Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ocean Modelling 11 (2006) 174–192
Ocean Modelling
www.elsevier.com/locate/ocemod

Using ensemble prediction methods to examine regional climate variation under global warming scenarios

J.C. Hargreaves *, J.D. Annan

Frontier Research Center for Global Change, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan
Proudman Oceanographic Laboratory, Joseph Proudman Building, 6 Brownlow Street, Liverpool L3 5DA, UK

Received 16 September 2004; received in revised form 16 December 2004; accepted 16 December 2004
Available online 18 January 2005

Abstract

The fate of the North Atlantic thermohaline circulation (THC) is of great significance for regional climate prediction. Research based on both numerical modelling and paleoclimate data has suggested that the THC might be intrinsically bistable, and could have the potential to switch rapidly between its stable modes. Using a low-resolution intermediate complexity model, we investigate the predictability of the response of the THC to anthropogenic forcing in the medium (100 years) and longer term. Using an ensemble Kalman filter we can efficiently tune the climate of ensemble members by varying multiple parameters simultaneously, and flux adjustments are not required to prevent unreasonable model drift. However, some biases remain, and we demonstrate that the common approach of subtracting the bias from a model forecast can result in substantial errors when the model state is close to a nonlinear threshold. Over 100 years of 1% per annum atmospheric CO_2 enrichment, the THC drops significantly but steadily by about 4 or 5 Sv, a result that appears robust over a wide range of scenarios. In the longer term, the THC can collapse entirely, or recover to its original state, and small changes in the present uncertainties can have a large effect on the future outcomes. We conclude that generating reliable forecasts over the next century should be achievable, but the long term behaviour remains highly unpredictable.

* Corresponding author. Address: Frontier Research Center for Global Change, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan. Tel.: +81 45 778 5577; fax: +81 45 778 5707.
E-mail address: jules@jamstec.go.jp (J.C. Hargreaves).

1463-5003/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2004.12.004

Keywords: Data assimilation; Thermohaline circulation; Climate science

1. Introduction

The climate of north–west Europe is strongly affected by the substantial northward transport of heat by the thermohaline circulation (THC) in the Atlantic Ocean, which is estimated to carry about 1.3 PW of heat northwards across the 25°N latitude line (Ganachaud and Wunsch, 2000). However, numerical modelling experiments (e.g. Manabe and Stouffer, 1988) suggest that this mode of ocean circulation is not unique, and that there is an alternative mode with much weaker THC which carries less heat, leading to a regional cooling of as much as 8 °C (Vellinga and Wood, 2002). Furthermore, paleoclimate records indicate that there have been rapid transitions between warm and cold conditions during both Dansgaard–Oeschger (Dansgaard et al., 1993) and Heinrich (Heinrich, 1988) events, and modelling shows that similar transitions can be triggered by changes in freshwater fluxes (Ganopolski and Rahmstorf, 2001).

It has been suggested that the anthropogenic forcing of the Earth's climate could generate changes in the strength of the THC, both through thermal forcing and from the resulting changes in the hydrological cycle. Experiments with some simpler models suggest that the THC could switch off in the longer term (Rahmstorf and Ganopolski, 1999) although other models indicate an initial drop followed by recovery even if atmospheric CO_2 levels are stabilised at 4× the present day value (Stouffer and Manabe, 2003).

Most projections of anthropogenically-forced climate change using more sophisticated models also show a reduction in the THC over the next 100 years, but there is a large range of uncertainty on these results, with the summary of results presented in the IPCC TAR ranging from a small increase to a decrease of more than 10 Sv over the next 100 years (Cubasch et al., 2001, Section 9.3.4.3). Although all of the model results presented there showed regional warming over that time interval, the rate of warming over NW Europe can be expected to depend heavily on any changes in the THC.

The research presented in the IPCC TAR is effectively an 'ensemble of opportunity', a compilation of integrations using a range of different models and scenarios. Therefore, it cannot be interpreted as a true probabilistic prediction, but merely a selection of scenarios which may cover only part of the range of possible outcomes and which provides no information as to their relative likelihoods. Recent developments in parameter estimation have now opened up the possibility of performing ensemble integrations of models which have been objectively tuned to climate observations, and which therefore have the potential to generate more meaningful probabilistic estimates of future climate.

Tuning the climate of a numerical model is a nonlinear multivariate parameter estimation problem, which has for some time been considered a rather intractable task. Early approaches to this problem have mostly been based around randomised or exhaustive multifactorial sampling of the multivariate parameter distribution (Forest et al., 2000; Knutti et al., 2002), but the cost of these methods is exponential in the number of parameters which are covaried, so they have limited practicality. The Ensemble Kalman filter (EnKF) Evensen (1994) has recently been developed applied to parameter estimation in climate models (Annan et al., 2004; Hargreaves et al., 2004). The method generates an ensemble of models with different parameter sets which sample the posterior probability distribution function defined by prior beliefs and observations of climate data.

This ensemble can then be used to make objective probabilistic estimates of climate change for different global warming scenarios. However, all models have error and their climatologies are regionally and even globally biased compared to reality. The GCMs which contribute to the global warming studies in the IPCC TAR actually have quite widely varying estimates of the present day climate (Lambert and Boer, 2001; Jia, 2003). In order to use such models for prediction, it is necessary to assume that the dominant component of model error is in the form of a fixed (or perhaps slowly drifting) bias. This may be one reason for the common practice of predicting future climate *change*, since the subtraction of the model's present day climate (or control run, which has a slow drift in addition to a fixed bias) from the future climate will eliminate this component of error. For example, the projections of the THC provided in Cubasch et al. (2001, Fig. 9.21) have all been rebased such that the model output is not plotted as absolute strength of overturning, but deviation from the models' average hindcast values over 1961–1990 (which vary by a factor of 3 from less than 10 Sv to greater than 30 Sv). The limitations of this approach do not appear to have been very well quantified but it does essentially represent the zeroth order approach to error correction.

Another method by which large biases are sometimes accounted for is via flux correction, when nonphysical terms (typically artificial sources and sinks of heat and moisture) are added to the model equations in order to prevent drift away from a desired climate state. However this approach is generally considered to be something of a last resort and is only used in cases where the basic model climatology would otherwise be wholly unrealistic. It is known that flux correction can qualitatively and substantially change model behaviour (e.g. Marotzke and Stone, 1995; Dijkstra and Neelin, 1995). When parameters in climate model are varied independently, in an attempt to explore the prior range of model behaviour, substantial imbalances often result which demand the imposition of flux corrections. However, our technique of multivariate optimisation ensures that the climate of the model is always plausible (within the constraints of the model's own limitations) and stable without any need for flux corrections. Although this source of error has therefore been eliminated, some biases remain in the model climate. The risks of assuming a fixed bias in climate predictions has not been so carefully investigated, and we will show how it may also lead to unreliable predictions.

In this paper we explore the predictability of the North Atlantic THC, and the significance of model biases, by applying the EnKF to a climate model of intermediate complexity and examining the changes in the North Atlantic THC under various global warming scenarios. The decline of the THC is a response partly to warming, and partly to changes in the hydrological cycle, and the relative significance of these factors differs among the model experiments previously performed. In this paper we conduct a number of experiments to estimate the range of results that might be expected from different hydrological responses. We use this intermediate complexity model because, being computationally fast it is suitable for work involving a large number of experiments. It is tuned to a realistic present day climatology based on ocean temperature and salinity, and surface air temperature and humidity.

In the next section, we briefly review the EnKF as applied to parameter estimation. We also show how it is possible to assimilate arbitrary climatological statistics into the ensemble in order to better initialise the model for a particular experiment. We further demonstrate the sensitivity of the ensemble THC response to different behaviours of the hydrological cycle, and how different initial states can affect the outcomes.

Although the model used in this research is dynamically rather stable and does not exhibit the internal variability characteristic of chaotic atmosphere–ocean dynamics, the EnKF has already been shown to work for highly chaotic models (Annan and Hargreaves, 2005; Annan et al., submitted for publication) so the methods adopted here should be highly relevant for future work in climate prediction using more complex models.

2. The ensemble Kalman filter and parameter estimation

The use of the EnKF for parameter estimation is described in Annan et al. (2004) and Hargreaves et al. (2004). Our implementation is identical to that used in these previous papers, so we will give only a brief overview.

The ensemble Kalman filter (Evensen, 1994) is an efficient Monte Carlo approximation to the Kalman filter equations (Kalman, 1960) which provide a method for optimal estimation in linear systems. It has been widely used in near-operational forecasting, especially for short-term numerical weather and ocean prediction. Evensen (2003) contains a full description of the theory and application of the method, along with a survey of recent applications, which need not be repeated here. Although the EnKF has generally been used for initial state estimation, parameter estimation can readily be included in the same framework, by the means of state space augmentation (Derber, 1989; Anderson, 2001). The idea here is that the parameters can be considered to be part of the model state alongside the conventional variables. Although this method has been known for some time, it does not appear to have been widely exploited, perhaps due to the added difficulties of combining parameter estimation with the initial value estimation problem of short-term forecasting using chaotic models (Pisarenko and Sornette, 2004). However, the method works well for climatological estimation where we are only interested in the long-term means rather than transient chaotic trajectories.

The state augmentation method also simplifies the assimilation of asynoptic and/or nonlinear observations, which includes not only observations of climatological mean values, but also the heat transport diagnostic which is used here. As in the case of parameters, the model state is augmented by its prediction of the measurement, and the ensemble samples the covariance matrix relating the measurement to all model variables. For an adjoint model, using such arbitrary observations would require the creation of an adjoint for the nonlinear observation operator.

Our previous application of the EnKF for parameter estimation in C-GOLDSTEIN is described in Annan et al. (2004) and Hargreaves et al. (2004). C-GOLDSTEIN is a low-resolution fully coupled global 2D atmosphere–3D ocean model, computationally efficient enough to allow multiple integrations over paleoclimatic time scales (Marsh et al., 2004; Rohling et al., 2004). A parallel super computer was used to integrate the ensemble members simultaneously, and a domain decomposition was implemented for the analysis step following the method of Keppenne (2000). An iterative process to converge to the (climatological) steady state solution was implemented. The system converges to the same solution as would be obtained in the linear case by a single standard analysis, but the iterative method appears to be more robust in nonlinear and high dimensional applications where the prior ensemble is very sparse and may with high probability contain no samples which are close to the posterior distribution.

3. Global warming experiments

In this paper we do not aim to make a confident prediction of the effect of anthropogenic forcing on the THC—indeed, given the limitations of the model and intractability of the problem, it will be clear that we cannot give such a prediction. Rather we wish to illustrate the range of behaviours present in an ensemble where all members are well-tuned to present day conditions and also we also aim to give some guide as to the magnitude of error in prediction that may be incurred by using an ensemble that is not well-tuned.

The basic global warming experiment used in this paper is a 100 year 1% p.a. increase of atmospheric CO_2 (to 2.7× present day) followed by a stabilisation at this fixed elevated CO_2 level for 5000 years. The radiative parameterisation was not varied or tuned in any way, so this experiment is really testing the response of the climate to a specified radiative forcing increase of 5.8 Wm^{-2} ($\Delta F_{2\times CO_2} = 4\ \mathrm{Wm}^{-2}$).

Later in this paper some further experiments are presented with other CO_2 scenarios, in order to test the limits of certain behaviours (2× and 4× CO_2 stabilisation experiments) and also for comparison with the work of other researchers and assessment of the affect of different CO_2 recovery rates (3.3× CO_2, with varying recovery rates). All the CO_2 increases for all experiments are at 1% p.a. All the ensembles used in this paper have 54 members for computational reasons.

4. An ensemble tuned to present day climatology

In Hargreaves et al. (2004) we used the method described in the previous section to assimilate climatological data collected over recent decades and derive an estimate of present day climatology. The data consisted of 3D fields of ocean temperature and salinity, from 1945 to 1998 (Levitus, 1998), and 2D NCEP/NCAR atmospheric reanalysis data (surface air temperature and humidity only) averaged between 1948 and 2002 (NCEP Reanalysis data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/). We varied 12 model parameters and produced an ensemble of 54 tuned models. The tuned ensemble looked surprisingly realistic, particularly in the ocean where, despite the limited resolution of the model, the large scale climatological diagnostics are at least as good as most, and better than many, of the much more computationally expensive models of the Coupled Model Intercomparison Project as analysed by Jia (2003). The atmospheric component of the model is a rather simple 2D energy–moisture balance model and does not perform so well compared to state of the art GCMs, but still has reasonable diagnostics over large scales. Fig. 1 shows sea surface temperature (SST) and salinity (SSS) and atmospheric temperature and humidity for the ensemble mean (a–d), and the differences between the ensemble mean and the data used in the assimilation (e–h), indicating the horizontal resolution of the model, as well as some of the features of the results. Basin-wide and global averages are well reproduced, but systematic errors at the regional scale remain. Most noticeably, the atmospheric variables are unrealistically

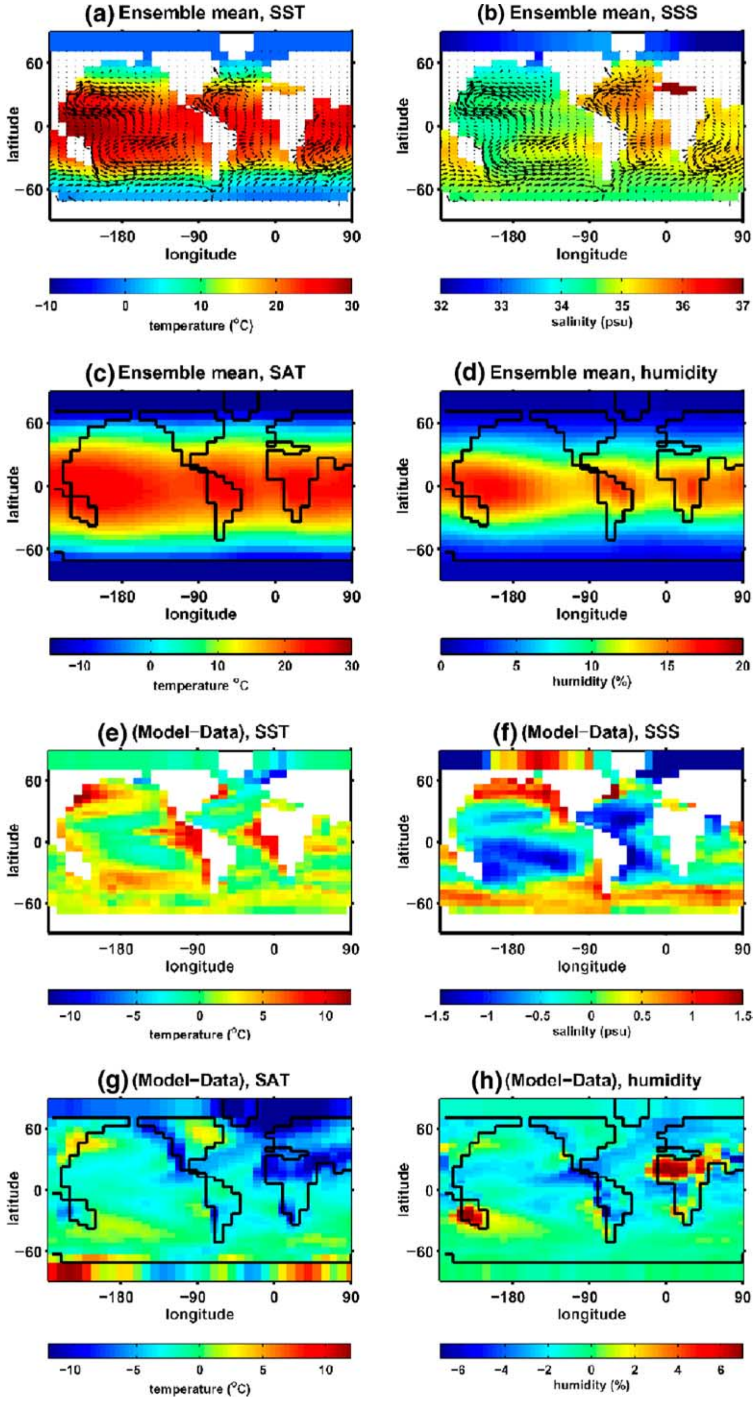

Fig. 1. (a–d) Ensemble mean surface ocean and atmosphere: SST = sea surface temperature; SSS = sea surface salinity. (e–h) Difference between the model ensemble mean and the data used for the assimilation, for the variables shown in the upper plots.

smooth. The SST shows a warm bias to the east of each ocean basin and the SSS has insufficient variability, with regional maxima being too low and regional minima too high compared to the data. These results are discussed in more detail in Hargreaves et al. (2004).

Fig. 2 shows some model diagnostics (meridional overturning and transport and upper and lower temperatures) at 25°N in the North Atlantic. For each diagnostic the ensemble produced by the EnKF has a spread which approximately matches the range of uncertainty of the observations. In contrast, the CMIP ensemble of opportunity has a much broader range. The right plot of Fig. 2 shows variables that were assimilated into the EnKF so the good agreement there is not unexpected. The diagnostics shown in the left plot were not directly used in the assimilation, however, and thus act as an independent validation. Although some model bias is clearly evident (in particular, the heat transport is too low, for reasons discussed below), our results are well within the range of those of the CMIP ensemble.

In addition to experiments with a tuned ensemble, we also performed an integration with an ensemble where the parameters were sampled directly from their prior estimates, with no tuning other than to discard members which had a negligible THC (<5 Sv). The output from this ensemble is not plotted but it has a similar spread in the diagnostics at 25°N in the North Atlantic to that of the CMIP ensemble. Outputs from a very similar ensemble are further described by Edwards and Marsh (2004). In contrast to the work presented here, their selection of the "good" runs from their ensemble did not substantially reduce the width of results, due to a more generous interpretation of what is considered "good". We use the results from this prior ensemble to aid comparison between the results of the tuned ensemble and those of the CMIP ensemble. Although

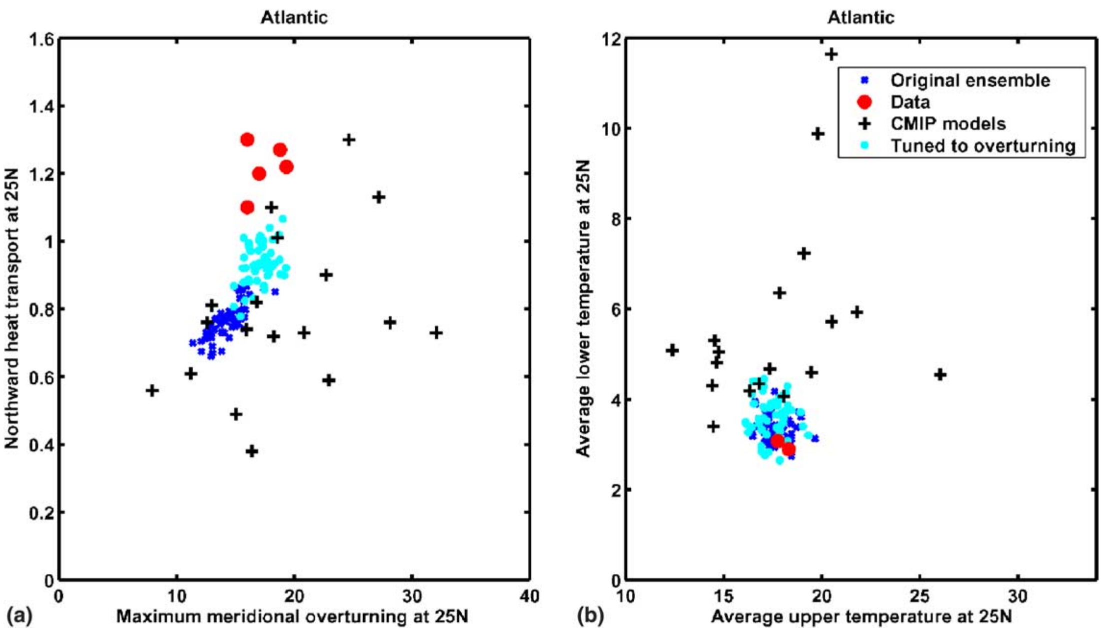

Fig. 2. Comparison with Coupled Model Intercomparison Project, Jia (2003). Dark blue crosses are the original tuned ensemble; cyan crosses are the ensemble with re-tuned THC; black crosses show the results from the CMIP project; red dots are the observational data (Jia, 2003).

our prior ensemble only samples parameter variability and does not include the structural differences between the CMIP models, there is some evidence that parameters alone can account for large differences. Murphy et al. (2004) have presented an ensemble of GCM simulations in which the variability due to parameter variation in a single model greatly exceeds that found in an ensemble of opportunity.

4.1. Results from the global warming experiment

We performed the basic (2.7× CO_2 followed by stabilisation) global warming experiment described in Section 3 on the tuned ensemble. As explained in Hargreaves et al. (2004), the parameters which we tuned are those which predominantly affect the global heat and mass transports (diffusion and friction). These have only a minor effect on the overall radiative balance of the Earth, and so even though the climate sensitivity of the model is very reasonable (a doubling of CO_2 giving rise to a steady-state globally-averaged temperature increase of 2.9 °C), the global temperature response is constrained to lie in an unrealistically tight range (1 standard deviation width of 0.1 °C). We also performed the 100 years CO_2 enrichment experiment on the prior ensemble. The average rise in temperature was very similar to the tuned ensemble, but the range in warming was somewhat wider, although still too narrow, at 0.25 °C.

Further developments of the model and system are planned to address this limitation, but at a regional level the response is already much more interesting, as shown by Fig. 3, which shows the change in the maximum North Atlantic overturning for each ensemble member of the tuned ensemble (hereafter OT). The results for ensemble OT shown in Tables 1 and 2 are the ensemble mean and standard deviations of the North Atlantic THC after 100 years of global warming and after a further 5000 years stabilisation respectively. The results from the prior ensemble are also shown in Table 1. The other results shown in Tables 1 and 2 are discussed later in the paper.

For ensemble OT, after 100 years of global warming, the THC has dropped substantially for all ensemble members. For the ensemble sampled directly from the prior, the range in the THC drop was twice as large as that for the tuned ensemble, with a 2 standard deviation range of 1–6.8 Sv (which is about half of the range in results shown in the IPCC TAR). The substantial difference between the spread of the THC drop in ensembles OT and Prior may give some indication of the magnitude of the errors that could hidden in the IPCC THC prediction by assuming the validity of bias subtraction.

Ensemble OT was run for a further 5000 years at the enhanced 2.7× CO_2 level. The ensemble results were dramatic compared to the 100 year result described in the last paragraph. During the following thousands of years, a large divergence of the ensemble members is apparent, with some undergoing a collapse in the THC and the remainder recovering to a strength close to, but systematically lower than, their original value. For those which recover, the spread of results is slightly wider than for the original ensemble. For those members where the THC has collapsed, the temperature locally in the North Atlantic is approximately 5 °C colder than for those where the THC recovers to near its original strength. We note that all the changes in THC are relatively slow compared to freshwater pulse experiments (Manabe and Stouffer, 1995; Dong and Sutton, 2002) with the shutdown occurring on the century time scale rather than decadal.

Ensemble OT has a width in initial values of the THC which is comparable with the spread of the observations. Therefore the result of the global warming experiment suggest that if the model is

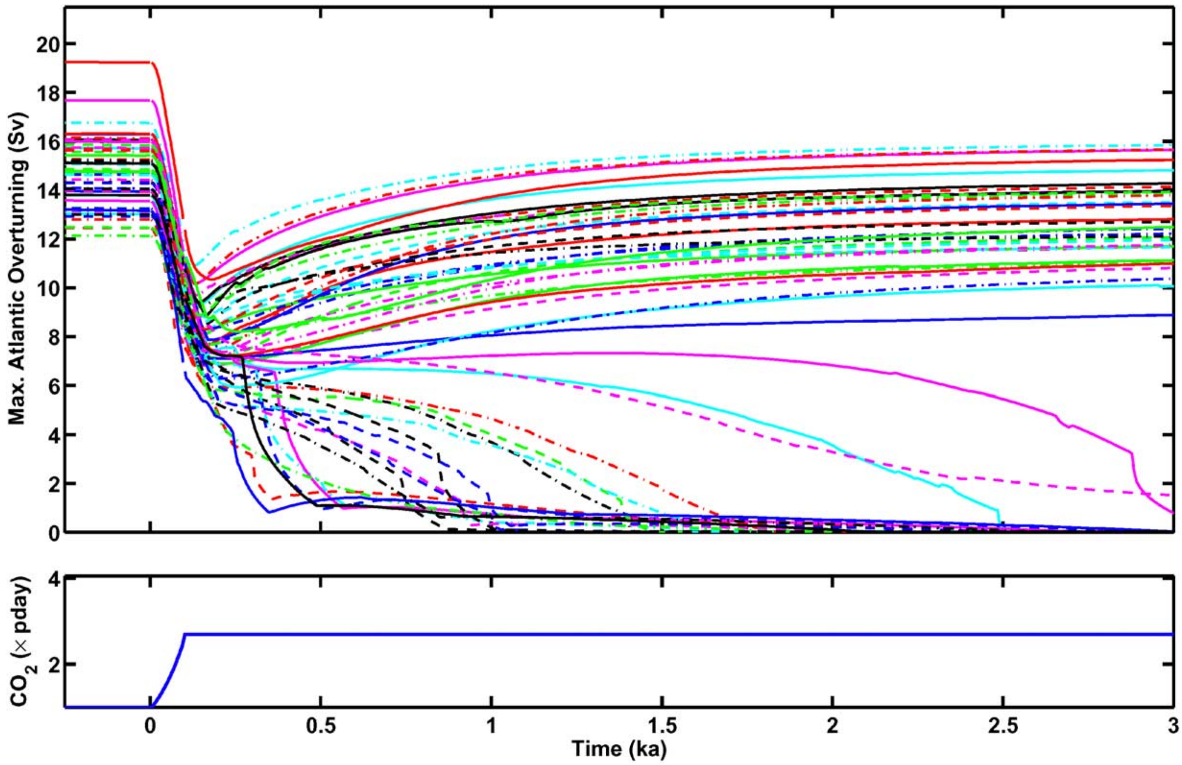

Fig. 3. Ensemble OT: THC Collapse and Recovery caused by a 2.7× increase in CO_2. After 5000 years, 37% of ensemble members have collapsed THC.

Table 1
Maximum North Atlantic overturning for 100 years global warming experiment

Ensemble	Maximum North Atlantic overturning		
	Initial (Sv)	After 100y (Sv)	Difference (Sv)
OT	14.65 ± 1.40	9.35 ± 1.30	5.30 ± 0.58
RT	17.74 ± 1.31	13.58 ± 1.37	4.16 ± 0.61
RTM	17.74 ± 1.31	12.63 ± 1.39	5.12 ± 0.64
RTP	17.74 ± 1.31	12.56 ± 1.39	5.18 ± 0.65
RTA	17.74 ± 1.31	13.05 ± 1.40	4.69 ± 0.65
Prior > 0	20.48 ± 6.78	17.05 ± 6.13	3.42 ± 1.23

Key to rows: OT: tuned to present day data; RT: retuned to the THC; RTA: RT plus additional FWF from tropical Atlantic to North Atlantic; RTP: RT plus additional FWF from North Pacific to North Atlantic; RTM: RT plus additional FWF from Greenland ice sheet to North Atlantic; Prior > 0: parameters chosen randomly from prior ranges, selecting runs with positive THC.

reasonably accurate, the state of the present day THC may be so close to the bifurcation point of its hysteresis curve that the observations that have been used to tune the model here may be insufficient for us to predict the long-term fate of the THC under a particular climate change scenario.

Table 2
Maximum North Atlantic overturning after 5000 years stabilisation at 2.7× CO_2

	Maximum North Atlantic overturning		
Ensemble	% collapsed	After 5000y, uncollapsed (Sv)	Difference after 5000y, uncollapsed (Sv)
OT	37	12.99 ± 1.62	4.95 ± 1.82
RT	0	17.08 ± 1.76	0.66 ± 1.04
RTM	54	13.52 ± 1.92	4.81 ± 1.51
RTP	56	12.81 ± 2.35	5.55 ± 1.90
RTA	11	16.24 ± 2.16	1.59 ± 1.31

Key to rows: same as Table 1.

It is, however, noticeable that those ensemble members for which the THC collapsed tend to be the ones with a lower initial value (80% of the members which collapse, start with their THC in the lower half of the ensemble range). There is, however, no correlation between initial THC value and the magnitude of the drop over the first 100 years. Since, as shown in Fig. 2, the initial North Atlantic overturning of ensemble OT appears to be systematically lower than the observations, it is possible that this bias in the initial state is influencing the results.

5. An ensemble tuned to observed North Atlantic overturning

From the previous experiment, it appears that those ensemble members with lower overturning were more likely to collapse than those with higher overturning. This raises the issue of whether the bias can reasonably be treated as an additive error, since an ensemble with lower initial bias might be less likely to reach the bifurcation point under external forcing. Therefore, we would like to investigate the extent to which altering the model bias can generate a qualitative change in model results.

In order to minimise the risk of the biased model state generating an artificial nonlinear response, it is reasonable to assume as a first step that the bias itself should be minimised. Of course the model error is already minimised in a global sense by the tuning procedure, but for a regional forecast it may be preferable to tune more closely to the particular region in question, and also perhaps to the processes that are thought to be most important for the forecast. This will necessarily result in a worsening of the model performance in other areas, but if they are distant and unrelated then this should not matter. In this case, a more realistic fit to the current THC is probably more important than the salinity structure in the North Pacific, for example.

Therefore, we now try tuning the THC directly so that it comes more closely into line with the observations described in Jia (2003). One useful feature of the EnKF method is that arbitrary (asynoptic and/or nonlinear) observations of climate diagnostics can be easily used to tune the ensemble. In this case, we simply augment the model state with a pair of variables that 'observe' the model's THC strength and northward heat transport. To avoid confusion, we emphasise that this results in a tuning of the steady state climate via parameter adjustment, and is not merely a case of initialising the existing ensemble members with a different model state field.

We note that the THC data are not fully independent of the climatological data which we are already using. Several of these estimates of the THC are based on model inversions, with the model initialised from (and also roughly tuned to) a data set which overlaps substantially with the global ocean observations which we are already using. In order to reduce the problem of overfitting the model and generating an unreasonably low-spread in the resulting ensemble (Allen et al., 2002), the closeness of the fit to the spatial data was weakened (by doubling the assumed observational error values) at the same time as the new diagnostic data were introduced. As a result, the overall fit of the ensemble members to the global data was worsened slightly (in the RMS sense) by about 5%.

The observed estimates of North Atlantic THC range between 16 and 21 Sv (Jia, 2003). The observed northward neat transport at 25° is around 1.2 PW. We added this information into the assimilation and re-ran the EnKF. The re-tuned ensemble is shown as the cyan crosses in Fig. 2 and results are labelled as RT in the Tables. The THC at 25°N is now consistent with the observations. The net northward heat transport has also increased in the re-tuned ensemble, but it is still slightly too low (although clearly better than most of the CMIP ensemble). As can be seen in Fig. 1, the model has a warm bias on the east side of the Atlantic ocean basin and this results in a negative contribution to the northward heat transport of the order of 0.1–0.2 PW due to the barotropic gyre. Therefore we consider that our somewhat low-value of 0.9 ± 0.1 PW is acceptable for this study which is focussing on the changes to the overturning component of the circulation. Since the model is constrained by its own dynamics to lie close to the line along which the combined ensemble results are distributed in the left hand plot of Fig. 2 (i.e. overturning in Sv $\simeq 20\times$ heat transport in PW), attempting to improve the heat transport further by weakening the fit to Levitus data inevitably results in an erroneously high overturning. Since the heat transport error can be directly attributed to a model weakness which has a relatively minor effect on the dynamics, we consider that the retuned ensemble shown is most appropriate for investigations into the fate of the THC.

5.1. Results from the global warming experiment for the retuned ensemble

The $2.7\times$ CO_2 global warming experiment outlined in Section 3 was run on the retuned ensemble. The results are shown in Tables 1–3 where the ensemble is labelled as ensemble RT. The global temperature results are almost identical to those for the OT ensemble so we immediately move on to discussing the THC results in more detail.

In the RT ensemble the initial North Atlantic overturning is considerably higher than in the OT ensemble. Over the first 100 years of the experiment the overturning decreases slightly less than in

Table 3
Fresh water flux (FWF) into the Atlantic, north of 50N for $2.7\times$ CO_2 experiment

Ensemble	Initial FWF (Sv)	ΔFWF_{100} (Sv)	ΔFWF_{5000} (Sv)
OT	0.161 ± 0.02	0.046 ± 0.004	0.061 ± 0.007
RT	0.145 ± 0.035	0.044 ± 0.006	0.062 ± 0.010
RTM	0.145 ± 0.035	0.095 ± 0.006	0.137 ± 0.012
RTP	0.145 ± 0.035	0.098 ± 0.007	0.138 ± 0.011
RTA	0.145 ± 0.035	0.096 ± 0.007	0.145 ± 0.010

Key to rows: as Table 1. Columns: FWF before CO_2 increased; change in FWF after 100 years of increasing CO_2; change in FWF after 5000 years of stabilisation at $2.7\times$ CO_2.

the OT run but the spread of results is similar for the two ensembles. There are, however, clear differences between runs OT and RT for the 2.7× CO_2 5000 year stabilisation part of the experiment. For ensemble RT the THC does not collapse in any of the ensemble members. As shown in Fig. 4 even for a 4× CO_2 experiment, this ensemble has a very stable THC, illustrating that the estimate of steady state climate in the model has moved away from the THC bifurcation point. For this 4× CO_2 experiment, although the THC recovers in all the ensemble members, the final range of maximum overturning is considerably greater than that of the initial RT ensemble, with some ensemble members recovering to higher than the initial values and others to lower. Although these results appear less dramatic than for ensemble OT it again shows evidence of an initial ensemble, with only small differences in the present day climatology of the ensemble members, producing a range of characteristically different results within a single global warming scenario.

These results illustrate the risks of assuming that a bias is fixed and can safely be subtracted. Although in many situations calculating a climate change in this way will give good results, there is a risk that it will be unrealistic in situations where the climate state is near to a threshold. In this case, the implicit assumption of a fixed bias or slow drift breaks down, and it is necessary to explicitly consider the likelihood of a nonlinear response. It is not clear which of our model experiments is really the most realistic, given the substantial limitations of the model (discussed further in the next section) but the experiments performed here certainly demonstrate the importance of the issue. For a model with significant faults, it seems to be necessary to consider which areas

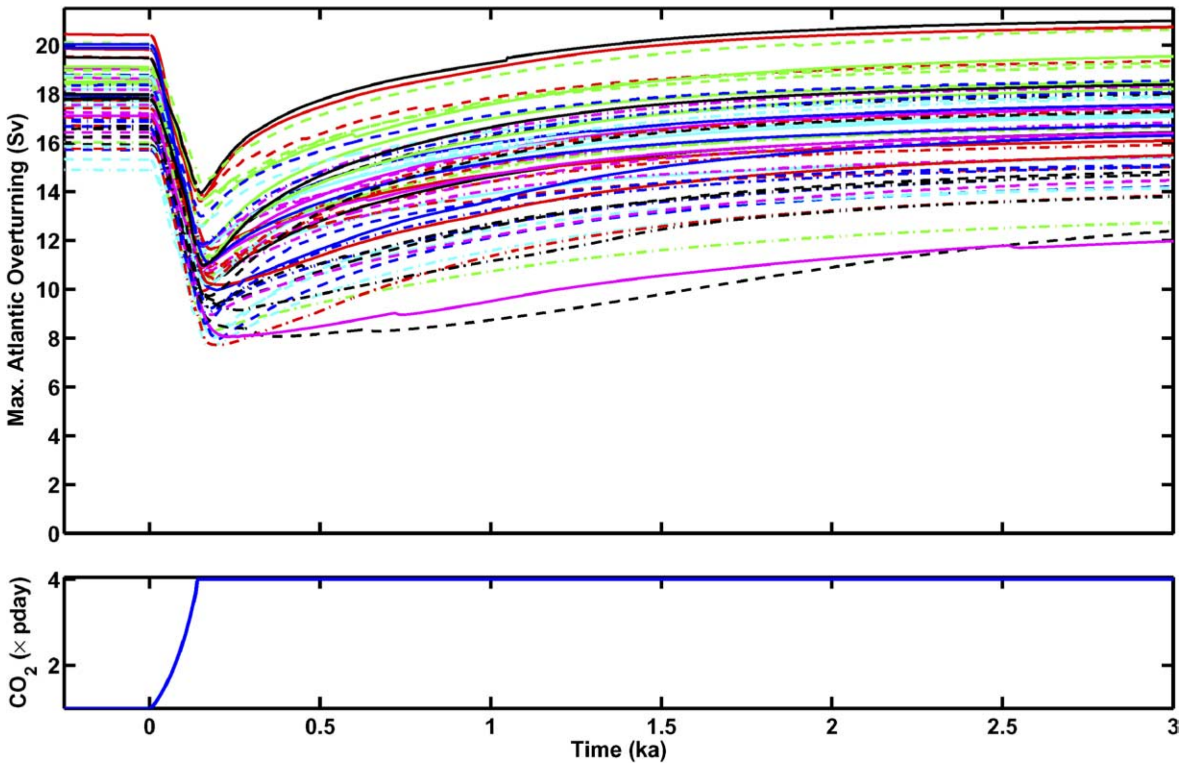

Fig. 4. THC of ensemble RT, at 4× CO_2 0% of ensemble members collapse.

(perhaps processes, or physical regions) should be more carefully tuned (by giving greater weight to the most relevant observations) at the expense of a less accurate representation in other areas. One interpretation of our results could be that both ensembles are individually too narrow, and a much wider meta-ensemble covering both sets is the only reasonable answer. It certainly seems that more work is needed in exploring how the results of climatological estimation problems depend on the prior assumptions.

6. Enhanced hydrological response

One significant weakness of atmospheric models of this simple type is that their moisture transport is weak (it is substantially augmented by fixed fluxes between the Atlantic to Pacific in our model) which means it may not respond realistically under the anthropogenic forcing. Table 3 shows the freshwater flux (FWF) into the Atlantic north of 50°N for the 2.7× present day CO_2 experiments discussed in this paper. It is thought that the hydrological cycle will change with global warming, although exactly what will happen is still a matter of much debate. Some models show an increase of freshwater flux from the Pacific to the North Atlantic which weakens the THC (Cubasch et al., 2001, Section 9.3.4.3) whereas in others this effect is counteracted by enhanced freshwater flux from the tropical Atlantic, which stabilises or even strengthens the THC (Latif et al., 2000). Meltwater from the Greenland ice sheet is another factor which may play a significant rôle. Since C-GOLDSTEIN exhibits very little change in FWF with change in temperature, we now explore this situation by introducing a modified hydrological response for each of the three sources of freshwater to the North Atlantic described above.

For these experiments we follow the work of Rahmstorf and Ganopolski (1999), hereafter RG99, who performed similar experiments with the CLIMBER model. RG99 added an additional freshwater flux ΔF into the Atlantic North of 50°N, given by $\Delta F = k\Delta T_{\mathrm{NH}}$, where ΔT_{NH} is the change in mean northern hemisphere temperature compared to pre-industrial conditions and k is an adjustable scaling factor. In our case we take 'pre-industrial' to mean the CO_2 level of our tuned ensemble and perform all our experiments as multiples of this value. The data used in the tuning of the ensemble spans the years 1945–2002, so our effective 'pre-industrial CO_2' model state may be equivalent to slightly more than the usual 280 ppm, but this error is small compared to the significant biases already present.

RG99 experimented with three sources of freshwater to the North Atlantic: from Greenland ice sheet melt; from the North Pacific; and from the tropical Atlantic. They tuned k in order to achieve a maximum flux into the North Atlantic of about 0.1 Sv under global warming conditions. They performed 2× and 4× CO_2 equilibrium experiments and also conducted an experiment where the CO_2 rose to approximately 3.3× CO_2 and then declined to about 1.5× CO_2 over about 800 years.

In order make our results comparable to those of CLIMBER we follow similar experiments. In addition we performed a 2.7× CO_2 experiment in order to provide a comparison with the results presented earlier in this paper. We set k to the same value for all these experiments yielding a flux of about 0.12 Sv into the North Atlantic at 3.3× CO_2. Since the actual amount of extra flux also depends on the amount of northern hemisphere warming, there are slight differences in the amount of extra fresh water flux (FWF) gained by each ensemble member. However, for the ensemble mean, the model has a natural response to global warming of 0.02 Sv/K. For these experiments we added $k = 0.02$ Sv/K to all ensemble members, giving a total model response of, on average, 0.04 Sv/K, which is equivalent to the middle of the three experiments performed by RG99. All the experiments start from the ensemble with re-tuned THC (ensemble RT).

6.1. Results from the global warming experiments including an enhanced hydrological response

The results for the 2.7× CO_2 global warming experiments with enhanced hydrological response are shown in Tables 1–3. The three ensembles are labelled in accordance with the source of the additional FWF into the North Atlantic. Ensemble RTA has a tropical Atlantic freshwater source, RTP has a Pacific source and for RTM all the FWF is assumed to be from Greenland meltwater. Here we mainly discuss the RTP (Pacific source) results and then comment on the differences between the results with different sources later in this section.

As shown in Table 3 the fresh water flux to the North Atlantic under global warming is significantly increased over that found in the ensembles discussed earlier in this paper. Over the first 100 years of global warming, the THC for ensemble RTP drops by more than that for the ensemble RT, and is more comparable with the drop observed for the original optimally tuned ensemble, OT. The longer term effect of the increased FWF to the North Atlantic is to destabilise the THC, and after 5000 years the THC has collapsed in 30 (56%) of ensemble members.

The left and right plots of Fig. 5 show the response of the North Atlantic overturning to 4× and 2× CO_2 experiments respectively for ensemble RTP. The overturning results from these experiments is summarised is Table 4. For the 4× CO_2 experiment all but one of the ensemble member collapse, but for the 2× CO_2 experiment only 8 members (15%) collapsed. The majority vote in each case is in agreement with those obtained by the CLIMBER model in RG99, where THC collapse and recovery were observed for the 4× and 2× CO_2 experiments respectively.

We also repeated (approximately) the global warming scenario illustrated in Fig. 2(a) of RG99. In our experiment the CO_2 rises to about 3.3× present day before declining slowly (330 years at −0.13% followed by 500 years at −0.07%) back to about 1.5× present day over about 800 years.

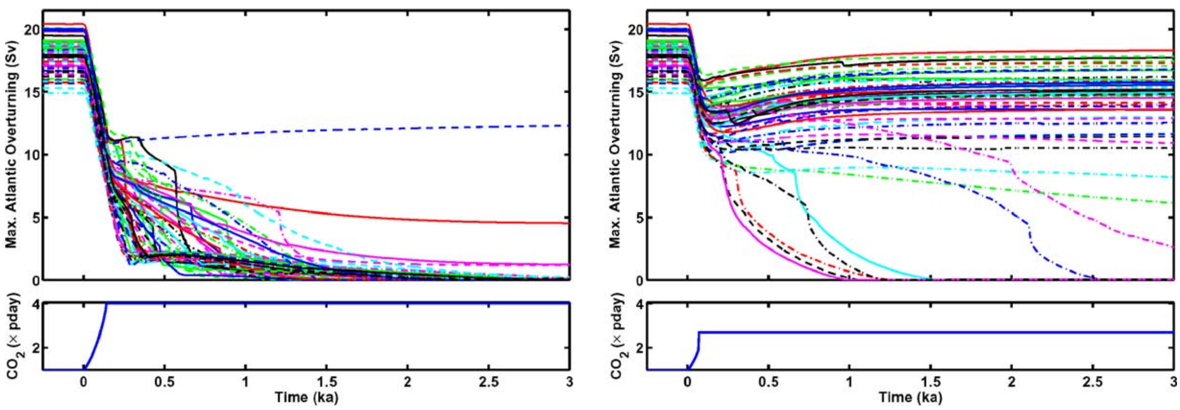

Fig. 5. Maximum Atlantic overturning for ensemble RTP for 4× CO_2 (98% collapse) and 2× CO_2 (15% collapse) experiments.

Table 4
The response of the North Atlantic overturning for ensemble RTP with enhanced FWF to the North Atlantic

Experiment	Maximum North Atlantic overturning			
	GW response (Sv)	5000 year response (Sv)	% collapse	RG99-res
2× CO_2	14.62 ± 1.26	14.69 ± 2.27	15	'recovery'
2.7× CO_2	12.56 ± 1.39	12.81 ± 2.35	56	
3.3× CO_2 fast	11.13 ± 1.50	16.76 ± 1.62	2	
3.3× CO_2 RG		16.00 ± 1.68	15	19 – 10 – 16 Sv
3.3× CO_2 flat		10.44 ± 3.00	87	
4× CO_2	9.75 ± 1.57	12.44	98	'collapse'

Initial max. N.Atl. THC is (see Table 1), 17.74 ± 1.31 Sv. Column RG99-res: results for similar experiments for the CLIMBER model described in RG99.

The results from ensemble RTP for this experiment are shown in Fig. 6 and Table 4. We observed a range of responses in the North Atlantic overturning with 15% of the ensemble members undergoing permanent THC collapse, despite the recovery of the CO_2 to 1.5× pre-industrial levels. After 3000 years of the experiment the ensemble members have settled into either a collapsed or

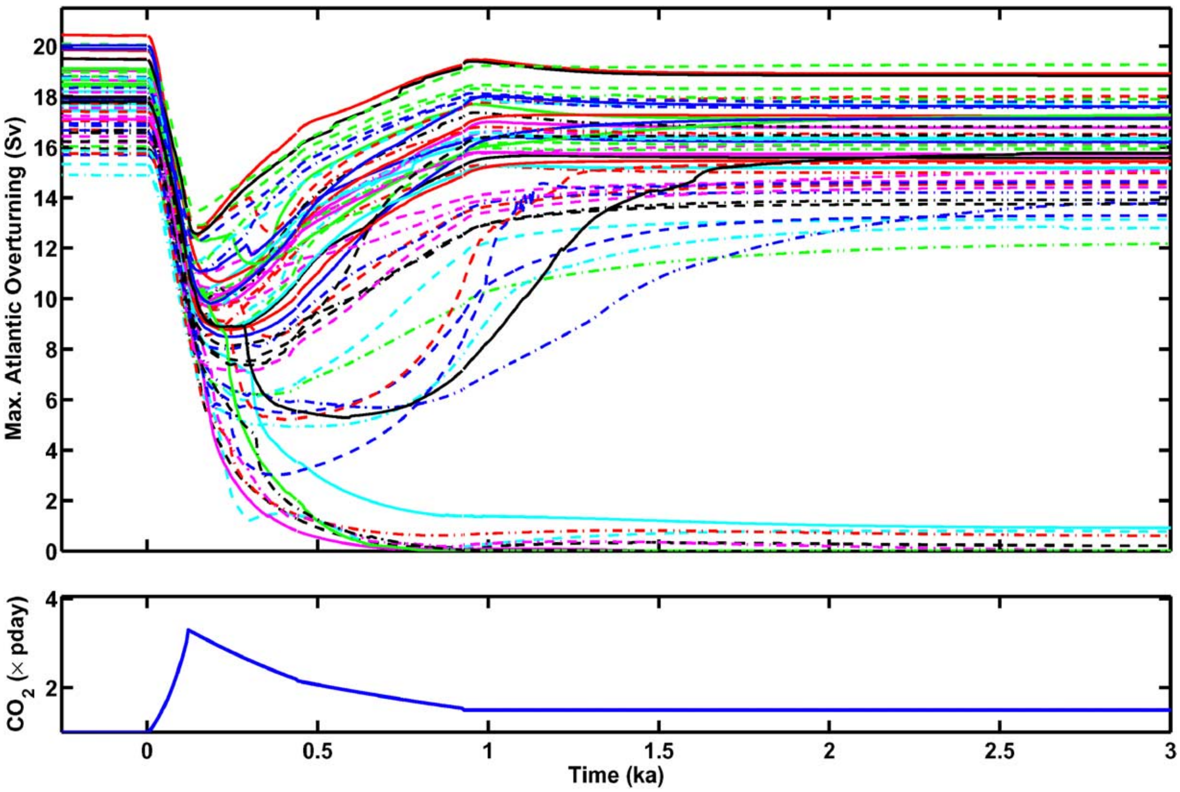

Fig. 6. Maximum Atlantic overturning for ensemble RTP for 3.3× CO_2 followed by CO_2 recovery over 800 years and stabilisation at 1.5× CO_2 thereafter.

largely recovered THC state, with the ensemble members which recover having a slightly lower overturning level than they had at the start of the experiment. Up to about 1000 years into the experiment, however, a wide range of intermediate THC states is apparent in the results.

In RG99, the experiment was slightly different since they started the run from a pre-industrial state with initially a more gradual increase in CO_2 for the first 200 years, and the switch over from rising CO_2 to falling CO_2 is also less abrupt. In their experiment with the CLIMBER model the initial North Atlantic overturning was approximately 19 Sv which fell to 10 Sv over the increasing CO_2 period and then recovered to 16 Sv about 1000 years into the experiment. The RG99 experiment was halted after the equivalent of about 1000 years of our experiment. It is clear from the RG99 results that the CLIMBER model is quite sensitive to the value of k and due to model differences, our implementation of the additional FWF is not completely identical to their description. Despite these caveats the RG99 results appear to be encouragingly consistent with the majority of the members of the RTP ensemble. However, 30% of the RTP ensemble exhibit qualitatively different behaviour to CLIMBER with either total THC collapse or a substantial collapse followed by a rather slow recovery, indicating the range of uncertainty that could surround a single scenario.

We conducted a further two experiments (not included in RG99) in order to assess the sensitivity of THC to rate of recovery of CO_2. We find that for an ensemble where some of the members collapse, that the eventual fate of the THC (recovery or collapse) is highly dependent on the rate of recovery of CO_2. Fig. 7 shows the 3.3× CO_2 experiment but with a relatively fast CO_2 recovery on the left plot and a 3.3× CO_2 stabilisation on the right plot. For the fast recovery, almost all the ensemble members collapse. For the stabilisation experiment many of the ensemble member collapse, while others stabilise at a considerably lower level to their initial level and none return to close to their initial value. This demonstrates the necessity of accurately modelling the carbon cycle as well as estimating realistic emissions scenarios, if prediction beyond the medium term by GCMs is to be of value.

So far we have only discussed the hydrological enhancement taking the additional FWF from the Pacific. The results for the meltwater ensemble (RTM) are almost identical to ensemble RTP.

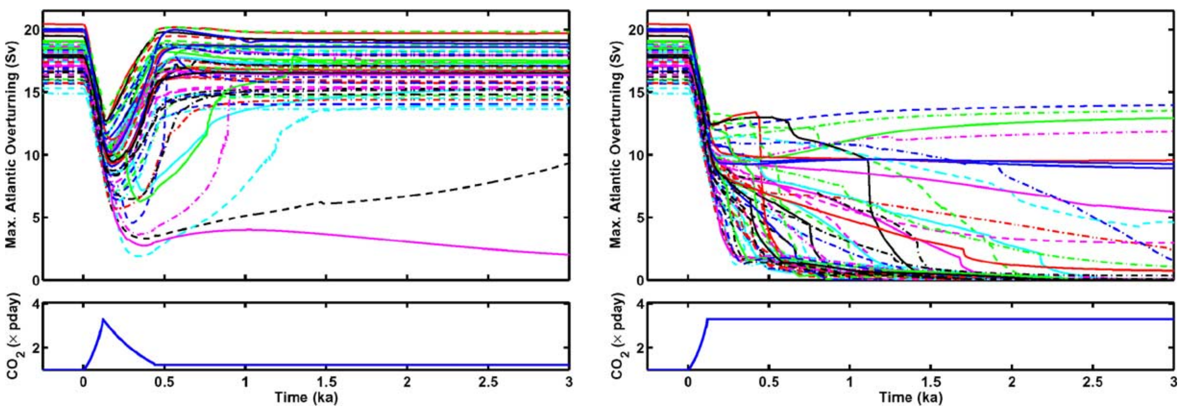

Fig. 7. Sensitivity of the fate of the THC to CO_2 recovery rate. Maximum Atlantic overturning for ensemble RTP for 3.3× CO_2: left plot—with fast CO_2 recovery; right plot—stabilisation at 3.3×CO_2.

The weakening of the THC in response to CO_2 increase is very slightly less. In the 2.7× CO_2 experiment, just one fewer ensemble member has a collapsed THC. The effect of taking the freshwater from the tropical Atlantic (ensemble RTA) is more pronounced and this has the effect of stabilising the THC relative to the RTP run. For the RTA ensemble and the 2.7× CO_2 experiment only 11% of ensemble members suffer a THC collapse compared with 55% of the RTP ensemble. RG99 reported that changing the source of the freshwater changed the critical FWF necessary for collapse by less than 25%. The significance of the source region would appear to be somewhat greater in our model, with a similar increase in FWF (for the tropical Atlantic source) only causing an increase in the probability of collapse to 31%, still some way short of the RTP value of 55%.

7. Conclusions

By performing a wide variety of ensemble integrations, we have been able to examine the response of the THC to anthropogenic forcing and its predictability in the medium (100 year) and longer term. The medium term response of the tuned ensembles is fairly well constrained with a drop in THC strength ranging from 4.2 ± 0.6 to 5.3 ± 0.6 Sv, over a wide range of different hydrological effects and two different initial tuning experiments. This value is by no means predetermined by the model structure, however, with the range of results from the untuned model ranging by a factor of 10 from a very small drop (0.7 Sv) to a rather large one (7.4 Sv). The more sophisticated and complex models used in the IPCC process also had a wide range of responses, but although much effort has been expended on their construction, they have not been subjected to objective optimal parameter tuning procedures. Possibly as a result of this, they also had a wide range of climate states, more comparable to our untuned ensemble sampled directly from the prior parameter ranges, than to the tuned output from the EnKF. This suggests that parameter variability can by itself account for a large part of the differences between the climates of various models even if they are also structurally different. Therefore, it seems reasonable to hope that model tuning could also improve the consistency and accuracy of forecasts using these more sophisticated models.

Over the longer term, the fate of the THC is very uncertain in our ensembles, and depends critically on the magnitude of the response of the hydrological cycle to anthropogenic forcing and the initial strength of the modelled THC. A predominance of our model results suggest a long-term collapse at 2.7× CO_2, but this is far from certain and is based on a rather large change in the hydrological cycle.

The common practice of subtracting the model bias has been shown to have substantial limitations. When the system is in a region of largely linear behaviour (such as the response of the THC over the first 100 years), such an approach may be valid, but in the neighbourhood of a non-linear threshold it leads to very unreliable results. With our first tuned ensemble (which had somewhat low THC, but well within the range of the CMIP models), a long-term shutdown resulted at 2.7× CO_2 even in the absence of hydrological forcing. When the model was retuned to have a more realistic THC, this shutdown was completely eliminated. The existence of two ensembles of similar skill but such widely differing forecasts raises the issue of whether both of them have an overly narrow spread, given the uncertainty in the remaining structural model error. Increasing the ensemble width could increase the robustness of the prediction but only by means of increasing the uncertainty still further. Thus, confident and accurate predictions of the long-term evolution of the THC seem some way off.

Acknowledgement

We are grateful to the GENIE team (http://www.genie.ac.uk), and especially Neil Edwards and Bob Marsh, for the use of the model.

References

Allen, M., Kettleborough, J., Stainforth, D., 2002. Model error in weather and climate forecasting. In: Proceedings of the 2002 ECMWF predictability seminar. ECMWF, Reading, UK, pp. 275–294.

Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. Monthly Weather Review 129, 2884–2902.

Annan, J.D., Hargreaves, J.C., 2005. Efficient parameter estimation for a highly chaotic system. Tellus A 56 (5), 520–526.

Annan, J.D., Hargreaves, J.C., Edwards, N.R., Marsh, R., 2004. Parameter estimation in an intermediate complexity Earth System Model using an ensemble Kalman filter. Ocean Modelling 8 (1–2), 135–154.

Annan, J.D., Lunt, D.J., Hargreaves, J.C., Valdes, P.J., submitted for publication. Parameter estimation in an atmospheric GCM. Nonlinear Proceses in Geophysics.

Cubasch, U., Meehl, A., Boer, G.J., Stouffer, R.J., Dix, M., Noda, A., Senior, C.A., Raper, S., Rap, K.S., 2001. Projections of future climate change. In: Houghton, J.T., Ding, Y., Griggs, D.J., Nouger, M., v. d. Linden, P.J., Dai, X., Maskell, K., Johnson, C.A. (Eds.), Climate Change 2001: The Scientific Basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 525–582.

Dansgaard, W., Johnsen, S.J., Clausen, H.B., Dahl-Jensen, D., Gundestrup, N.S., Hammer, C.U., Hvidberg, C.S., Steffensen, J.P., Sveinbjrnsdottir, A.E., Jouzel, J., Bond, G., 1993. Evidence for general instability of past climate from a 250-kyr ice-core record. Nature 364, 218–220.

Derber, J., 1989. A variational continuous assimilation scheme. Monthly Weather Review 117, 2437–2446.

Dijkstra, H.A., Neelin, J.D., 1995. Ocean–atmosphere interaction and the tropical climatology, Part I: The dangers of flux corrections. Journal of Climate, 1343–1359.

Dong, B.W., Sutton, R.T., 2002. Adjustment of the coupled ocean–atmosphere system to a sudden change in the thermohaline circulation. Geophysical Research Letters 29 (15) (article number 1728).

Edwards, N.R., Marsh, R., 2004. Uncertainties due to transport-parameter sensitivity in an efficient 3-D ocean-climate model. Climate Dynamics.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research 99 (C5), 10143–10162.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics 53, 343–367.

Forest, C.E., Allen, M.R., Stone, P.H., Sokolov, A.P., 2000. Constraining uncertainties in climate models using climate change detection techniques. Geophysical Research Letters 27, 569–572.

Ganachaud, A., Wunsch, C., 2000. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. Nature 408, 453–457.

Ganopolski, A., Rahmstorf, S., 2001. Simulation of rapid glacial climate changes in a coupled climate model. Nature 409, 153–158.

Hargreaves, J.C., Annan, J.D., Edwards, N.R., Marsh, R., 2004. An efficient climate forecasting method using an intermediate complexity Earth System Model and the ensemble Kalman filter. Climate Dynamics 23 (7–8), 745–760.

Heinrich, H., 1988. Origin and consequences of cyclic ice rafting in the northeast Atlantic during the past 130,000 years. Quaternary Research 29, 143–152.

Jia, Y., 2003. Ocean heat transport and its relationship to ocean circulation in CMIP. Climate Dynamics 20, 153–174.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering D 82, 33–45.

Keppenne, C.L., 2000. Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. Monthly Weather Review 128, 1971–1981.

Knutti, R., Stocker, T.F., Joos, F., Plattner, G.-K., 2002. Constraints on radiative forcing and future climate change from observations and climate model ensembles. Nature 416, 719–723.

Lambert, S.J., Boer, G.J., 2001. CMIP1 evaluation and intercomparison of coupled climate models. Climate Dynamics 17, 83–106.

Latif, M., Roeckner, E., Mikolajewicz, U., Voss, R., 2000. Tropical stabilisation of the thermohaline circulation in a greenhouse warming simulation. Journal of Climate 13, 1809–1813.

Levitus, S., 1998. World Ocean Atlas. Available from http://www.cdc.noaa.gov/.

Manabe, S., Stouffer, R.J., 1995. Simulation of abrupt climate-change induced by fresh-water input to the North-Atlantic ocean. Nature 378 (6553), 165–167.

Manabe, S., Stouffer, R.J., 1988. Two stable equilibria of a coupled ocean-atmosphere model. Journal of Climate 1 (9), 841–866.

Marotzke, J., Stone, P.H., 1995. Atmospheric transports, the thermohaline circulation, and flux adjustments in a simple coupled model. Journal of Physical Oceanography 25 (6), 1350–1364.

Marsh, R.J., Yool, A., Lenton, T.M., Gulamali, M.Y., Edwards, N.R., Shepherd, J.G., Krznaric, M., Newhouse, S., Cox, S.J., 2004. Bistability of the thermohaline circulation identified through comprehensive 2-parameter sweeps of an efficient climate model. Climate Dynamics 23 (7–8), 761–777.

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768–772.

Pisarenko, V.F., Sornette, D., 2004. Statistical methods of parameter estimation for deterministically chaotic time series. Physical Review E 69, 036122.

Rahmstorf, S., Ganopolski, A., 1999. Long-term global warming scenarios computed with an efficient coupled climate model. Climatic Change 43, 353–367.

Rohling, E.J., Marsh, R., Wells, N.C., Siddall, M., Edwards, N.R., 2004. Similar meltwater contributions to glacial sea level changes from Antarctic and northern ice sheets. Nature 430, 1016–1021.

Stouffer, R.J., Manabe, S., 2003. Equilibrium response of thermohaline circulation to large changes in atmospheric CO_2 concentration. Climate Dynamics 20, 759–773.

Vellinga, M., Wood, R.A., 2002. Global climatic impacts of a collapse of the Atlantic thermohaline circulation. Climatic Change 54 (3), 251–267.