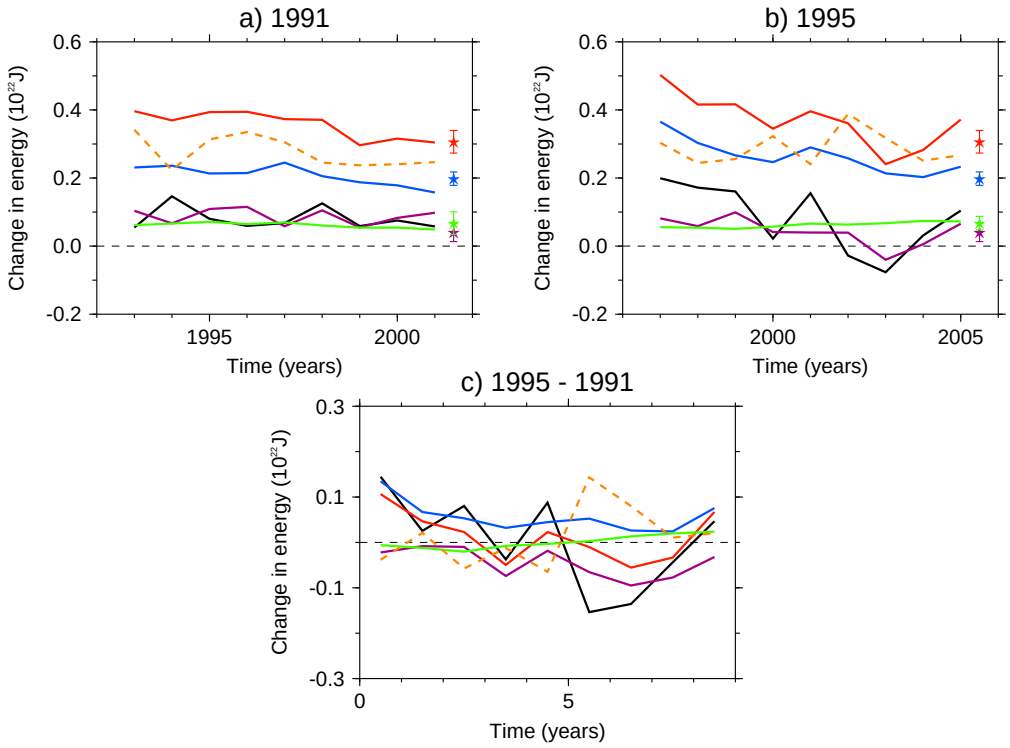


Figure 7.18: a) shows the integrated change in annual mean temperature (ΔT, expressed as energy) of the Atlantic ocean between 50° and 60°N in black for the new climatology December 1991 hindcast. The change in energy in the box that is due to the total northward heat transport (H, red), and the contribution due to the overturning (H_O, blue), gyre (H_G, purple) and diffusive (H_D, green) is also shown. The stars represent the climatological values, and the 95% confidence intervals on the estimate of the climatologies of H and its components. The atmospheric flux (orange dash) out of the ocean is calculated as the residual ($A_f = O_C - \Delta T$) and represents the atmospheric flux out of the ocean volume. If ΔT is zero then the ocean component and the atmospheric component of the energy budget should be equal. b) shows the same but for the **new climatology** December 1995 hindcast. c) shows the difference between the two energy budgets (i.e. b - a).

in H_O will likely have an important effect on the SPG temperature in HadCM3.

When examining the change in the heat content in terms of energy, it becomes immediately clear that the convergence of H is initialised stronger than its climatological value (represented by the stars in fig's. 7.18 a and b), especially in the December 1995 hindcast. The increase in the convergence of H, relative to the climatology, in the first few years of the December 1995 case is certainly large enough energy to bring about the warming seen SPG. When looking at the important components, the increase in the convergence of H (red) in the first years of the hindcasts, is dominated by the increase in H_O (blue).

In contrast, H_G (purple) seems to be much less important for the increase in H in the first few years of the hindcasts. Over the course of both the hindcasts the amount of energy supplied by the ocean to the SPG, H decreases due to a decrease in H_O.

Figure 7.18, therefore, suggests that the warming of the SPG, especially in the 1995 hindcast, was caused by an increase in the convergence of the oceanic heat transport, and especially the heat transport that is associated with the overturning. However, the hindcasts differ in their rate of warming. The December 1995 hindcast warms rapidly, and then slows in line with the changes in the convergence of H_O (see fig. 7.18 b). However, the December 1991 hindcast continues to warms at the a slower rate for the whole length of the hindcast with out such a large increase in H_O (see fig. 7.18 a). What is causing the difference between the two hindcasts? One aspect of interest is the atmospheric forcing; the analysis of the cooling of the SPG by the atmosphere is rather crude here. Thus, is it possible to conclude that the warming was caused by the ocean heat transport, or rather, was it just caused by continuous weak flux to the atmosphere?

To elucidate the mechanisms that are important for the difference in the warming in the December 1991 and December 1995 hindcast figure 7.18 c) shows the difference in the energy budget (i.e. 1995-1991). Pleasingly, figure 7.18 c) shows that the ensemble mean atmospheric residual (A_f, orange dash) is approximately equal in both the 1991 and 1995 predictions, which provides confidence that the atmospheric component is correctly calculated. The difference in A_f in year 5 and 6 seems to be due to two of the members of the December 1995 hindcast having positive NAO-like circulations (not shown). Thus, the warming in the 1991 hindcast is less likely to be caused only by weak atmospheric fluxes away from the ocean; the heat transport of the ocean must be important. The similarity of A_f also gives confidence that that the rapid warming of the SPG in the first few years of the 1995 hindcast is due to the changes in the oceans H, and in particular, H_O.

After the first few years, H_G decreases in the December 1995 hindcast relative to the December 1991 hindcast, which, therefore, also contributes to the reduction of H. Further analysis revealed that the reduction in H_G in the 1995 hindcast is due to a reduction in the gyre volume transport that occurs after the warming (not shown). A reduction of the gyre

circulation is seen in the real SPG after the warming event in 1995/1996 (*Hakkinen and Rhines* (2004)). Further analysis of the hindcasts, and comparison with the experiments of *Lohmann et al.* (2009), may help to explain the mechanism of the warming further. In contrast, the gyre remains strong in the 1991 hindcast, and even increases in strength slightly. Thus, the stronger gyre in the December 1991 hindcast seems to be at least part of the reason why the SPG continues to warm through out the length of the hindcast.

When examining the energy budget of each member of the ensembles separately it becomes apparent that some years warm rapidly that cannot be explained by fluctuations in the ocean heat transport. A simple analysis of the pressure anomalies in the hindcasts showed that the atmospheric flux away from the ocean is somewhat correlated with the North Atlantic Oscillation (not shown). Warming (cooling) of the individual members associated with negative (positive) NAO-like atmospheric pressure anomalies, as expected. Therefore, a deeper analysis is needed in the future to understand the relation of the atmospheric circulation and the warming of the subpolar gyre in HadCM3. For example, what is the relative importance of the negative NAO compared to the increase in the heat transport for the warming of the SPG. Nevertheless, the years where there is a large warming associated with changes in the atmospheric flux are relatively rare (i.e. less than once per decade) and appear at random times in the hindcasts. Thus, negative NAO events cannot adequately describe the ensemble mean warming of the 1995 hindcast, which further suggests that the warming of the SPG was inevitable before the negative NAO event in the winter of 1995/96.

So far, it has been shown that the ensemble mean rapid warming of the North Atlantic SPG in the December 1995 **new climatology** hindcast is associated with an increase in the northward heat transport that is associated with the overturning in the Atlantic. However, as discussed previously, the heat transport diagnostics in HadCM3 do not discriminate between anomalies in the heat transport brought about by departures away from the time mean velocity, or temperature (i.e. V' or T'). Therefore, the question remains, is the warming of the SPG mainly caused by the assimilation, and then advection, of warm temperature anomalies into the SPG region. Or is it caused by a dynamical change in the circulation?

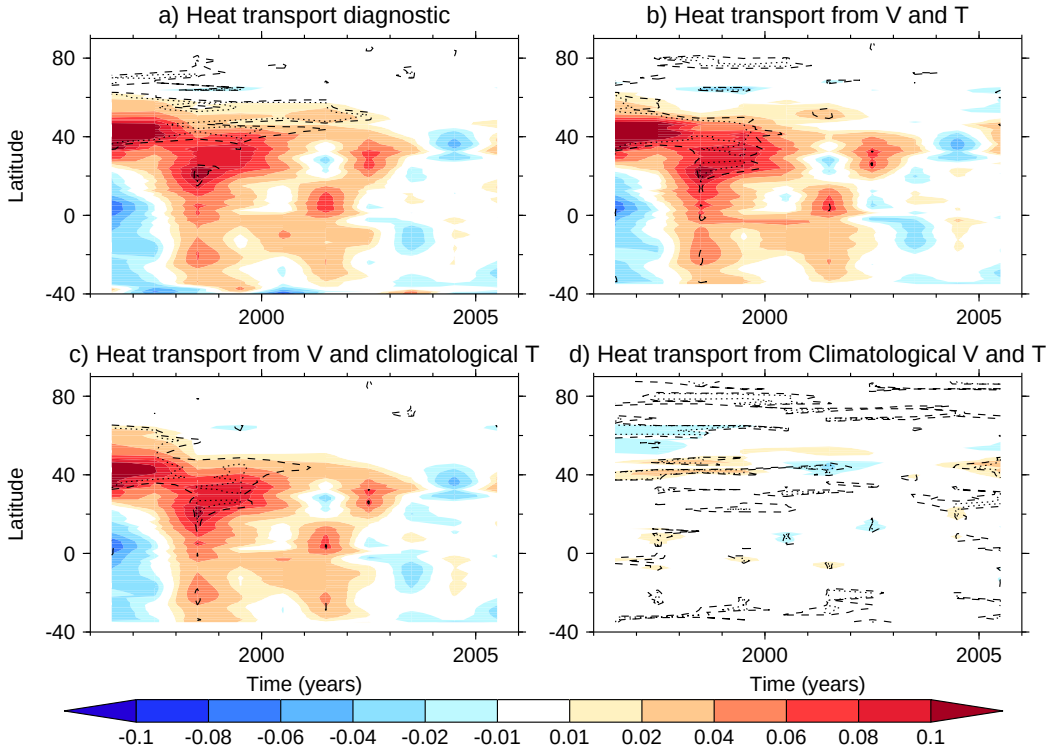


Figure 7.19: a) shows the anomalous annual mean overturning component of the northward heat transport (H_O) in the Atlantic for the December 1995 **new climatology** hindcast calculated directly from the model (PW). b) shows the annual mean H_O anomalies that are calculated off line using the seasonal mean meridional velocity (V) and temperature (T). c) shows the heat transport anomaly that is calculated when using the hindcasts time dependent V and the climatological T. Finally, d) shows the anomalous heat transport that is found when using the hindcasts time dependent T, but a climatological V. All anomalies are expressed relative to their own climatology, i.e. a climatology calculated from all the **new climatology** hindcasts as before, see text for details. The black contours show where the anomaly is significantly different from the climatology at the 5% (dashed) and 1% (dotted) levels calculated with a T-test.

The contributions of both changes in circulation and changes in temperature to the *overturning component*, H_O, of the heat transport have been estimated by using equation 7.5 and the V and T output from the model. The velocity and temperature data from the hindcasts is only in seasonal time scales, presenting a potential challenge for comparisons with the heat transport diagnostics, which are calculated at every model time step. However, the variability of H_O anomalies calculated from the model directly, and off line from using the V and T, are largely comparable in distribution and magnitude (see fig. 7.19 a and b). There are some small differences between the patterns, but overall the comparison gives confidence that the northward heat transport calculated off line is

sensible.

Figure 7.19 c) shows clearly that when using the climatological temperature (i.e. $\overline{T}$) in the calculation, a similar anomaly pattern to using both V and T emerges. Therefore, figure 7.19 c) clearly suggests that it was the increase in the overturning its self that causes an increase in the northward heat transport. The conclusion that V was more important than T is further supported when the northward heat transport is calculated using climatological V (which is again defined as the mean of the last 5 years of all the **new climatology** hindcasts), and the temperature from the hindcasts (see fig. 7.19 d). Although still contributing a significant proportion of the total overturning anomaly, especially north of 40°N, the contribution due to the temperature alone is much smaller than the contribution of the velocity. Therefore, it was the initialisation of the stronger AMOC that was important for the warming of the North Atlantic SPG in DePreSys.

In summary, the northward heat transport of the Atlantic ocean does deliver enough energy to cause the warming of the subpolar gyre in the DePreSys hindcasts, and did seem to play a part in the warming in the early 1990s hindcasts. An anomalous strengthening of the northward heat transport seems to be especially important for the rapid warming in the first few years of the December 1995 hindcast. In the December 1995 **new climatology** hindcast, the increase in the northward heat transport was dominated by the overturning contribution, which its self seems to be dominated by the velocity component. Therefore, the warming was not just a result of assimilating warm anomalies to the south of the SPG that were simply advected north, but involved a non-local dynamical response of the AMOC to the assimilated density anomalies. However, the analysis of the atmospheric fluxes still remains rather simple, and should be extended in the future to give more confidence to the proposed mechanism of the warming.

7.4 Discussion and implications

The sensitivity of the hindcasts to the new climatology, in particular, the too strong AMOC, and the potential reduction of drift in the North Atlantic, suggests that future decadal predictions based on the anomaly assimilation method should try to construct a

model climatology from an ensemble of spun up transient integrations. However, such a climatology highlights an important issue for anomaly assimilation. The creation of a spun up climatology is computationally expensive, requiring more than 2000 years of control integration to remove the drift for HadCM3. Many transient runs are also required to ensure that the low frequency variability of the deep ocean is sampled correctly. Such investments in computational resources may not be appealing, or even realistic, for the application of the anomaly assimilation technique to higher resolution models such as HiGEM (*Shaffrey et al.* (2009)), which are significantly more expensive than HadCM3.

With only two initial conditions, it is also impossible to safely conclude a great deal about the drift when using the new climatology, and, therefore, caution is greatly encouraged. However, it is clear that the large drift in the North Atlantic that dominates DePreSys, DePreSys2, and the unperturbed hindcasts, is not seen when using the new climatology. Hence, some comments upon the drift are possible. The use of an improved climatology (i.e., a climatology that is spun up and constructed from an ensemble of transient runs) seems to have reduced the drift in the North Atlantic. It is also clear that some large adjustments or drifts, such as in the Southern Ocean, are still present when using the new climatology, suggesting that imbalance remains in the DePreSys initial conditions. Thus, it is clear that more work is required on quantifying the drift and adjustment and understanding its origin, particularly when the common drift is likely to be different for every climatology because of the relaxation to mean temperature and salinity fields. It should be noted that large drifts would also occur when using a full field assimilation scheme because of model biases. Model biases in complex coupled climate models will be present for the foreseeable future. Therefore, it seems likely that some form of anomaly assimilation may be required for decadal climate prediction until model biases are improved. However, it seems important to objectively asses the arguments for and against the use of anomaly and full field assimilation for decadal climate prediction. Regardless of the assimilation approach that will be used in the future, more research is required into balanced initialisation for decadal prediction, especially for the AMOC, to minimise imbalance in the initial conditions.

Although the salinity perturbations do have a significant effect upon the evolution of the hindcasts, it is perhaps clear that making large perturbations to the salinity content

could affect the future evolution of the hindcasts in an un-predictable, non-linear manner. For example, a positive salinity perturbation placed at the surface, and cooled by the atmosphere during a hindcast, could become much denser in the perturbed hindcasts than in the unperturbed hindcast. Therefore, is it this difference caused primarily by the stochastic forcing of the atmosphere that affects the evolution of the hindcasts? Regardless of the non-linear nature of density anomalies in forecast mode, the minimisation of the density errors significantly affects the assimilation run, and is, therefore, important for creating the hindcasts initial conditions. If the anomaly assimilation approach is to be pursued further for decadal prediction, further research will be needed into suitable methods to minimise the effect of the error in the assimilated density anomalies. The use of the scheme used in this thesis will need improvement before it could be used in an operational system. For example, there has been no effort to ensure that the salinity perturbations will create a balanced initial condition that is consistent with the model. Hence, such a scheme to minimise density errors would need significant further testing before it could be used in an operational system.

7.5 Chapter Conclusions

This chapter has tried to understand why there are sometimes large errors in the DePreSys hindcasts of the rapid warming of the North Atlantic subpolar gyre (SPG); in particular, why do many of the hindcasts in the early 1990s predict that the rapid warming occurs years earlier than was observed. The two hypotheses proposed in chapters 5 and 6 were tested. More specifically, the hypotheses were tested using prediction system experiments of the rapid warming of the SPG in which perturbations were made to the initial conditions to test the hypotheses. The two hypotheses tested are as follows:

Hypothesis 1:- DePreSys hindcasts can be sensitive to the model climatology that is used in the assimilation.

Hypothesis 2:- Errors in the assimilated density anomalies caused a spin up of the overturning that was too large or too early, which caused the rapid warming's that are seen in the DePreSys hindcasts.

To test **Hypothesis 1** the climatology used in the assimilation was changed to an improved climatology constructed from an ensemble of 6 transient runs, initialised from a spun up part of the HadCM3 control run. **Hypothesis 2** was tested by perturbing the assimilated salinity anomalies to ensure that density anomalies were similar to the observed. The main results of the experiments are as follows:

- Hindcasts made using an anomaly assimilation technique *are* sensitive to the climatology used as the background state of the assimilation run. The strength of the Atlantic Ocean circulation and ocean heat transport have been shown to be sensitive to the climatology, which changes the evolution of sea surface temperature (SST) and heat content anomalies in the Atlantic.

- The use of the new climatology improved the prediction of heat content anomalies in the North Atlantic subpolar gyre (SPG) region for both the 1991 and 1995 hindcasts by reducing the rapidity and magnitude of the warming. The reduction in the rate, and magnitude, of the warming in the new climatology hindcasts seems to be due to a reduction in the overturning strength, and hence the northward heat transport.

- The presence of positive errors in the assimilated density anomalies, due to the nonlinear equation of state for sea water, did not cause the early warming of the SPG in the December 1991 hindcast. However, the assimilation run, and the hindcasts, are sensitive to the errors in the assimilated density anomalies. The strength of the AMOC, and hence northward heat transport, is sensitive to the presence of such errors, which impacts on the evolution of heat content anomalies in the North Atlantic. There are also effects on the evolution of heat content anomalies elsewhere around the globe, particularly in the Pacific.

- Using the new climatology for the background state of the assimilation seems to reduce the large drift seen in the North Atlantic density and AMOC, suggesting that a large portion of the drift in the North Atlantic was directly a result of using C_1, as hypothesised in chapter 5. However, the new climatology hindcasts still develop a low density signal in the GIN and Arctic seas that also fills the North Atlantic, as well as a reduction in the AMOC strength, albeit less dramatically than in the

original DePreSys and DePreSys2 hindcasts. Also, the drift in the Southern Ocean does *not* change when using the new climatology. Therefore, it seems likely that other imbalances are present in the initial conditions of the DePreSys hindcasts as postulated in chapter 5.

- Further analysis of the SPG warming in the new climatology hindcasts showed that an increase in the northward heat transport of the ocean was important for the warming seen in the DePreSys hindcasts, especially the December 1995 hindcast. The increase in the northward heat transport was dominated by an increase in the overturning component in the first few years, which was it's self dominated by changes in the circulation, i.e. not just by the temperature anomalies. Therefore, the warming in the DePreSys hindcasts was not just caused simply by the advection of warm temperature anomalies assimilated to the south of the SPG, but was due to a dynamical change. However, the analysis of the atmospheric forcing of the SPG still remains simple, and more analysis is needed to understand the effect of the atmosphere on the SPG in HadCM3.

The need for climatologies with a well spun up deep ocean will require long control runs, and many transient runs are required to ensure that low frequency variability is properly sampled. Therefore, creating such a climatology will be computationally expensive, especially if the anomaly assimilation method used for DePreSys is applied to higher resolution models.

Although the new climatology seemed to reduce the drift in the North Atlantic, the lack of change of the drift in the Southern Ocean highlights that the drift, has not been removed by using an improved climatology. Indeed, we expect drift, or adjustments, to be inevitable when using an anomaly assimilation technique, particularly, because of the relaxation to the density climatology calculated from the mean temperature and salinity, and not the mean density climatology (as discussed in chapter 5). Also, the creation of errors in the assimilated density anomalies is a fundamental limitation of the anomaly assimilation technique that significantly alters the evolution of the assimilation run, and the hindcasts. Therefore, the errors in the assimilated density anomalies should be considered in the future. Research is required to investigate strategies to minimise the errors

in the assimilated density anomalies, therefore, minimising their effect on the hindcasts.

Significant model error will be present in coupled climate models for the foreseeable future. Therefore, it seems anomaly assimilation may have an important role to play in decadal climate prediction. Thus, further research is required to quantify the drifts and adjustments caused by the anomaly assimilation technique, and to understand its origins. Research on balanced initialisation for decadal predictions is vital to ensure that predictions are not corrupted by drift or adjustments, and can, therefore, reach their full potential.

Chapter 8

Conclusions

8.1 Introduction

This thesis has focussed on understanding the performance of UK Met Office's Decadal Prediction System (DePreSys, *Smith et al.* (2007)), especially in the Atlantic sector. In particular, it has tried to gain a deeper understanding of how DePreSys improves predictions, and what might be limiting the ability of DePreSys to successfully predict changes that we believe to be potentially predictable at decadal time scales. Finally, this thesis has tried to exhibit the benefit of directly validating initialised climate predictions at the process level for case studies. Understanding when and, more importantly, why the prediction system succeeds, or fails, could guide research and investments to improve both the model, and the prediction system itself.

A brief overview of the analyses performed in this thesis, and the main conclusions, is presented in section 8.2. Recommendations for future work are presented in section 8.3. Finally, some concluding remarks are presented in section 8.4.

8.2 Thesis overview and main conclusions

8.2.1 Observed decadal variability in the Atlantic Ocean: characteristics and mechanisms

An examination of observed ocean analyses showed that the North Atlantic had undergone a large warming in the mid 1990s. The warming was especially clear in the North Atlantic subpolar gyre (SPG) region, which warmed rapidly in the winter of 1995/96. Comparison with an idealised modelling study of *Lohmann et al.* (2009) showed that the warming of the North Atlantic could be understood as a lagged response to the persistently positive North Atlantic Oscillation (NAO) that preceded the warming, and was, therefore, potentially predictable in advance. The mechanism of warming involved increased deep convection, especially in the Labrador Sea, and an increase in deep water formation in response to the positive NAO conditions. The spreading of cold and dense anomalies southward along the deep western boundary caused a temporary increase in the strength of the Atlantic Meridional Overturning Circulation (AMOC), and a subsequent surge in the northward heat transport of the Atlantic Ocean.

8.2.2 Mean skill and bias of DePreSys and NoAssim

The mean skill of DePreSys, relative to NoAssim, was repeated after an error was discovered in the construction of the NoAssim hindcast anomalies that were used in *Smith et al.* (2007). The analysis showed that when using an anomaly assimilation method, extreme caution should be used when analysing the mean skill of initialised climate predictions (in this case DePreSys), relative to a prediction system that does not assimilate information (NoAssim). The results can be sensitive to a great many choices that are made, such as the climatology used, and the climatological period. It was shown that the choice of climatology used in *Smith et al.* (2007) affected the apparent skill of DePreSys relative to NoAssim; in particular, it resulted in DePreSys appearing more skillful in the Southern Ocean.

The Analysis of mean skill in this thesis showed that much of the improvement in skill of DePreSys, compared to NoAssim, for both surface air temperature and ocean heat content, is in the first couple of years of the hindcasts, and is located primarily in the North Atlantic, the North Pacific, and the Tropical Pacific, which is consistent with expectations from potential predictability studies (*Collins* (2002); *Boer* (2004); *Boer and Lambert* (2008)).

Both DePreSys and NoAssim exhibit a mean bias that changes with lead time. The bias has a number of causes including: the model's inability to correctly simulate the patterns of variability; errors in the radiative forcing; errors in the model's response to radiative forcing (i.e. errors in the models climate sensitivity, missing feedbacks, or process errors); the drift of the transient runs (especially transient run 1). The multidecadal variability of the observations (B^{obsvar}, see chapter 4) will also contribute to the mean bias.

There are many key choices that affect the time dependent bias, and hence, the skill of both DePreSys and NoAssim. These are summarised here:

- The sampling uncertainty.
  - Time period of the climatology (i.e. its length, and start and end dates), which interacts with the low frequency variability of the observations.
  - Time period of hindcasts (i.e. duration and timing, e.g. with respect to volcanoes) relative to that of the climatology.
  - Number and distribution of hindcast start dates (i.e. every year / every 5 years).
  - Ensemble size of DePreSys and NoAssim hindcasts.
- Drift in the transient model integrations and the number and independence of such integrations.
- Rapid adjustment (i.e. faster than drift) of DePreSys hindcasts.
- Whether NoAssim has been shifted to remove it's initial bias.

For the particular hindcast period (1982-2001) NoAssim's bias is larger than DePreSys at all lead times, but crucially, DePreSys's has zero bias at zero hindcast lead time (by construction), but NoAssim's does not. Removal of the initial bias for each individual NoAssim hindcast, or particularly, the mean bias as a function of lead time, suggests that DePreSys can only outperform NoAssim for just over a year for predictions of the global mean temperature. Therefore, much of the improvement at long lead times in forecasts of *global mean* temperature seems to be due to the correction of the *initial bias,* not improved prediction of the variability. At a *regional* level, DePreSys does improve the predictions of temperature for a number of years ahead, especially in the North Atlantic and North Pacific, suggesting that DePreSys may be predicting variability in these regions. However, more work is needed to understand the mechanisms that improve the skill in these regions. Therefore, more research is required into objectively understanding the origin of the improved skill, and also on the effect of the mean bias on the mean skill.

8.2.3 Mean drift of DePreSys and NoAssim

An analysis of the mean evolution of DePreSys and DePreSys2 hindcasts showed that DePreSys is undergoing significant common drift - a surprise as the anomaly assimilation method was designed to remove common drift. The drift is especially large in the North Atlantic where there is a large and rapid weakening of the AMOC strength, and large changes in density. The large drifts are likely to limit the skill of the DePreSys hindcasts in the North Atlantic.

NoAssim exhibits a similar mean drift to that seen in DePreSys. Thus, the drift in the AMOC of the North Atlantic was, at least in part, due to the drift in transient run 1 that was used to construct the climatology used in the assimilation. However, the drift in DePreSys is often larger, and occurs faster than that seen in NoAssim, which suggests that the initial conditions for DePreSys are not balanced and that rapid adjustments are taking place. The imbalance could be caused by the sub-sampling of the low frequency variability in the climatology (the climatology for DePreSys is constructed from only one transient run), which could lead to a common evolution or drift of any hindcast that is initialised from near the climatological state. Also, the imbalance could be caused by the

relaxation to a density field that is constructed from mean temperature and salinity rather than relaxing to the mean density field (i.e. $\overline{\rho}(T,S) \neq \rho(\overline{T},\overline{S})$); this issue appears to be a fundamental limitation of the anomaly assimilation method. Also, the strong relaxation (i.e. short relaxation times) could dominate the assimilation, making it hard for the model to adjust into balance.

The possible sensitivity of the hindcasts to the drift in the original transient runs shows that the DePreSys hindcasts could be sensitive to the climatology used in the assimilation. In other words, the evolution of the hindcasts is not just dependent on the assimilation of the observed anomalies, but also depends largely on the climatology that is used. This led to the hypothesis:

Hypothesis 1:- DePreSys hindcasts can be sensitive to the model climatology that is used in the assimilation.

8.2.4 Case study in decadal prediction: the rapid warming of the North Atlantic in the mid 1990s

DePreSys is able to capture the rapid warming of the SPG in the mid 1990s with remarkable skill just prior to the warming. However, many hindcasts warmed too early with very little ensemble spread, which degrades the average skill of DePreSys in the North Atlantic. Analysis of the density anomalies highlighted another limitation of the anomaly assimilation method: it assumes that placing the same anomalies of temperature and salinity onto a different climatology will produce the same anomalies of density. This assumption of linearity was shown to produce large errors in the assimilated density field, and concerningly, in areas which may be sensitive to density anomalies, such as the North Atlantic.

A significant overestimation of density anomalies was present in the North Atlantic SPG region immediately prior to the rapid warming of the SPG in the early 1990s. The strength of the AMOC in HadCM3 is sensitive to density anomalies in this region. The knowledge of the errors in the assimilated density anomalies led to the hypothesis:

Hypothesis 2:- Errors in the assimilated density anomalies caused a spin up of the overturning that was too large or too early, which caused the early rapid warmings that are seen in the DePreSys hindcasts.

8.2.5 Improving the Decadal Prediction System: Hypothesis testing experiments

The two hypotheses were tested by using prediction system experiments by altering the initial conditions of particular hindcasts; that is, hindcasts with perturbed and unperturbed initial conditions, for a small set of start times, were compared. Using an improved climatology, constructed from an ensemble of 6 spun-up transient runs, as the background state in the assimilation stage, showed that DePreSys hindcasts *are* sensitive to the climatology as postulated in **Hypothesis 1**. Using the improved climatology also improved a hindcast that predicted the rapid warming of the North Atlantic SPG occurred years before it was observed. The improvement seems to be because the AMOC, and hence, the northward heat transport, was initialised too strongly when using the original climatology; i.e. the reduced northward heat transport leads to a reduced rate of warming.

The errors in the assimilated density anomalies did *not* cause the early warming of the SPG in a hindcast that was initialised in December 1991 (when positive density errors are present in the North Atlantic), as postulated in **Hypothesis 2**. However, the presence of these errors did significantly affect the evolution of SST and heat content anomalies, and also altered the circulation of the Atlantic. Therefore, research is required on strategies to minimise the errors in the assimilated density anomalies.

The use of the new climatology seemed to reduce the drift in the North Atlantic, although with only two initial conditions it is impossible to draw firm conclusions. However, it is clear that the drift was not altered every where when using the new climatology. For example, the drift in the Southern Ocean is not changed. This suggests that the drift in the original transient run was not the only cause of the drift, but that there is another source of imbalance that causes rapid adjustments.

Making use of the DePreSys hindcasts initialised from the new climatology, it was shown that the warming of the SPG in DePreSys did involve a surge in the northward heat transport of the ocean. The increase in the northward heat transport was dominated by an increase in the overturning circulation (i.e. the AMOC) and not simply the assimilation of warm anomalies to the south of the SPG. Therefore, in the DePreSys hindcasts at least, the initialisation of the strong AMOC was vital for the prediction of the rapid warming of the SPG in 1995/1996.

8.3 Recommendations and future work

The analysis of DePreSys presented in this thesis is not exhaustive. Decadal (or initialised) climate prediction is in its infancy. Thus, there is a huge array of research questions to address. However, it will be increasingly important to understand initialised climate projections. This is especially because of the inclusion of "near-term" (initialised) climate prediction in the fifth Coupled Model Intercomparison Project (CMIP5), which will be used to inform the Intergovernmental Panel on Climate Change (IPCC) fifth Assessment Report (AR5). This section presents a brief discussion of the recommendations and future work needed for DePreSys, and then also for decadal climate prediction in general.

8.3.1 Recommendations for the future development of DePreSys

As shown in chapter 4 the mean bias of NoAssim and DePreSys can affect the skill of DePreSys relative to NoAssim, because the biases are different. Therefore, for an objective evaluation of the initialisation of decadal climate predictions, it is essential that steps are taken to minimise, and more reliably quantify, the impact of the bias of DePreSys and NoAssim on the skill (RMSE). The steps include:

- Attention to reducing and characterising the drift present in the transient runs used to generate the model climatology and the NoAssim initial conditions.

- Steps should also be taken to try to reduce the rapid adjustment of DePreSys. More work on the nudging method should be pursued to minimise the drift. *Changes to the strong relaxation to allow the model more freedom are recommended*, especially in areas that are poorly observed, or could be prone to imbalance. It is also recommended that an ensemble of transient integrations is used to construct the climatology.

- Attention should be given to the sampling uncertainty of the bias as a function of lead time, and therefore, to how it affects the skill (RMSE). Hindcast experiments should be designed so that they give a fair comparison between DePreSys and NoAssim. For example, making the climatological and hindcast period the same should reduce the initial bias of NoAssim to zero.

Work to better understand the sensitivity of the hindcasts to the climatology should also be considered to understand the origin of imbalance, so that strategies to minimise it can be designed. Understanding the imbalance using hindcast experiments would require a large data set, and hence would prove costly. Therefore, it seems as though more idealised investigation into the imbalance would be desirable. One simple way to examine the mean drift would be to *only* assimilate climatological states. Another alternative would be to assimilate anomalies from sub-sampled *model states* (e.g. using a realistic observational density) to try to predict the models future evolution. Changes to the assimilation could be tested, and crucially, understood more easily in such idealised studies.

The nudging method that is currently used in DePreSys is very simple, thus, there is much potential for improvement. Therefore, in the longer-term, investment may be needed into more complex assimilation strategies. A possible next step would be to use an analysis method that assimilates observational anomalies directly into a dynamical model (i.e. similarly to the ECMWF ORA-S3 (*Balmaseda et al.* (2008)) so that initial conditions which are more consistent with the model could be created. If a more complicated assimilation method is chosen, research into effective balance increments will also be needed to update the variables which are not directly assimilated (e.g. as in *Weaver et al.* (2005)). An investment into a more sophisticated assimilation method for DePreSys will prove costly in terms of time and computational resources. Therefore, any future

upgrades will require careful consideration. Fortunately, the wide range of data assimilation schemes that will be used in the "near term" climate projections for CMIP5 may inform the development of an improved assimilation scheme for DePreSys.

Finally, the non-linear density errors created when using the anomaly assimilation method need to be addressed. One possible approach would be to devise a scheme similar to that used in chapter 7, which perturbs the initial conditions so that the density is minimised. The scheme used in this thesis only perturbed salinity, but different choices could be made to compromise between perturbing temperature and salinity to create the initial conditions.

8.3.2 Recommendations for the future development of initialised decadal predictions

The challenge of achieving balanced initial conditions for decadal climate predictions should be an important research aim of initialised climate prediction in the future. There has been some work on providing balance in the tropical Pacific near the surface (*Burgers et al.* (2002); *Bell et al.* (2004)), but there seems to have been little work into providing balance for the initialisation of the AMOC. Hence, research into methods for balanced AMOC initialisation will be an important priority, especially due to the sensitivity of the AMOC to relatively small density perturbations (*Marotzke and Klinger* (2000)). An objective evaluation of the assimilation strategies should also be a priority. For example, understanding the arguments for and against the use of anomaly, or full field assimilation, is an important consideration as both assimilation strategies have limitations.

The sensitivity of the DePreSys skill to a number of choices that affect the mean bias of the system, suggests that great care should be taken when evaluating the mean skill of decadal prediction systems. Multiple skill scores should also be used in the evaluation of decadal hindcasts, relative to un-initialised predictions. The standard deviation of the error should be included to highlight regions where the mean bias is important, and anomaly correlation should also be considered for comparison. *The results from this thesis suggest that future evaluations of the mean skill of initialised climate predictions*

should also include a shifted NoAssim for comparison. However, the results of chapter 4 showed that Shifted NoAssim may be too sensitive to the interannual variability. Therefore, more work is needed to improve the construction of Shifted NoAssim to ensure that the results are not sensitive to the construction. Other predictions, such as damped persistence, or other statistical predictions (e.g. *Hawkins and Sutton* (2009b)), should also be routinely used as a comparison. Lastly, the evaluation of skill in this thesis suggests that the analysis of skill in decadal climate predictions should focus on regional validation, and not on the global mean.

Another important consideration for decadal climate prediction is the density of observations in the ocean. The deployment of Argo floats has ensured that the ocean is now observed at unprecedented detail. However, the ocean in the past was, in comparison, very poorly observed, especially for salinity. Decadal predictions of the past observed changes will, therefore, always be somewhat limited in their ability to predict the observed changes, and will not give a fair indication of the potential of decadal climate prediction. Thus, it seems as though significant work with pseudo-observations (i.e. taken from sub-sampled model states) is needed to understand the potential of decadal climate prediction.

Questions also concern where limited observational capacity should be focussed to minimise uncertainty in predictions, which the pseudo-observations will also help to explore. For example, are observations of the Atlantic deep ocean, or deep currents such as the Deep Western Boundary Current, crucial for the initialisation and prediction of the AMOC? If such observations are deemed to be necessary to achieve skillful predictions of the AMOC, then where should limited observations be placed to improve the predictions? The design of ensembles that efficiently sample the uncertainty is also a priority, and is likely to help highlight regions where observations will be important. One such method is to create optimal perturbations (*Hawkins and Sutton* (2009b, 2010)) for decadal climate predictions, which could highlight regions where the predictions are most sensitive to observational errors, i.e. highlighting where observations would most improve the prediction.

To understanding the low frequency variability exhibited in the real world, and for ini-

tialised climate predictions to give skillful predictions, it is crucial that coupled climate models are improved. One way to ensure that the models are improved is to validate hindcasts of specific case studies, such as the warming of the North Atlantic in the mid 1990s, at the process level. Although the problems encountered with initialising predictions (i.e. the adjustments in the ocean) may limit the current reliability of the hindcasts, huge potential remains for using initialised climate predictions to understand the variability in the real world at the process level, and improve the models. The testing also allows the exploration of whether processes missing from the model are important for the predictions. For example, does including a stratosphere, or improving the sea ice model, improve the predictions of the case studies?

8.4 Concluding remarks

Decadal (or initialised) climate prediction is an area of climate science that is in its infancy. DePreSys represents one of the first attempts at building an operational decadal prediction system, and has been remarkably successful at showing that assimilating observed information into a climate model can yield more skillful predictions. However, DePreSys remains rather crude in its construction and a large array of questions remain regarding the construction, and delivery, of decadal climate forecasts. Thus, there is huge scope for future improvement, especially regarding the assimilation of information and, also, in the quality of the model used to make predictions. Therefore, it seems that DePreSys should be taken to represent the lower bound of what is possible in decadal climate prediction, assuming the high *potential* predictability on decadal time scales seen in model's is not just an artifact of their low resolution.

Appendix A

Description of Ocean Analyses used in this thesis

This appendix gives a brief description of the ocean analyses that are used in this thesis.

A.1 Levitus

The world Ocean Atlas-05 (also known as Levitus or WOA-05 (*Levitus et al.* (2005); *Locarnini et al.* (2006))) is a series of independent statistical analyses upon a $1^\circ \times 1^\circ$ grid. The analysis method is based upon the successive corrections weighted-average interpolation scheme of *Barnes* (1964), which updates a background (first guess) field with the observations at each analysis time. The background field is the zonal mean climatology at each basin for each latitude band and depth level. At each analysis time, the observations in a window centred on that time (one year for annual mean temperature) are used to update the background field. An observation at a particular grid point has a sphere of influence over which surrounding grid points have knowledge of that data, which is determined by the response function (in other words a decorrelation length scale). There are three iterations for the Levitus ocean analysis, which use progressively smaller decorrelation length scales of 888km, 666km and 444km. For each subsequent iteration, the background field is taken to be the analysed field from the previous iteration. The use

of the progressively smaller decorrelation length scales is to resolve small scale features in the ocean, as the analysed large scale pattern is not sensitive to the use of the smaller decorrelation length scale in later iterations (*Daley* (1991)).

A.2 Original Met Office Ocean analysis

The decorrelation length scale in the real ocean is unlikely to be constant with space (or in fact time). For example, temperature anomalies in the Tropical Pacific are likely to be correlated over much longer spatial scales than anomalies in features such as Western Boundary Currents. *Smith and Murphy* (2007) were concerned that the use of a single decorrelation length scale, as in Levitus, could create sub-optimal ocean analyses. Attempting to produce decorrelation length scales that are representative of the processes in the ocean, *Smith and Murphy* (2007) uses covariances constructed from a coupled climate model. More specifically a 50 year integration of HadCM3 (*Gordon et al.* (2000)), which includes representation of the major historical radiative forcings, is used to calculate a global covariance matrix for each grid box.

The global covariance matrix could be used to create the analysis. However, *Smith and Murphy* (2007) were concerned that model errors will create unrealistic covariances that would degrade the ocean analysis. Therefore, the influence of the covariances restricted to a defined box around each particular grid box. The size of the box is calculated by comparing the covariance matrix for the models Sea Surface Temperature (SST) with the covariance matrix calculated from HadISST (*Rayner et al.* (2003)). The zonal and meridional extent of the boxes are usually of the order of 40° and 20° respectively, although boxes in the western Tropical Pacific are up to 100° and 50° respectively (*Smith and Murphy* (2007)). Information is also spread through depth using model derived covariances, but similar restrictions are created to limit the covariances influence. The covariances are restricted globally in depth to the top 80m at the surface, to 1500m near the bottom of the ocean.

The analysis is performed upon anomalies, relative to 1941-1996, for each month from 1951-2006 (although 1951-1955 has been lost due to data corruption). To increase the

number of observational data for each monthly window, lagged correlations of the previous 6 months of data is used, with observations closer to the analysis time given more weight. Also to increase available information of salinity, the modelled covariance between the temperature and salinity is used to inflate the observational density of salinity.

A.3 Updated Met Office Ocean analysis

The update Met Office Ocean analysis is based upon the same design as the Original Met Office ocean analysis (*Smith and Murphy* (2007)), but has some key differences:

- The ability of the *Smith and Murphy* (2007) method to successfully recreate the real state of the global ocean is partly dependent upon HadCM3's ability to recreate the real world's covariance matrices. To reduce the model error in the derivation of oceanic covariances, an ensemble of 9 versions of HadCM3 are used to generate the covariance matrices, which all have different parameters in their atmospheric parameterisations (*Collins et al.* (2006b)).
- The horizontal boxes that limit the covariances influences are removed. Global covariances are used to generate the Updated Met Office ocean analysis.
- The analysis is iterated. The first analysis is created using only the model covariances. That first analysis is then used to create a new covariance matrix, which is used to create a the next iteration of the analysis. Therefore, the new covariance matrix uses the observed information as well as the model covariances. The iterations could be continued until the covariance matrix converges, but the iteration is computationally expensive. The analysis used here has had two iterations from the model covariance only analysis.

Unfortunately an error was present in the Updated Met Office ocean analysis so that when summed over the climatological period, the anomalies of temperature and salinity did not sum to zero. Figure A.1 shows the average of seasonal temperature anomalies over the climatological period (i.e. 1951-2006) from the Updated Met Office ocean analysis for

each season separately. It is clear that there is a significant offset in the temperature anomalies, which is different for each climatology.

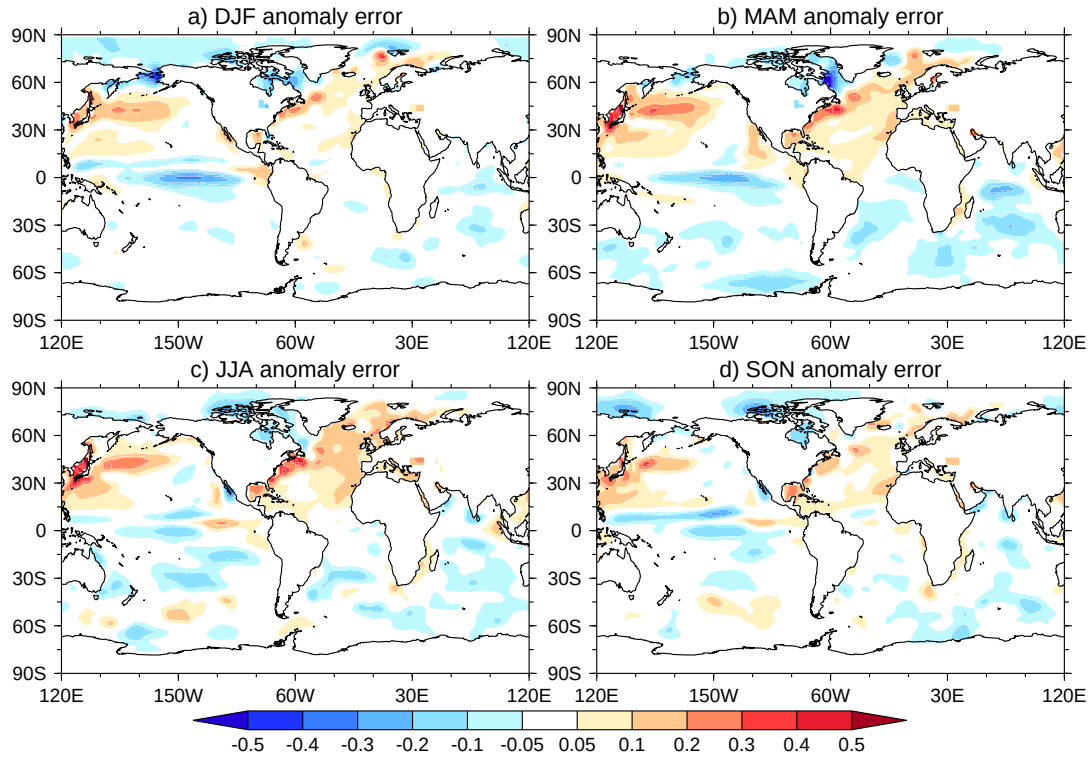


Figure A.1: a) shows the DJF 0-113m average temperature anomalies when averaged over the climatological period. The average of anomalies taken over the period they were made relative to, should be equal to zero. b), c) and d) show the same but for MAM, JJA and SON seasons respectively. Note the error is different for each season.

For the observational analysis presented in chapter 2, and the evaluation of mean skill presented in chapter 4, the error was *removed* by expressing the anomalies from a modified climatology. Unfortunately, the error was not known before we performed the hindcasts to test the hypotheses in chapter 7. The hindcasts in chapter 7, the DePreSys2 hindcasts, and the DePreSys PPE hindcasts were, therefore, initialised with an extra bias equal to that seen in figure A.1. Care must be taken in the future to ensure that such errors do not corrupt the skill of the hindcasts.

A.4 ECMWF Ocean Reanalysis System 3

The ECMWF Ocean Reanalysis system 3 (ORA-3; *Balmaseda et al.* (2008)) is a fully dynamical reanalysis, based upon a modified version of the Hamburg Ocean Primitive Equation (HOPE) model, and provides the oceanic initial conditions for the seasonal forecast system. The model has a $1^\circ \times 1^\circ$ grid with equatorial refinement of 0.3° in the meridional direction. The model has 29 vertical levels 11 of which occupy the top 140m. The first guess field is produced by integrating HOPE to the analysis time. The model is forced at the surface by daily fluxes of momentum, heat and fresh water. The fluxes are supplied from ERA-40 (*Uppala et al.* (2005)) from 1957 until June 2002, after June 2002 the ocean model is forced by the operational analyses.

Sea surface height, temperature and salinity are assimilated into the model sequentially. Temperature is assimilated into the top 2000m only using a 3D optimal interpolation scheme. The optimal interpolation scheme uses a Gaussian decorrelation length scale that varies with latitude. Within 4° of the equator, the zonal land meridional length scales are 1000km and 150km respectively. The decorrelation length scale evolves smoothly so that pole ward of 15° from the equator, the decorrelation length scale is 300km in all directions.

Bibliography

AchutaRao, K. M., B. D. Santer, P. J. Gleckler, K. E. Taylor, D. W. Pierce, T. P. Barnett, and T. M. L. Wigley (2006), Variability of ocean heat uptake: Reconciling observation and models, *J. Geophys. Res.*, *111*, doi:10.1029/2005JC003136.

AchutaRao, K. M., M. Ishii, B. D. Santer, P. J. Gleckler, K. E. Taylor, K. E. Barnett, P. D. W., R. J. Stouffer, and T. M. L. Wigley (2007), Simulated and observed variability in ocean temperature and heat content, *Proceedings of the National Academy of Sciences*, *104*.

Anderson, D., F. Doblas-Reyes, M. Balmaseda, and A. Weisheimer (2009), Decadal variability: processes, predictability and prediction, *ECMWF Tech. Memo. 591*.

Anthes, R. (1974), Data assimilation and initialization of hurricane prediction models, *Journal of the Atmospheric Sciences*, *31*(3), 702–719.

Balmaseda, M., A. Vidard, and D. Anderson (2008), The ECMWF Ocean Analysis System: ORA-S3, *Monthly Weather Review*, *136*, 3018–3034, doi: 10.1175/2008MWE2433.1.

Bao, J., and R. Errico (1997), An adjoint examination of a nudging method for data assimilation, *Monthly Weather Review*, *125*(6), 1355–1373.

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

Bell, M., M. Martin, and N. Nichols (2004), Assimilation of data into an ocean model with systematic errors near the equator, *Quarterly Journal of the Royal Meteorological Society*, *130*(598), 873–894.

Bellucci, A., and K. Richards (2006), Effects of NAO variability on the North Atlantic Ocean circulation, *Geophysical Research Letters*, *33*(2), L02,612.

Boer, G. (2000), A study of atmosphere-ocean predictability on long time scales, *Climate Dynamics*, *16*(6), 469–477.

Boer, G. (2004), Long time-scale potential predictability in an ensemble of coupled climate models, *Climate Dynamics*, *23*, 29–44, doi:10.1007/s00382-004-0419-8.

Boer, G., and S. Lambert (2008), Multi-model decadal potential predictability of precipitation and temperature, *Geophysical Research Letters*, *35*(5), L05706, doi: 10.1029/2008GL033234.

Böning, C., M. Scheinert, J. Dengg, A. Biastoch, and A. Funk (2006), Decadal variability of subpolar gyre transport and its reverberation in the North Atlantic overturning, *Geophys. Res. Lett*, *33*, L21S01, doi:10.1029/2006GL026906.

Bower, A., M. Lozier, S. Gary, and C. Böning (2009), Interior pathways of the North Atlantic meridional overturning circulation, *Nature*, *459*(7244), 243–247.

Brauch, J. P., and R. Gerdes (2005), Response of the northern North Atlantic and Arctic oceans to a sudden change of the North Atlantic Oscillation, *J. Geophys. Res.*, *110*, doi:10.1029/2004JC002436.

Brohan, P., J. Kennedy, I. Harris, S. Tett, and P. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, *J. Geophys. Res*, *111*(D12).

Bryan, K. (1969), Climate and the ocean circulation, *Monthly Weather Review*, *97*(11), 806–827.

Burgers, G., M. Balmaseda, F. Vossepoel, G. van Oldenborgh, and P. van Leeuwen (2002), Balanced ocean-data assimilation near the equator, *Journal of Physical Oceanography*, *32*(9), 2509–2519.

Burgman, R., A. Clement, C. Mitas, J. Chen, and K. Esslinger (2008), Evidence for atmospheric variability over the Pacific on decadal timescales, *Geophys. Res. Lett*, *35*.

Collins, M. (2002), Climate predictability on interannual to decadal time scales: the initial value problem, *Climate Dynamics*, *19*(8), 671–692.

Collins, M., and B. Sinha (2003), Predictability of decadal variations in the thermohaline circulation and climate, *Geophys. Res. Lett*, *30*, 1306, doi:10.1029/2002GL016504.

Collins, M., S. F. B. Tett, and C. Cooper (2001), The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments, *Climate Dynamics*, *17*, 61–81, doi:10.1007/s003820000094.

Collins, M., et al. (2006a), Interannual to Decadal Climate Predictability in the North Atlantic: A Multimodel-Ensemble Study, *Journal of Climate*, *19*, 1195–1203, doi: 10.1175/JCLI3654.1.

Collins, M., B. Booth, G. Harris, J. Murphy, D. Sexton, and M. Webb (2006b), Towards quantifying uncertainty in transient climate change, *Climate Dynamics*, *27*(2), 127–147.

Cummins, P., G. Lagerloef, and G. Mitchum (2005), A regional index of northeast Pacific variability based on satellite altimeter data, *Geophysical Research Letters*, *32*(17), L17,607.

Daley, R. (1991), *Atmospheric data analysis*, Cambridge Univ Pr.

Delworth, T., S. Manabe, and R. Stouffer (1993), Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model, *Journal of Climate*, *6*(11), 1993–2011.

Deser, C., and M. Blackmon (1993), Surface climate variations over the North Atlantic Ocean during winter: 1900-1989, *Journal of Climate*, *6*(9), 1743–1753.

Deser, C., M. Alexander, and M. Timlin (1996), Upper-ocean thermal variations in the North Pacific during 1970-1991, *Journal of Climate*, *9*(8), 1840–1855.

Deser, C., A. Phillips, and J. Hurrell (2004), Pacific interdecadal climate variability: Linkages between the tropics and the North Pacific during boreal winter since 1900, *Journal of Climate*, *17*(16), 3109–3124.

Deshayes, J., and C. Frankignoul (2008), Simulated variability of the circulation in the North Atlantic from 1953 to 2003, *Journal of Climate*, *21*, 4919–4933, doi: 10.1175/2008JCLI1882.1.

Doblas-Reyes, F., et al. (2009), Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts, *qjrms*, pp. 1538–1559, doi:10.1002/qj.464.

Dong, B., and R. Sutton (2002a), Adjustment of the coupled ocean–atmosphere system to a sudden change in the thermohaline circulation, *Geophysical Research Letters*, *29*(15), 1728.

Dong, B., and R. Sutton (2002b), Variability in North Atlantic heat content and heat transport in a coupled ocean-atmosphere GCM, *Climate Dynamics*, *19*(5), 485–497.

Dong, B., and R. T. Sutton (2005), Mechanism of Interdecadal Thermohaline Circulation Variability in a Coupled Ocean-Atmosphere GCM., *Journal of Climate*, *18*, 1117–1135, doi:10.1175/JCLI3328.1.

D'Orgeville, M., and W. Peltier (2009), Implications of both statistical equilibrium and global warming simulations with CCSM3. Part II: On the multidecadal variability in the North Atlantic basin, *Journal of Climate*, *22*, 5298–5318.

Eden, C., and J. Willebrand (2001), Mechanism of Interannual to Decadal Variability of the North Atlantic Circulation., *Journal of Climate*, *14*, 2266–2280, doi:10.1175/1520-0442(2001)014.

Elmore, K., M. Baldwin, and D. Schultz (2006), Field significance revisited: Spatial bias errors in forecasts as applied to the Eta model, *Monthly Weather Review*, *134*, 519–531.

Folland, C., et al. (2001), Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett*, *28*(13), 2621–2624.

Folland, C., J. Renwick, M. Salinger, and A. Mullan (2002), Relative influences of the Interdecadal Pacific Oscillation and ENSO on the South Pacific Convergence Zone, *Geophysical Research Letters*, *29*(13), 21–1.

Frankignoul, C., and K. Hasselmann (1977), Stochastic climate models. II- Application to sea-surface temperature anomalies and thermocline variability, *Tellus*, *29*, 289–305.

Frankignoul, C., P. Müller, and E. Zorita (1997), A Simple Model of the Decadal Response of the Ocean to Stochastic Wind Forcing, *Journal of Physical Oceanography*, *27*(8), 1533–1546.

Gibson, J., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano (1997), ERA description. ECMWF Re-Analysis Project Report Series 1, ECMWF, *Reading, UK*, p. 77.

Gordon, C., C. Cooper, C. Senior, H. Banks, J. Gregory, T. Johns, J. Mitchell, and R. Wood (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Climate Dynamics*, *16*, 147–168.

Gouretski, V., and K. P. Koltermann (2007), How much is the ocean really warming?, *Geophys. Res. Lett.*, *34*, l01,610, doi:10.1029/2006GL027834.

Gregory, J., and P. Forster (2008), Transient climate response estimated from radiative forcing and observed temperature change, *Journal of Geophysical Research-Atmospheres*, *113*(D23), D23,105.

Gregory, J., et al. (2005), A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO 2 concentration, *Geophys. Res. Lett*, *32*, L12,703.

Gregory, J. M., H. T. Banks, P. A. Stott, J. A. Lowe, and M. D. Palmer (2004), Simulated and observed decadal variability in ocean heat content, *Geophys. Res. Lett.*, *31*, doi: 10.1029/2004GL020258.

Griffies, S. M., and K. Bryan (1997a), Predictability of North Atlantic multidecadal climate variability, *Science*, *275*(5297), 181.

Griffies, S. M., and K. Bryan (1997b), A predictability study of simulated North Atlantic multidecadal variability, *Climate Dynamics*, *13*, 459–487, doi: 10.1007/s003820050177.

Haines, K., L. Hermanson, C. Liu, D. Putt, R. Sutton, A. Iwi, and D. Smith (2009), Decadal climate prediction (project GCEP), *Philosophical Transactions A*, *367*(1890), 925.

Hakkinen, S., and P. Rhines (2004), Decline of Subpolar North Atlantic Circulation During the 1990s, *Science*, *304*, 555–559, doi:10.1126/science.1094917.

Hakkinen, S., and P. Rhines (2009), Shifting surface currents in the northern North Atlantic Ocean, *Journal of Geophysical Research*, *114*(C4), C04,005.

Hall, A., and S. Manabe (1997), Can local linear stochastic theory explain sea surface temperature and salinity variability?, *Climate Dynamics*, *13*(3), 167–180.

Hasselmann, K. (1976), Stochastic climate models, *Tellus*, *28*(6), 31–762,896.

Hátún, H., A. B. Sandø, H. Drange, B. Hansen, and H. Valdimarsson (2005), Influence of the Atlantic Subpolar Gyre on the Thermohaline Circulation, *Science*, *309*, 1841–1844, doi:10.1126/science.1114777.

Hátún, H., M. Payne, G. Beaugrand, P. Reid, A. Sandø, H. Drange, B. Hansen, J. Jacobsen, and D. Bloch (2009), Large bio-geographical shifts in the north-eastern Atlantic Ocean: From the subpolar gyre, via plankton, to blue whiting and pilot whales, *Progress in Oceanography*, *80*, 149–162.

Hawkins, E., and R. Sutton (2007), Variability of the Atlantic thermohaline circulation described by three-dimensional empirical orthogonal functions, *Climate Dynamics*, pp. 1–18.

Hawkins, E., and R. Sutton (2009a), The potential to narrow uncertainty in regional climate predictions, *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107.

Hawkins, E., and R. Sutton (2009b), Decadal Predictability of the Atlantic Ocean in a Coupled GCM: Forecast Skill and Optimal Perturbations Using Linear Inverse Modeling, *Journal of Climate*, *22*(14), 3960–3978.

Hawkins, E., and R. Sutton (2010), Estimating climatically relevant singular vectors for decadal predictions of the Atlantic Ocean, *submitted*.

Hayes, S., D. Behringer, D. Hansen, M. Blackmon, and N. Lau (1986), The equatorial Pacific Ocean climate studies (EPOCS) plans 1986-1988, *EOS*, *67*, 442–444.

Hegerl, G., F. Zwiers, P. Braconnot, N. Gillett, Y. Luo, J. Orsini, N. Nicholls, J. Penner, and P. Stott (2007), Understanding and attributing climate change, *Climate change 2007*, pp. 665–745.

Hermanson, L., and R. Sutton (2009), Case studies in interannual to decadal climate predictability, *Climate Dynamics*, pp. 1–21.

Hoke, J., and R. Anthes (1976), The initialization of numerical models by a dynamic-initialization technique, *Monthly Weather Review*, *104*(12), 1551–1556.

Holland, D., R. Thomas, B. de Young, M. Ribergaard, and B. Lyberth (2008), Acceleration of Jakobshavn Isbræ triggered by warm subsurface ocean waters, *Nature Geoscience*, *1*, 659–664, doi:10.1038/ngeo316.

Hurrell, J. (1995), Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation, *Science*, *269*, 676–679, doi:10.1126/science.269.5224.676.

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

Ishii, M., M. Kimoto, and M. Kachi (2003), Historical ocean subsurface temperature analysis with error estimates, *Monthly Weather Review*, *131*(1), 51–73.

Johns, T., et al. (2003), Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios, *Climate Dynamics*, *20*(6), 583–612.

Jones, P., T. Osborn, K. Briffa, C. Folland, E. Horton, L. Alexander, D. Parker, and N. Rayner (2000), Adjusting for sampling density in grid box land and ocean surface temperature time series, *Journal of Geophysical Research-Atmospheres*, *106*(D4).

Keenlyside, N., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, *453*, 84–88, doi:10.1038/nature06921.

Kerr, R. (2000), A North Atlantic climate pacemaker for the centuries, *Science*, *288*(5473), 1984.

Knight, J., R. Allan, C. Folland, M. Vellinga, and M. Mann (2005), A signature of persistent natural thermohaline circulation cycles in observed climate, *Geophys. Res. Lett*, *32*, doi:10.1029/2005GL024233.

Knight, J. R., C. K. Folland, and A. A. Scaife (2006), Climate impacts of the Atlantic Multidecadal Oscillation, *Geophys. Res. Lett.*, *33*, L17706, doi: 10.1029/2006GL026242.

Krahmann, G., M. Visbeck, and G. Reverdin (2001), Formation and Propagation of Temperature Anomalies along the North Atlantic Current*, *Journal of Physical Oceanography*, *31*, 1287–1303.

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

Kushnir, Y., W. Robinson, I. Blade, N. Hall, S. Peng, and R. Sutton (2002), Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation, *Journal of Climate*, *15*, 2233–2256.

Laepple, T., S. Jewson, and K. Coughlin (2008), Interannual Temperature Predictions using the CMIP3 multi-model ensemble mean, *Geophysical Research Letters*, *35*(10), L10,701.

Latif, M., D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider (1998), A review of the predictability and prediction of ENSO, *Journal of Geophysical Research*, *103*(14), 375–14.

Latif, M., et al. (2004), Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with sea surface temperature, *Journal of Climate*, *17*(7), 1605–1614.

Latif, M., M. Collins, H. Pohlmann, and N. Keenlyside (2006), A Review of Predictability Studies of Atlantic Sector Climate on Decadal Time Scales, *Journal of Climate*, *19*, 5971–5987, doi:10.1175/JCLI3945.1.

Lazier, J., R. Hendry, A. Clarke, I. Yashayaev, and P. Rhines (2002), Convection and restratification in the Labrador Sea, 1990–2000, *Deep-Sea Research Part I*, *49*(10), 1819–1835.

Leadbetter, S. J., R. G. Williams, E. L. McDonagh, and B. A. King (2007), A twenty year reversal in water mass trends in the subtropical North Atlantic, *Geophys. Res. Lett.*, *34*, L12608, doi:10.1029/2007GL029957.

Levitus, S., and T. Boyer (1994), *World ocean atlas 1994, volume 4: Temperature*, PB–95-270112/XAB, National Environmental Satellite, Data, and Information Service, Washington, DC (United States).

Levitus, S., J. Antonov, J. Wang, T. Delworth, K. Dixon, and A. Broccoli (2001), Anthropogenic warming of Earth's climate system, *Science*, *292*(5515), 267.

Levitus, S., J. Antonov, and T. Boyer (2005), Warming of the world ocean, 1955-2003, *Geophys. Res. Lett.*, *32*, doi:10.1029/2004GL021592.

Levitus, S., J. Antonov, T. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov (2009), Global ocean heat content 1955-2008 in light of recently revealed instrumentation problems, *Geophys. Res. Lett.*, *36*, L07608, doi:10.1029/2008GL037155.

Livezey, R., and W. Chen (1983), Statistical field significance and its determination by Monte Carlo techniques, *Monthly Weather Review*.

Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, and H. E. Garcia (2006), *World Ocean Atlas 2005, Volume 1: Temperature*, NOAA Atlas NESDIS 61, U.S. Government Printing Office, Washington, D.C.

Lohmann, K., H. Drange, and M. Bentsen (2009), Response of the North Atlantic subpolar gyre to persistent North Atlantic oscillation like forcing, *Climate Dynamics*, *32*, 273–285, doi:10.1007/s00382-008-0467-6.

Lozier, M. S., S. Leadbetter, R. G. Williams, V. Roussenov, M. S. C. Reed, and N. J. Moore (2008), The Spatial Pattern and Mechanisms of Heat-Content Change in the North Atlantic, *Science*, *319*, 800–803, doi:10.1126/science.1146436.

Lyman, J. M., J. K. Willis, and G. C. Johnson (2006), Recent cooling of the upper ocean, *Geophys. Res. Lett.*, *33*, 18,604, doi:10.1029/2006GL027033.

Mann, H., and D. Whitney (1947), On a test of whether one of two random variables is stochastically larger than the other, *The Annals of Mathematical Statistics*, *18*(1), 50–60.

Mantua, N., and S. Hare (2002), The Pacific decadal oscillation, *Journal of Oceanography*, *58*(1), 35–44.

Mantua, N., S. Hare, Y. Zhang, J. Wallace, and R. Francis (1997), A Pacific interdecadal climate oscillation with impacts on salmon production, *Bulletin of the American Meteorological Society*, *78*(6), 1069–1079.

Marotzke, J., and B. Klinger (2000), The dynamics of equatorially asymmetric thermohaline circulations, *Journal of Physical Oceanography*, *30*(5), 955–970.

Marshall, J., et al. (2001a), North Atlantic climate variability: phenomena, impacts and mechanisms, *International Journal of Climatology*, *21*(15), 1863–1898.

Marshall, J., H. Johnson, and J. Goodman (2001b), A Study of the Interaction of the North Atlantic Oscillation with Ocean Circulation., *Journal of Climate*, *14*, 1399–1421, doi:10.1175/1520-0442(2001)014.

Masina, S., N. Pinardi, and A. Navarra (2001), A global ocean temperature and altimeter data assimilation system for studies of climate variability, *Climate dynamics*, *17*(9), 687–700.

Meehl, G., et al. (2007), Global climate projections, *Climate change 2007*.

Michaelsen, J. (1987), Cross-validation in statistical climate forecast models, *Journal of Applied Meteorology*, *26*(11), 1589–1600.

Munk, W., and C. Wunsch (1998), Abyssal recipes II: energetics of tidal and wind mixing, *Deep-Sea Research Part I*, *45*(12), 1977–2010.

Münnich, M., M. Latif, S. Venzke, and E. Maier-Reimer (1998), Decadal oscillations in a simple coupled model, *Journal of Climate*, *11*, 3309–3319.

Murphy, J., D. Sexton, D. Barnett, G. Jones, M. Webb, M. Collins, and D. Stainforth (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, *430*(7001), 768–772.

Nakicenovic, N., et al. (2000), *Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change*, PNNL-SA-39650, Cambridge University Press, New York, NY (US).

Newman, M., G. Compo, and M. Alexander (2003), ENSO-forced variability of the Pacific decadal oscillation, *Journal of Climate*, *16*(23), 3853–3857.

Pickart, R., M. Spall, M. Ribergaard, G. Moore, and R. Milliff (2003), Deep convection in the Irminger Sea forced by the Greenland tip jet, *Nature*, *424*(6945), 152–156.

Pierce, D., T. Barnett, N. Schneider, R. Saravanan, D. Dommenget, and M. Latif (2001), The role of ocean dynamics in producing decadal climate variability in the North Pacific, *Climate Dynamics*, *18*(1), 51–70.

Pohlmann, H., M. Botzet, M. Latif, A. Roesch, M. Wild, and P. Tschuck (2004), Estimating the decadal predictability of a coupled AOGCM, *Journal of Climate*, *17*(22), 4463–4472.

Pohlmann, H., J. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke (2009), Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic, *Journal of Climate*, *22*(14), 3926–3938.

Power, S., T. Casey, C. Folland, A. Colman, and V. Mehta (1999), Inter-decadal modulation of the impact of ENSO on Australia, *Climate Dynamics*, *15*(5), 319–324.

Rahmstorf, S. (2000), The thermohaline ocean circulation: A system with dangerous thresholds?, *Climatic Change*, *46*(3), 247–256.

Ramamurthy, M., and F. Carr (1987), Four-dimensional data assimilation in the monsoon region. Part I: Experiments with wind data, *Monthly Weather Review*, *115*(8), 1678–1706.

Rayner, N., D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research-Atmospheres*, *108*(D14), 4407.

Rhines, P., and R. Schopp (1991), The wind-driven circulation: Quasi-geostrophic simulations and theory for nonsymmetric winds, *Journal of Physical Oceanography*.

Robock, A. (2000), Volcanic Eruptions and Climate, *Reviews of Geophysics*, *38*, 191–219.

Saenko, O., J. Gregory, A. Weaver, and M. Eby (2002), Distinguishing the influence of heat, freshwater, and momentum fluxes on ocean circulation and climate, *Journal of Climate*, *15*, 3686–3697.

Schlesinger, M., and N. Ramankutty (1994), An oscillation in the global climate system of period 65–70 years, *Nature*.

Schneider, E., B. Huang, Z. Zhu, D. DeWitt, J. Kinter, B. Kirtman, and J. Shukla (1999), Ocean data assimilation, initialization, and predictions of ENSO with a coupled GCM, *Monthly Weather Review - USA*, *127*, 1187–1207.

Schneider, N., and B. Cornuelle (2005), The forcing of the Pacific decadal oscillation, *Journal of Climate*, *18*(21), 4355–4373.

Schneider, N., A. Miller, and D. Pierce (2002), Anatomy of North Pacific decadal variability, *Journal of climate*, *15*(6), 586–605.

Schuster, U., and A. J. Watson (2007), A variable and decreasing sink for atmospheric CO_2 in the North Atlantic, *J. Geophys. Res.*, *112*, doi:10.1029/2006JC003941.

Shaffrey, L., 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.

Smith, D., and J. Murphy (2007), An objective ocean temperature and salinity analysis using covariances from a global climate model, *Journal of Geophysical Research*, *112*, doi:10.1029/2005JC003172.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model, *Science*, *317*, 796–799, doi:10.1126/science.1139540.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller (Eds.) (2007), *Climate Change 2007, The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK.

Stauffer, D., and N. Seaman (1990), Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data, *Monthly Weather Review*, *118*(6), 1250–1277.

Stockdale, T. (1997), Coupled ocean–atmosphere forecasts in the presence of climate drift, *Monthly Weather Review*, *125*(5), 809–818.

Stott, P., S. Tett, G. Jones, M. Allen, J. Mitchell, and G. Jenkins (2000), External Control of 20th Century Temperature by Natural and Anthropogenic Forcings.

Sutton, R., and M. Allen (1997), Decadal predictability in Gulf Stream sea surface temperatures, *Nature*, *388*, 563–567.

Sutton, R. T., and D. L. R. Hodson (2005), Atlantic Ocean Forcing of North American and European Summer Climate, *Science*, *309*, 115–117, doi:10.1126/science.1109496.

Thompson, D., J. Kennedy, J. Wallace, and P. Jones (2008), A large discontinuity in the mid-twentieth century in observed global-mean surface temperature, *Nature*, *453*(7195), 646–649.

Thorpe, R. B., J. M. Gregory, T. C. Johns, R. A. Wood, and J. F. B. Mitchell (2001), Mechanisms Determining the Atlantic Thermohaline Circulation Response to Greenhouse Gas Forcing in a Non-Flux-Adjusted Coupled Climate Model., *Journal of Climate*, *14*, 3102–3116.

Timmermann, A., M. Latif, R. Voss, and A. Grötzner (1998), Northern Hemispheric interdecadal variability: A coupled air–sea mode, *Journal of Climate*, *11*(8), 1906–1931.

Trenberth, K., and J. Caron (2001), Estimates of meridional atmosphere and ocean heat transports, *Journal of Climate*, *14*(16), 3433–3443.

Trenberth, K., and J. Hurrell (1994), Decadal atmosphere-ocean variations in the Pacific, *Climate Dynamics*, *9*(6), 303–319.

Trenberth, K., and D. Shea (2006), Atlantic hurricanes and natural variability in 2005, *Geophysical Research Letters*, *33*, L12,704.

Uppala, S., et al. (2005), The ERA-40 re-analysis, Quart, *JR Meteorol. Soc*, *131*, 2961.

Vellinga, M., and P. Wu (2004), Low-Latitude Freshwater Influence on Centennial Variability of the Atlantic Thermohaline Circulation., *Journal of Climate*, *17*, 4498–4511.

Visbeck, M., H. Cullen, G. Krahmann, and N. Naik (1998), Ocean model's response to North Atlantic Oscillation-like wind forcing, *Geophysical Research Letters*, *25*(24), 4521–4524.

Wallace, J., E. Rasmusson, T. Mitchell, V. Kousky, E. Sarachik, and H. Von Storch (1998), On the structure and evolution of ENSO-related climate variability in the tropical Pacific: Lessons from TOGA, *Journal of Geophysical Research*, *103*(C7), 14,241–14,259.

Weaver, A., C. Deltel, E. Machu, S. Ricci, and N. Daget (2005), A multivariate balance operator for variational ocean data assimilation, *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3605–3626.

Webster, P., G. Holland, J. Curry, and H. Chang (2005), Changes in tropical cyclone number, duration, and intensity in a warming environment, *Science*, *309*(5742), 1844.

Wijffels, S., J. Willis, C. Domingues, P. Barker, N. White, A. Gronell, K. Ridgway, and J. Church (2008), Changing Expendable Bathythermograph Fall Rates and Their Impact on Estimates of Thermosteric Sea Level Rise, *Journal of Climate*, *21*(21), 5657–5672.

Wilcoxon, F. (1945), Individual comparisons by ranking methods, *Biometrics Bulletin*, *1*(6), 80–83.

Wilks, D. (1995), *Statistical methods in the atmospheric sciences: an introduction*, Academic press San Diego.

Wilson, S. (2000), Launching the Argo armada, *OCEANUS*, *42*(1), 17–19.

Wunsch, C., and R. Ferrari (2004), Vertical mixing, energy, and the general circulation of the oceans, *Annual Review of Fluid Mechanics*.

Yashayaev, I., M. Bersch, and H. Van Aken (2007), Spreading of the Labrador Sea Water to the Irminger and Iceland basins, *Geophysical Research Letters*, *34*(10), 10,602.

Zhang, R. (2007), Anticorrelated multidecadal variations between surface and subsurface tropical North Atlantic, *Geophys. Res. Lett*, *34*.

Zhang, R., and T. Delworth (2007), Impact of the Atlantic multidecadal oscillation on North Pacific climate variability, *Geophysical Research Letters*, *34*(23), L23,708.