Collapse of the Atlantic Meridional Overturning described by Langevin dynamics

Jelle van den Berk[1], Sybren Drijfhout[1,2], and Wilco Hazeleger[2]

[1]Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
[2]Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands

Correspondence: J. van den Berk (jelle.van.den.berk@knmi.nl)

Abstract. Using a machine learning technique, collapse trajectories of the Atlantic Meridional Overturning Circulation from climate models of intermediate complexity are fitted to a simple model based on the Langevin equation. A total of six parameters are sufficient to quantitatively describe the collapses seen in these simulations under a freshwater forcing. Reversing the freshwater forcing results in asymmetric behaviour that is less well captured and would require a more complicated model.

1 Introduction

The Atlantic Meridional Overturning Circulation (AMOC) is an important circulation in the Atlantic ocean. It is also an important part of the climate system overall due to the heat it transports from the South Atlantic to the North Atlantic (Ganachaud and Wunsch, 2000; Vellinga and Wood, 2002). The AMOC therefore has a substantial influence on the (western) European climate and a weakening of the AMOC might cause changes in the European climate and weather–see Weijer et al. (2019) for a review. The AMOC has also been identified as one of Earth's 'tipping elements' where a rapid change on markedly faster times scales could take place in the (near) future (Lenton et al., 2008). The Atlantic Meridional Overturning Circulation (AMOC) is partly buoyancy driven by the deep water formations in the North Atlantic subpolar gyre which produces the North Atlantic Deep Water (NADW) (e.g. Rahmstorf (2000)). The AMOC might be bistable in nature which means it admits an 'off' state, with little or no transport from north to south, as a counterpart to its current 'on' state (Broecker et al., 1985).

Palaeoclimate records of the last glacial maximum and early holocene show a rapid switching of temperature, which might be associated with the presence/absence of a vigorous AMOC as exists today (Dansgaard et al., 1993). The possibility of a bistable AMOC being the cause of these rapid changes has been noted (Broecker et al., 1990). With the current climate warming rapidly, the stability of the AMOC is of particular interest (Collins et al., 2013) and climate modelling projections indicate the AMOC strength will decrease under an increase of CO_2. Recent measurements show the AMOC has decreased in strength (Smeed et al., 2018). An understanding of the possibly bistable nature of the AMOC is therefore relevant to understand the consequences of climate change.

The Langevin equation has been posited before as appropriate to capture the essential dynamics of an AMOC collapse (Ditlevsen and Johnsen, 2010; Berglund and Gentz, 2002). It has also been used elsewhere as the basis for describing the dynamics of climate sub-systems (Livina et al., 2010) and the AMOC in particular (Kleinen et al., 2003; Held and Kleinen,

25 2004). A fourth order potential function is used in Ditlevsen and Johnsen (2010); Berglund and Gentz (2002) because it is the
minimum required for having three distinct solutions (double wells). This potential function has two parameters which are–
presumed to be–functions of the freshwater forcing. Variation in the freshwater forcing is assumed to directly drive changes in
AMOC strength by changing the potential function in the Langevin equation.

Though the Langevin equation has played a role in the conceptual picture of bistability and tipping points in the climate,

30 but it has not been used to actually fit the parameters to a (simulated) AMOC collapse. Here, we attempt to construct a simple
model based on the Langevin equation and fit its dynamics to salt-advection driven collapse trajectories of the AMOC seen
in climate models (Rahmstorf et al., 2005). The result is a set of parameters that quantitatively describe the AMOC collapse
process.

Section 2 sketches the theoretical background of the Langevin equation and of the salt-advection mechanism. In Section 3 we

35 fit the proposed Langevin model to the AMOC collapse trajectories seen in a set of climate models of intermediate complexity
(EMICs) taken from Rahmstorf et al. (2005). We end with a discussion and conclusions in Section 4.

2   The Langevin model

An increase in surface air temperatures, or an increase in P-E (precipitation - evaporation) of freshwater surface flux, will
decrease the buoyancy in the shallow layer of the deep water formation regions in the North Atlantic subpolar gyre. The deep

40 water formation is reduced, and the southward meridional flow reduced. In principle, this mechanism can reduce the AMOC
to zero gradually if fully buoyancy-driven. A salt-advection feedback mechanism that leads to a bimodal AMOC was proposed
by Stommel (1961). In this mechanism, the deep southward flow couples to the surface return flow via upwelling or other
mechanisms in the Southern Ocean, affecting the salinity in the North Atlantic subpolar gyre. A reduction in salinity decreases
buoyancy, and this positive feedback accelerates the process of AMOC weakening and a collapse results on relatively short

45 timescales.

Fig. 1 shows a conceptual picture of the two stable AMOC (index) states. The AMOC is a zero-dimensional variable arrived
at by integrating the overturning transport and selecting its maximum value (typically located in the subtropical North Atlantic).
In red, the upper branch is drawn up to the collapse point where a bifurcation occurs. The real AMOC in the current climate
moves along this branch from the left, to the right, towards its (assumed) collapse point. The branch in blue is the counterpart

50 of the upper branch and represents the off state of the AMOC and ends in another bifurcation point to the left where the AMOC
jumps back to full strength. The dashed line (repellor) separates the two basins of attraction associated with the two stable
branches (attractors). At the bifurcation point one of the two basins of attraction ceases to exist and a qualitative change takes
place in the potential function (the number of solutions for a given value of the freshwater forcing μ goes from 2 to 3).

Below we will derive a model based on the Langevin equation that captures the essential dynamics of a bimodal AMOC
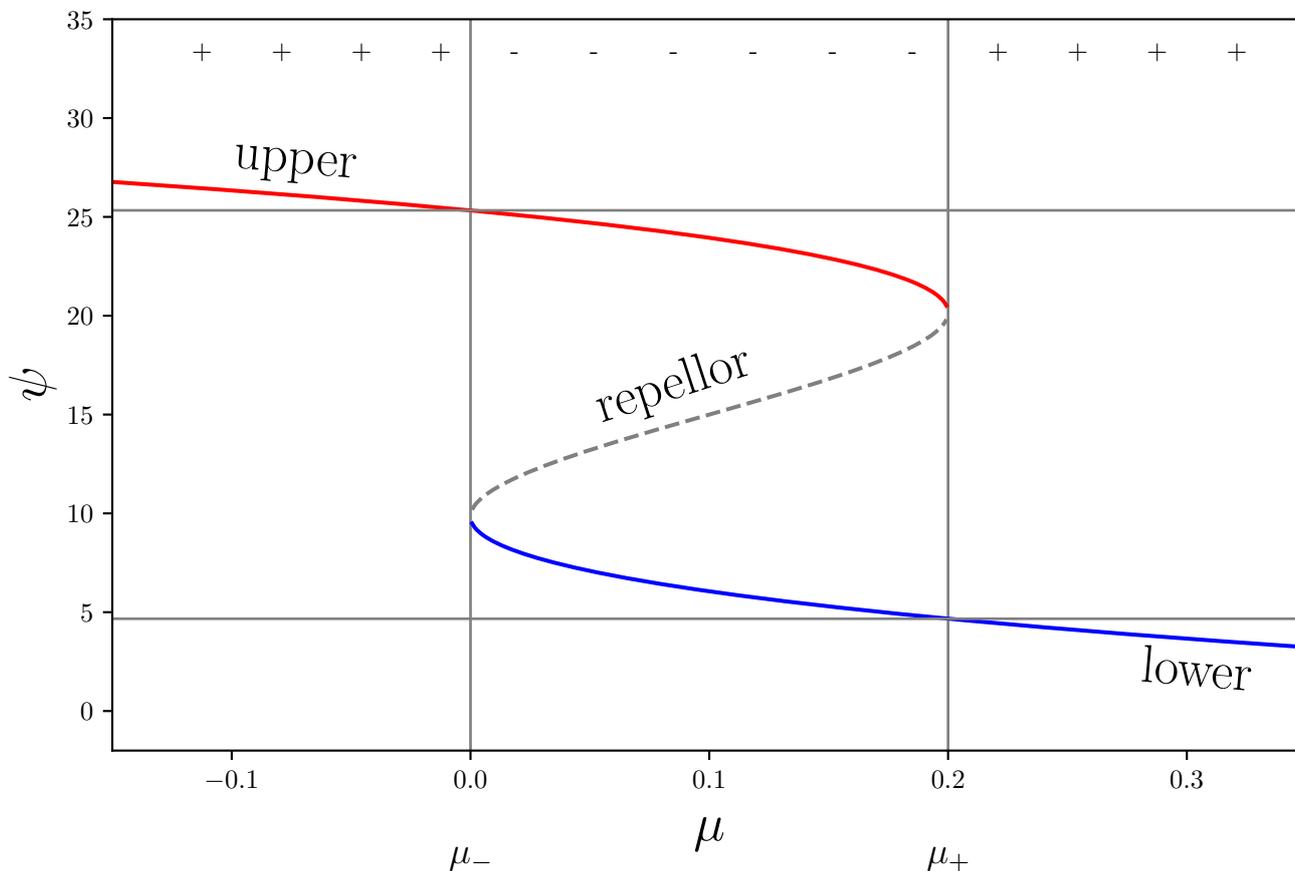
55 under a freshwater forcing μ.

Figure 1. Example trajectory of the AMOC (Ψ) in response to a control variable μ. The red branch is the on-state (upper), blue the off-state (lower). The trajectory deforms when closer to the bifurcation points which are connected though the repellor that forms the trench of the distribution (dashed line). The two bifurcations points are indicated as μ_+ (collapse point) and μ_- (re-invigouration point). Top $\pm$ symbols indicate unimodal ($+$) or bimodal ($-$) regime.

2.1 Multiple stable AMOC states

The conceptual picture of the AMOC being a zero-dimensional variable that is driven by stochastic forces trapped in a potential is similar to that of a particle's motion described by Langevin dynamics (Lemons et al., 1908). The Langevin equation (Gardiner, 2004; Ditlevsen and Johnsen, 2010),

60
$$\dot{x} = -\partial_x U_\mu(x) + \sigma \eta \tag{1}$$

describes the position of a noise-driven particle (x) trapped in a potential function U. The stochastic term is a white noise process (η) scaled with an intensity parameter σ. At first we will ignore the stochastic nature of the AMOC collapse process and focus on the deterministic behaviour.

The (deterministic) bistability seen in Fig. 1 has been studied mainly in a qualitative way (within catastrophe theory). In
65 Poston and Stewart (1978) an extensive treatment is given why only two parameters are sufficient to describe the bistability. These two parameters describe the critical behaviour, not just locally near the critical points, but the entire trajectory under a suitable transformation. (The behaviour at small scale is fundamentally tied to the global behaviour.) A direct consequence is that only partial information, in the form of a piece of the trajectory, should suffice to describe the entire trajectory (the full hysteresis loop).

70 The potential function takes the form (Gardiner, 2004; Ditlevsen and Johnsen, 2010)

$$-U(x) = -\frac{1}{4}x^4 + \frac{\beta}{2}x^2 + \alpha x. \tag{2}$$

The two parameter are functions of the freshwater forcing μ. The AMOC state variable Ψ requires an affine transformation (Cobb, 1980),

$$\alpha = \alpha(\mu)$$
75
$$\beta = \beta(\mu)$$
$$x = (\Psi - \lambda)/\nu.$$

To fit the model trajectories we need to find expressions for α and β, and suitable values for the transformation parameters λ and ν. The value of ν is roughly the distance in Ψ between the bifurcation point on the top branch to the bifurcation point on the lower branch. The value of λ is roughly the Ψ value between the bifurcation points at $\mu_\pm$. The transformation uses λ to
80 shift the trajectory and ν to scale it. Below we describe the potential visually and state additional constraints that follow from the demand that the freshwater forcing is the only variable that determines the dynamical behaviour.

2.2 Potential description

In Fig. 2 an overview of the qualitatively different forms of potential are shown ($-U(x)$, right panels) together with their derivative functions ($-\partial_x U$, left panels). Dots indicate the location of critical points and are related to the number of wells
85 in the potential. The top panels show the typical bimodal form (I) with two stable states and one unstable one in the middle.

Below these are the three possible unimodal states (E). These occur for forcing values to the left of μ_- and to the right of μ_+. The panels B_1 and B_2 are the submanifolds that separates the unimodal regime from the bimodal regime. These two meet in the cusp point P, as shown in the bottom panels. See Poston and Stewart (1978) for further details.

In Fig. 3 the stability landscape is shown where the areas indicated are those with qualitatively different behaviour seen in Fig. 2. See also Poston and Stewart (1978) for similar diagrammes. The cusp point P is the singular point where no proper solution can exist because only the trivial solution is allowed here (both collapse points $\mu_\pm$ and AMOC strength are at zero). The two parameters are α and β and are the two coefficients in the potential function. Their values change because of their dependency on the forcing value (μ).

Our aim is to arrive at a description that matches a track of μ values across the stability landscape. The two parameters α, β are independent but can be parametrised by other variables that map them to observations. In the literature α is referred to as the normal factor, and β the splitting factor (Poston and Stewart, 1978). If parametrised by a single variable the track across the stability surface is one-dimensional, as suggested by the AMOC index. On one side of the cusp point, along the splitting axis (β), only a unimodal regime exists, while on the other side two regimes exist with the modes at relative distances apart.

2.3 Constraints

With a varying α there exist two critical points ($\alpha_\pm$) in between which the distribution is bimodal and unimodal outside that interval. Because the AMOC trajectory is 1-dimensional, there must be a relation between α and β that reduces dimensionality from two to one dimensions. When passing through the critical point α_+, the number of potential wells goes from two to one. Similarly, moving through α_- changes the number of wells from one to two (for given $\mu_\pm$). The two critical points of $\partial_x U$, $\mu_\pm$, can be found analytically for $\mu_\pm$ real and being degenerate solutions. It can be shown (Birkhoff and Mac Lane, 1970, p. 106) that the discriminant $D = 27\alpha^2 - 4\beta^3 = 0$ (real solutions) needs to be solved for α to obtain the two critical solutions that relate α and β. It is at these solutions that the number of critical points changes at forcing values $\mu_\pm$. (When $D < 0$ there are three distinct real solutions which corresponds to the bimodal regime, when $D > 0$ there is only one distinct real solution, which corresponds to the unimodal regime.) When any two of the roots are the same, the number of extrema goes from 3 to 2 (or 1 if all are the same) and the solutions become degenerate (this occurs at $B_{1,2}$ in Fig. 3).

Solving for α gives two solutions that are the critical values as functions of β,

$$\alpha_\pm = \pm \frac{2\sqrt{3}}{9} (\beta)^{3/2} \quad \text{or} \quad \alpha_\pm = \mp \frac{2\sqrt{3}}{9} (\beta)^{3/2},$$

with $\beta \geq 0$ for real solutions. The points $\alpha_\pm$ correspond to where the lines $B_{1,2}$ in Fig. 3 are passed when moving across the stability surface.

For $\alpha_+ < 0 -U(1) < 0$. This corresponds with the AMOC undergoing a collapse at μ_+ from an on state to an off state, and the correct choice of sign is

$$\alpha_\pm = \mp \frac{2\sqrt{3}}{9} (\beta_\pm)^{3/2}, \tag{3}$$
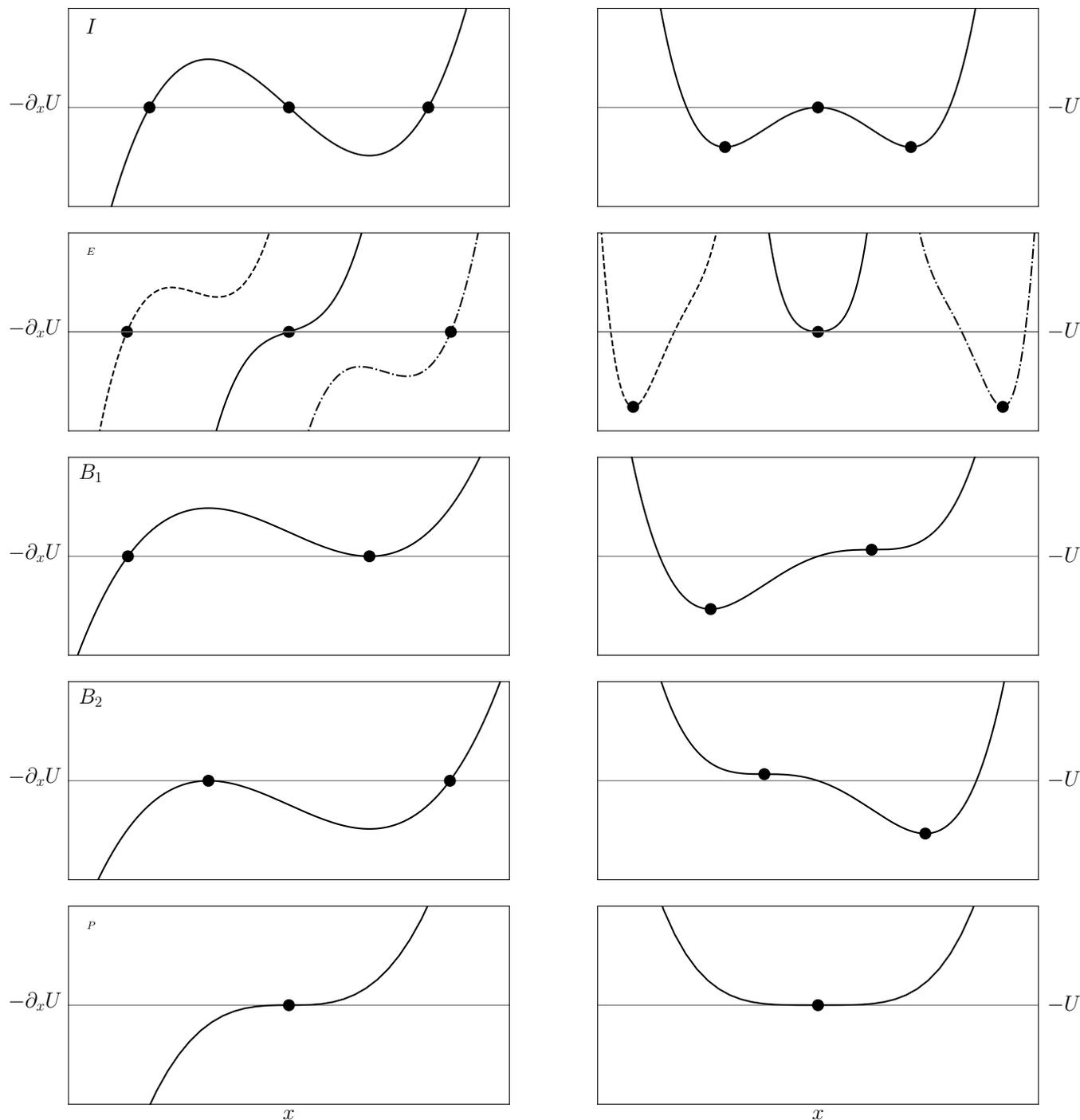
Figure 2. Sample potentials (right) and their derivatives (left) for (top to bottom) the three possible varieties of bimodal state (I), three types of unimodal state (E), the two pathological cases where $D = 0$ (B_1 and B_2), and the cusp catastrophe point (P). Dots indicate the critical points. (Scaling is not uniform between panels. Note the choice of negative sign of the potential U.)
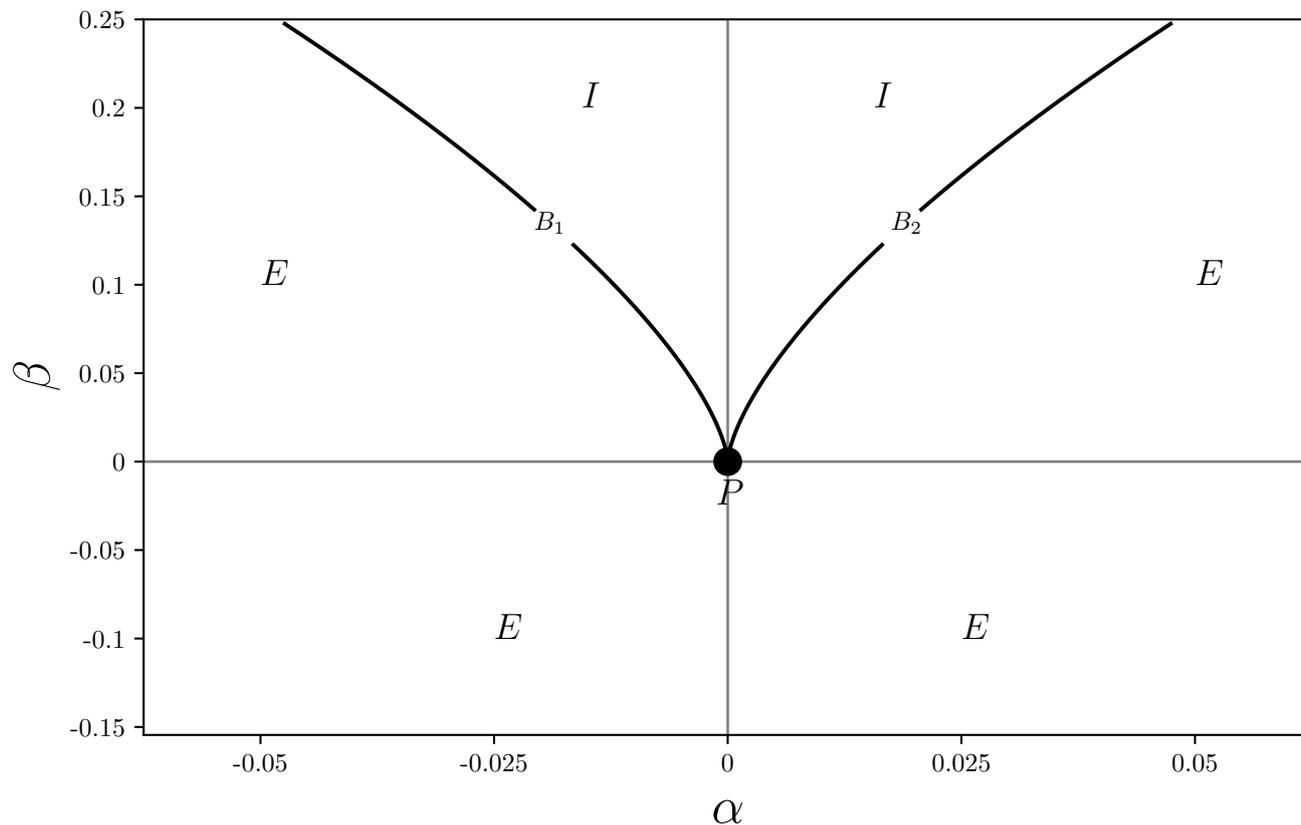
Figure 3. Discriminant determining the stability and number of critical points. The splitting factor β and normal factor α describe the stability landscape. The bimodal regime (I) is separated from the unimodal regime (E) by two lines ($B_{1,2}$) which meet in the point P.

with $\alpha_\pm$ and $\beta_\pm$ the values corresponding to $\mu_\pm$. The track across the stability landscape is from curve B_2 to curve B_1 and Eq. 3 relates the two stability parameters α and β at the two critical forcing values $\mu_\pm$.

2.3.1 Linear functions α, β

120 The value of β does not need to be fixed (to $\alpha_\pm$ there is a corresponding $\beta_\pm$ at the respective critical points) and a varying β corresponds to a slanted track across the stability landscape in Fig. 3. We assume linear functions for α and β,

$$\alpha(\mu) = \alpha_0 + \mu\,\delta\alpha$$
$$\beta(\mu) = \beta_0 + \mu\,\delta\beta,$$

reducing the dependency to these four parameters. Linear functions are the most simple non-trivial dependencies, while adding

125 non-linear parameters introduces further unknowns, making this the most parsimonious parametrisation that captures the first order behaviour. Also, intuitively we can understand the pair (α, β) as the angle under which the system moves to the bifurcation

point ($B_{1,2}$) in Fig. 3), which locally only requires the values of α and β up to first order. Poston and Stewart (1978, p. 59) also remark that the system's local behaviour is essentially the same between critical points, which means a linear expansion should suffice for fitting the upper branch. From this parametrisation we can determine the offset α_0 and rate $\delta\alpha$ in terms of β_0 and $\delta\beta$,

$$\alpha_+ = \alpha_0 + \mu_+ \delta\alpha = -\frac{2\sqrt{3}}{9}(\beta_+)^{3/2} \quad \text{and}$$

$$\alpha_- = \alpha_0 + \mu_- \delta\alpha = +\frac{2\sqrt{3}}{9}(\beta_-)^{3/2}$$

gives

$$\delta\alpha = -\frac{2\sqrt{3}}{9}\frac{(\beta_+)^{3/2} + (\beta_-)^{3/2}}{\mu_+ - \mu_-} \tag{4}$$

$$\alpha_0 = \alpha(\mu = 0) = \frac{\sqrt{3}}{9}\left[-(\beta_+)^{3/2} + (\beta_-)^{3/2}\right] - \frac{1}{2}\delta\alpha(\mu_+ + \mu_-). \tag{5}$$

This constrains the values of α, leaving only β as a free variable, which is then parametrised by β_0 and $\delta\beta$. Note that only solutions with $\beta_\pm > 0$ are valid. Also, values for β_0 and $\delta\beta_0$ that result in a track that crosses B_2 in another point besides β_- are unsuitable. (The curves $B_{1,2}$ are each intersected by a straight line in at most two points, and we require intersection at a single point only.)

2.4 Stochastic interpretation

With the deterministic framework in place, the stochastic nature can be reintroduced. We obtain a distribution needed to fit the parameters of the potential function. As shown by Cobb (1978), this distribution belongs to the exponential family.

For a polynomial function as the potential, the distribution obtained is from the exponential family, which is a generalisation of the exponential distribution (Balakrishnan and Nevzorov, 2004) where any function can determine the exponent value. The potential, we had already supposed to be described by a fourth-order polynomial in the previous section, gives the probability distribution

$$P(x,\alpha,\beta) = Ce^{-2/\sigma^2 U(x)} = Ce^{2/\sigma^2(-1/4x^4 + \beta/2x^2 + \alpha x)}. \tag{6}$$

Note that $C = C(\alpha,\beta)$ and does not have a (known) analytical expression for the general case. The value of σ is a measure of intrinsic variation in the AMOC. See Gardiner (2004) for a derivation of this distribution using the Fokker-Planck equation, from which also the Langevin equation can be derived. Note that $\sigma \rightarrow \sigma/\nu$ because to the scaling with ν we introduced in Section 2.1.

A sample collapse trajectory with corresponding distribution is shown in Fig. 4. The grey lines indicate the (fixed) positions of the bifurcation points. The dashed grey line marks the positions of the unstable solution (repellor) in between the two attractor branches which separates the two basins of attraction. Note that the bifurcation points are extremal in the sense that no bimodality can exists beyond them. With the trajectories being noisy and driven along the attractor, there is (always) some probability of a 'noise-induced' transition. The state shifts from one basin of attraction to the other, crossing the repellor, and
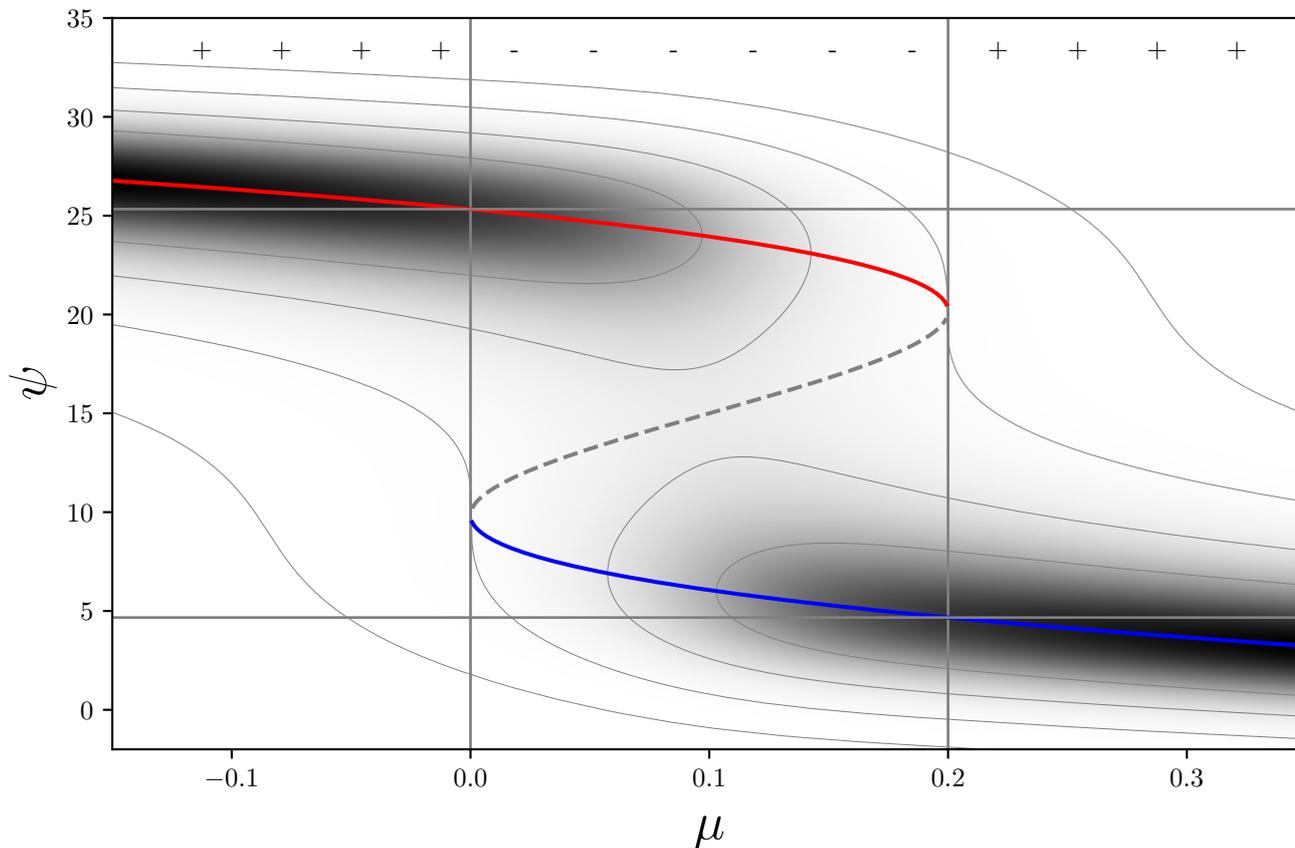
Figure 4. Example trajectory with corresponding distribution. Parametrised by $\lambda = 15$, $\nu = 20$, $\sigma = 0.12\nu$, $\mu_+ = 0.2$, $\mu_- = 0$, $\beta_0 = 0.2$, $\delta\beta = 0$; α under constraints in Eqs 4 and 5. The distribution of one of the attractor branches (red: on state, blue: off state) deforms when closer to the bifurcation points which are connected though the repellor that forms the trench of the distribution (dashed line). Top $\pm$ symbols indicate unimodal ($+$) or bimodal ($-$) regime based on the discriminant value (D). The value of σ is relatively large and is chosen for clarity.

the AMOC rapidly moves from one attractor to the other. For this reason, the sampled bimodality region might be larger than is apparent from a particular sample AMOC trajectory.

The distributions in Figs 5 show that qualitatively distinct behaviour occurs when α or β are varied. For both parameters,
160   a change from a unimodal to a bimodal distribution can be seen. These changes correspond directly to the potential functions in Fig. 2. In principle, a change in on only one of the two structural parameters (α and β) can move the distribution between unimodal and bimodal forms.

The normalisation of the family of distributions depends on the values of the parameters. Therefore, we are required to calculate the normalisations factors for each parameter set. This cannot be done analytically, but can be done accurately with
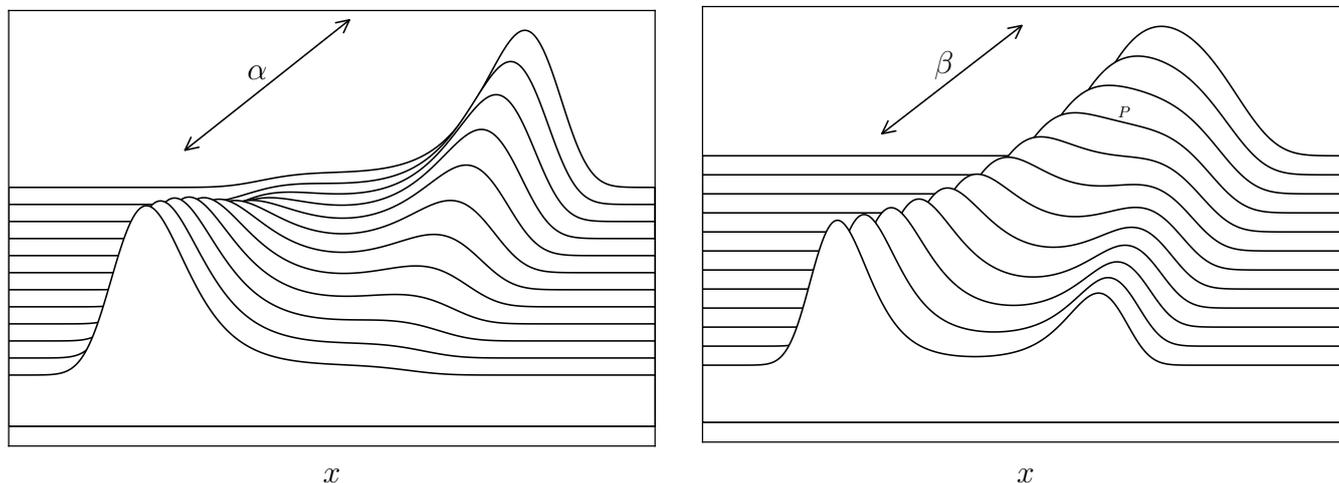165   an adaptive quadrature method (Piessens et al., 2012), though it suffers from numerical limitations.

Figure 5. Left: Distributions from the exponential family (Eq. 6) where the parameter β is kept at a fixed value and α is varied. The distribution transforms from unimodal (back), to bimodal (middle), to a different unimodal distribution (front). The unimodal states have distinct singular maxima. The bimodal states have a dominant and a weak mode, depending on the position within the bimodal regime; in the middle and inversion from weak to dominant takes place. The relative strength between dominant and weak depends on σ. Right: Distributions from the exponential family (Eq. 6) where the parameter α is kept at a fixed value and β is varied. A broad unimodal state (at the back) splits into distinct bimodal states (to the front). In the middle a critical point exists, called the cusp (point P in Fig. 3) where the split occurs.

We are now in a position to apply the above to collapse trajectories from climate models.

3   Simulated AMOC collapse parameter estimation

We describe how to find an optimal solution under the framework arrived at in the previous section. Using a Bayesian opti-misation procedure, estimated values of β_0 and $\delta\beta$ can be found, together with the scaling parameters ν and λ. We will also estimate the values for $\mu_\pm$, resulting in a six parameter list that describes (the upper branch) of an AMOC collapse.

The parameters β_0 and $\delta\beta$ are independent to each other, but need to cross the curves $B_{1,2}$ in Fig. 3) to match the correspond-ing values for $\mu_\pm$. This constraint is satisfied by the resulting values for α_0 and $\delta\alpha$. (This can still lead to solution candidates that are not suitable for the collapse trajectories and are eliminated in the sampling process below.) The scaling parameters are not fully independent because $\lambda < \nu$ and knowing where the upper and lower branches are located already gives a rough estimate.

3.1   Parameter estimation

Cobb (1978) was able to fit the distribution in Eq. 6 using optimisation techniques (which were numerically unstable and not very flexible). Though the estimates for the scaling parameters λ and ν can be quite good with this approach, estimating the

trajectory parameters β_0 and $\delta\beta$ requires a more flexible method. Knowing which distribution to use, we can fit its parameters

180 under some measure of goodness-of-fit by machine learning. Specifically, we can use Bayes' rule to maximise the likelihood of a parametrised Langevin model $L_\sigma(\nu, \lambda, \beta_0, \delta\beta, \mu_\pm)$ given a trajectory $\Psi(\mu)$ (Bolstad, 2010),

$$P(\nu, \lambda, \beta_0, \delta\beta, \mu_\pm \mid \Psi),$$

to arrive at the (linearised) posterior distributions of ν, λ, β_0, $\delta\beta$, $\mu_\pm$ under the observed $\Psi(\mu)$. Bayes' rule tells us the probability of a given observation Ψ given the probability of the parameters (marginal on the left, or posterior) is proportional

185 to the probability given the parameters (marginal on the right, or prior) and the full distribution (likelihood),

$$P(\nu, \lambda, \beta_0, \delta\beta, \mu_\pm \mid \Psi) \propto P(\Psi \mid \nu, \lambda, \beta_0, \delta\beta, \mu_\pm) \cdot P(\nu, \lambda, \beta_0, \delta\beta, \mu_\pm).$$

(The right-hand side of Bayes's rule is called the Bayes factor and can be normalised by the probability of the observed trajectory $P(\Psi)$ (called the evidence) to obtain an equality.)

Sampling different values from the parameters' prior distributions will give corresponding values for the posterior distribu-

190 tions. These resultant posterior distributions can, in turn, be used as prior distributions, yielding a chain of sampled parameter vectors. A Bayesian sampler chooses successive values that tend towards greater likelihood of the model, given the observed trajectory, and will converge towards an optimal fit. This is roughly what an MCMC (Markov chain Monte-Carlo) optimiser does (Bolstad, 2010). A widely used sampling algorithm is the Metropolis algorithm (Hastings, 1970), which we also use here.[1]

The models can be fit with uninformative priors, but the sampling process is time consuming because the evaluation of the

195 potential (to calculate $P(\Psi \mid \nu, \lambda, \beta_0, \delta\beta, \mu_\pm)$) requires numerical integration (using a quadrature method), which is costly to evaluate (the exponential family of distributions cannot, in general, be evaluated analytically).

3.1.1 Prior distributions

The prior distribution of a parameter represents all the information known about that parameter before confrontation with the observed values (Bolstad, 2010). With ν and λ introduced earlier, the state variable x undergoes an affine transformation and

200 normalises the polynomial. These transform the AMOC state variable (Ψ) with a shift (λ) and a scaling (ν). The shift λ cannot exceed the normalisation ν, giving an upper bound on λ. Also, we note the lower limit of the lower branch, meaning λ must be larger than this minimum value. Similarly, the scaling ν cannot be larger than the maximum value of the AMOC on the upper branch. We expect the linear parametrisation of α and β introduced in the previous section to be $\mathcal{O}(1)$.

We are nonetheless still faced with infinite support on the coefficients of the expansion of the parameters (β_0, $\delta\beta$). We

205 therefore transform β_0 and $\delta\beta$, with support $(-\infty, \infty)$, using the arctan function to map to $(-\pi/2, \pi/2)$. After such a transformation, we can sample from the flat prior distribution on that interval with most of the probability mass on 'reasonable' values (i.e. $\mathcal{O}(1)$). For β_0 and $\delta\beta$ this transformation will be used and α will follow from the constraints in Eqs 5 and 4. An overview

---

[1]This algorithm has been implemented in many software packages.

of priors is:

$$\nu = U(\min(\text{AMOC}), \max(\text{AMOC}))$$

$$\lambda = U(\min(\text{AMOC}), \nu)$$

$$\mu_+ = U(\mu_{\text{S+}}, \mu_{\text{UP}})$$

$$\mu_- = U(\mu_{\text{DN}}, \mu_{\text{S-}})$$

$$\tan(\beta_0) = U(-\pi/2, \pi/2)$$

$$\tan(\delta\beta) = U(-\pi/2, \pi/2),$$

with $\min(\text{AMOC})$ and $\max(\text{AMOC})$ is the minimum/maximum values in an observed collapse trajectory. U is the uniform distribution on indicated intervals. The intervals values of the collapse points $\mu_\pm$ we stipulate as being bounded by where the trajectories merge (μ_{UP} and μ_{DN}) and the inner values ($\mu_{\text{S-}}$ and $\mu_{\text{S+}}$) observed in the trajectories (within which bimodality is demanded, see Fig. 6). [2]

3.2 Fitting EMIC collapse trajectories

An AMOC collapse was induced in six models of intermediate complexity in Rahmstorf et al. (2005) by applying a freshwater forcing to the North-Atlantic subtropical gyre region that reduced the salinity in the subpolar gyre to its north. In Fig. 6 the trajectories of those collapses are reproduced (right column, the freshwater flux has been labelled μ here) together with their numerical derivatives (left column). Both the AMOC strength and the forcing value have units Sv ($=10^6 \, \text{m s}^{-1}$). Note that the bifurcation points ($\mu_\pm$) must lie within the range where the trajectories appear bimodal.

model	σ	μ_-	μ_+	present day
UVic	0.260	[-0.020, 0.010]	[0.188, 0.225]	(0.080, 25.0)-
MOM iso	0.216	[-0.010, 0.020]	[0.150, 0.210]	(0.050, 22.8)-
MOM hor	0.526	[-0.010, 0.010]	[0.130, 0.200]	(0.110, 20.0)-
C-GOLDSTEIN	0.122	[-0.100, 0.035]	[0.115, 0.190]	(-0.100, 29.0)+
Bremen	0.181	[-0.018, 0.010]	[0.120, 0.220]	(0.070, 18.8)-
ECBilt-CLIO	0.176	[-0.044, 0.030]	[0.115, 0.210]	(-0.110, 18.2)+

11.3

Table 1. Overview of models, the estimated standard deviation with the upper branch fitted to a linear function (note that the original trajectories had already been smoothed), the ranges of $\mu_\pm$, the location of present day in the models, and whether the present day value is in the unimodal regime (+) or not (-). All values in units of Sv.

---

[2]To ensure the signs of the discriminants are the same, we added an additional constraint as a sharp-peaked likelihood on the discriminant that follows from the proposed solutions and that follows from the observed trajectory. This helps the fitting process by explicitly excluding invalid solutions with incorrect modality.

Earth System
Dynamics
Discussions

225   The trajectories are taken from the numerical Earth System Models (EMICs) Rahmstorf et al. (2005, Fig. 2, bottom panel)
by extracting the data points directly from the graphic in the electronic publication. The numerical derivatives show where the
AMOC changes quickest as a response to the change in freshwater forcing. Each model has two peaks where the changes are
largest, one for each change between stable branches. These peaks are located at the repellor in between the two attractors (the
stable branches). At the repellor only unstable solutions exist and the AMOC is driven to a more stable solution, away from
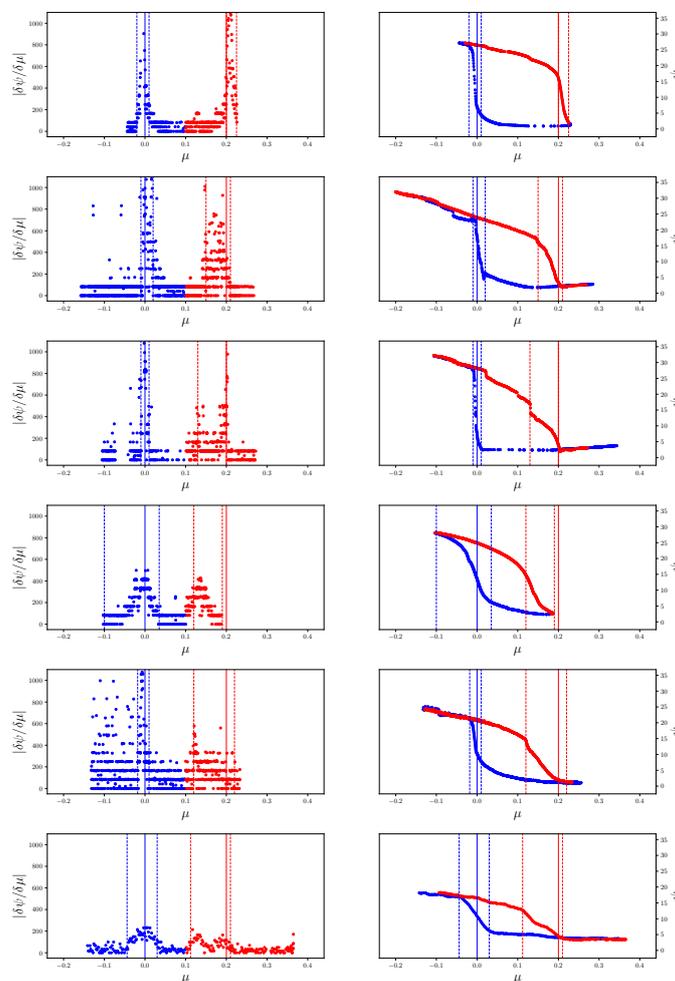230   these states.

Figure 6. Absolute values of numerical derivatives (left) from the trajectories of AMOC strength as function of freshwater forcing to the
right (taken from Rahmstorf et al. (2005, Fig. 2, bottom panel)). In red the upper branch, blue the lower branch. Top to bottom: UVic, MOM
iso, MOM hor, C-GOLDSTEIN, Bremen, ECBilt-CLIO. Vertical solid lines mark $\mu = 0$ (blue) and $\mu = 0.2$ (red); vertical dashed lines mark
the chosen boundary values for $\mu_\pm$. All values in units of Sv.

If no other mechanisms apart from the salt advection are important we expect the bifurcation points to lie beyond the observed transition points because a noise-induced transition pushes the AMOC into the off-state sooner. (Note that although the collapse points are expected to lie before these peaks, low levels of noise will obscure this effect.) The dashed lines indicate the regions where we will search for the optimum values of $\mu_\pm$. These differ from the fixed 0 and 0.2 values chosen by (Rahmstorf et al., 2005), who also shifted the trajectories to align on these values. The dragged-out descent to the lower branch (e.g. the model Bremen) indicates that the salt-advection mechanism does not necessarily result in an abrupt collapse in the trajectory.

Before fitting, the upper and lower branches were extended to the left and right to fill the space of $-0.2 < \mu < 0.4$. A linear fit was use to produce additional values of the corresponding branches (at the same density of those points already present). All models then occupy the same freshwater forcing space. This is desirable because not all models have a lower branch that is fully sampled (specifically, UVic). The lower branch was extended with a negative rate of increase if the lower branch was moving upwards with increasing μ (MOM hor and MOM iso).

Because we want to model the collapse of the AMOC, only the behaviour of the upper branch and the lower branch beyond μ_+ is of interest. We expect these points of the trajectory do be dominated by the salt-advection mechanism. The hysteresis loop is exemplified in the trajectory of C-GOLDSTEIN. This model shows a smooth path with a relatively rapid transition region–but not a collapse as such by appearance.

We start by identifying some characteristic points in the trajectories in Tab. 1. The σ (variance of the process) of the models is not given in Rahmstorf et al. (2005) or elsewhere in the literature, but was estimated as the deviation with a fitted function to (part of) the top branch (note that smoothing was already applied in Rahmstorf et al. (2005), lowering the variance of the trajectories). Note that the 'off-state' of the AMOC in these models is not 0, but ~ 2Sv of AMOC strength. If the salt-advection mechanism were the only operative effect, we expect this value to be ≤ 0. If a reverse advection cell emerges as the lower hysteresis branch, this value is negative.

The model trajectories are apparently driven by the forcing value, which means we should be able to explain the behaviour of the hysteresis using this variable as the only driver applied to the Langevin model.

In Fig. 7 fitted distributions are shown for linearly parametrised sample paths through the stability space (also tabulated in Tab. 7). The β parameter changes linearly and α follows from the constraints in Eqs. 4 and 5. Blue and red lines indicate the prior bounds for μ_- and μ_+, respectively. The dashed grey line marks the positions of the unstable solution (repellor) in between the two attractor branches which separates the two basins of attraction.

The fits with a linear track through the (α, β) parameter space result in a mismatch between the behaviour seen on lower branches and that on the upper branches. In particular, the upper branch shows a non-linear degradation, where the fitted distributions do not. This is less obvious for UVic and ECBilt-CLIO, but especially apparent for the two MOM models.

4 Discussion and conclusion

We derived a simple model of AMOC collapse based on Langevin dynamics (Eq. 1) with a changing freshwater forcing (μ) and applied this to EMIC simulated collapse trajectories taken from Rahmstorf et al. (2005). The collapse occurs at a bifurcation
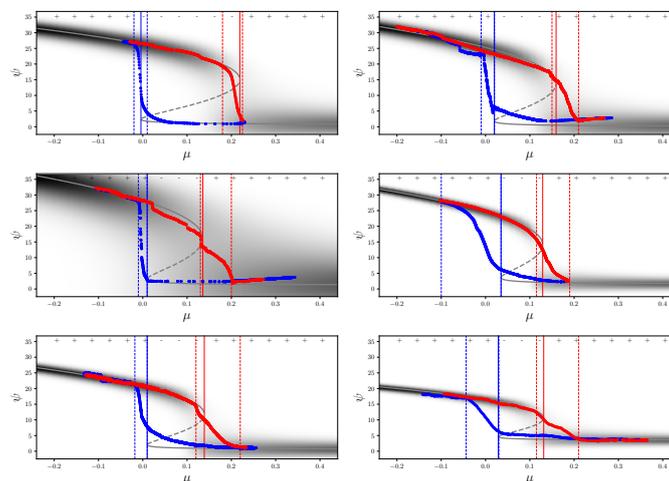
Earth System
Dynamics
Discussions

Figure 7. Estimated distributions under changing μ. Top-left to bottom-right: UVic, MOM iso, MOM hor, C-GOLDSTEIN, Bremen, ECBilt-CLIO. Vertical dashed lines mark the chosen boundary values for $\mu_\pm$, with solid lines the fit values. Grey dashed line indicates the local minimum in the distribution (trench). Top $\pm$ symbols indicate the sign of the discriminant D for the fitted distribution (+ for unimodal, − for bimodal). Distribution spreads have been inflated with a factor $\nu/2$ to make them visible. All values in units of Sv.

	model	ν	λ	β_0	$\delta\beta$	μ_-	μ_+
	UVic	23.32	10.39	0.3523	-1.051	-0.004	0.219
	MOM iso	25.38	9.395	0.293	-1.259	0.020	0.159
11.3	MOM hor	27.93	11.36	0.2719	-1.339	0.010	0.136
	C-GOLDSTEIN	23.99	10.12	0.2715	-1.458	0.035	0.129
	Bremen	20.75	8.184	0.2778	-1.339	0.010	0.139
	ECBilt-CLIO	14.1	8.369	0.2497	-1.303	0.029	0.131

Table 2. Parameters corresponding to the fitted functions in Fig 7.

point μ_+ which appears smaller than given in (Rahmstorf et al., 2005). A corresponding bifurcation point μ_- relates an abrupt

265 transition back to the on-state. Additionally, a linear parameterisation through state space couples to the freshwater applied to the North Atlantic subtropical gyre region. The AMOC also requires an offset and scaling parameter to be fitted (λ and ν). These six parameters are sufficient to describe the abrupt collapse/re-invigouration of the AMOC that leads to a hysteresis loop under varying freshwater forcing.

The AMOC collapse and re-invigouration seen in these models cannot be completely fitted with Langevin dynamics due to

270 the asymmetry in the lower vs the upper branch. It is, however, possible to fit the change in the upper branch of the AMOC–the 'on-state'–as it moves towards a critical point and the dominant salt-advection feedback mechanism breaks down.

We note that Rahmstorf et al. (2005) determine the AMOC strength as the maximum of the meridional volume transport in the North Atlantic and might explain the asymmetry between the two branches. If for a reverse overturning cell the wrong metric has been used then the lower branch location is not correct. It is conceivable that the Langevin model results in better fits if Rahmstorf et al. (2005) had sampled $|\max(\Psi)|$ instead of $\max(\Psi)$, which would have resulted in a better metric of the lower branch. With the metric used it is not apparent whether a reversed overturning cell was present or not because it was not sampled if the AMOC had taken on a negative value. Therefore, there is no obvious way to model the asymmetry between the two branches, and obtain a full description. The two branches could be separated by associating each with a different overturning cell. The upper branch is identified with the NADW-driven cell, while a reverse cell is responsible for the lower branch. If indeed a reverse overturning cell (as described in e.g. Yin and Stouffer (2007)) dominates the lower AMOC branch, two separate overturning cells are responsible for the observed trajectories, and the two branches then cannot be expected to fit with the same parameter set.

However, another possible explanation is that (two) separate mechanisms are responsible for the upper and lower branch dependency on μ. Possible mechanisms include possible mechanisms include the influence of wind-stress, SPG convective instability (Hofmann and Rahmstorf, 2009), or other pathways of deep water formation (Heuzé, 2017). Also, changes in the ITCZ (inter-tropical convergence zone) due to ocean-atmosphere feedbacks are possible (Green et al., 2019); these can, in turn, can affect the salinity of the North Atlantic subtropical gyre region. However, Mecking et al. (2017) showed that for a high-resolution model the salt-advection feedback was nevertheless stronger than the ITCZ effects. Also, Gent (2018) notes that the EMICs in Rahmstorf et al. (2005) have reduced air-sea interaction feedbacks compared to more modern and more complicated fully coupled ocean-atmosphere models; stronger feedbacks lead to greater AMOC stability. Other wind coupling can occur further south through a coupling with the ACC (Antarctic Circumpolar Current) which is based on the thermal wind relation (Marshall and Johnson, 2017).

A third explanation is that deep water formation is a local process, and as a result an asymmetry is to be expected between the two branches. Local convection can, however, be subject to global controls and be associated with a sinking branch which occurs in conjunction with deep convection, but is not directly driven by it, see Spall and Pickart (2001) for detailed discussion. The AMOC could develop a reverse cell where the overturning is driven by Antarctic Intermediate Water (AAIW), which is not part of the conceptual picture presented here (Yin and Stouffer, 2007; Jackson et al., 2017). The reverse cell introduces an asymmetry in the collapse trajectories because the driver of deep water formation is not in the North Atlantic, and might break our assumption that both the on and off branches are controlled by the same process. It is therefore difficult to estimate the return path of the AMOC if the lower branch has additional drivers from the dominant salt-advection mechanism of the upper branch. Forcing values appropriate for the lower branch might be different than those found for the upper branch.

Furthermore, the methodology used in this paper comes with difficulties in the numerical implementation. The fit procedure requires the normalisation of each distribution in the μ timeseries. Because no analytic solution exist a numerical approach is needed. The numerical integration adds to the computational costs of the fits. The Markov chain method is prone to find local optima. Also, the cost of numerical integration necessitates stopping the fits at shorter chains than (perhaps) are needed, an analytic formulation of the integrand would alleviate this but none exists to our knowledge. Modern sampling algorithms allow

for gradient information to be used, which is effective when sampling a higher dimensional parameter space (the Metropolis algorithm used in this paper has greater difficulty as the dimensionality of the parameter space increases). It is possible for non-admissible solutions to be generated; in which case, the sampler is effectively reduced to a random searcher, till its finds a

310 solution subspace to optimise under. (Tighter constraints on the prior distributions could be beneficial here.)

As stated in Rahmstorf et al. (2005), the EMIC trajectories had already been smoothed, resulting in a smaller variance; a smaller variance leads to distributions that are more sharply peaked. This increases the computational cost of integrating the distributions numerically. Smoothing can also add to the inertia seen in the collapses, but might be due to other reasons such as stopping the EMIC simulations before equilibration of the AMOC collapse, leaving the AMOC in a winding-down

315 state. Also, it is not clear how long the models in Rahmstorf et al. (2005) were integrated per freshwater forcing value. If the integrations were done for an insufficient amount of time, the AMOC collapse is incomplete, leaving the measured value out of equilibrium. The intermediate points in the collapse trajectories beyond the bifurcation points indicate that either the sample points are inaccurate or other processes are involved in the AMOC.

Finally, the fitted collapse trajectories were done on an ensemble of EMICs, which arguably are not sufficiently representative

320 of the real climate. The question arises to what extent the procedure outlined in this paper can be applied to more complicated models such as those in the CMIP archives (Taylor et al., 2012). These models do not show a full collapse trajectory like those in Rahmstorf et al. (2005), which means no sample points of the lower branch are available. If it is indeed possible to use direct numerical stochastic integration of the Langevin equation, no lower branch needs to be sampled. Sampling from the prior distributions and optimising under the observed upper branch should also lead to robust estimates. Another advantage

325 is that the probability distribution can be assumed Gaussian when on the upper branch. Using a simple box model, transition probabilities for an AMOC collapse have been determined by Castellana et al. (2019). From the CMIP ensemble a similar estimate might be obtained using a direct numerical stochastic integration approach. Provided the CMIP models accurately capture the behaviour of the real AMOC and the freshwater forcing counterpart (our μ) can be identified, an estimate can be made of the distance of the current climate state to the collapse point. It is therefore still an open question how probable an

330 AMOC collapse is in more realistic models, and reality, but with the method outlined in this paper a first step could be made in answering those questions.

References

Balakrishnan, N. and Nevzorov, V. B.: A primer on statistical distributions, John Wiley & Sons, 2004.

Berglund, N. and Gentz, B.: Metastability in simple climate models: pathwise analysis of slowly driven Langevin equations, Stochastics and
Dynamics, 2, 327–356, 2002.

Birkhoff, G. and Mac Lane, S.: A survey of modern algebra, Macmillan New York, 1970.

Bolstad, W. M.: Understanding computational Bayesian statistics, vol. 644, John Wiley & Sons, 2010.

Broecker, W. S., Peteet, D. M., and Rind, D.: Does the ocean-atmosphere system have more than one stable mode of operation?, Nature, 315,
21–26, 1985.

Broecker, W. S., Bond, G., Klas, M., Bonani, G., and Wolfli, W.: A salt oscillator in the glacial Atlantic? 1. The concept, Paleoceanography,
5, 469–477, 1990.

Castellana, D., Baars, S., Wubs, F. W., and Dijkstra, H. A.: transition probabilities of noise-induced transitions of the Atlantic ocean circula-
tion, Scientific Reports, 9, 1–7, 2019.

Cobb, L.: Stochastic catastrophe models and multimodal distributions, Systems Research and Behavioral Science, 23, 360–374, 1978.

Cobb, L.: Estimation theory for the cusp catastrophe model, 1980.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe,
M., Tebaldi, C., Weaver, A. J., and Wehner, M.: Long-term climate change: Projections, commitments and irreversibility, pp. 1029–1136,
Cambridge University Press, Cambridge, UK, https://doi.org/10.1017/CBO9781107415324.024, 2013.

Dansgaard, W., Johnsen, S., Clausen, H., Dahl-Jensen, D., Gundestrup, N., Hammer, C., Hvidberg, C., Steffensen, J., Sveinbjörnsdottir, A.,
Jouzel, J., et al.: Evidence for general instability of past climate from a 250-kyr ice-core record, Nature, 364, 218, 1993.

Ditlevsen, P. D. and Johnsen, S. J.: Tipping points: Early warning and wishful thinking, Geophysical Research Letters, 37,
https://doi.org/10.1029/2010GL044486, l19703, 2010.

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

Gardiner, C. W.: Handbook of stochastic methods for physics, chemistry and the natural sciences, vol. 13 of *Springer Series in Synergetics*,
Springer-Verlag, third edn., 2004.

Gent, P. R.: A commentary on the Atlantic meridional overturning circulation stability in climate models, Ocean Modelling, 122, 57–66,
2018.

Green, B., Marshall, J., and Campin, J.-M.: The 'sticky' ITCZ: ocean-moderated ITCZ shifts, Climate dynamics, 53, 1–19, 2019.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57, 97–109,
https://doi.org/10.1093/biomet/57.1.97, 1970.

Held, H. and Kleinen, T.: Detection of climate system bifurcations by degenerate fingerprinting, Geophysical Research Letters, 31, 2004.

Heuzé, C.: North Atlantic deep water formation and AMOC in CMIP5 models, Ocean Science Discussions, 2017, 1–22,
https://doi.org/10.5194/os-2017-2, 2017.

Hofmann, M. and Rahmstorf, S.: On the stability of the Atlantic meridional overturning circulation, Proceedings of the National Academy
of Sciences, 106, 20 584–20 589, 2009.

Jackson, L., Smith, R. S., and Wood, R.: Ocean and atmosphere feedbacks affecting AMOC hysteresis in a GCM, Climate Dynamics, 49,
173–191, 2017.

Kleinen, T., Held, H., and Petschel-Held, G.: The potential role of spectral properties in detecting thresholds in the Earth system: application
375   to the thermohaline circulation, Ocean Dynamics, 53, 53–63, 2003.

Lemons, D., Gythiel, A., and Langevin's, P.: paper "Sur la théorie du mouvement brownien [On the theory of Brownian motion]", CR Acad.
Sci.(Paris), 146, 530–533, 1908.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate
system, Proceedings of the national Academy of Sciences, 105, 1786–1793, 2008.

380   Livina, V. N., Kwasniok, F., and Lenton, T. M.: Potential analysis reveals changing number of climate states during the last 60 kyr., Climate
of the Past, 6, 2010.

Marshall, D. P. and Johnson, H. L.: Relative strength of the Antarctic Circumpolar Current and Atlantic Meridional Overturning Circulation,
Tellus A: Dynamic Meteorology and Oceanography, 69, 1338 884, https://doi.org/10.1080/16000870.2017.1338884, 2017.

Mecking, J., Drijfhout, S., Jackson, L., and Andrews, M.: The effect of model bias on Atlantic freshwater transport and implications for
385   AMOC bi-stability, Tellus A: Dynamic Meteorology and Oceanography, 69, 1299 910, 2017.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C., and Kahaner, D.: Quadpack: A Subroutine Package for Automatic Integration, Springer
Series in Computational Mathematics, Springer Berlin Heidelberg, 2012.

Poston, T. and Stewart, I.: Catastrophe theory and its applications, Surveys and reference works in mathematics, Pitman, 1978.

Rahmstorf, S.: The thermohaline ocean circulation: A system with dangerous thresholds?, Climatic Change, 46, 247–256, 2000.

390   Rahmstorf, S., Crucifix, M., Ganopolski, A., Goosse, H., Kamenkovich, I., Knutti, R., Lohmann, G., Marsh, R., Mysak, L. A., Wang, Z.,
et al.: Thermohaline circulation hysteresis: A model intercomparison, Geophysical Research Letters, 32, 2005.

Smeed, D., Josey, S., Beaulieu, C., Johns, W. E., Moat, B., Frajka-Williams, E., Rayner, D., Meinen, C., Baringer, M., Bryden, H., et al.: The
North Atlantic Ocean is in a state of reduced overturning, Geophysical Research Letters, 45, 1527–1533, 2018.

Spall, M. A. and Pickart, R. S.: Where does dense water sink? A subpolar gyre example, Journal of Physical Oceanography, 31, 810–826,
395   2001.

Stommel, H.: Thermohaline convection with two stable regimes of flow, Tellus, 13, 224, 1961.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological
Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Vellinga, M. and Wood, R. A.: Global climatic impacts of a collapse of the Atlantic thermohaline circulation, Climatic change, 54, 251–267,
400   2002.

Weijer, W., Cheng, W., Drijfhout, S. S., Fedorov, A. V., Hu, A., Jackson, L. C., Liu, W., McDonagh, E. L., Mecking, J. V., and Zhang,
J.: Stability of the Atlantic Meridional Overturning Circulation: A review and synthesis, Journal of Geophysical Research: Oceans, 124,
5336–5375, 2019.

Yin, J. and Stouffer, R. J.: Comparison of the Stability of the Atlantic Thermohaline Circulation in Two Coupled Atmosphere–Ocean General
405   Circulation Models, Journal of Climate, 20, 4293–4315, https://doi.org/10.1175/JCLI4256.1, 2007.