Early detection of changes in the North Atlantic meridional overturning circulation: Implications for the design of ocean observation systems

Klaus Keller 1,*, Curtis Deutsch 2,
Matthew G. Hall 3, and David F. Bradford 4,5,6,##

1 Department of Geosciences, The Pennsylvania State University.
2 School of Oceanography, University of Washington.
3 Brookings Institution.
4 Department of Economics, Princeton University.
5 Woodrow Wilson School of Public and International Affairs, Princeton University.
6 School of Law, New York University.
## deceased.

running head:
Keller et al., Early detection of MOC trends

Revised and Resubmitted to the Journal of Climate

* Corresponding author:
208 Deike Building
University Park, PA, 16802-2714
Tel.: (814) 865-6718, Fax : (814) 863-7823
email: kkeller@geosc.psu.edu
ABSTRACT

Many climate models predict that unabated greenhouse gas emissions may cause a threshold response of the North Atlantic meridional overturning circulation (MOC). These model predictions are, however, uncertain. Reducing this uncertainty can have an economic value, as it would allow for the design of more efficient risk management strategies. Early detection (i.e., before past actions have committed future societies to a threshold response) would be especially valuable in such situations. Here we show that an MOC observation system based on infrequent (decadal-scale) hydrographic observations may well fail in the task of early MOC change detection. This is because this system observes too infrequently and the observation errors are too large. More frequent observations and reduced observation errors would result in earlier detection. A simple cost-benefit analysis suggests that the economic value of information derived from such an improved ocean observation system can far exceed the costs. One open challenge is to identify an observation system that would enable a confident and early MOC prediction across the range of possible MOC responses.

KEY WORDS: climate thresholds, uncertainty, economic value of information, climate change detection, Bayesian analysis, North Atlantic meridional overturning circulation.
1 Introduction

Anthropogenic greenhouse gas emissions may trigger climatic changes with major implications for human welfare [Smith et al., 2001]. The possibility of abrupt and extreme events has increasingly been a focus of climate research. Abrupt climate change is an example of a potential high impact event that has received considerable attention [Nordhaus, 1994; Lempert et al., 1994; Keller et al., 2000; Alley et al., 2003; Keller et al., 2005]. One example of a potential abrupt change is a collapse of the North Atlantic meridional overturning circulation (MOC).

An MOC collapse would likely have a considerable impact on the continental climates of the North Atlantic basin [Manabe and Stouffer, 1999; Vellinga and Wood, 2002], where the heat transported by the MOC is an important component of regional climates [Ganachaud and Wunsch, 2000]. The resulting climate change may be abrupt and potentially disruptive of livelihoods and economies in these regions [Tol, 1998; Alley et al., 2003]. In addition, climate management strategies designed to reduce the risk of an MOC collapse require higher CO$_2$ abatement levels than strategies that neglect a possible MOC collapse [Stocker and Schmittner, 1997; Keller et al., 2000; Zickfeld and Bruckner, 2003]. It is, however, uncertain whether and how anthropogenic greenhouse gas emissions will affect the MOC [Cubasch and Meehl, 2001; Gregory et al., 2005]. Reducing the uncertainty about the MOC response can have economic value, because it can result in an improved design of climate management strategies.

Estimating the economic value of information for various environmental monitoring and prediction systems is an area of active research. Most of this work has focused on the 10-day time scale of operational weather forecasting [Johnson and Holt, 1997]. The seasonal to annual time scale has received some attention, especially with regard to El Niño-Southern Oscillation (ENSO) prediction [Solow et al., 1998]. Longer time scales of a decade or more have received markedly less at-
tention, in part because prediction uncertainties increase considerably [Knutti and Stocker, 2002; Sutton and Hodson, 2005; Pohlmann et al., 2004]. The economic value of information about long term climate change has been the focus of relatively few studies, even though it can be substantial because of the large spatial and temporal scales involved [Katz and Murphey, 1997; Nordhaus and Popp, 1997]. In considering the value of an integrated, sustained ocean observation system, Adams et al. [2000] conclude that such an observation system might yield an early warning sign of an MOC collapse. They go on “it is unknown whether any actions could be taken in response to such a warning, but if actions were possible, the value of that information would be very high” (p. 37). This is the problem we address in this study.

Here we take a step towards an improved design of an MOC observing system by considering the costs and benefits of monitoring the MOC in a conceptual study. In a scientific analysis, we present a simple Bayesian method for analyzing the future development of our knowledge about the MOC strength. We account for the effects of observation error, observation frequency, and random variability in the MOC intensity. We argue that a continuation of past practices of MOC observation based on infrequent hydrographic observations is unlikely to detect a model-predicted MOC collapse with high confidence within a century. This raises the possibility that we will detect MOC changes only after we have passed a forcing threshold. Earlier detection can be achieved by increasing the observation frequency or reducing the observation error.

We put our scientific analysis into an economic perspective by comparing order of magnitude estimates of the costs and the benefits of an improved ocean observation system. We estimate the additional costs and benefits associated with a decision to monitor the MOC and examine them in a simple decision analysis framework [Katz and Murphey, 1997]. We conclude that a confident early warning
signal of a potential MOC collapse can provide an economic value of information that is orders of magnitude larger than costs of typical ocean observation systems. As a result, investments into an ocean observation system that would deliver such an early warning signal can pass a cost-benefit test. These conclusions are subject to several important caveats (discussed below).

2 The North Atlantic Meridional Overturning Circulation

The present distribution of climate zones is strongly affected by ocean circulation [Siedler et al., 2001]. The meridional overturning circulation (MOC), which transports large amounts of heat from the tropics to the polar regions, consists of a fairly shallow wind-driven component and a deeper, buoyancy-driven component known as the thermohaline circulation (THC). The THC is driven by fluxes of heat and fresh water across the air-sea interface, which alter the density of poleward flowing surface waters. While heat lost to the atmosphere promotes the formation of deep water in the THC, net fresh water inputs tend to inhibit this circulation by reducing surface density. In the North Atlantic, poleward flowing surface waters can become sufficiently dense (due to heat loss to the atmosphere) that they sink, becoming a key branch of global deep water circulation and its associated heat transport.

Ocean modeling studies suggest that the strength of the MOC — measured in million m$^3$ per second (or Sverdrups [Sv]) and typically estimated as passing through the 1500 m depth level — could change dramatically because of anthropogenic greenhouse gas emissions [Manabe et al., 1991; Schiller et al., 1997; Stocker and Schmittner, 1997; Stouffer and Manabe, 1999; Gregory et al., 2005]. In these models, the collapse occurs through two main mechanisms. First, a warmer climate results in warmer, less dense surface waters. Second, warmer climates gen-
eraly intensify the hydrological cycle, including increased freshwater input into the North Atlantic [Schmittner and Stocker, 1999]. The increased input of heat and freshwater into the North Atlantic reduces surface water densities and therefore the deep-water formation rates. The relative importance of heat and fresh water forcing in driving the MOC changes varies from model to model [Gregory et al., 2005]. In addition, not all models show a change in MOC strength in scenarios of global warming [Latif et al., 2000].

Observation-based estimates of the MOC strength are crucial for resolving these differences. These estimates are typically inferred from hydrographic measurements [Broecker, 1991; Macdonald and Wunsch, 1996; Peacock et al., 2000; Ganachaud and Wunsch, 2000; Smethie and Fine, 2001] that are sampled sparsely compared to the relevant spatial and temporal variability [Wunsch, 1992; Gruber et al., 2000]. MOC variability and the uncertainty in observational estimates makes detecting a change difficult [Rahmstorf, 1999; Rahmstorf, 2000]. There is some evidence that the ocean circulation has changed in the North Atlantic over the last three to four decades [Schlosser et al., 1991; Dickson et al., 2002; Bryden et al., 2005]. Whether the reported circulation changes in the North Atlantic are caused by anthropogenic greenhouse gas emissions or are just part of the natural variability is unclear at this time [Rahmstorf, 2000]. Although it is the THC, and not the wind-driven component of the MOC, that has been predicted to change as a result of anthropogenic climate change, observational estimates of the strength of the overturning circulation do not readily distinguish between thermohaline and wind-driven components. We therefore refer to changes in the MOC rather than in the THC.
3 Methods

The ability of decision-makers to account for a potential response of the MOC to anthropogenic climate change depends on the time at which the information becomes available. Here we present a simple method to simulate our ability to detect a trend in the MOC based on observations. We analyze a model scenario in which MOC changes do occur and investigate the following question: how does our ability to learn about a collapse of the MOC depend on observation frequency and observation error? We focus on the example of MOC observations, but the approach could also be applied to other time series.

We account for two key sources of uncertainty. First, a single estimate of the true state of the MOC strength involves considerable observation uncertainty. Second, even in the absence of climate forcing, the strength of the MOC is not constant in time but instead exhibits temporal variability. We adopt a Bayesian approach to represent these uncertainties and to explore the temporal evolution of the belief about the state of the MOC. In the Bayesian approach, a prior belief about the probability of a hypothesis is modified in an objective manner by new information. At any time, one may quantify the belief in a hypothesis \( H_0 \) by the probability \( P_{\text{prior}}(H_0) \). Then, given an estimate of a property \( \psi \), having some bearing on the hypothesis, we can calculate the likelihood, or the conditional probability \( P(\psi|H_0) \), that the value \( \psi \) is consistent with the hypothesis \( H_0 \). The new (posterior) belief in the hypothesis \( P_{\text{post}}(H_0) \) depends on the initial belief and the conditional probability, according to Bayes theorem:

\[
P_{\text{post}}(H_0) = \frac{P(\psi|H_0)P_{\text{prior}}(H_0)}{\sum_i P(\psi|H_i)P_{\text{prior}}(H_i)},
\]

(1)

where the denominator is the sum over a set of exhaustive and mutually exclusive hypotheses \( H_i \).

Here we wish to discriminate between two alternative hypotheses. The first
hypothesis \((H_c)\) is that the MOC maintains a constant long-term mean overturning strength. We approximate this hypothesis with the model

\[
H_c : \psi(t) = \psi_s + \epsilon,
\]

where \(\psi(t)\) is the MOC strength through time and \(\psi_s\) is its mean value in an unforced climate system. For simplicity, we approximate the errors in eqn. 2 as normally and independently distributed with a zero mean and variance of \(\sigma\) (see below, eqn. 7). We will return to the potential effects of this approximation in the section “caveats” (below).

The second hypothesis \((H_s)\) is that the MOC is sensitive and responds to greenhouse gas forcing by declining from its stable climatological strength at some long-term linear rate, \(\alpha\). This approximation is expressed in a simple linear model

\[
H_s : \psi(t) = \psi_s - \alpha t + \epsilon.
\]

This linear model is a reasonable approximation over the century time scale of interest in this analysis (cf. Figure 1, below). We limit this analysis to a simplified description with only two possible MOC models, one in which the MOC decreases at a rate defined by the “fingerprint” of a model in question (discussed below) and one in which it remains stable. This description is, of course, a simplification of the range of possible MOC trajectories, but can be extended easily to more complex settings. In essence, this approach is equivalent to Bayesian Model Averaging [Hoeting et al., 1999] with two models. The current setup allows us to derive some general insights from a simple and transparent framework. Applying Bayes theorem to each hypothesis yields the ratio of the probability of the two hypotheses

\[
\frac{P_{\text{post}}(H_c)}{P_{\text{post}}(H_s)} = \frac{P(\psi|H_c) P_{\text{prior}}(H_c)}{P(\psi|H_s) P_{\text{prior}}(H_s)},
\]

which provides a measure of the relative probability of one hypothesis against the alternative given the observations, the priors, and the structural assumptions.
To complete the formulation of the Bayesian updating procedure, we need to specify the likelihood of measuring a value $\psi$, conditional on a given hypothesis. This likelihood depends on the that is not resolved by the model ($\sigma_s$) as well as the uncertainty associated with each individual observation ($\sigma_{\text{obs}}$) of $\psi$. We assume, for simplicity, that the effective error is reasonably approximated by independently, identically, and normally distributed errors. In this case, the likelihood of measuring a particular value $\psi_i$ given the hypotheses $H_s$ and $H_c$ is given by

$$P(\psi_i|H_s) \propto \exp \left( \frac{(\psi_i - \psi_s)^2}{2\sigma^2} \right), \text{ and}$$

$$P(\psi_i|H_c) \propto \exp \left( \frac{(\psi_i - \psi_c)^2}{2\sigma^2} \right),$$

where $\psi_s$ and $\psi_c = \psi_s - \alpha t$ are the expected values of $\psi$ under the respective hypotheses, and $\sigma$ is the total effective error of a single observation, defined by

$$\sigma^2 = \sigma_{\text{obs}}^2 + \sigma_s^2.$$  

Bayes theorem thus provides a way to modify the belief in a changing MOC after each new measurement. Given the model structure, an initial belief, and a time-series of observed overturning strength, this framework allows us to simulate the evolution of the belief in a changing MOC by iteratively applying eqn. 4 after each new observation. Here we are interested in the future development of the belief in an unchanging MOC strength. Since future observations are not available, we adopt a set of simulated values that represents one possible trajectory of the MOC. Using the simulated response of the MOC to anthropogenic CO$_2$ emissions, we can compute the evolution of the ratio of the probabilities for the alternate hypotheses following each new simulated observation.

Our analysis is limited in scope and based on several approximations. We simply want to assess how quickly one might learn about a changing MOC if one particular model that predicts an MOC collapse turns out to be correct. We will return to a discussion of key caveats after the results and discussion section.
4 Results and Discussion

4.1 Scientific Analysis

We now apply this detection method to the response of the MOC in a coupled atmosphere-ocean climate model. Using a single model is, of course, vulnerable to the effects of structural model uncertainty [Draper, 1995] (cf. discussion in the section caveats, below). The results of this detection study depend on the nature of the assumed MOC trajectory. Detecting an MOC change will occur faster for higher signal-to-noise ratios (i.e., the ratio between the MOC changes due to anthropogenic forcing and the internal variability). We would like to determine whether a given MOC observation system would enable a detection of MOC changes even under circumstances that are relatively favorable for detection. We therefore choose a model with a relatively high signal-to-noise ratio, i.e., a model with a relatively large MOC response and a relatively low internal variability compared to the suite of available models (cf. [Cubasch and Meehl, 2001; Gregory et al., 2005]).

The GFDL coupled ocean-atmosphere model in the setup described in Manabe and Stouffer [1994] is one potentially useful choice in this regard, as it exhibits an MOC collapse that is relatively fast compared to those of other models [Cubasch and Meehl, 2001]. In addition, the internal variability in the GFDL model is relatively small, compared to other models (e.g., [Santer et al., 1995]). The GFDL coupled ocean-atmosphere model as described in [Manabe and Stouffer, 1994] consists of an atmospheric model using the spectral element method with nine vertical levels. This atmospheric model is coupled to an ocean model with a resolution of 4.5 °latitude and 3.7 °longitude, 12 finite difference vertical levels, and isopycnal mixing following Bryan [1987]. The model assumes that ice sheets do not melt. The model is initialized using observed oceanic conditions [Levitus, 1982]. The “control” scenario is without anthropogenic greenhouse gas forcing. In the consid-
ered forced scenario, atmospheric CO$_2$ increases at 1 percent per year and stabilizes at four times the preindustrial level (4xCO$_2$). In the control scenario, the modeled MOC has a deep water formation rate of 18 Sv with a relatively small interannual variability characterized by a standard deviation of about 1 Sv. In the 4xCO$_2$ forced scenario, the MOC reduces to approximately 6 Sv after 150 years and stabilizes at approximately 3 Sv after approximately 250 years (Figure 1). The least-squares estimate of the linear MOC slope for the 4xCO$_2$ forced scenario is -0.08 Sv per year over first 150 years. This slope is used as the “fingerprint” (parameter $\alpha$) for the hypothesis of a sensitive MOC (eqn. 3).

We apply the detection method to this model output with a simulated monitoring system representing the observation period and estimation error of recent ocean measurements. The two previous global scale oceanographic expeditions on which estimates of MOC strength have been based were the Geochemical Ocean Sections Study (GEOSECS) in the 1970s and the World Ocean Circulation Experiment (WOCE) in the 1990s. These expeditions collected very different types of data, and estimates of MOC strength were made with different methods. Nonetheless, we take a 20 year observation period to represent the status quo for established measurement programs capable of estimating the MOC strength. As discussed in Baehr et al. [2005], hydrographic transects provide a snapshot of the MOC that is ‘blurred’ on time-scales of months to years. Hydrographic transects may hence provide limited information on variability on shorter time-scales. Observation systems based on relatively sparse observations face additional challenges due to potential changes in the spatial structure of the MOC [Wood et al., 1999].

Global inverse calculations of the circulation from the resulting hydrographic observations typically yield a standard error in the MOC intensity of roughly 2 to 5 Sv [Broecker, 1991; Peacock et al., 2000; Ganachaud and Wunsch, 2000; Smethie and Fine, 2001]. We adopt an observation error of 3 Sv to characterize
MOC estimates based on hydrographic observations and perform sensitivity studies with respect to this parameter spanning the range of 1 to 5 Sv. We illustrate the updating process with an uninformative prior where the two hypotheses are initially (i.e., before the observations are considered) assumed to be equally likely. As can be seen from eqn. 4, the two priors then cancel each out for the first observation and the resulting likelihood ratio is equivalent to the one derived using a Frequentist approach.

By sampling the model data at 20-year intervals and superimposing a random observation error ($\sigma_{\text{obs}}$) of 5 Sv, we can estimate the development of the posterior belief over time (sometimes referred to as a learning curve) in the hypothesis of a stable MOC (Figure 2A). With each observation (denoted by a circle) a previous prior is updated to account for the new information, reflected in the the posterior belief (denoted by a square). The posterior belief then becomes the prior for the next iteration at the next observation time. Following Santer et al. [1995], we refer to the time at which 95 percent certainty is reached as the detection time. The detection time in this situation is a random variable, because it depends on random observation errors. For this example, the change in the MOC is initially quite small and the posterior does not change much compared to the initial prior of 0.5. For an observation frequency of every 20 years and an observation error of 5 Sv, the detection time exceeds a century (Figure 2A). For an observation system with an increased observation frequency of every 5 years and a decreased observation error of 3 Sv (Figure 2B), detection occurs earlier, within approximately 70 years.

We now ask whether detection could occur early (i.e., before past actions have committed future societies to a threshold response). To address this question, we must first examine the timing of the MOC threshold. Using a simplified ocean-atmosphere model, Stocker and Schmittner [1997] derive critical stabilization levels of equivalent atmospheric CO$_2$ concentration beyond which the MOC shuts down
and does not return for centuries. These critical CO$_2$ concentrations depend on several factors, including climate sensitivity and the growth rate of atmospheric CO$_2$. We refer to the time at which this critical CO$_2$ concentration is crossed as the point of no return. While the point of no return represents a threshold of the physical system, the decision must be made earlier. This is because a strategy to reduce the risk of an MOC collapse has to account for two additional effects. First, the social and technological systems exhibit inertia. For example, a rapid change in CO$_2$ control would make much of the energy infrastructure obsolete before the end of its useful lifetime [Grubler, 1991; Wigley et al., 1996]. Second, small changes in CO$_2$ emissions over an extended time are cheaper than large (and very costly) changes in CO$_2$ emissions over a short time. As a result, the decision regarding preservation of the MOC has to be made before the point of no return. We refer to the time when the decision about the climate management strategy has to be made as the decision point. Based on the economic analysis (discussed below), we adopt a decision point at roughly 75 years.

Thus, an MOC monitoring program continuing current practice might well fail to detect an MOC change before the decision point. Detecting an MOC change before the decision point is, however, possible through various combinations of observation frequency and observation error. The detection time associated with the current system could be reduced through more precise and/or more frequent observation.

### 4.2 Economic Analysis

The implications of a potential MOC collapse on economically efficient climate management have been the subject of numerous analyses (e.g., [Lempert et al., 1994; Nordhaus, 1994; Toth et al., 1998; Keller et al., 2004]). The economic damages associated with an MOC collapse are very uncertain, but arguably non-
negligible [Tol, 1998; Keller et al., 2000]. Avoiding an MOC collapse would imply a considerable increase in greenhouse gas control relative to a strategy neglecting an MOC collapse [Keller et al., 2000; Zickfeld and Bruckner, 2003]. This implies that significant losses could result from either of two errors in choosing a strategy: (i) choosing to invest too little in CO\textsubscript{2} control and thus failing to prevent an impending MOC collapse or (ii) choosing to invest too much in CO\textsubscript{2} abatement in order to prevent an MOC collapse that would not have occurred. Either error is costly, so detecting the true MOC sensitivity with some confidence has potential economic benefits. The benefits derived from improved information about an underlying decision problem are referred to as the economic value of information (VOI) [Bradford and Kelejian, 1977; Manne and Richels, 1991].

We examine the economic value of information about the MOC in five steps: (i) we estimate the costs of CO\textsubscript{2} abatement required to avoid an MOC collapse in a simplistic model of the system, (ii) we derive rough bounds for the economic losses associated with an MOC collapse based on published values, (iii) we approximate the strategy choice as a binary decision problem, (iv) we calculate the economic value of information that could be obtained from an MOC observation system yielding early detection, and (v) we estimate the order of magnitude of the benefit-cost ratio of investments in such an MOC observation system.

First, we estimate the cost $C$ of preserving the MOC using a model of economically optimal climate management. We adopt the DICE model [Nordhaus, 1994] as a simple description of the economic trade-offs between consumption and CO\textsubscript{2} abatement. The DICE model couples simple sub-models of economically optimal growth (cf. Ramsey [1928]), climate change, the related economic damages, and the costs of reducing anthropogenic CO\textsubscript{2} emissions. In the DICE model, abating CO\textsubscript{2} emissions imposes costs but reduces environmental damages. The trade-off between abatement costs and environmental damages is evaluated by
the effects on a weighted sum of the consumption of goods by present and future
generations. In this framework, an economically optimal strategy of CO₂ control
maximizes a weighted sum of the utilities of per capita consumption. The DICE
model has been used in a wide variety of economic studies addressing global cli-
mate change, and the results are broadly consistent with more complex economic
models [Dowlatabadi, 1995]. The DICE model has the advantages of simplicity
and transparency, but has clear limitations. The DICE model lacks, for example, an
explicit representation of technological inertia [Grubler, 1991; Wigley et al., 1996]
or learning-by-doing [Argote and Epple, 1990].

We represent a possible MOC collapse in the DICE model by imposing an ad-
ditional constraint on the equivalent atmospheric CO₂ concentrations (cf. [Keller
et al., 2000]). Specifically, we require the stabilization of equivalent atmospheric
CO₂ concentrations below the critical CO₂ level estimated by Stocker and Schmit-
tner [1997] to preserve the MOC (cf. Figure 4, panel B). Satisfying the MOC
constraint requires a stronger reduction of industrial CO₂ emissions, and hence ad-
ditional costs, than the optimal strategy without the MOC constraint (Figure 4, panel
C). An optimal economic path in the model, with the additional constraint of MOC
preservation, requires additional CO₂ abatement costs with a total net present value
of approximately 0.7 percent of gross world product (GWP). (All GWP numbers
are relative to the arbitrary reference year 1995.)

Higher investments into reducing CO₂ emissions (Figure 4, panel C) result in a
smaller increase in atmospheric equivalent CO₂ concentrations (Figure 4, panel B)
and in a smaller increase in globally averaged surface temperatures (Figure 4, panel
A). The divergence between the cost paths with and without an MOC preservation
constraint is used to estimate the location of the decision point. In the model, the
optimal investments in CO₂ abatement with and without an MOC preservation con-
straint start to diverge considerably around the year 2075 (panel C), well before the
CO$_2$ concentration in the unconstrained case exceeds the CO$_2$ threshold (panel B). The divergence in optimal strategy before the threshold is reached is driven, in part, by a nonlinearity in the costs of CO$_2$ abatement as more and more CO$_2$ emissions have to be cut back. In the model, smaller relative reductions in CO$_2$ emissions are more than proportionally cheaper than larger relative reductions. This is because an economically efficient strategy uses the cheapest option before it moves on to add more expensive options. One effect of the increasing marginal abatement costs is that delaying a large CO$_2$ abatement can increase the net-present value of the costs (imposing the same atmospheric CO$_2$ constraint). Thus an economically optimal strategy to preserve the MOC in this framework requires taking action well before the physical threshold (the point of no return) is reached.

Second, we derive a rough order of magnitude estimate of the economic damages $L$ due to an MOC collapse by reviewing previous studies. This problem is an area of active research, and the results so far are deeply uncertain. Previous studies suggest that MOC collapse might cause considerable economic damages [Tol, 1998; Keller et al., 2000]. According to Tol’s [1998] estimate, a thermohaline circulation collapse may temporarily increase the climate damage by up to 3 percent of gross domestic product in Western Europe. Keller et al. [2000], for example, discuss the economic impacts of decreased oceanic CO$_2$ uptake, decreased fishery yields, and changes in atmospheric surface temperature, and estimate the order of magnitude for a subset of MOC specific damages between 0 and 3 percent of GWP. Here we adopt a uniform probability distribution over this range with a mean of 1.5 percent of GWP per year. (It is important to note that the current estimates of MOC specific economic impacts are incomplete. Adopting higher estimates of the MOC specific economic impacts would result in higher economic values of information and would strengthen our forthcoming conclusions.) We impose this additional damage in the economic model whenever the equivalent CO$_2$ exceeds the critical
CO₂ level. The resulting expected net present value of MOC specific losses, \( L \), is equivalent to approximately 0.9 percent of GWP (Table 1).

Third, we analyze a specific decision-making problem characterized by a binary decision at a single decision point in time. In this case, we can characterize the decision problem by a 2x2 decision matrix that summarizes the possible outcomes and their costs and probabilities [Katz and Murphey, 1997]. For the analyzed problem, the decision matrix consists of two possible future MOC states: (a) an MOC collapse or (b) no MOC collapse. These MOC states are combined with two available decisions: (i) to preserve or (ii) not to preserve the MOC. The task of the decision-maker in this stylized problem is then to choose the strategy that maximizes the net expected benefit among all the possible outcomes in the decision matrix. This optimal strategy minimizing the overall expected costs of climate change and climate control, where the expected net cost for either strategy is the sum of the costs and losses weighted by the probability that the corresponding MOC state occurs. Of course, many other decision-making frameworks and rules are possible (e.g., Kahneman and Tversky [1979], Lempert and Schlesinger [2000], or McInerney and Keller, 2005), that would likely result in different outcomes.

Given this strategy, we can draw a number of conclusions from the estimated costs and losses shown in Table 1. First, for an impossible MOC collapse \((p = 0)\), this stylized decision-maker faced with this simplified problem would not choose the preservation strategy because the additional costs of maintaining atmospheric CO₂ below the threshold value would exceed the zero losses. For a certain MOC collapse (given a no-stabilization strategy) \((p = 1)\), the optimal strategy would be to stabilize CO₂, because the losses \( L \) due to an MOC collapse exceed the costs of the preservation strategy. Thus, the optimal strategy under the risk of an MOC collapse depends strongly on the belief in the likelihood of such a collapse.

A risk-neutral decision-maker maximizing the net benefit in this simplified
problem would choose the strategy that minimizes expected costs (\(\min [C, p \cdot L]\)). Thus, for \(p > \frac{C}{L}\) (where \(\frac{C}{L}\) is the cost-loss ratio), the optimal decision is to invest in climate control to reduce the risk of an MOC collapse. For \(p < \frac{C}{L}\) the optimal decision in this framework would be to choose the strategy of not stabilizing CO\(_2\) and thus taking the risk of an MOC collapse. (Whether such a risk-taking strategy is a reasonable description of the behavior of the real decision-making process is an open question.) The optimal strategy changes from the choice of risking an MOC collapse to the choice of avoiding an MOC collapse as the probability of an MOC collapse rises above the cost-loss ratio, or roughly 80 percent.

Fourth, we identify the effect of obtaining additional information about the MOC on the expected costs of both considered strategies. Ocean observations can be used to revise the probabilities upon which the decision-maker acts. Because information about the true state of the MOC is expected to reduce the likelihood of the costly strategy errors discussed above, ocean monitoring programs can have economic value. The value of information obtained through a monitoring program depends on the quality of that information.

Following Katz and Murphey [1997], we define the quality of information obtained through monitoring as

\[
q = \frac{P - P_o}{1 - P_o},
\]

where \(P\) is the probability of an MOC collapse at the decision point (equal to \(P_{post}(H_s)\)), and \(P_0\) is the initial belief before monitoring begins. This definition of quality measures the change in the belief in an MOC collapse as a result of monitoring. A monitoring system that leaves our prior belief unchanged (\(P = P_0\)) has zero quality, whereas one that achieves certainty has \(q = 1\). An observation system with detection properties shown in Figure 2 B (i.e., changing a prior of 0.5 to a correct belief in a sensitive MOC of 0.95 at the detection time) has a quality of \(q = 0.9\).
The value of information (VOI) can be computed as the decrease in expected net present value of the costs that results from learning about the probability of MOC collapse. It is obtained by subtracting the expected costs incurred by a policy with information from the expected costs incurred by a policy operating from prior belief only. The VOI is related to the quality of information, or equivalently, to the level of belief that can be obtained by the decision point. As derived by Katz and Murphey [1997], the value of information for the case of \( \frac{C}{L} < P < 1 \) is

\[
VOI(q) = P_o \{ (P_o + (1 - P_o)q)L - C \}.
\] (9)

This expression satisfies the intuitive expectation that monitoring schemes that achieve greater certainty (and therefore have higher quality) have more value than those that leave the initial uncertainty unchanged. For the considered costs, losses and prior, an observation system with a quality of 0.9 has an economic value of information on the order of 0.1 percent of GWP, a number in the tens of billions of U.S. dollars.

One might ask whether investment in an MOC observation system would pass a simple cost-benefit test. Addressing this question is subject to numerous caveats and uncertainties (discussed below). Yet, a simple order-of-magnitude estimate may provide some robust insights. The observation costs of one possible observation system that would detect an MOC change before the decision time might be estimated by extrapolating the costs from WOCE. According to the U.S. WOCE Office, a single transatlantic cruise with hydrographic measurements and dense station sampling costs roughly U.S.$ 0.5 million [Piers Chapman, personal communication]. The model inversions discussed above rely, however, on multiple transatlantic cruises. Following WOCE, we assume about 6 cruises, placing the total cost at around U.S.$ 3 million per estimate of MOC strength. The results shown in Figure 3 suggest that an observation system characterized by observations every 2 years and an observation uncertainty of 3 Sv would result in a median detection
time of approximately 60 years. Implementing such an observation system over 60 years implies a net present value of costs in the tens of millions of U.S. $, orders of magnitude less than the economic value of information derived above. Hence, the investments in this expanded observation system would pass this simple cost-benefit test.

5 Caveats and Open Questions

Our conceptual study should be interpreted as a first step towards the economic analysis of a possible MOC observation system and is adorned with numerous caveats. Here we briefly outline a subset of potentially important refinements.

First, this study analyses a single ocean model, similar to previous analyses of ocean observation systems [Kohl and Stammer, 2004; Baehr et al., 2004; Baehr et al., 2005]. While understanding the detection problem for a single model is a logical first step, our analysis is silent on the effects of model uncertainty [Gregory et al., 2005], considering a larger number of ensemble runs [Baehr et al., 2005], or systematic observation errors [Baehr et al., 2005]. It may well be that considering these effects would result in more stringent requirements (i.e., higher observation frequency and lower observation error) for the observation system to result in early detection. The recent study by Baehr et al. [2005] may provide some relevant information regarding the effects of structural uncertainty. This study uses a different Atmosphere Ocean General Circulation Model (AOGCM) (ECHAM5 / MPI-OM) driven by a more realistic radiative forcing than the current study to analyze the detection time using virtual observations at 26 °N. They also use a different detection method based on a Frequentist approach and account for the effects of potential autocorrelation of the MOC time-series. Baehr et al. [2005] derive median detection times for annual observations and an observation error of 3 Sv around 50 years.
This is broadly consistent with our estimate of a median detection time of approximately 60 years for observations every two years and an observation error of 3 Sv. Thus, our conclusion that more frequent MOC observations with lower observation error can result in early MOC change detection seems to be somewhat robust with respect to choosing an alternative statistical method, a different AOGCM, and a more realistic forcing. Of course, a sample of just two AOGCMs does not cover the much larger range of available model runs (cf. [Gregory et al., 2005]). A more robust assessment of the detection capabilities of a given MOC observation system would require to analyze the performance of this system across a sufficiently large sample of structural and parametric uncertainty.

Second, we neglect the effects of potential temporal autocorrelation in unresolved MOC variability (term $\sigma_s$ in eqn. 7). Accounting for autocorrelation increases the uncertainty in the estimated slope [Zellner and Tian, 1964]. The effects of autocorrelated errors in $\sigma_s$ on the likelihood functions (eqns. 5 and 6) are, however, reduced by the considerable observation error (term $\sigma_{\text{obs}}$ in eqn. 7) that is assumed to be independently distributed. As discussed above, the results of Baehr et al. [2005] suggest that accounting for the effects of temporal autocorrelation in the unresolved MOC variability (in addition to other model refinements) does not change the main conclusion of our detection study.

Third, the future detection time is a stochastic variable and we analyze so far only the medium value. This poses the questions how different levels of reliability affect the design of the observation system as well as the economic analysis [Baehr et al., 2005; McInerney and Keller, 2005]. Addressing these questions will require a more refined characterization of the decision-making problem than the one adopted in this study.

Fourth, the economic analysis neglects important uncertainties (e.g., in the expected environmental damages or the climate sensitivity) [Tol, 2005; McInerney
An analysis based on a higher estimated of the climate sensitivity would result in lower critical CO$_2$ concentrations \cite{Stocker1997} and potentially change the timing of the decision point (Figure 4). The analysis furthermore approximates the abatement and monitoring decisions as discreet choices.

Fifth, an MOC change detection is equivalent in our simple analysis to an MOC change prediction. This is because we consider only two MOC responses (one with an MOC response equal to the unforced variability). An accurate MOC prediction with an increased range of MOC responses can impose more stringent requirements on an observation system compared to MOC change detection \cite{Keller2006}. Note, however, that we use a rather simple MOC observation and detection system. Using more refined methods based on spatial or spectral fingerprints \cite{Hasselmann1998, Banks2002, Held2004, Kleinen2003}, continuous observations \cite{Hirschi2003}, and analyzing additional tracers affected by ocean circulation changes \cite{Keller2002, Joos1996, Matear2000} may well lead to an earlier detection and/or prediction of potential MOC changes.

\section{Conclusions}

We analyze when an MOC observation system based on infrequent (decadal scale) hydrographic observations would detect anthropogenic MOC changes. Specifically, we adopt a Bayesian statistical framework to estimate the time required to detect the weakening MOC signal in a specific model simulation. We conclude that infrequent hydrographic observation system may well fail at the task of early detection (i.e., before the system is committed to an MOC threshold response). We estimate the costs and benefits of implementing MOC monitoring schemes that would improve
the odds of early MOC change detection. Subject to several caveats, the benefits of an improved MOC observation system can far exceed the necessary investments. One key outstanding challenge is to identify a design of an MOC observation system that could provide a reliable early prediction of a potential MOC threshold response across the range of parametric and structural uncertainties.

**Acknowledgments** We thank J. Baehr, S. G. H. Philander, R. S. J. Tol, and three anonymous reviewers for helpful feedback. Special thanks go to R. Stouffer for providing the GFDL model results and for valuable discussions. This study was partially supported by the National Science Foundation (SES #0345925). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We gratefully acknowledge additional support from the Carbon Mitigation Initiative at Princeton University and the Environmental Consortium at the Pennsylvania State University.

This paper is dedicated to the memory of David F. Bradford.
References


Table 1: Expense matrix for the simplified decision model applied to an MOC collapse.  $C$ is the net present value of the costs of stabilizing CO$_2$ below the critical value given by Stocker and Schmittner [1997] and a climate sensitivity of 3.6 °C [Tol and de Vos, 1998] minus the ancillary benefits of reduced climate damages related to the reduced warming.  $C$ is estimated as 0.7 percent of gross world product (GWP) in the DICE model of Nordhaus [1994].  All GWP numbers are given for the reference year of 1995 and in units of 1989 U.S. dollars.  For reference, the 1995 GWP in the DICE model is around 24 trillion U.S. dollars (Nordhaus [1994], page 83).  $L$ represents the net present value of additional damage caused by the MOC collapse.  $L$ would be 0.9 percent of GWP if the additional MOC damages cause a stream of 1.5 percent GWP loss per year after the MOC collapsed using the model of Nordhaus [1994].  The decrease in the considerable damages of 1.5 percent of GWP in the future to a present value of only 0.9 percent of GWP arises from the application of the monetary discount rate.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>MOC collapsed</th>
<th>MOC not collapsed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stabilize CO$_2$</td>
<td>Costs = 0.7 percent of GWP</td>
<td>Costs = 0.7 percent of GWP</td>
</tr>
<tr>
<td>Do not stabilize CO$_2$</td>
<td>Losses = 0.9 percent of GWP</td>
<td>Base case</td>
</tr>
</tbody>
</table>
Figure 1: Comparison of the meridional overturning (a measure of the MOC strength) for control and the forced (4xCO$_2$) scenarios of the GFDL Atmosphere Ocean General Circulation Models (AOGCM) [Manabe and Stouffer, 1994]. The plotted values are the maximum of the meridional streamfunction in the North Atlantic. See Gregory et al. [2005] for an overview of other MOC simulation results.
Figure 2: Development of the belief in a stable MOC ($H_c$) when the observations are actually derived from the alternative model ($H_s$). The simulations are for a prior belief in $H_c$ of 0.5. The upper panel (A) mimics an approximate continuation of the currently implemented MOC observation system. The lower panel analyses a possible improvement based on an increased observation frequency and a decreased observation error.
Figure 3: Sensitivity of the detection time to observation error and observation period. Shown are the median values of the detection time based on 100 Monte Carlo samples.
Figure 4: Estimation of the additional costs of avoiding an MOC collapse. The calculations are based on the DICE-94 model of Nordhaus [1994] with the additional constraint to keep the equivalent CO$_2$ concentrations below the critical to preserve the MOC in the model of Stocker and Schmittner [1997]. The effects of the reduced atmospheric CO$_2$ concentrations (panel B) on globally averaged surface temperatures are shown in panel A. The dashed vertical line illustrates the time when the investments in CO$_2$ abatement (panel C) start to diverge considerably. This time is the adopted decision point in this specific example.