Basin bifurcations, oscillatory instability and rate-induced thresholds for AMOC in a global oceanic box model

Hassan Alkhayuon\textsuperscript{1}, Peter Ashwin\textsuperscript{1}, Laura C Jackson\textsuperscript{2}, Courtney Quinn\textsuperscript{1,3} and Richard A Wood\textsuperscript{2}

\textsuperscript{1}Department of Mathematics, University of Exeter, Exeter EX4 4QF, U.K.
\textsuperscript{2}Met Office Hadley Centre, Exeter, U.K.
\textsuperscript{3}CSIRO Oceans and Atmosphere, Hobart, TAS, Australia

The Atlantic Meridional Overturning Circulation (AMOC) transports substantial amounts of heat into the North Atlantic sector, and hence is of very high importance in regional climate projections. The AMOC has been observed to show multi-stability across a range of models of different complexity. The simplest models find a bifurcation associated with the AMOC “on” state losing stability that is a saddle node. Here we study a physically derived global oceanic model of Wood \textit{et al} with five boxes, that is calibrated to runs of the FAMOUS coupled atmosphere-ocean general circulation model. We find the loss of stability of the “on” state is due to a subcritical Hopf for parameters from both pre-industrial and doubled CO\textsubscript{2} atmospheres. This loss of stability via subcritical Hopf bifurcation has important consequences for the behaviour of the basin of attraction close to bifurcation. We consider various time-dependent profiles of freshwater forcing to the system, and find that rate-induced thresholds for tipping can appear, even for perturbations that do not cross the bifurcation. Understanding how such state transitions occur is important in determining allowable safe climate change mitigation pathways to avoid collapse of the AMOC.

© The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.
1. Introduction

Since Stommel’s realisation [1] that competing thermal and haline circulation effects may give rise to multiple stable states in the Atlantic Meridional Overturning Circulation (AMOC), several studies have found evidence of this across a wide range of scales in the model hierarchy: see for example [2–7]. Dansgaard-Oeschger (DO) events give strong palaeoclimatic evidence of repeated and significant changes to the transport properties of AMOC between an “on” state that transports significant amounts of heat to high latitudes and an “off” state that does not. In the palaeoclimate record, the time of residence in the states can be a matter of 1000s of years while the switching can be relatively rapid and may be associated with Heinrich events where there were large inputs of fresh water into the Atlantic from glaciers [8,9]. It is possible that fresh water input at high latitudes (i.e. changes to precipitation and evaporation patterns or ice sheet melting) could give a perturbation that can move the AMOC over a tipping point from “on” to “off” state.

Tipping points are significant changes to the solutions of a system in response to a small perturbation, and there has been a concerted effort to understand their appearance and predictability [10,11]. They are often thought of in terms of bifurcations for some slowly varying nonlinear system, though [12] point out that, in addition to such bifurcation-induced tipping (B-tipping), noise fluctuations (N-tipping) and sufficiently rapid shifts of a parameter can give rise to similar phenomena, where the tipping is associated with critical rates [13] (R-tipping) rather than specific bifurcations.

For the climate, a change in external forcing (e.g. atmospheric greenhouse gases or fresh water input from ice sheets) would be expected to change both the current equilibrium state (“on”) of the AMOC and its basin of attraction. In some cases the forcing could be strong enough that the “on” state is no longer a valid equilibrium. In that case the AMOC would be expected to collapse to the “off” state (B-tipping), although if the forcing were reversed quickly enough it could recover [7] (with some “resilience time” [15]). For weaker forcing, the “on” state may remain a viable equilibrium. If the forcing changes sufficiently slowly, the AMOC will keep up with the changing equilibrium and remains within its basin of attraction, whereas if the change in forcing is too fast, the current AMOC state may lie outside the basin of the new “on” equilibrium. In this latter case the AMOC may be expected to transition to an “off” equilibrium, even if the “on” state still exists and is stable. Such rate dependent tipping of the AMOC has already been observed in a an intermediate-complexity climate model [16]: when atmospheric carbon dioxide was increased to a new, steady concentration value slowly, the AMOC remained on, but when it was increased more quickly to the same steady concentration, the AMOC collapsed.

For the AMOC, one hopes that monitoring fluctuations in its strength [17], or other related ocean variables [18], will enable us to spot early warning signals of an approaching tipping point due to changing parameters in the system. These changing parameters are typically idealised as changes in freshwater input or “hosing” of regions of the North Atlantic. We note that some more complex models (e.g. [19]) can simulate repeated switching between these AMOC states as slow and spontaneous relaxation oscillations without the need for any freshwater input.

For simple ocean box models such as Stommel’s [1], the ocean is divided into a small number of boxes and continuity of salt and heat fluxes is used to derive a number of dynamic equations. For these models it is possible to perform a fairly complete analysis of the dynamics and find steady bifurcation points where changes to model parameters give changes in the qualitative set of solutions, although even for these models there can be surprises, for example in the susceptibility to noise [20]. These have been extended by some authors to include multiple-box models of the oceans [21,25] where the increased dimension of phase space means that bifurcations can involve dynamically more complex states. In particular, [25,26] find the destabilisation of the “on” state may involve Hopf and homoclinic bifurcations. At the other end of the modelling scale, the most realistic models use detailed global atmosphere-ocean general circulation models (AOGCMs) with realistic geometry, and earth system models of intermediate complexity (EMICs) which are similar to AOGCMs but use simplified physical schemes to speed up computation. A few studies
have shown the existence of multiple AMOC equilibria and transitions between the states, usually in response to idealised “hosing” experiments in which additional fresh water input is applied to the North Atlantic. This is either slowly increased up to and beyond a critical value (e.g. [4,7,27]), or applied as a large perturbation for a limited time (e.g. [15,28–30]). A few AOGCM studies have begun to explore the transient AMOC dynamics as thresholds are crossed, including the resilience time when fresh water forcing is temporarily increased to a level where the AMOC “on” state is unsustainable [15,22,23,31,32].

As mathematical models of AMOC across scales may have quite different structures and parametrizations, it can be a challenge to make direct quantitative comparisons between the models at the ends of the scale. Analysis of AOGCM solutions requires a substantial reduction of complexity, for example by considering time- and spatially-averaged flows. In an attempt to make direct quantitative comparison between the most complex and simple models, a simple but physically-based five box model for global ocean circulation that exhibits AMOC switching has been developed by Wood et al. [18]. They show that the dynamics of the AMOC in a hosing experiment with the FAMOUS AOGCM [4,27,33] are well described by the dynamics of the box model, when the boxes are calibrated to the water masses in the unperturbed AOGCM state. This allows direct quantitative comparisons between states and fluxes for the AOGCM runs and the much simpler five box model, so that the box model can be used to provide insights into the AOGCM’s behaviour.

In this study we investigate in detail the dynamics of this box model of [18], using our analysis to gain insight into possible tipping behaviours of the AMOC and hence to interpret the behaviours seen in more complex AOGCM experiments. We use a timescale separation that further reduces the five box model of [18] to a three box model that seems to include the qualitative and quantitative behaviour of interest for the five box model. We find notable departures from the dynamics of the two-box model [1], some of which have been found in other multi-box models [21,25,26]. The differences are:

- The first loss of stability of the on-state is via a subcritical Hopf bifurcation rather than a fold bifurcation. This distance between the subcritical Hopf and fold bifurcations increases on increasing CO$_2$. Although there is a fold nearby, it represents the meeting of two solutions that are both unstable: cf [25,26].
- There can be an oscillatory relaxation to the on-state found both in three box model and the AOGCM runs: cf [21].
- The basin of attraction of the “on” state goes from semi-infinite size to finite size at a homoclinic bifurcation.

These changes in the basin of attraction of the “on” state mean that the response of the system to changes in forcing can vary with time in a non-trivial manner. We find rate-dependent effects where thresholds involve interaction between the attractors of the system and the unstable states. In climate terms, these insights into the dynamics of the model system can potentially lead to observable indicators of when the AMOC is approaching, or has crossed, critical stability thresholds; for example trajectories of the subtropical and subpolar Atlantic salinity undergo a qualitative change in character when the current, strong AMOC becomes unsustainable. The possibility of rate-dependent tipping has important implications for determining safe emissions pathways to reach climate stabilisation targets.

The paper is organized as follows. In Section 1(a) we briefly review box models of the AMOC and their bifurcations. Section 2 introduces the five box model of [18] and presents an empirical three box reduction to fit numerical observations that two of the boxes show very little in the way of dynamic variation. We perform a comparative bifurcation analysis of these models for parameters from [18] that are fit to FAMOUS runs with pre-industrial (1×CO$_2$) and double (2×CO$_2$), and find good agreement in both cases. Section 3 discusses some of the implications for tipping points caused by time dependent changes to the hosing that may lead to collapse of AMOC. We compare these time-dependent perturbations to similar AOGCM studies [15,18].
In particular, Section 3(c) demonstrates that crossing the threshold need not inevitably lead to tipping: there may be temporary resilience as long as and crossing is reversed sufficiently rapidly, and we quantify how fast this reversal must be. Section 3(d) shows that on the other hand, too rapid a change in forcing can destabilise the system, even if the bifurcation is not crossed. In Section 4 we relate the results of our box model analysis to relevant AOGCM experiments. Section 5 is a further discussion of implications of this work on the stability of the AMOC for rate-dependent perturbed CO$_2$ scenarios.

(a) Box models of the AMOC

The two-box model of Stommel [1] can be written in non-dimensional form as

$$\frac{dT}{dt} = \eta_1 - T(1 + |T - S|),$$

$$\frac{dS}{dt} = \eta_2 - S(\eta_3 + |T - S|),$$

and it is an idealisation of the dynamical interaction between non-dimensional equator-to-pole temperature gradient $T$ and a corresponding salinity gradient $S$ [3]. There are three non-negative dimensionless parameters: $\eta_1$ and $\eta_2$ represent the relative strength of thermal and freshwater forcing and $\eta_3$ is the ratio of thermal to freshwater surface restoring times. Note that even though this system is composed of two coupled ODEs where each expression is quadratic, there remain fundamental unsolved problems about the dynamics of quadratic systems on the plane - in particular Hilbert’s 16th problem is unsolved - this conjectures that there can only be finitely many limit cycles for such a system [34].

As recognised already by Stommel [1], the system (1.1) allows two stable equilibrium states, and bifurcations (possibly non-smooth, due to the terms $|T - S|$ in (1.1) that are continuous but not differentiable at $T = S$ [35,36]) that go between these states on varying a parameter (typically $\eta_3$). On varying $\eta_3$ in (1.1) we can find a bifurcation diagram as shown in Figure 1 which shows a region of hysteresis with two stable states. Note that the linear stability of any equilibrium $(T^*, S^*)$ for (1.1) with $T^* \neq S^*$ and $\phi = \text{sgn}(T^* - S^*)$ is determined by the Jacobian

$$J = \begin{bmatrix} -1 + [-2T^* + S^*]\phi & T^*\phi \\ -S^*\phi & -\eta_3 + [2S^* - T^*]\phi \end{bmatrix}.$$ 

(1.2)

Note that

$$\text{Tr}(J) = -1 - \eta_3 + [-T^* + S^*]\phi = -1 - \eta_3 - 3[T^* - S^*].$$

The fact that this trace is negative for all positive $\eta_3$ implies that there can be no Hopf bifurcation of an equilibria for (1.1): the only smooth local bifurcations that occur will have one dimensional centre manifold, in this case saddle-node. There are also non-smooth saddle-node bifurcations of equilibria with $T^* = S^*$ [36]. Figure 1 shows a typical bifurcation diagram on varying $\eta_3$ for this model: note that complex eigenvalues can appear at equilibria even though there is no Hopf bifurcation.

Modifications of this two-box model can be reduced to an Itô SDE for one variable $y(t)$ [3, Equations 10.19, 10.20] of the form

$$dy = -V'(y, F)dt + \sigma dW_t,$$

(1.3)

where

$$V(y, F) = -FY + y^2/2 + \mu^2(y^4/4 - 2y^3/3 + y^2/2)$$

is a potential with two wells corresponding to AMOC “on” and “off” states, $F$ represents an (possibly time-dependent) injection of a freshwater flux in the North Atlantic box corresponding to a “hosing” of this with fresh water, and $W_t$ is a noise process with amplitude $\sigma$ that represents a stochastic component to the freshwater flux [37]. The model (1.3) for $\sigma = 0$ effectively produces the same behaviour as (1.1) and a bifurcation diagram similar to Figure 1, namely a region...
Figure 1: Bifurcation diagram showing $\Psi = T - S$ plotted against $\eta_2$ for $\eta_1 = 3.0$ and $\eta_3 = 0.3$ for the two-box model (1.1); see for example [3, Figure 10.4]. Observe the upper and lower branches correspond to AMOC “on” and “off” states. Note the saddle node bifurcations at the end of each branch, the lower of which is a non-smooth saddle node.

of hysteresis between the “on” and “off” states, $X_{on}$ and $X_{off}$, terminated by saddle-node bifurcations with the unstable separating state $X_{saddle}$. More complex models, such as the multiple box models studied in [21,24–26] show richer dynamics including Hopf bifurcations.

2. Bifurcations of an AMOC box model

By examining the geometry and behaviour of ocean currents and salinity for the experiments of Hawkins et al [4] using the FAMOUS AOGCM [33], Wood et al [18] propose a realistic box model with five boxes that models the balance of salinity in the main ocean water masses. They consider North Atlantic ($N$) and Tropical Atlantic ($T$) boxes, as well as an Indo-Pacific ($IP$) box, and assume that the temperature of the North Atlantic box is dependent on the AMOC strength. These boxes are coupled together via boxes corresponding to the Southern Ocean ($S$) and the bottom waters ($B$). Figure 2 schematically illustrates the salinity fluxes between boxes.

The five box model is as follows. We assume that the AMOC flow $q$ is proportional to the density gradient between the $N$ and $S$ boxes (which in turn is determined by the temperatures and salinities of the boxes):

$$q = \frac{\lambda [\alpha (T_S - T_0) + \beta (S_N - S_S)]}{1 + \lambda \alpha \mu}$$

For $q \geq 0$ conservation of salt for the five boxes gives:

$$\begin{align*}
V_N \frac{dS_N}{dt} &= q(S_T - S_N) + K_N(S_T - S_N) - F_NS_0, \\
V_T \frac{dS_T}{dt} &= q(\gamma S_S + (1 - \gamma)S_{IP} - S_T) + K_S(S_S - S_T) + K_N(S_N - S_T) - F_TS_0, \\
V_S \frac{dS_S}{dt} &= \gamma q(S_B - S_S) + K_{IP}(S_{IP} - S_S) + K_S(S_T - S_S) + \eta(S_B - S_S) - F_SS_0, \\
V_{IP} \frac{dS_{IP}}{dt} &= (1 - \gamma)q(S_B - S_{IP}) + K_{IP}(S_S - S_{IP}) - F_{IP}S_0, \\
V_B \frac{dS_B}{dt} &= q(S_N - S_B) + \eta(S_S - S_B),
\end{align*}$$

(2.1)
Figure 2: The flows between the five boxes representing (from top left) Southern Ocean, Tropical Atlantic, North Atlantic, Indo-Pacific and bottom waters, in the global ocean box model of [18] (cf their Fig. 1). There are freshwater surface fluxes \( F_X \) and exchange of salinity via the indicated arrows. The \( K_X \) indicate wind-driven fluxes while the advective flux due to the AMOC, \( q \), is driven by a balance of salinity and temperature gradients. There is an \( S/B \) mixing parameter \( \eta \). A proportion \( \gamma \) of the flux \( q \) follows a cold water path via the Southern Ocean, while the remainder follows the warm water path via the Indo-Pacific.

While for \( q < 0 \) we have

\[
\begin{align*}
V_N \frac{dS_N}{dt} &= |q|(S_B - S_N) + K_N(S_T - S_N) - F_NS_0, \\
V_T \frac{dS_T}{dt} &= |q|(S_N - S_T) + K_S(S_S - S_T) + K_X(S_N - S_T) - F_TS_0, \\
V_S \frac{dS_S}{dt} &= \gamma|q|(S_T - S_S) + K_{IP}(S_{IP} - S_S) + K_S(S_T - S_S) + \eta(S_B - S_S) - F_S S_0, \\
V_{IP} \frac{dS_{IP}}{dt} &= (1 - \gamma)|q|(S_T - S_{IP}) + K_{IP}(S_S - S_{IP}) - F_{IP} S_0, \\
V_B \frac{dS_B}{dt} &= \gamma|q|S_S + (1 - \gamma)|q|S_{IP} - |q|S_B + \eta(S_S - S_B).
\end{align*}
\]

(2.2)

If we consider the total salt content \( C \),

\[
C = V_NS_N + V_TS_T + V_SS_S + V_{IP}S_{IP} + V_BS_B,
\]

(2.3)

then

\[
dC/dt = -(F_N + F_T + F_S + F_{IP})S_0.
\]

So if the total surface fresh water fluxes satisfy

\[
F_N + F_T + F_S + F_{IP} = 0,
\]

(2.4)

then \( C \) is constant on trajectories of (2.1,2,2). This means that we can eliminate one of the equations.

Note that if we assume \( C \) is constant and solve (2.3) to give \( S_B \) in terms of the other variables this corresponds to the solutions of the original system as long as (2.4) is satisfied. If we impose (2.3) but (2.4) is not satisfied, then we are effectively solving the system with an additional flux \( F_B \) to the \( V_B \frac{dS_B}{dt} \) equation, such that \( F_B = -F_N - F_S - F_T - F_{IP} \).

In this paper we consider values of the model parameters that are calibrated to the FAMOUS_b AOGCM model version [33] with either pre-industrial \( \text{CO}_2 \) concentrations (“1 \times \text{CO}_2”, shown in Appendix A, Table 3 and Table 4), or with doubled \( \text{CO}_2 \) concentrations (“2 \times \text{CO}_2”, shown in Appendix A, Table 5 and Table 6). See [18] for details.

(a) A three box reduction of the five box model

We consider an empirical three box reduction as follows. We note from [18, Figure 2] that the variation of \( S_S \) and \( S_B \) are apparently small and slow compared to \( S_N \) and \( S_T \). Hence we fix \( S_S \)
and $S_B$ and consider only $S_N$, $S_T$ and $S_{IP}$ as dynamic variables. Conservation of salt (2.3) means we can solve for one of the variables (we choose $S_{IP}$) to give

$$
\begin{align*}
V_N \frac{dS_N}{dt} &= q(S_T - S_N) + K_N(S_T - S_N) - F_NS_0, \\
V_T \frac{dS_T}{dt} &= q[\gamma S_T + (1 - \gamma)S_{IP} - S_T] + K_S(S_S - S_T) + K_N(S_N - S_T) - F_T S_0,
\end{align*}
$$

for $q \geq 0$ and

$$
\begin{align*}
V_N \frac{dS_N}{dt} &= |q|(S_B - S_N) + K_N(S_T - S_N) - F_NS_0, \\
V_T \frac{dS_T}{dt} &= |q|(S_N - S_T) + K_S(S_S - S_T) + K_N(S_N - S_T) - F_T S_0,
\end{align*}
$$
for $q < 0$, where we fix $S_S$, $S_B$ and use (2.3) to determine the remaining variable $S_{IP}$. The baseline parameters (we refer to as $1 \times CO_2$ as they are tuned to FAMOUS runs with pre-industrial atmospheric $CO_2$) are given in Appendix A. For the numerical studies presented in the following sections we will use scaled salinities to enable more accurate numerical computation: $\tilde{S}_i = 100(S_i - S_0)$ for $i \in \{N,T,S,IP,B\}$ (see Appendix B for the modified equations). To convert to the commonly used units of parts per thousand, $S_{ppt} = 35.0 + 10\tilde{S}$.

(b) Bifurcations of the five box model

We consider variation of the fresh water fluxes as in [18]: we consider a “hosing” forcing $H$ which may depend on time and that affects all of the fresh water fluxes simultaneously. The pattern of hosing represents a surface fresh water input into the North Atlantic over 20-50°N (into parts of both $N$ and $T$ boxes) and removal elsewhere in order to preserve fresh water content so that (2.4) is satisfied, as in [4]. Table 1 shows the surface fluxes corresponding to hosing $H$, for the $1 \times CO_2$ parameter set.

We present xppaut [38] and COCO [39] continuation and bifurcation analysis of (2.2,2.1) for the $1 \times CO_2$ scenario. Similarly to the two-box model, there is a hysteresis between two branches of stable equilibria $X_{on}$ and $X_{off}$, separated by saddles $X_{saddle}$, on varying $H$ and imposing $C$ constant (i.e. using (2.3) to specify $S_{IP}$), as shown in Figure 3. Figure 4 schematically shows changes to the system on varying $H$. The behaviour is qualitatively the same for either $1 \times CO_2$ or $2 \times CO_2$ parameters. The hysteresis in this model is not as straightforward as appears in the two-box model but is more like that found in [25,26].

In particular, although the lower branch is destroyed at a saddle-node bifurcation, the upper branch loses stability at a subcritical Hopf bifurcation before it is destroyed. In addition there is a homoclinic bifurcation (see Figure 3) that leads to a sudden collapse in the size of the basin of attraction before the Hopf bifurcation occurs.

(c) Bifurcations of the three box model

The bifurcation structures for the five box model are qualitatively the same as those of the three box reduction (2.5,2.6) on varying $H$, and both are qualitatively the same for both $1 \times CO_2$ and $2 \times CO_2$ parameters. Indeed, Table 2 compares the locations of the four bifurcations for the two systems. With the assistance of computer algebra (Maple) it is possible to find all branches for $q \geq 0$ and $q < 0$ respectively to arbitrary precision and the local (i.e. saddle-node and Hopf) bifurcation points for the three box model: these parameter values are included on Table 2. The homoclinic
Figure 3: The bifurcation diagram of the five box model (2.1,2.2) on varying $H$ ($1 \times CO_2$ parameters). The stable equilibria are shown in blue, unstable in green. The green circles indicate saddle node bifurcations, and the brown diamonds indicate Hopf bifurcations. The dotted pink line shows the minimum and maximum of the unstable periodic orbit. A zoom of the upper bifurcation region is shown for $S_N$. Note that 0.1 scaled salinity units as shown here corresponds to a salinity difference of 1 ppt in oceanographic units, with a scaled salinity of 0 corresponding to an absolute value of 35ppt.

bifurcations are approximated by following a large but fixed period orbit near the end of the branch that bifurcates from the Hopf bifurcation.

Comparing the bifurcations of the five and three box models on varying $H$, note there is close agreement qualitatively (and quantitatively for $1 \times CO_2$): see Table 2. The bifurcation scenario is robust: for example on changing the parameter $\gamma$ we find the Hopf curve meets the upper saddle node bifurcations on reducing $\gamma$; see Figure 5. The two curves meet at a Bogdanov-Takens (BT) point at $(H, \gamma) = (0.2268, 0.1559)$ that can be determined (with the aid of Maple) by simultaneous solution of an equilibrium of (2.5) with $q > 0$ whose Jacobian has both determinant and trace zero. Note that the BT point represents a convergence not just of Hopf and Saddle node bifurcations but also we can infer [40] that a homoclinic bifurcation continues to the BT point.

Table 2 and Figure 6 also show the bifurcations for the 3-box model with the $2 \times CO_2$ parameter set. We see that for $2 \times CO_2$ there is a larger separation (in $H$) between the Hopf and upper saddle node bifurcations. For this reason we use the $2 \times CO_2$ parameters from here on, for clarity of exposition. We expect qualitatively similar results for the $1 \times CO_2$ parameters.

(d) Basin bifurcations

Figures 4 and 7 shows changes that take place in the basin of attraction of the upper branch $X_{on}$ on varying $H$ (3-box model, $2 \times CO_2$ parameters). We observe that (similarly to [25,26])
Figure 4: Schematic of the bifurcation diagram (top left) and phase portraits on varying $H$ through various points on the bifurcation diagram of five box and three box models (2.5,2.6) for both $1 \times CO_2$ and $2 \times CO_2$ parameters. The shaded region shows the basin of attraction for the upper branch - this shrinks from being semi-infinite size for $H < H_{hom}$ to being finite size for $H > H_{hom}$. As in Figure 3, blue branches are stable, red are unstable and green are saddles. Note that the upper branch loses stability before the saddle node. The dotted lines represent minimum and maximum values of periodic orbits.

<table>
<thead>
<tr>
<th></th>
<th>Five box numerical</th>
<th>Three box numerical</th>
<th>Three box Maple</th>
<th>Five box numerical</th>
<th>Three box numerical</th>
<th>Three box Maple</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{bin}$</td>
<td>-0.07996</td>
<td>-0.05446</td>
<td>-0.05445</td>
<td>-0.4056</td>
<td>-0.3795</td>
<td>-0.3792</td>
</tr>
<tr>
<td>$H_{hom}$</td>
<td>0.2165</td>
<td>0.2128</td>
<td>NA</td>
<td>0.4510</td>
<td>0.3555</td>
<td>NA</td>
</tr>
<tr>
<td>$H_{Hopf}$</td>
<td>0.2191</td>
<td>0.2134</td>
<td>0.2133</td>
<td>0.4789</td>
<td>0.3895</td>
<td>0.3888</td>
</tr>
<tr>
<td>$H_{usn}$</td>
<td>0.2214</td>
<td>0.2139</td>
<td>0.2138</td>
<td>0.511</td>
<td>0.4236</td>
<td>0.4225</td>
</tr>
</tbody>
</table>

Table 2: Comparison of the locations of bifurcations for $H$ for the five box (2.1,2.2) and the three box (2.5,2.6) model. Note that $H_{bin/usn}$ are the saddle-node (fold) bifurcations on the lower and upper branches respectively, $H_{hom/Hopf}$ are respectively the homoclinic and Hopf bifurcations on the upper branch: see Figure 4. The numerical solutions by path-following use COCO, apart from the homoclinic bifurcations which found using xppaut. The last column gives the analytically found values for the three box model using Maple (not available for $H_{hom}$).
Figure 5: Two parameter plot showing the location of $H_{\text{Hopf}}$ (red) and $H_{\text{usn}}$ (green) for the three box model ($1 \times \text{CO}_2$ parameters). Other details are as in Figure 6 on varying $\gamma$ away from the default value 0.39. The two curves meet at a Bogdanov-Takens point at $(H, \gamma) = (0.2268, 0.1559)$ while the Hopf and upper saddle node bifurcations move apart for larger $\gamma$ and smaller $H$.

Figure 6: The bifurcation diagram of the three box model (2.5,2.6) when parameters are tuned to atmospheric conditions of doubled preindustrial CO$_2$ ($2 \times \text{CO}_2$). Observe an equivalent set of bifurcations but a larger area of instability before the saddle-node as compared to the $1 \times \text{CO}_2$ case shown in Figure 3.

- The approach to the upper branch is oscillatory: for the entire stable part of the upper branch there is a complex pair of slowly attracting eigenvalues for the linearisation. These become real at a spiral to node transition between Hopf and Saddle-node points.
- The basin of the upper branch shrinks from unbounded to a small and finite size at a homoclinic bifurcation just before the Hopf bifurcation.

This has some counter-intuitive effects for perturbations, for example it predicts an oscillatory overshoot of $S_N$ at a perturbation that switches the system from the lower to upper branch. As the basin of the upper branch shrinks to a small size, small perturbations in any direction can take the system out of the basin of the upper branch.
Figure 7: The basins of attraction of the two stable equilibria for different values of $H$, system (2.5,2.6) ($2 \times CO_2$ parameters). The colour map represents the duration (in years) to get as close as $10^{-3}$ to the equilibria. The values of $H$ are (a) 0, (b) 0.32678, (c) 0.34696 and (d) 0.35503. The integration for basins of attraction use 4th order Runge-Kutta with stepsize 10 years. Supplementary Movie MOC_basin_2co2.mp4 shows an animation of this figure.

3. Tipping in response to time-dependent forcing

In this section we illustrate the potential tipping behaviour of the AMOC in response to time-dependent forcing. We use the 3-box model with the $2 \times CO_2$ parameter set and integrate using a 4th order Runge-Kutta scheme with stepsize between 1 and 2 years.

(a) Time-dependent hosing

We consider the influence of time-dependent perturbations of (2.5,2.6) where the freshwater fluxes $F_X$ are varied via $H(t)$ according to a given protocol of variation that allows exploration of various scenarios of moving around the bifurcation diagram. The time-dependent hosing can be considered as an idealised representation of different future climate change mitigation scenarios, or palaeo-climatic periods.

If $H(t)$ is varied slowly enough, standard arguments of adiabatic reduction [41] can be used to predict the behaviour of the nonautonomous perturbed system. However, it is not always clear how slow this needs to be, and in many cases there are scientifically important questions that need exploration of a range of fast or slow perturbations; for example current anthropogenic forcing of the climate system is on a similar or faster timescale than the inherent timescales of AMOC adjustment.
Figure 8: Schematic diagram showing a piecewise linear time-dependent perturbation \( H(t) = H_{\text{pwl}}(t - t_0) \) where \( t_0 = 50 \), \( T_{\text{rise}} = 10 \), \( T_{\text{pert}} = 140 \), \( T_{\text{fall}} = 30 \), \( H_0 = 0 \) and \( H_{\text{pert}} = 0.1 \). Whether the system tips from one state to the other depends not only on \( H_{\text{pert}} \) but also the other parameters. In the special case \( T_{\text{rise}} = T_{\text{fall}} = 0 \) we refer to the perturbation as a “press”.

The particular protocol we consider here is inspired by the question of how quickly changes to hosing need to be reversed to avoid tipping of the upper branch onto the lower branch [42]. This represents an idealisation of a scenario in which climate forcing temporarily overshoots its final value.

To this end, we consider the effect of piecewise linear (PWL) perturbations \( H(t) = H_{\text{pwl}}(t - t_0) \) (visualized in Figure 8, ) on the system, where

\[
H_{\text{pwl}}(t) = \begin{cases} 
H_0 & t < 0, \\
\alpha(t) & t \in [0, T_{\text{rise}}], \\
H_{\text{pert}} & t - T_{\text{rise}} \in [0, T_{\text{pert}}], \\
\beta(t) & t - T_{\text{rise}} - T_{\text{pert}} \in [0, T_{\text{fall}}], \\
H_0 & t \geq T_{\text{rise}} + T_{\text{pert}} + T_{\text{fall}}.
\end{cases}
\] (3.1)

where \( \alpha(t) \) and \( \beta(t) \) are linear functions such that \( H \) is continuous. This depends on constants \( T_{\text{rise}}, T_{\text{pert}}, T_{\text{fall}} \), and levels \( H_0 \) and \( H_{\text{pert}} \). If we define rise and fall rates

\[
\begin{aligned}
r_{\text{rise}} &= \frac{|H_0 - H_{\text{pert}}|}{T_{\text{rise}}}, \\
r_{\text{fall}} &= \frac{|H_0 - H_{\text{pert}}|}{T_{\text{fall}}},
\end{aligned}
\] (3.2)

then note that \( \alpha(t) = r_{\text{rise}} t \) and \( \beta(t) = r_{\text{fall}} (t - T_{\text{rise}} - T_{\text{pert}}) \). Similar PWL forcing experiments to the AMOC are explored in [22,23] where \( T_{\text{fall}} \) is always very large.

As an example of the sort of critical transition that can arise, Figure 9 shows the effect of two similar PWL perturbations that differ only in \( T_{\text{fall}} \). As can be seen in the phase plane (a) and time series (b) these lead to quite different outcomes at the end of the perturbation.

(b) Resilience time: B-tipping and avoided B-tipping for instantaneous forcing changes

A special case of the PWL perturbation (3.1) is when \( T_{\text{rise}} = T_{\text{fall}} = 0 \): an instantaneous change from \( H_0 \) to the perturbed \( H_{\text{pert}} \) and then back. In all cases we start from the \( X_{\text{on}} \) state for \( H = 0 \). This type of perturbation has been studied extensively in ecological settings [43] where it is called
Figure 9: Trajectories of (2.5,2.6) for $2 \times \text{CO}_2$ parameters and piecewise linear time-dependent parameter $H(t)$ defined by (3.1), the parameter values are $T_{\text{pert}} = 200$, $H_0 = 0$, $H_{\text{pert}} = 0.5$, $T_{\text{rise}} = 200$, $T_{\text{fall}} = 100$ for the black trajectory, and $T_{\text{fall}} = 200$, $T_{\text{rise}} = 200$ for the dark red trajectory. (a) shows the phase portrait and (b) the time profile of $S_N$ (for time profile of $S_T$ see Figure 12). In all cases we start at an initial condition on $X_{on}$.

a “press”. We explore the behaviour of this case when varying $H_{\text{pert}}$ and $T_{\text{pert}}$, fixing $H_0 = 0$: an example is depicted in Figure 10. The black line shows the initial departure from the equilibrium state for $H = H_0$. All trajectories follow this path in the $H = H_{\text{pert}}$ phase plane for the time $T_{\text{pert}}$ corresponding to the line colour. When $t > T_{\text{pert}}$ trajectories then make an excursion back to one of the two stable equilibria $X_{on/off}$ for $H = H_0$. If a trajectory returns to $X_{on}$ we say it does not tip, while if it approaches $X_{off}$ then we say it tips.

The critical situations that divide these “tipping” and “non-tipping” responses can be estimated as follows. Comparing Figures 10a and 7 we note that the divergence corresponds to where the black trajectory intersects with the basin boundary for $H = 0$ (also shown on Figure 10). This critical situation corresponds precisely to the situation where a trajectory starting at $X_{on}$ is on the stable manifold of the saddle $X_{saddle}$ when the perturbation turns off. This corresponds to there being a “sink-saddle connection” for the perturbation we consider. We illustrate this critical value (“resilience time”) for different values of $H_{\text{pert}}$ (mostly $> H_{\text{Hopf}}$) in the case $T_{\text{rise}} = T_{\text{fall}} = 0$ and $H_0 = 0$ in Figure 11. Note that there is not an exact relationship between whether tipping occurs and the total amount of fresh water input $H_{\text{pert}} \times T_{\text{pert}}$. In some cases a short, intense pulse can cause tipping while a longer but weaker pulse containing the same total fresh water does not (compare red and black dashed curves in the Figure). Physically this is because for the long, weak pulse the circulation (including the gyre diffusion term $K_N(S_T - S_N)$) has longer to flush the extra fresh water away from the North Atlantic (N box) (see also [15]).

We now consider some illustrative examples of the effect of more general perturbations of the form (3.1) starting at $H_0 = 0$ and the upper branch.

(c) Resilience time: B-tipping and avoided B-tipping for gradual forcing changes

For more general perturbations of the type $H_{\text{pwl}}$ illustrated in Figure 8 we highlight that not only the perturbation and duration spent there, but also the timescales of getting there and returning will influence whether tipping will occur. In particular, even if $H_{\text{pert}} > H_{\text{Hopf}}$ B-tipping may be avoided: there can be a temporary resilience of the system even after crossing a bifurcation point if the return is fast enough, as already seen in Section (b).
Figure 10: Trajectories of (2.5, 2.6) (2 × CO₂ parameters) with \( H \) defined by (3.1) for the “press” perturbation with \( H_{\text{pert}} = 0.5 \) and \( T_{\text{rise}} = T_{\text{fall}} = 0 \). The initial hosing value is fixed at \( H_0 = 0 \), and \( T_{\text{pert}} \) (in years) is indicated by the line colour. An initial condition is taken close to \( X_{\text{on}} \) for \( H = H_0 \). Note that longer durations result in switching to \( X_{\text{off}} \), and there is a critical duration, 234 years, that gives a connection to the saddle state \( X_{\text{saddle}} \) (not shown). The grey dashed line represents the stable manifold of the saddle equilibrium (off the top of the figure) for \( H = 0 \).

Figure 11: Final state of (2.5, 2.6) starting at \( X_{\text{on}} \) for \( H = 0 \) subjected to the perturbation \( H_{\text{pwl}} \) for a range of \( T_{\text{pert}} \) and \( H_{\text{pert}} \), for the “press” perturbation with \( T_{\text{rise}} = T_{\text{fall}} = 0 \) (2 × CO₂ parameters). White indicates eventual return to \( X_{\text{on}} \) (not tipping), while grey indicates tipping to \( X_{\text{off}} \) after the perturbation. The red line indicates the threshold where there is a source-saddle connection induced by the perturbation (“resilience time” for a given perturbation strength). The black dashed line represents a constant perturbation volume, i.e. \( H_{\text{pert}} \times T_{\text{pert}} = 175 \).
Figure 12: Bifurcation-induced tipping for system (2.5,2.6) for two trajectories: one tips (dark red) and the other does not (black) with $2 \times$ CO$_2$. Figures (a) and (b) show the time series of $S_N$ and $S_T$ respectively. The dashed blue lines represent the two stable equilibria for $H = H_0$. Figure (c) shows the $S_N$ trajectory plotted over the bifurcation diagram. Figure (d) shows the piecewise linear forcing (3.1) against time with $H_0 = 0$, $H_{\text{pert}} = 0.5$ and $T_0 = 100$, $T_{\text{pert}} = 200$. The other parameter values are: $T_{\text{rise}} = T_{\text{fall}} = 200$ for the dark red trajectory and $T_{\text{rise}} = 200$, $T_{\text{fall}} = 100$ for the black trajectory. Note that the parameter values of this figure are the same as in Figure 9. Supplementary Movies MOCphas_Btip.mp4 and MOCphas_Btrack.mp4 show animations corresponding to the red and black trajectories respectively.

In the case that $H_0 < H_{\text{Hopf}} < H_{\text{pert}}$, the bifurcation diagram in Figure 6 for $2 \times$ CO$_2$ parameters shows that the lower branch is the unique attractor for $H = H_{\text{pert}}$. This means that regardless of $T_{\text{rise}}$ and $T_{\text{fall}}$ we expect to see tipping onto the lower branch if $T_{\text{pert}}$ is large enough. However, if some combination of $T_{\text{rise}}$, $T_{\text{fall}}$ and $T_{\text{pert}}$ is sufficiently small we may avoid B-tipping. Figure 12 illustrates these two scenarios, the black trajectory shows the avoiding tipping behaviour while the dark red trajectory shows B-tipping.

For fixed $T_{\text{pert}}$ and $H_{\text{pert}} = 0.5$ we characterise in Figure 13 the regions of $(T_{\text{rise}}, T_{\text{fall}})$ where we get tipping onto the lower branch as a result of the perturbation. For large enough $T_{\text{pert}}$ we find only tipping. However for $T_{\text{pert}} < 235$ we find that a rapid enough combination of rise and fall can avoid tipping. Again, whether tipping occurs is not solely determined by the total amount of fresh water added, with shorter, more intense perturbations being more effective in inducing tipping.
Figure 13: Results of applying the piecewise linear perturbation $H_{\text{pert}}(t)$ starting from $X_{\text{on}}$ with $H_{\text{pert}}$ beyond the Hopf bifurcation, with $H_0 = 0$, $H_{\text{pert}} = 0.5$, $T_{\text{pert}} = 100$ (left), $150$ (middle), $200$ (right) and $2 \times \text{CO}_2$ parameters. Otherwise as in Figure 11: white indicates an eventual return to $X_{\text{on}}$ (does not tip), while grey ends at the $X_{\text{off}}$ state. The threshold is indicated in red. Observe that rapid enough rates (white regions of figures) gives temporary resilience, i.e. the system avoids tipping in spite of crossing the threshold.

(d) R-tipping, avoided R-tipping and basin boundaries

The bifurcation diagram Figure 6 might seem to suggest that a perturbation from $H_0$ to $H_{\text{pert}} < H_{\text{Hopf}}$ will not result in tipping to the lower branch for any choice of $T_{\text{rise}}, T_{\text{fall}}$ and $T_{\text{pert}}$. Indeed, Figure 14 shows the effect of a perturbation that increases and then decreases $H$ fast enough for there to be no tipping from the upper branch. On the other hand, for a less rapid decrease (the dark red trajectory in Figure 14) we can find situations that switch to the lower branch. For fixed $T_{\text{pert}}$ and $H_{\text{pert}} = 0.37$ we characterise in Figure 15 the regions of $(T_{\text{rise}}, T_{\text{fall}})$ where we get tipping onto the lower branch as a result of the perturbation. For large enough $T_{\text{pert}}$ we find tipping if $T_{\text{rise}}$ is small enough, independent of $T_{\text{fall}}$.

We can interpret this behaviour as follows: for $T_{\text{pert}}$ large enough, tipping depends on whether we are in the basin of attraction of the lower branch at the end of the rise. Because the basin for $H = 0.125$ is bounded, a fast enough rise can leave the system outside the upper branch basin.

4. Relationship to AMOC tipping in complex climate models

The analysis of Sections 2 and 3 provides insight into the bifurcation structure and tipping behaviour of the five box model and its three box reduction. Given that the box model has been shown to capture the dynamics of AMOC hysteresis and tipping in the FAMOUS AOGCM [18], it is natural to ask how relevant our insights are to the AMOC behaviour in complex climate models. For the present study we simply ask how the behaviour of the box model relates to existing GCM and EMIC studies. Future work will exploit the insights from the box model to target experimental design for further AOGCM studies, and hence make best use of limited computational resources.

A particular feature of the transient box model solutions can be seen in the $(S_N, S_T)$ phase trajectories shown in idealised form in Figure 4: trajectories in the vicinity of the upper branch (strong AMOC) equilibrium oscillate in a clockwise sense in phase space (whether they are approaching or diverging from the upper branch), but trajectories that are eventually destined for the lower branch (weak/reversed AMOC) eventually show anticlockwise curvature (or a sharp left turn at the middle branch), and approach the lower branch in a non-oscillatory way. This suggests a simple and in principle observable diagnostic in a more complex model (or the real world): anticlockwise trajectories in phase space suggest that the system is on the attractor of the lower branch.

We use the above diagnostic to interpret the “hosing hysteresis” experiments described by [4,18]. In these experiments hosing $H$ is slowly increased from 0 to $15\text{ Sv}$ over 2000 years, then decreased again. We know from [18] that the five box model reproduces the dynamics of the FAMOUS AOGCM for these experiments, especially for the “ramp up” phase where $H$ is
Figure 14: Rate-induced tipping for system (2.5,2.6) with $2 \times \text{CO}_2$ parameters. Two trajectories are shown, where one tips (dark red) and the other does not (black). (a) and (b) show the time series of $S_N$ and $S_T$ respectively the dashed blue lines represent the equilibria $X_{\text{on/off}}$. (c) shows $S_N$ plotted over the bifurcation diagram. Finally, (d) shows the piecewise linear forcing against time with $H_0 = 0$ and $H_{\text{pert}} = 0.37$ which is less than the $H_{\text{Hopf}}$ where the upper branch loses its stability. The other parameter values are: $T_{\text{rise}} = 100$, $T_{\text{fall}} = 320$, $T_0 = 100$ and $T_{\text{pert}} = 400$ for the dark red trajectory. $T_{\text{rise}} = 100$, $T_{\text{fall}} = 310$, $T_0 = 100$ and $T_{\text{pert}} = 400$ for the black trajectory. Supplementary Movies MOC_phas_Rtip.mp4 and MOC_phas_Rtrack.mp4 show animations corresponding to the red and black trajectories respectively.

Figure 15: Eventual behaviour of the system as in Figure 11 for $H_{\text{pwl}}$ with $H_{\text{pert}}$ before the homoclinic, $H_0 = 0$, $H_{\text{pert}} = 0.37$, $T_{\text{pert}} = 400$ (left), 600 (middle), 800 (right), and $2 \times \text{CO}_2$ parameters. As before, white: does not tip, grey: tips, red: threshold. Note that too rapid a rise of the perturbation can give rate-induced tipping, even when the perturbation is held for a long time and the return is very slow.
increasing. The usual assumption in such experiments is that the forcing change is slow enough that the system remains close to its equilibrium state throughout. However, [4] noted that the AMOC appeared to remain “on” for long after $H$ had passed the critical value where no “on” equilibrium existed, suggesting significant transient resilience. Looking at Figure 7 we see that for states just outside the basin of the “on” state, the time to reach the “off” state equilibrium is of order hundreds to thousands of years; since this timescale coincides with the timescale of the forcing increase it is not surprising that the AMOC persists long after the forcing threshold $H_{\text{usn}}$ has been passed.

Figure 16 shows the $(S_N, S_T)$ phase plane for the ramp-up phase of the hosing hysteresis experiment in the five box model (red). Points on the trajectory are joined by blue/green lines to the equilibrium point corresponding to their instantaneous hosing value (estimated by re-running the hysteresis experiment with a ten times slower rate of increasing $H$, black, cf. [22]). We see that initially the transient solution remains close to its corresponding upper branch equilibrium. However when $H$ increases beyond $H_{\text{usn}}$ (around 0.22 Sv for this parameter setting, Table 2) and the upper branch ceases to exist, the transient solution remains in a strong AMOC state (large $S_N$, $S_T$), even though its equilibrium is now a weak AMOC state (low $S_N$, $S_T$). This change occurs at a point in the trajectory where $S_T$ starts to increase, while $S_N$ continues to decrease, an indication that a positive salinity advection feedback has started to dominate between the $T$ and $N$ boxes, reducing the amount of salty water carried northwards and so causing further MOC weakening([22], [18]). The trajectory for the corresponding FAMOUS experiment is qualitatively similar (orange curve): again $S_T$ starts to increase before the trajectory makes a sharp left turn towards the lower branch state. The change from clockwise to anticlockwise curvature indicates an approach to the lower branch state, but the critical value of hosing may have already been passed by the time the curvature changes; the earlier change from decreasing to increasing $S_T$ may be a better indicator of when $H_{\text{usn}}$ has been passed, but this result may be scenario-dependent as it does not appear to apply for the “press” scenario (see Figures 10, 17).

We now consider the “press” forcing scenario studied for the box model in Section (b). Jackson and Wood [15] have recently studied resilience times in a CMIP6-generation AOGCM, HadGEM3-GC2 under such a scenario, for a range of values of $H_{\text{pert}}$ and $T_{\text{pert}}$. Figure 18 shows the resilience times for the box model (with $1 \times \text{CO}_2$ parameters) and for HadGEM3-GC2. We do not expect quantitative comparison as HadGEM3-GC2 has different bifurcation points from FAMOUS (generally showing AMOC collapse at lower $H$ values). However, we do note that the forms of the two curves are the same, and the result that short, intense perturbations are more likely to cause tipping than long, weak perturbations is consistent across the models [15]. Figure 17 shows the phase space trajectories for the various HadGEM3-GC2 experiments of [15]. The AOGCM trajectories compare qualitatively with those for the box model in Figure 10: cases for which the AMOC clearly recovers after the forcing is removed show the same clockwise trajectories, while the cases where there is not a clear recovery (shown in red) mostly appear to be close to the dividing trajectory between recovery and collapse (cf. dashed line in Figure 10). In these cases the AOGCM AMOC appears to be in an intermediate state and it was not clear to [15] whether that state would persist, revert to a strong AMOC or collapse (see Figure 1 of [15]). Comparison to Figure 10 with a press scenario suggests that the AOGCM trajectories are not inconsistent with the box model but that the runs would need to be continued for longer to resolve this.

Overall the box model dynamics appears to provide insights into the behaviour and design of AOGCM hosing experiments. However as a caveat we note that recently Haskins et al. [44] have analysed a “press” scenario in the FAMOUS AOGCM, in a case which shows an oscillatory AMOC recovery after the forcing is removed. While this oscillatory recovery looks qualitatively similar to our box model solutions, it appears to be driven by variations in the South Atlantic density, a processes that is not captured in the three box, and possibly not the five box model. Other processes may operate in AOGCMs that produce qualitatively similar looking dynamics to the box model solutions.
Figure 16: Trajectories in \((S_N, S_T)\) phase space for the “ramp-up” phase of the hysteresis experiment \((H\) slowly increasing from 0 to 1 Sv), for the FAMOUS\(_{B_1\times CO_2}\) AOGCM experiment of [18] (orange) and for the corresponding five box model experiment (red). Symbols show the instantaneous state when the hosing \(H\) reaches different values (intervals of 0.05 Sv hosing or 100 years in time). The black curve shows a repeat of the box model hysteresis experiment but with a 10 times slower timescale for increasing \(H\), providing an estimate of the equilibrium state associated with that \(H\) value. Transient states from the red trajectory are joined to the corresponding equilibrium states by lines. (a) Full ramp-up phase, (b) zoom in on early part (top right hand corner of (a)).
Finally we discuss the rate-dependent tipping observed in Section 3(d). The concept of rate-dependent tipping in response to fresh water hosing was explored in a two-dimensional Boussinesq model in [22,23], where it was shown that fast perturbations can destabilize the typical advective (negative) feedback of the AMOC. Additionally, Stocker and Schmittner [16] found rate-dependent AMOC tipping in an intermediate-complexity climate model (EMIC) in response to increasing greenhouse gases. Faster CO₂ increase could result in tipping, while a more gradual increase toward the same steady state value did not. Although the CO₂ forcing used by [16] is not the same as the fresh water forcing we use, we note from Figure 15 that similar behaviour can be seen in the box model, supporting the view that in order to predict the future stability of the AMOC we need to understand not just the sustainability of an equilibrium “on” state under

Figure 17: Trajectories in (Sn, St) phase space for the “press” experiments using the AOGCM HadGEM3-GC2 [15], with various values of (Hpert, Tpert). The hosed phase of the run is shown in blue. Black trajectories show cases where the AMOC clearly recovers towards its initial state after hosing stops, red trajectories show cases where the AMOC remains weak. In all the red cases, further integration of the AOGCM would be needed to determine the final state.
climate forcing, but also the rate of forcing change (i.e. the emissions pathway, not just the final greenhouse gas levels).

5. Discussion

In this paper, we investigate the global ocean box model of [18] and demonstrate, similarly to [25], that the switching from the "on" state is associated with subcritical Hopf rather than saddle-node bifurcation. We interpret time-dependent perturbations in terms of the bifurcation diagrams and show in particular that there can be a mixing of rate- and bifurcation-tipping.

While the forcing scenarios we investigate are highly idealised, they illustrate the possibility of rate-dependent tipping in the climate system, which is of particular interest as the international community considers options to limit dangerous climate change under the Paris climate accord. Our results reveal that tipping can occur in response to a rapid increase in forcing, even if slower increase to the same level of forcing does not result in tipping. Conversely if a forcing threshold is passed such that the AMOC is ultimately unsustainable, tipping may still be avoided if the forcing increase is reversed quickly enough. These results suggest a framework for the design of safe climate stabilisation pathways to avoid AMOC tipping. The characteristic direction and curvature of the phase trajectories in $(S_N, S_T)$ space is potentially observable in the real ocean and so could provide early warning that the AMOC is on the attractor of an "off" state.

Our work needs to be followed up using more realistic scenarios and models, but the theoretical understanding developed here can be used to guide experimentation with more comprehensive (but computationally expensive) models to address the questions "Under what circumstances might the AMOC switch to an off state?", and "What would we need to observe to give the earliest possible warning of such a switch?"

We highlight that because the bifurcation to a finite basin occurs before loss of stability, this means that even perturbations that are not large enough to cause tipping may result in significant and long deviations, even if there is eventually a return. For example, in Figure 7 (c) if the state of the system is perturbed such that it gets close enough to the boundary of the basin of $X_{on}$, it needs more than 2000 years to get back to equilibrium. For this reason it seems plausible that the present day AMOC may still retain significant memory of major perturbations that it underwent hundreds of years ago.
For our investigations we only consider time-dependent perturbation of the “hosing” parameter $H$ where it is assumed that fluxes are balanced and so total ocean volume and salt content are preserved. Over longer timescales or for larger perturbations (e.g. glacial-interglacial transitions) this approximation will not necessarily be valid. In particular (2.3) will no longer be constant and the associated dimension reduction not valid. Potentially this will give rise to new dynamical effects.

Finally we highlight the close relationship between R-tipping and nonautonomous stability: [45,46] show that R-tipping can be understood as a change in properties of a local pullback attractor and identify criteria for the appearance of R-tipping in simple systems. R-tipping can also appear in the presence of noise, for example [47,48] have studied the probability of the occurrence of R-tipping in the presence of noise, as well as some potential early warning indicators of such a phenomenon. Other recent work has examined in detail the crossing “points of no return” for systems with R-tipping [42,49], but in general the science of how reliable early warning signals may be for R-tipping is still to be worked out.

Data Accessibility. The simulation code used to generate figures for the box models, and the movies, are available via the code repository https://github.com/peterashwin/AAJQW_AMOC.

Authors’ Contributions. All authors contributed to the design of the study and the writing of the manuscript. HA, PA and CQ contributed the theoretical and numerical analysis of the box models, LJ and RW contributed the interpretation and analysis of the complex climate models.

Competing Interests. There are no competing interests.

Funding. HA was supported by the Higher Committee For Education Development in Iraq (HCED Iraq) via grant agreement No D13436. CQ was supported by the CRITICS Innovative training Network for funding via the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 643073. LJ and RW were supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. All authors thank the Past Earth Network (EPSRC grant number EP/M008363/1) and ReCoVER (EPSRC grant number EP/M008495/1) for supporting a workshop in 2017 where this work started.

Acknowledgements. Thanks also to Ulrike Feudel, Felix Mendelssohn, Paul Ritchie, Jan Sieber, Sebastian Wieczorek for inspiration and insightful comments.

References

<table>
<thead>
<tr>
<th>Region</th>
<th>Volume</th>
<th>Salinity</th>
<th>Flux</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Atlantic</td>
<td>$V_N = 0.3261\times10^{17}$ m$^3$</td>
<td>$S_N = 0.034912$</td>
<td>$F_N = 0.384$ Sv</td>
</tr>
<tr>
<td>Tropical Atlantic</td>
<td>$V_T = 0.7777\times10^{17}$ m$^3$</td>
<td>$S_T = 0.035435$</td>
<td>$F_T = -0.723$ Sv</td>
</tr>
<tr>
<td>Southern Ocean</td>
<td>$V_S = 0.8897\times10^{17}$ m$^3$</td>
<td>$S_S = 0.034427$</td>
<td>$F_S = 1.078$ Sv</td>
</tr>
<tr>
<td>Indo-Pacific</td>
<td>$V_{IP} = 2.2920\times10^{17}$ m$^3$</td>
<td>$S_{IP} = 0.034668$</td>
<td>$F_{IP} = -0.738$ Sv</td>
</tr>
<tr>
<td>Bottom Ocean</td>
<td>$V_B = 8.6490\times10^{17}$ m$^3$</td>
<td>$S_B = 0.034538$</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: The volumes and standard (baseline, $1 \times$ CO$_2$) values for the AMOC models, based on the FAMOUS$_B$ runs [33] that are used in [18]. Note that for the five box model all Salinities vary with time so these are initial conditions while for the three box model the Salinities $S_T, S_N$ and $S_B$ vary with time; the others are fixed parameters. Note that the fluxes are assumed to balance: $F_N + F_T + F_S + F_{IP} = 0$.


### A. Model constants and parameters

The default constants and parameters for the five box model are listed in Tables 3 and 4.

### B. The scaled five box model

In order to improve computational efficiency (particularly for the continuation software), we rescale both time and state variables of the models. The time unit $t$ of the original model is in seconds; we will consider a time unit of years $\tau = tY^{-1}$ where $Y = 3.15\times10^7$. In addition we will consider the state variables as scaled perturbations from a background state, namely

$$\tilde{S}_i = 100(S_i - S_0), \quad i \in \{N, T, S, IP, B\}. \quad (A1)$$
Table 5: The volumes and standard (baseline) values for the AMOC models with doubled atmospheric CO$_2$ (2 x CO$_2$), based on the FAMOUS$_B$ runs [33] that are used in [18]. Note that for the five box model all Salinities vary with time so these are initial conditions while for the three box model the Salinities $S_T, S_N$ and $S_B$ vary with time; the others are fixed parameters. Note that the fluxes are assumed to balance: $F_N + F_T + F_S + F_{IP} = 0$.

<table>
<thead>
<tr>
<th>Volume</th>
<th>Salinity</th>
<th>Flux</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Atlantic</td>
<td>$V_N = 0.3683 \times 10^{17} \text{ m}^3$</td>
<td>$S_N = 0.034912$</td>
</tr>
<tr>
<td>Tropical Atlantic</td>
<td>$V_T = 0.5418 \times 10^{17} \text{ m}^3$</td>
<td>$S_T = 0.035435$</td>
</tr>
<tr>
<td>Southern Ocean</td>
<td>$V_S = 0.6097 \times 10^{17} \text{ m}^3$</td>
<td>$S_S = 0.034427$</td>
</tr>
<tr>
<td>Indo-Pacific</td>
<td>$V_{IP} = 1.4860 \times 10^{17} \text{ m}^3$</td>
<td>$S_{IP} = 0.034668$</td>
</tr>
<tr>
<td>Bottom Ocean</td>
<td>$V_B = 9.9250 \times 10^{17} \text{ m}^3$</td>
<td>$S_B = 0.034538$</td>
</tr>
</tbody>
</table>

Table 6: Parameters used in the model with doubled atmospheric CO$_2$ (2 x CO$_2$), taken from FAMOUS$_B$ runs [33] in [18]. Note that 1 Sv = $10^6 \text{ m}^3\text{s}^{-1}$.

<table>
<thead>
<tr>
<th>Name</th>
<th>Default value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.12</td>
<td>kg m$^{-3}$ °C$^{-1}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>790.0</td>
<td>kg m$^{-3}$</td>
</tr>
<tr>
<td>$S_0$</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>$T_S$</td>
<td>7.919</td>
<td>°C</td>
</tr>
<tr>
<td>$T_0$</td>
<td>3.870</td>
<td>°C</td>
</tr>
<tr>
<td>$\eta$</td>
<td>33.264</td>
<td>Sv</td>
</tr>
<tr>
<td>$K_N$</td>
<td>5.456</td>
<td>Sv</td>
</tr>
<tr>
<td>$K_S$</td>
<td>5.447</td>
<td>Sv</td>
</tr>
<tr>
<td>$K_{IP}$</td>
<td>96.817</td>
<td>Sv</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>2.79x10$^7$</td>
<td>m$^6$ kg$^{-1}$s$^{-1}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>5.5</td>
<td>°C$^{-1}$m$^{-3}$sx10$^{-8}$</td>
</tr>
</tbody>
</table>

This leads to a modified equation for $q$,

$$q = \lambda[\alpha(T_S - T_N) + \frac{\beta}{100}(\tilde{S}_N - \tilde{S}_S)],$$

and the following ODEs: For $q \geq 0$ we have

$$\begin{align*}
\frac{d\tilde{S}_N}{dt} &= q(\tilde{S}_T - \tilde{S}_N) + K_N(\tilde{S}_T - \tilde{S}_N) - 100F_N S_0, \\
\frac{d\tilde{S}_S}{dt} &= q(\tilde{S}_S + (1 - \gamma)\tilde{S}_{IP} - \tilde{S}_T) + K_S(\tilde{S}_S - \tilde{S}_T) + K_N(\tilde{S}_N - \tilde{S}_T) - 100F_T S_0, \\
\frac{d\tilde{S}_{IP}}{dt} &= \gamma(\tilde{S}_B - \tilde{S}_S) + K_{IP}(\tilde{S}_B - \tilde{S}_S) + K_S(\tilde{S}_T - \tilde{S}_S) + \eta(\tilde{S}_B - \tilde{S}_S) - 100F_S S_0, \\
\frac{d\tilde{S}_B}{dt} &= (1 - \gamma)(\tilde{S}_B - \tilde{S}_{IP}) + K_{IP}(\tilde{S}_S - \tilde{S}_{IP}) - 100F_{IP} S_0, \\
\frac{d\tilde{S}_N}{dt} &= q(\tilde{S}_N - \tilde{S}_B) + \eta(\tilde{S}_S - \tilde{S}_B). \\
\end{align*}$$

(A2)
while for $q < 0$ we have

$$
\begin{align*}
\frac{V_N}{\tau} \frac{d\tilde{S}_N}{dt} &= |q|(\tilde{S}_B - \tilde{S}_N) + K_N(\tilde{S}_T - \tilde{S}_N) - 100F_NS_0, \\
\frac{V_T}{\tau} \frac{d\tilde{S}_T}{dt} &= |q|(\tilde{S}_N - \tilde{S}_T) + K_S(\tilde{S}_S - \tilde{S}_T) + K_N(\tilde{S}_N - \tilde{S}_T) - 100F_TS_0, \\
\frac{V_S}{\tau} \frac{d\tilde{S}_S}{dt} &= \gamma |q|(\tilde{S}_T - \tilde{S}_S) + K_{IP}(\tilde{S}_{IP} - \tilde{S}_S) + K_S(\tilde{S}_T - \tilde{S}_S) + \eta(\tilde{S}_B - \tilde{S}_S) - 100F_SS_0, \\
\frac{V_{IP}}{\tau} \frac{d\tilde{S}_{IP}}{dt} &= (1 - \gamma)|q|(\tilde{S}_T - \tilde{S}_{IP}) + K_{IP}(\tilde{S}_S - \tilde{S}_{IP}) - 100F_{IP}S_0, \\
\frac{V_B}{\tau} \frac{d\tilde{S}_B}{dt} &= \gamma |q|\tilde{S}_S + (1 - \gamma)|q|\tilde{S}_{IP} - |q|\tilde{S}_B + \eta(\tilde{S}_S - \tilde{S}_B).
\end{align*}
$$

Note that the total salt content is still conserved with

$$
C = \frac{1}{100}(V_N\tilde{S}_N + V_T\tilde{S}_T + V_S\tilde{S}_S + V_{IP}\tilde{S}_{IP} + V_B\tilde{S}_B) + S_0(V_N + V_T + V_S + V_{IP} + V_B).
$$

We have only shown the scaled version of the five box model but the three box model is scaled equally.

\section{C. Supplementary Movies}

Supplementary Movie MOC\_basin\_2co2.mp4 shows animations of Figure 7. Supplementary Movies MOC\_phas\_Btip.mp4 and MOC\_phas\_Btrack.mp4 show animations corresponding to the red and black trajectories respectively for Figure 12. Supplementary Movies MOC\_phas\_Rtip.mp4 and MOC\_phas\_Rtrack.mp4 show animations corresponding to the red and black trajectories respectively for Figure 14.