RESEARCH ARTICLE | OCTOBER 16 2024

Investigating ocean circulation dynamics through data assimilation: A mathematical study using the Stommel box model with rapid oscillatory forcings FREE

Special Collection: Nonautonomous Dynamics in the Climate Sciences

Nathaniel Smith ; Anvaya Shiney-Ajay ; Emmanuel Fleurantin ; Ivo Pasmans

Chaos 34, 103131 (2024)
https://doi.org/10.1063/5.0215236

Investigating ocean circulation dynamics through data assimilation: A mathematical study using the Stommel box model with rapid oscillatory forcings

Cite as: Chaos 34, 103131 (2024); doi: 10.1063/5.0215236
Submitted: 23 April 2024 · Accepted: 26 September 2024 ·
Published Online: 16 October 2024

Nathaniel Smith,[1,a] Anvaya Shiney-Ajay,[2,b] Emmanuel Fleurantin,[3,c] and Ivo Pasmans[4,d]

AFFILIATIONS

[1]Department of Mathematics, Miami University, Oxford, Ohio 45056, USA
[2]School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455, USA
[3]Department of Mathematical Sciences, George Mason University, Fairfax, Virginia 22030, USA
[4]Department of Meteorology, University of Reading, RG6 6ET, Reading, United Kingdom

Note: This paper is part of the Focus Issue, Nonautonomous dynamics in the climate sciences.
[a]**Author to whom correspondence should be addressed:** nismith33@gmail.com
[b]ajayx006@umn.edu
[c]efleuran@gmu.edu
[d]i.c.pasmans@reading.ac.uk

ABSTRACT

We investigate ocean circulation changes through the lens of data assimilation using a reduced-order model. Our primary interest lies in the Stommel box model, which reveals itself to be one of the most practicable models that has the ability of reproducing the meridional overturning circulation. The Stommel box model has at most two regimes: TH (temperature driven circulation with sinking near the north pole) and SA (salinity driven with sinking near the equator). Currently, the meridional overturning is in the TH regime. Using box-averaged Met Office EN4 ocean temperature and salinity data, our goal is to provide a probability that a future regime change occurs and establish how this probability depends on the uncertainties in initial conditions, parameters, and forcings. We will achieve this using data assimilation tools and DAPPER within the Stommel box model with fast oscillatory regimes.

The Atlantic Meriodional Overturning Circulation (AMOC), a crucial element of the global ocean circulation system, is primarily influenced by the temperature and salinity gradients between the north pole and the equator. As climate change progresses, it becomes increasingly important to understand its potential impacts on the direction and behavior of this current. To investigate these effects, we will simulate the AMOC with the aid of the simplified Stommel box model. We show that the Stommel model has no periodic solutions. Instead, depending on its model parameters, either a solution with sinking near the pole, representing the current circulation, or near the equator is retrieved. Using a twin experiment, it is shown that data assimilation can be used to retrieve estimates for the model parameters. Finally, by assimilating the MetOffice EN4 dataset, realistic estimates for the model parameters are produced, which are then used to estimate the probability of an AMOC direction reversal under different possible future climate change scenarios.

I. INTRODUCTION

Thermohaline circulation is the global current driven by density variations in the oceans. Its branch in the Atlantic ocean is known as the Atlantic Meridional Overturning Circulation

(AMOC). In the northern part of the Atlantic, it transports warm, salty water from the subtropics to the Arctic. Here, it comes into contact with colder (sub)polar waters and sink forming the North Atlantic Deep Water and flows southward in the deep ocean.[1] It plays a crucial role in the climate system, especially for the climate in northwestern Europe. Based on a simple two-box model,[2] it is suggested that AMOC can be described as a bistable system characterized by two distinct equilibrium states: TH (thermally driven), representing the current circulation regime with sinking near the pole, and SA (salinity driven), signifying reversed circulation with equatorial sinking. The existence of multiple equilibria has since been confirmed by both reversal visible in the paleo-record as well as more sophisticated climate models; see, e.g., Refs. 3–5.

In the region where sinking occurs, temperature and salinity exert opposing effects on density. On one hand, heat exchange with the cold (sub)polar environment cools down the subtropical water that is being advected northward increasing its density. On the other, contact with melt water freshens it, decreasing its density. Currently, the former effect dominates. However, it is feared that climate change driven increases in temperature, and ice melt and precipitation can tip the balance to the latter, thus reversing the circulation. The transition between the circulation regimes can occur through various mechanisms.[6–8] Beyond a critical point, parameter values may lead to bifurcation-induced tipping (B-tipping). Alternatively, the rate of parameter changes might trigger rate-induced tipping (R-tipping). Furthermore, system variability due to inherent noise could result in noise-induced tipping (N-tipping). Previous research, as documented in Refs. 9–16, has established thresholds for tipping occurrence. In Ref. 17, the Stommel box model is coupled to a sea-ice component in the polar box and displays rate-induced tipping.

It is still an open question how close to such a tipping point the world is.[9] In 2023 alone, two studies on the topic[18,19] have received considerable attention in the press, but their results have been criticized.[20] The aim of this study is twofold. The first is to explore whether it is possible to use observations of ocean temperature and salinity in combination with a reduced-order model in a process called data assimilation to estimate model parameters. The second is to determine how the addition of forcing, both seasonal and climatological, to the Stommel model affects this probability of an AMOC reversal.

It is important to note that while one might assume that a two-box model would have limitations in fully describing tipping behavior, our approach, incorporating high-frequency oscillations, parameter variability, and data assimilation, brings us close to realistic scenarios. This methodology also sheds light on the feasibility of assimilating data to further understand the qualitative dynamics of our model. It offers a window of validity for our predictions, considering factors, such as the strength of the forcing.

A. Leveraging conceptual models for parameter estimation and predictive analysis

The use of conceptual models, such as the Stommel box model, for parameter estimation and prediction of complex systems, such as the AMOC, is motivated by several important considerations. First, reduced-complexity models often provide clearer insights into the fundamental mechanisms driving the system's behavior, which can be obscured in more complex models.[21–23] This interpretability is vital for understanding the key processes governing AMOC dynamics.

In light of these advantages, our objective is to model the AMOC and examine its behavior under various conditions. We aim to assess the feasibility of estimating temperature diffusivity, salinity diffusivity, and advection parameters by assimilating ocean temperature and salinity data. Additionally, we seek to forecast the probability of circulation regime changes. The simplicity of the chosen conceptual model makes this analysis possible as we are able to initialize large ensembles at low computational cost. It also serves as a baseline in complexity with more detailed versions of the model available for further studies on higher order dynamics.[24,25]

Last, these models serve as useful benchmarks against which more complex models can be evaluated, helping to assess whether increased complexity leads to improved predictive skill.[12,26–29] While we acknowledge the limitations of conceptual models, their use in conjunction with data assimilation techniques offers a powerful approach to understanding AMOC dynamics and predicting potential tipping points.

B. Outline of the paper

The outline of the paper is as follows. First, we introduce the autonomous and non-autonomous Stommel model in Sec. II and study the qualitative behavior of the solutions as time goes to infinity. In Sec. III, we explain our data assimilation method and the assimilated observations. In Sec. IV, we conduct a twin experiment to investigate the potential of data assimilation to estimate model parameters. In Sec. V, observations based on the Met Office Hadley Centre EN4 data set are assimilated in a non-autonomous Stommel model, and the impact of the different forcings on the tipping probability is investigated. In Sec. VI, we present and analyze the results of our data assimilation experiments, examining the model's behavior under various climate change scenarios and assessing the likelihood of AMOC regime shifts. Finally, in Sec. VII, we discuss our findings.

II. MATHEMATICAL ANALYSIS OF THE STOMMEL BOX MODEL: AUTONOMOUS AND NON-AUTONOMOUS PERSPECTIVES

A. The Stommel box model

The Stommel model comprises two fully mixed water containers (shown in Fig. 1). One represents the region of the North Atlantic just north of the equator with temperature and salinity denoted as (T_{e*}, S_{e*}), while the other symbolizes the polar region with temperature and salinity as (T_{p*}, S_{p*}). (The asterisk subscripts here denote variables with a unit.) These containers are interconnected by both a capillary tube and an overflow area, facilitating water circulation while preserving a constant volume in each container. The relative salinities and temperatures of each box determine the flow of water, and when the flow of water reverses, we consider that to be a regime change.

Initially, then, this scenario is described with four equations, which are described below:

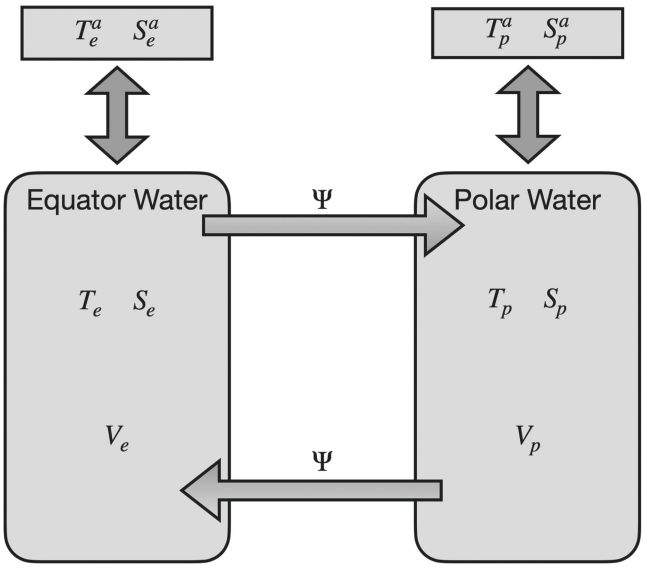

FIG. 1. A conceptual depiction of how Henry Stommel modeled thermohaline circulation. The circulation between the two connected bodies of well-mixed water in the Stommel model is driven by density gradients between the two boxes.

$$\begin{aligned}
\Delta x\Delta z\Delta y_e \frac{dT_{e*}}{dt_*} &= k_T\Delta x\Delta y_e\left(T_e^a - T_{e*}\right) + |\Psi_*|\left(T_{p*} - T_{e*}\right),\\
\Delta x\Delta z\Delta y_e \frac{dS_{e*}}{dt_*} &= k_S\Delta x\Delta y_e\left(S_e^a - S_{e*}\right) + |\Psi_*|\left(S_{p*} - S_{e*}\right),\\
\Delta x\Delta z\Delta y_p \frac{dT_{p*}}{dt_*} &= k_T\Delta x\Delta y_p\left(T_p^a - T_{p*}\right) + |\Psi_*|\left(T_{e*} - T_{p*}\right),\\
\Delta x\Delta z\Delta y_p \frac{dS_{p*}}{dt_*} &= k_S\Delta x\Delta y_p\left(S_p^a - S_{p*}\right) + |\Psi_*|\left(S_{e*} - S_{p*}\right) - m^* S_{p*},
\end{aligned} \tag{1}$$

with Δz, Δx being the depth and the zonal width of the boxes and Δy_p, Δy_e the meridional width of the polar and equatorial box, respectively, k_T the temperature diffusion coefficient, and k_S the salinity diffusion coefficient. The model incorporates the exchange of heat and salt within each box as a response to surface forcing, employing relaxation toward specified ocean surface temperature and salinity values (T^a, S^a). Next to this, the model accommodates for the addition of fresh water supply to the polar box at a discharge rate of m^*. In particular, this quantity represents the melting of icebergs due to climate change and the release of additional fresh water that comes from that.

The flow rate is given by the following equation:

$$\Psi_* = \gamma\,\Delta x\Delta z\left(\rho_{p*} - \rho_{e*}\right)/\rho_0, \tag{2}$$

where γ is the advection coefficient and ρ_0 is a reference density. Since those are both constants, the flow rate is linearly related to the difference in the water density between the two boxes. We note that if the flow is due to denser water in the polar box, as it is for the current TH regime, the flow rate is taken to be positive.

The Stommel system also assumes the following linear equation of state:

$$\rho_* = \rho_0(1 - \alpha_T(T_* - T_0) + \alpha_S(S_* - S_0)), \tag{3}$$

where a "0" subscript indicates a reference value, α_T is the thermal expansion coefficient, and α_S the haline coefficient.

However, by instead looking only at the temperature and salinity relative to one another, we can reduce this down to a two variable system by subtracting the equations for the pole box from the equatorial box in the case that $m^* = 0$,

$$\begin{aligned}
\frac{dT}{dt} &= \eta_1 - T(1 + |T - S|),\\
\frac{dS}{dt} &= \eta_2 - S(\eta_3 + |T - S|),\\
\Psi &= T - S,
\end{aligned} \tag{4}$$

where

$$\begin{aligned}
T &= \frac{\Delta z\gamma\alpha_T}{\overline{\Delta y}k_T}(T_{e*} - T_{p*}),\\
S &= \frac{\Delta z\gamma\alpha_S}{\overline{\Delta y}k_T}(S_{e*} - S_{p*}),\\
\Psi &= \frac{1}{\overline{\Delta y}\Delta x k_T}\Psi_*,\\
t &= \frac{k_T}{\Delta z}t_*,\\
\overline{\Delta y} &:= \frac{\Delta y_p\Delta y_e}{\Delta y_p + \Delta y_e}.
\end{aligned}$$

Here, η_1, η_2, and η_3 are parameters independent of the ocean state. Their full values are as follows:

$$\eta_1 = \frac{\Delta z\gamma\alpha_T}{\overline{\Delta y}k_T}\left(T_e^a - T_p^a\right), \tag{5}$$

$$\eta_2 = \frac{k_S}{k_T}\frac{\Delta z\gamma\alpha_S}{\overline{\Delta y}k_T}\left(S_e^a - S_p^a\right), \tag{6}$$

$$\eta_3 = \frac{k_S}{k_T}. \tag{7}$$

Setting $\frac{dT}{dt} = 0$, $\frac{dS}{dt} = 0$ in Eq. (4), we get the following equilibrium solutions:

$$T = \frac{\eta_1}{1 + |\Psi|} \quad \text{and} \quad S = \frac{\eta_2}{\eta_3 + |\Psi|}.$$

In the TH regime, we have $\Psi = |\Psi|$. In this case, the effect of the temperature difference dominates the effect of the salinity difference and the system is thermally driven. We find that

$$\eta_2 = -\Psi^2 - \eta_3\Psi + \eta_1\left(\frac{\eta_3 + \Psi}{1 + \Psi}\right). \tag{8}$$

On the other hand, in the SA regime, we have $\Psi = -|\Psi|$. In this case, the transport due to the salinity difference dominates over the

transport due to the temperature difference. We find that

$$\eta_2 = \Psi^2 - \eta_3\Psi + \eta_1\left(\frac{\eta_3 - \Psi}{1 - \Psi}\right). \tag{9}$$

The solutions from Eqs. (8) and (9) for fixed η_1, η_3 can be shown together as a function of the dimensionless transport. From this, it is clear that for a range of η_2, three steady-state solutions for the transport are possible since the vertices of both parabolas described above lie in the area where $\psi > 0$, meaning that the TH regime provides two solutions for certain values of η_2 and the SA regime provides one.

B. The non-autonomous Stommel model

We extend the Stommel model to include seasonal variations, represented by periodic changes in surface temperatures and salinities. We also know that any oscillatory behavior will not arise from the autonomous Stommel model (see Appendix A) and, thus, will solely emerge due to the fact that η_1 and η_2 are no longer constant in time but oscillate, i.e.,

$$\eta_1 = \eta_{1,auto} + B\sin(\Omega t), \tag{10}$$

$$\eta_2 = \eta_{2,auto} + \hat{B}\sin(\Omega t), \tag{11}$$

with $\eta_{\cdot,auto}$ representing the values used in the autonomous system in Eq. (4). Here, B and $\hat{B}$ are distinct, but we define $A = B - \hat{B}$ for future use. Using that, our definition of Ψ, and the overall Stommel system, we can write

$$\begin{aligned}\frac{d}{dt}\Psi = \frac{d}{dt}(T - S) &= \frac{dT}{dt} - \frac{dS}{dt}\\ &= (\eta_{1,auto} + B\sin(\Omega t) - T(1 + |T - S|))\\ &\quad - (\eta_{2,auto} + \hat{B}\sin(\Omega t) - S(\eta_3 + |T - S|))\\ &= \eta_{1,auto} - \eta_{2,auto} + \eta_3 S - T - (T - S)|T - S|\\ &\quad + (B - \hat{B})\sin(\Omega t)\\ &= \eta_{1,auto} - \eta_{2,auto} + \eta_3 S - T - \Psi|\Psi| + A\sin(\Omega t),\end{aligned}$$

then we substitute in $S = T - \Psi$, which ultimately gets us

$$\begin{aligned}\frac{d}{dt}\Psi &= \eta_{1,auto} - \eta_{2,auto} + (T - \Psi)\eta_3 - T - \Psi|\Psi| + A\sin(\Omega t),\\ \frac{d}{dt}T &= \eta_{1,auto} + B\sin(\Omega t) - T(1 + |\Psi|)\end{aligned} \tag{12}$$

for a system without a climate change, in particular, for a system in which $m^* = 0$.

In order to get a full picture of the dynamics of Eq. (12), it is, thus, fitting to analyze its Ψ-vs.-$\eta_{2,auto}$ bifurcation diagram. It is worth noting that the discontinuity surface $\Sigma_0 = \{(T, S) \in \mathbb{R}^2_+ | T = S\}$ gives rise to a Non-Smooth Fold (NSF) bifurcation, black star in Fig. 2, where the stable equilibrium branch terminates at $\eta_{2c} \equiv \eta_1\eta_3$.

Understanding NSF bifurcations can be helpful in predicting critical transitions or tipping points in complex systems, aiding in risk assessment and management. This paper examines the non-autonomous Stommel box model as a foundation for studying transitions in thermohaline circulation, while providing insights into the dynamical impact of NSF bifurcations.

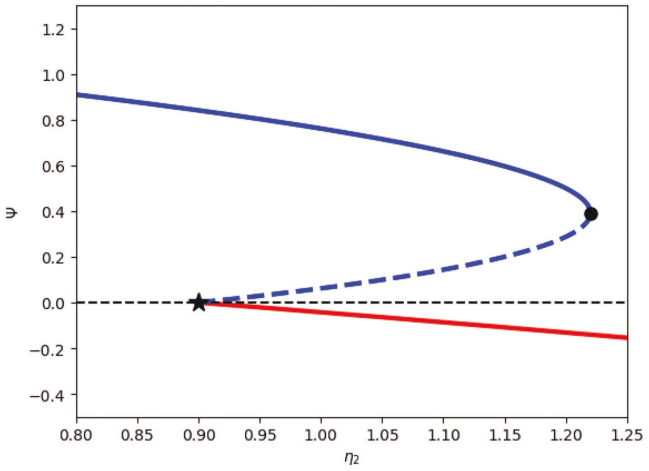

FIG. 2. A Ψ-vs.-$\eta_{2,auto}$ bifurcation diagram for the system in Eq. (12) with $\eta_{1,auto} = 3$, $\eta_3 = 0.1$, and A=B=0. The solid red and blue lines are stable equilibria, while the dotted blue line represents an unstable equilibrium. The solid black circle indicates the system's saddle-node bifurcation, while the black star indicates the NSF bifurcation. The dotted black line emphasizes the discontinuity set Σ_0.

In the context of our model, the NSF arises when the unstable equilibrium solution (dotted blue curve) and the focus (solid red curve) intersect with the discontinuity surface. This tipping point corresponds to the rapid transition from solutions near the temperature-dominated branch of nodes (solid blue curve) to the salinity-dominated branch of focus points.

For the non-autonomous Stommel model [Eq. (12)], under high-frequency forcing, we observe amplification or attenuation of patterns in ψ, which depends linearly on the amplitude-to-frequency ratio (A/Ω). That is to say, solutions starting near the NSF bifurcation point may naturally shift away toward higher values of η_2 or be driven toward lower values under the effect of data assimilation. The foundational work for this analysis can be found in a paper by Budd *et al.*[30] Their study provides valuable insights into the expected behaviors of the parameters η_1 and η_2, as well as the equilibria, when incorporating seasonal variations. Building upon their findings, we extend the investigation by introducing data assimilation techniques to further enhance our understanding of these complex dynamics. By assimilating observational data into the model, we aim to explore how the combination of seasonal variations and data assimilation influences the system's behavior and equilibria.

III. INVERSION OF THE STOMMEL BOX MODEL

The solution of the Stommel box model in (1) depends on the surface diffusion parameters k_T and k_S as well as the advection parameter γ. Given the simplified nature of the Stommel box model,

it is not possible to estimate appropriate values for these parameters from first principles. Data assimilation allows one to reconstruct plausible values for these parameters from comparisons between model output and observations. In this section, we will introduce the data assimilation methodology and the assimilated observations. The described model and methodology are implemented in the DA Python package DAPPER.[31]

A. Ensemble Kalman filter

The true state of the system, which in our case consists of the temperature and salinity in the ocean boxes as well as the advection and diffusion coefficients, is not known. However, based on, e.g., the model output, it is possible to provide an estimate of the true value in the form of a probability distribution. Data assimilation (DA) is a procedure that combines this *a priori* distribution with imperfect observations (see Sec. III B) to produce a posteriori distribution with smaller spread around the truth. The most-likely value for the truth based on the *a priori* distribution is referred to as the forecast, while the most-likely value of the *a posteriori* distribution is known as the analysis. In this work, we use a particular DA method known as the augmented ensemble transform Kalman (ETKF)[32] in which both errors in the observations as well as the aprior distribution are assumed to be Gaussian. The mean and covariance of the latter estimated from an ensemble of model runs. A more detailed description of the augmented ETKF approach used can be found in Appendix B.

B. Observations and model parameters

Geometry of the two boxes in the model, box-average temperature and salinity, and observation error variances and box-surface-averaged temperature and salinity necessary for the model's forcing are compiled based on the Met Office's EN4 quality controlled objective analysis[33] with bathymetric bias corrections.[34] The EN4 dataset is a dataset of several types of *in situ* potential temperature and salinity measurements interpolated onto a 1×1 degree grid of the globe at 42 non-equally space depths on a monthly basis. From this dataset, the vertical profiles in the North Atlantic between 23.5 °N and 89.0 °N and between 90.0 °W and 90.0 °E that are at least 1.5 km deep are selected. These boundaries were chosen, and such a model includes the North Atlantic north of the intertropical convergence zone (5–15 °N) as well as Greenland and Norwegian. The southern boundary was chosen because precipitation in the former leads to freshening of the ocean surface that the simple Stommel box model geometry cannot accommodate. The northern boundary was chosen because overflow from the Nordic sea over the Iceland–Schotland ridge has been shown to be the dominant source for North Atlantic Deep water (NADW),[35] which makes up the lower branch of the AMOC circulation. Only the part of the profiles above a depth of 3.5 km is used as this corresponds to the lower limit of occurrence of the NADW. Data from the period 2004–2022 are used. This selection is based on the fact that during this period, extensive subsurface measurements were available thanks to the Argo program[36] as well as that they exclude a sharp spike of surface temperatures at the beginning of the century probably caused by processes our simple Stommel box model cannot represent.

The processes behind formation of North Atlantic Deep Water (NADW) in the Arctic that makes up the deep-ocean part of the AMOC are numerous, complex, and their relative importance is disputed. They include the overflow of intermediate waters formed during the winter from the Nordic and Irminger seas over the Greenland–Scotland ridge,[35,37] downward convection in the Labrador Sea driven by heat-loss[38] and compensation for wind-driven upward mixing in the Atlantic.[39,40] Of these, only the former mechanism is incorporated in the Stommel model. Therefore, we divide the vertical profiles between the boxes as follows: those profiles in which surface temperature drops below the depth-averaged temperature for on average 1 month per year or more, and, consequently, convection due to heatloss can take place, are assigned to the polar box. The other profiles are assigned to the equator box. This choice was not only motivated by the physical formation process of NADW, but also on more practical grounds. The data assimilation system needs to be able to reduce the average temperature of the deep ocean if it is found to exceed observed values. This definition of the polar box ensures that a mechanism to release heat from the ocean is present. This choice is not entirely unproblematic (see Sec. VII) and deviates from the one used by Refs. 41–43. Their boxes extend to 23 °S, cover the whole Arctic Ocean, not only the Nordic seas. Furthermore, they simple define the equatorial box to be the part of the Atlantic Ocean between 30 °S and 40 °N as they base deep-ocean temperatures and salinities on fit to a long-term climate model output. Hence, overestimation of deep-ocean temperature can be dealt with by lowering initial deep-ocean temperatures, and no heat release mechanism is necessary. Furthermore, the fit in Chapman *et al.*[43] is based solely on temperature and salinity differences between the boxes, and hence, the box model does not need to provide a deep-ocean heat release mechanism. In other box model studies,[44] no attempt is even made to match boxes and their values with actual observations.

Volume V of each of the boxes is calculated by summing the volume of all grid cells in the box. Similarly, surface area A is calculated by summing the area of the EN4 grid cells in the top

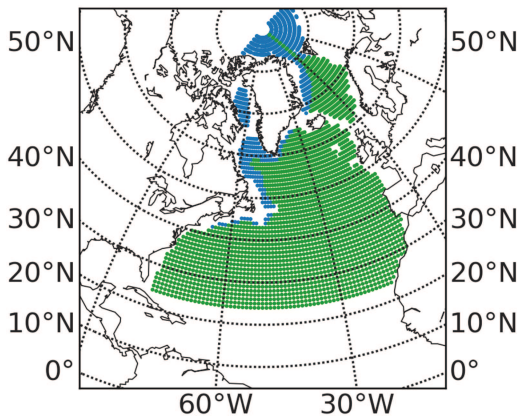

FIG. 3. An example of an EN4 profile assignment in the year 2010. Green points signify profiles assigned to the equatorial box, and blue points signify profiles assigned to the polar box.

TABLE I. The zonal width, the meridional width, and the depth of the two boxes used in the Stommel model based the EN4 data.

	Polar	Equator
Δx (km)	7274	7274
Δy (km)	324	2262
Δz (m)	3148	3148

layer in both boxes. The depth of both boxes is then calculated as $\Delta z = \frac{V_p+V_e}{A_p+A_e}$. The cross section A_c between boxes is derived by summing the vertical area between profiles assigned to the polar box and those assigned to the equatorial box. An example assignment is shown in Fig. 3. From this, zonal and meridional size of the boxes is defined as $\Delta x = \frac{A_c}{\Delta z}$ and $\Delta y_p = \frac{A_p}{\Delta x}$, $\Delta y_e = \frac{A_e}{\Delta x}$. The values resulting from this exercise are shown in Table I.

For each month t and each of the two boxes, the box-averaged temperature is calculated as

$$T_{\cdot *}(t) = \left(\sum_i V_i T_{i,*}(t)\right)\left(\sum_i V_i\right)^{-1}, \tag{13}$$

where i runs over all EN4 grid cells in the box below the surface layer, V_i is the volume of the grid cell, and $\sigma_{T,i}$ is the uncertainty in temperature as specified in the EN4 data set. The variance in the value is calculated as

$$\sigma^2_{T*}(t) = \left(\sum_i V_i \sigma^2_{T_{i,*}}(t)\right)\left(\sum_i V_i\right)^{-1}. \tag{14}$$

These calculations are repeated for salinity S. The values after January 2004 are used to create the observations to be assimilated in $\mathbf{d}(t) = \left[T_{p*}(t), T_{e*}(t), S_{p*}(t), S_{e*}(t)\right]^{\mathrm{T}}$. The 4×4 observational error covariance matrix Σ_d is assumed to be diagonal. The values on the diagonal are chosen to be the maximal values for $\sigma^2_{T*}(t)$ and $\sigma^2_{S*}(t)$ in each box. The resulting time-average values for the observations as well as the observational error standard deviation found this way are shown in Table II.

The calculations in Eqs. (13) and (14) are repeated but now with i running over all grid cells in the surface layer of a box. The signal is then represented as

$$T^a(t) \approx T^a_0 + T^a_{\sin}\sin\left(\frac{2\pi t}{\tau}\right) + T^a_{\cos}\cos\left(\frac{2\pi t}{\tau}\right), \tag{15}$$

TABLE II. Time-average of the assimilated temperature and salinity observations together with the observational standard deviation, i.e., the square-roots of the diagonal of Σ_d.

	Time-averaged obs.	Obs. error std. dev.
T_{p^*} (°C)	1.2	0.3
T_{e^*} (°C)	5.5	0.5
S_{p^*} (ppt)	34.83	0.07
S_{e^*} (ppt)	35.15	0.07

TABLE III. Regression coefficients found by fitting model Eq. (15) to the surface observations for temperature [a] and salinity [a] in polar $\cdot_p$ and $\cdot_e$ boxes. Also shown is the seasonal amplitude $\sqrt{\cdot^2_{\cos} + \cdot^2_{\sin}}$.

Field	$\cdot_0$	$\cdot_{\cos}$	$\cdot_{\sin}$	Seasonal amplitude
T^a_p (°C)	1.5	−1.5	−1.1	1.9
T^a_e (°C)	16.7	−2.4	−2.3	3.4
S^a_p (ppt)	33.05	0.22	0.32	0.39
S^a_e (ppt)	35.77	0.04	0.05	0.07

with $\tau = 1$ year and T_0, $T_{\cos}$, $T_{\sin}$ determined by a weighted linear regression against the box-averaged surface values using $\sigma^{-2}_{T*}(t)$ as weights. The regression coefficient for surface temperature and their equivalents for surface salinity are given in Table III.

The regression coefficients in Table III can be related to the perturbations added to η_1 and η_2 in Eqs. (10) and (11). These relations can be found in Table IV.

The numerical scheme for the Stommel model uses a finite volume approach with an explicit fourth-order Runge–Kutta scheme with a time step of 1 month. Initial conditions for the ocean temperatures and salinities in the different ensemble members are drawn from a Gaussian distribution having the EN4 box-averaged temperatures and salinities for January 2004 as mean and Σ_d as covariance. Initial values for the logarithm of the diffusion and advection parameters are estimated by minimizing, using SciPy's Nelder–Meat method,

$$\frac{\left(\Delta T_* - \Delta T_{eq}\right)^2}{\sigma^2_{T,e} + \sigma^2_{T,p}} + \frac{\left(\Delta S_* - \Delta S_{eq}\right)^2}{\sigma^2_{S,e} + \sigma^2_{S,p}} + \frac{\left(Q_* - Q_{eq}\right)^2}{(2.5\,\mathrm{Sv})^2}, \tag{16}$$

with ΔT_*, ΔS_* being the difference between equatorial and polar box's initial ocean temperature and salinity, $\sigma_{T,\cdot}$, $\sigma_{S,\cdot}$ the observational standard deviations in the third column of Table II, $Q = 18$ Sv the observed meridional transport, and ΔT_{eq}, ΔS_{eq} Q_{eq} equilibrium values for temperature, salinity difference, and transport that depend on the model parameters; i.e., we pick parameters such that our initial conditions are close to an equilibrium value. The logarithms are perturbed by adding realizations from a Gaussian with a standard deviation of 0.26. This corresponds to the assumption that the relative error in these model parameters is 30%. These model

TABLE IV. Relation between the coefficients Ω, A, B, $\hat{B}$ in Ref. 30 and the surface forcing obtained from the EN4 observation data.

	Dimensional	Scale	Nondimensional
Ω	2π rad yr^{-1}	$\frac{k_T}{\Delta z}$	3564
B	1.51 °C	$\frac{k_T}{\alpha_T \gamma \Delta z}\frac{\Delta y_p \Delta y_e}{\Delta y_p + \Delta y_e}$	1.18
$\hat{B}$	0.32 ppt	$\frac{k_T}{k_S}\frac{k_T}{\alpha_S \gamma \Delta z}\frac{\Delta y_p \Delta y_e}{\Delta y_p + \Delta y_e}$	0.66
A	…	…	0.52

TABLE V. Parameters, forcings, initial conditions, and their sources.

Data	Values	Source	Details
Temperature (ocean temperature)	Pole: $T^*_p = 1.23\,^\circ$C, Equator: $T^*_e = 5.4\,^\circ$C		Initial values from EN4 data
Salt (ocean salinity)	Pole: $S^*_p = 34.82$ ppt, Equator: $S^*_e = 35.15$ ppt		Initial values from EN4 data
temp_diff (surface heat mixing coefficient)	$k_T = 3.7 \times 10^{-6}\,\mathrm{m\,s^{-1}}$		Equation (16)
salt_diff (surface salinity mixing coefficient)	$k_S = 1.2 \times 10^{-6}\,\mathrm{m\,s^{-1}}$		Equation (16)
ρ_0 (reference density)	$1027\,\mathrm{kgm^{-3}}$	45	
S_0 (reference salinity)	35 ppt	45	
T_0 (reference temperature)	$10\,^\circ$C	45	
α_T (thermal expansion coefficient)	$\frac{0.15\,^\circ\mathrm{C}^{-1}}{\rho_{ref}}$	45	
α_S (haline expansion coefficient)	$\frac{0.78\,\mathrm{ppt}^{-1}}{\rho_{ref}}$	45	
Initial estimate $Q_{overturning}$ (meridional overturning flux)	18 Sv	46	
γ (advective transport coefficient)	$2.0\,\mathrm{ms^{-1}}$		Equation (16)
$T^a(t)$	Oscillates over time		Calculated from the EN4 dataset; see Table III
$S^a(t)$	Oscillates over time		Calculated from the EN4 dataset; see Table III
V_ice	$2.9 \times 10^6\,\mathrm{km^3}$		
Greenland ice sheet melt period t_{melt}	10 000 y	47–49	
Equatorial warming rate	$0.03\,^\circ\mathrm{C\,yr^{-1}}$	50–52	
Polar warming rate	$0.06\,^\circ\mathrm{C\,yr^{-1}}$	52 and 53	
Model time step	1 month		

parameters are assumed to be constant in time. More specifically, their values are held constant between DA corrections during the prediction step (see Appendix B. However, the parameters for each ensemble member might still change over time due to the DA corrections reflecting that the probability density distribution for the true value of the model parameters is not necessarily stationary. Values for the other parameters used in the model can be found in Table V. Note that this table includes an equatorial and a polar warming rate, and that the polar rate is twice the equatorial. The choice to use two warming rates was made to account for Arctic amplification. Arctic amplification is the name for an effect observed over many decades in which the Arctic warmed faster than other parts of the globe.[54] The precise magnitude of the difference is disagreed upon, but many studies report that the poles warm around twice the global average;[55,56] therefore, we used this ratio to inform our model.

IV. TWIN EXPERIMENTS

In this section, we will evaluate whether the data assimilation system is capable of correcting ocean temperatures and salinities as well as reconstructing the Stommel model parameters. In addition to this, we want to see what happens with the AMOC circulation if the effects of climate change are added to the model. This is done by running an Observing System Simulation Experiment (OSSE) or a twin experiment.[57] In such an experiment, one of the ensemble members is set apart as "truth" and withheld from the DA system forming one of the twins. Artificial observations are created from the "truth" by adding perturbations drawn from a Gaussian distribution with covariance Σ_d with these perturbations mimicking observation errors. After assimilating these observations into the experiment, the forecasts and analyses, making up the other twin, can be compared with the "truth" to quantify performance of the DA system.

We conduct simulations under four distinct scenarios. In the first scenario, initial conditions and model parameters are perturbed in the different ensemble members, but no data assimilation is used. The second scenario mirrors the first, but now ocean temperatures and salinities are assimilated each month during the period 2004–2022. For the third scenario, the effects of global warming are taken into account. In this study, we take two processes related to climate change into account. First, that the average temperature increases faster near the pole ($0.06\,\frac{^\circ\mathrm{C}}{\mathrm{year}}$ for a polar box) than at lower latitudes ($0.03\,\frac{^\circ\mathrm{C}}{\mathrm{year}}$ for an equatorial box). Second, that the Greenland ice sheet melts add a constant rate $m^* = \frac{V_{ice}}{t_{ice}}$ with V_{ice} and t_{ice} given in Table V. In order not to interfere with the DA, warming and melting only start after 2022, i.e., after the end of the DA period, and is applied uniformly to all ensemble members. The fourth scenario matches the third and includes data assimilation. The model time step is 1 month and DA is carried out monthly. Figure 4 shows

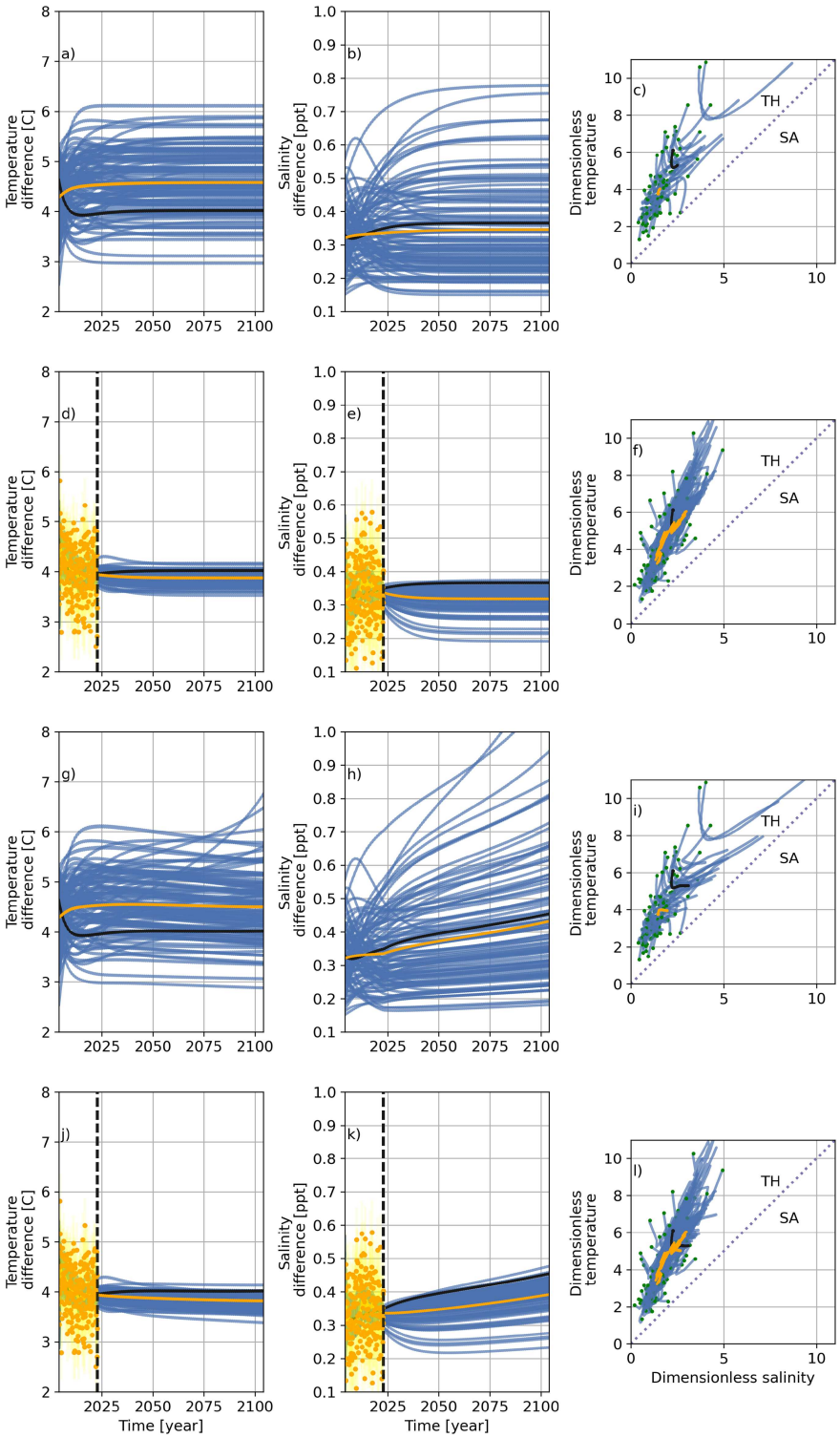

FIG. 4. Forecast time series of temperature and salinity differences between equatorial and pole boxes for 100 ensemble members, with corresponding dimensionless salinity-vs-temperature phase portraits. Green dots: initial conditions; dashed blue line: TH-SA regime boundary; black paths: synthetic truth; orange paths: ensemble mode. Rows show different experiments: (a)–(c) nCC-nDA, (d)–(f) nCC-yDA, (g)–(i) yCC-nDA, and (j)–(l) yCC-yDA. (nCC, no climate change; yCC, with climate change; nDA, no data assimilation; yDA, with data assimilation). Columns: temperature difference, salinity difference, dimensionless temperature vs salinity phase portrait.

the model ocean temperatures and salinities over 100 years for 100 ensemble members together with its most-likely value (the ensemble mean) and the artificial truth. For times at which an *update step* is carried out (see Appendix B) and two solutions are available, the solution after the DA correction is shown. The timeseries obtained in this way is referred to as the forecast solution. Temperatures and salinities in the third column are non-dimensionalized using the scaling constants in Sec. II A. Blue curves indicate the thermohaline (TH) regime, while red curves represent the salinity-driven (SA) regime. As shown by (4), the solution to the Stommel equations can be written solely as a function of the difference in temperature and salinity between the equatorial and polar box. Therefore, the results in Fig. 4 are depicted as a function of these differences only.

As seen in Fig. 4, none of the ensemble members predict a change in the flow direction in the experiment with no climate change and no data assimilation. This is notwithstanding the large uncertainties in the model parameters. One can see that there is a significant spread between trajectories in the time plots, but the ensemble remains monomodal; i.e., the ensemble members do not form different clusters. This is a necessary condition for the EnKF to function properly. Comparison of Figs. 4(a), 4(b), 4(g), and 4(h) with Figs. 4(d), 4(e), 4(j), and 4(k), respectively, shows that assimilating temperatures and salinities for the period 2004–2022 reduces the spread bringing all ensemble members closer to the truth. Trajectories become and stay tightly connected in the time and phase plots. This indicates that DA is successful in correcting the temperatures and salinities. When the effects of climate change are applied in Figs. 4(g), 4(h), 4(j), 4(k), the solutions no longer converge toward an equilibrium value. Instead, temperature and salinity keep increasing reflecting the introduced changes in forcing. Apart from this, the results are qualitatively the same as in the first two scenarios: the circulation remains in a TH mode and the DA reduces the errors in the ensemble members.

In order to verify whether the DA system is capable of correcting the model parameters, the temperature diffusivity parameter, the salinity diffusivity parameter, and the advection parameter are shown in Fig. 5 as a function of time during the DA period together with their true value. We see that the spread is sound as so far as the truth lies within the range of the ensemble. Comparing the most-likely estimate with the truth (in green), we see that initially DA can increase the error in the model parameters. However, by the end of the period, errors have reduced to values below the initial error.

To quantify the behavior of the ensembles in Fig. 4 in more detail, the ensemble standard deviation, or the ensemble spread, and RMSE for the different experiments are shown in Fig. 6 for the different state variables. The first thing that stands out is that ensemble spread and RMSE after the DA period generally stay constant or decrease, even in the absence of DA. The second noticeable point is that DA is successful in reducing the error in the model parameters. Errors for the advection parameter [Fig. 6(g)] and the temperature diffusion parameter [Fig. 6(e)] are decimated, while that of the salinity parameter is reduced but not to the same extent [Fig. 6(f)]. This reflects that salinity diffusion parameters are smaller. Consequently, it takes the system more time to equilibrate with surface salinity forcing and consequently, the DA is less sensitive to changes in the salinity diffusion coefficient than in the advection and the temperature diffusion coefficient. The fact that by the end of the DA period, the spread in the salinity diffusion parameter is still decreasing [Fig. 6(l)], while that in the temperature diffusion parameter [Fig. 6(j)] and the advection parameter [Fig. 6(n)] has leveled off suggest that further RMSE reduction for the former could be achieved by extending the DA period. Finally, for a properly calibrated ensemble Kalman filter, the expectation value of the ensemble spread and the root mean-square error (RMSE) should be equal. Comparison of the RMSEs and ensemble standard deviations at the end of the DA period shows that they are of a similar magnitude with the spread slightly overestimating the uncertainty. Thus, confirming that DA is performing as expected.

All ensemble members in Fig. 4 are in the TH circulation mode. In order to see whether the setup is capable of producing solutions with the SA circulation, the impacts of the climate change are scaled up. More precisely, the warming rate in the

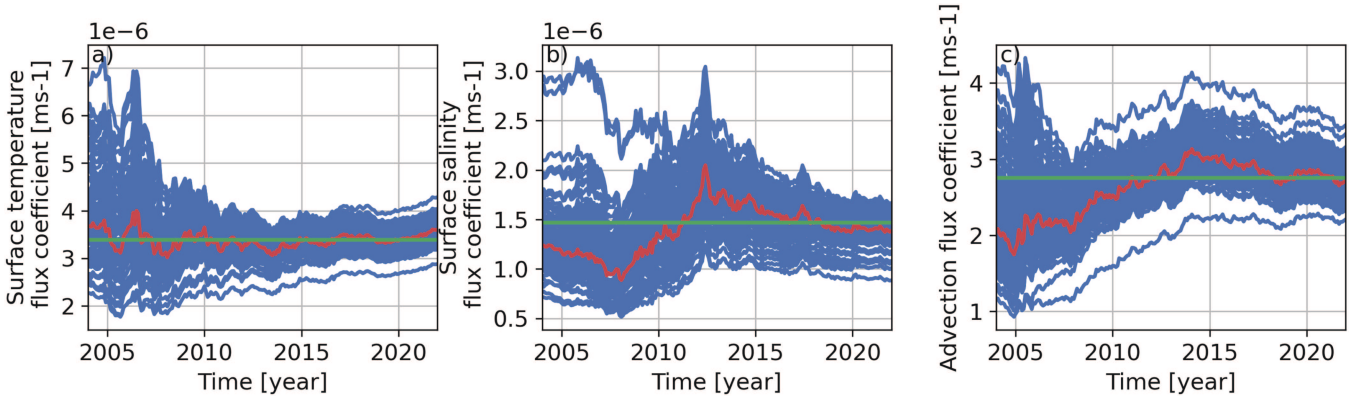

FIG. 5. (a) Temperature (κ_T), (b) salinity (κ_S), and (c) advective flux coefficients (γ) of the ensemble plotted over the assimilation period for an experiment assimilating 18 years of synthetic data without any global warming effects (the experiment with global warming is effectively identical). The red curve represents the ensemble mode, while the green line represents the true coefficient value.

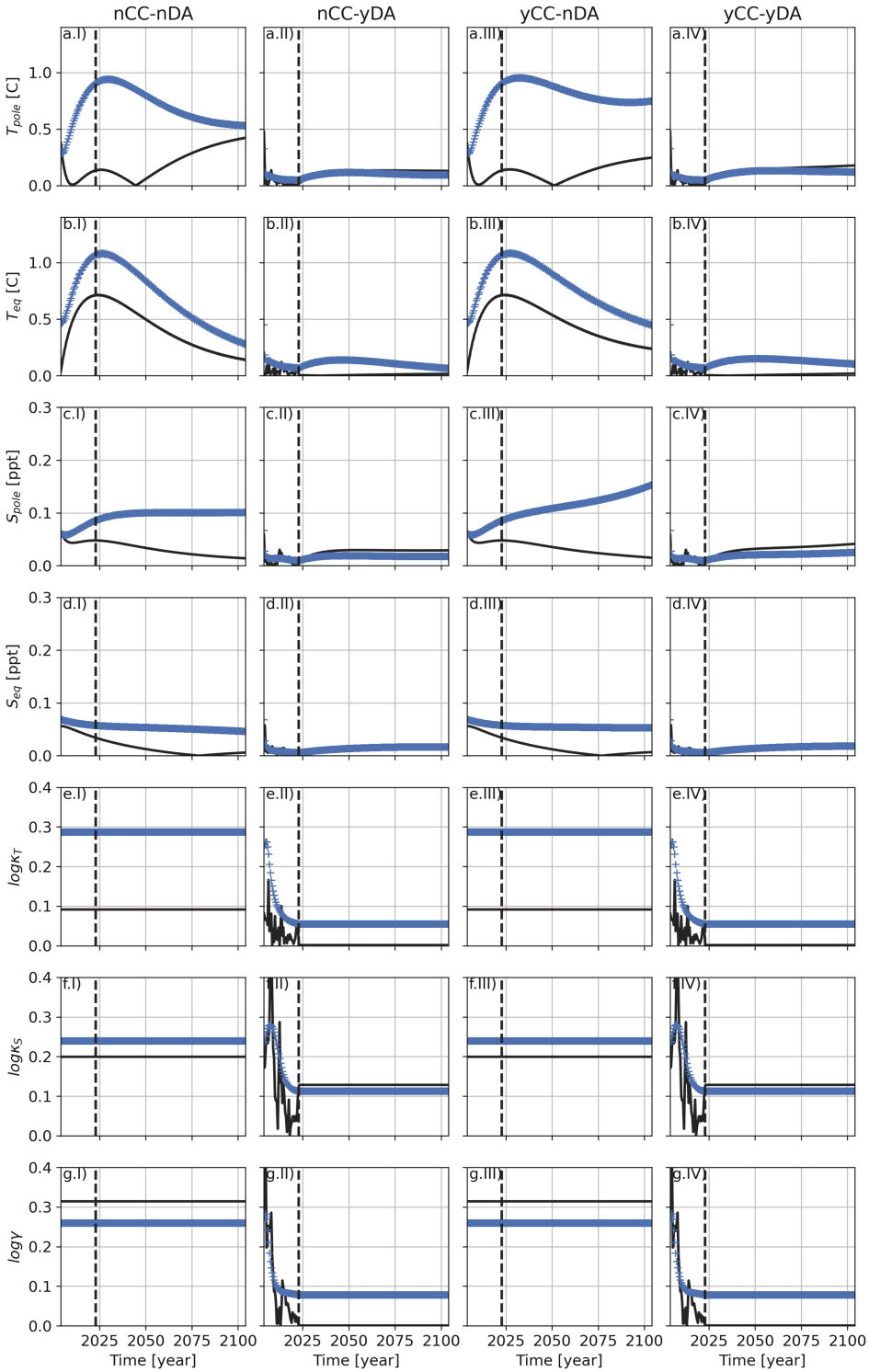

FIG. 6. (Black solid lines) RMSE and (blue pulses) ensemble standard deviation of (rows) different variables as a function of time for (columns) the four different experiments in Fig. 4. For visibility, only every sixth month is shown.

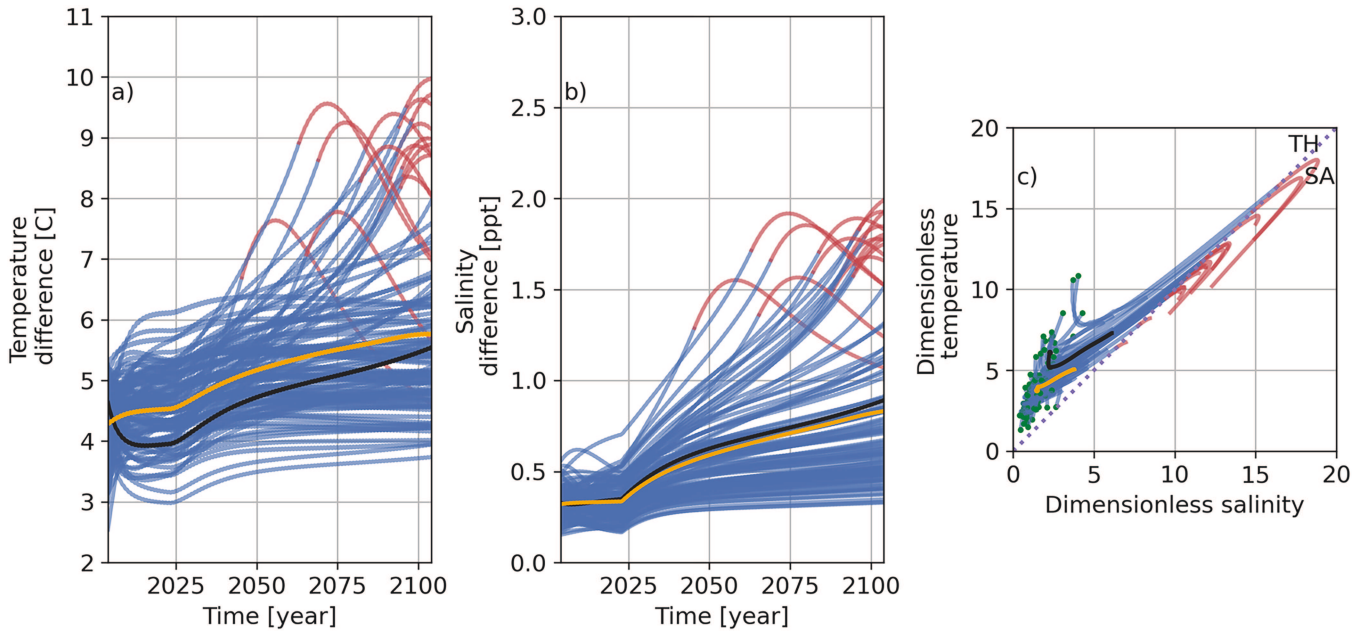

FIG. 7. (a) Temperature and (b) salinity time series, as well as (c) a phase portrait for an experiment with yearly warming of 0.03 °C per year in the equator (0.06 °C per year at the pole) and a melt period of 1000 years without DA. Trajectories turning red indicate predicted flipping of the flow direction. The black trajectories represent the synthetic truth values, and the orange represents the ensemble mode.

equatorial box is increased to 0.07 °C per year (the polar box warming rate is raised to 0.14 °C per year), and the melting period for the Greenland ice sheet is reduced from 10 000 to 1000 years. Once again, these climate change driven effects are only switched on in 2022; otherwise, the DA system would try to correct errors in the forcing by modifying the model parameters. In Fig. 7(b), the salinity between-box difference in all ensemble members increases due to the effects of climate change. This reduces the circulation and the temperature difference starts to increase as a consequence of the temperature difference present in the surface. Eventually, Figs. 7(a) and 7(b) show that depending on initial conditions and the model parameters of the ensemble member, there is a rapid increase in the temperature and the salinity between-box difference, resulting in a reversal of the AMOC circulation direction. Therefore, the setup is capable of producing solutions with a circulation change.

To summarize, in this section, we tested the ability of our setup to predict the truth using a twin experiment. We find that the setup produces an ensemble that covers the truth and can, with the sufficient forcing, produce solutions both with the TH as well as the SA circulation. DA is successful in reducing the spread and errors in both temperatures, salinities, diffusivity parameters, and the advection parameter. The 18-year DA period used is sufficient to reduce the uncertainty in temperature diffusivity and the advection parameter to the lowest value attainable with this setup, but the uncertainty in the salinity diffusivity could have been reduced further by extending the DA period. Having established that our DA setup has the ability to recover the model parameters from temperature and salinity observations, we will use those observations to find realistic values for these parameters in Sec. V.

V. ASSIMILATING BOX-AVERAGED OBSERVATIONS IN THE NON-AUTONOMOUS STOMMEL MODEL

In this section, we run simulations using (12) while assimilating box-average observations compiled from the EN4 dataset as described in Sec. III B. In this section, we pursue a twofold aim. First, we attempt to find realistic values for the salinity diffusivity k_S, temperature diffusivity k_T, and advection parameter γ as well as the non-dimensional values η_1, η_2, and η_3, which depend on these parameters. Second, having found realistic values, we continue running the box model forward in time to determine the likelihood of a circulation reversal happening in this century both for the case with climate change and the case without.

A. Interfered model parameters

The three different model parameters are shown in Fig. 8 as a function of time. As the effects of climate change are never applied during the DA period (see Sec. IV), the figure will look the same for both experiments with and without climate change. Notice that we are still assuming that within the conceptual constraints of the Stommel box model, the true value of these parameters is fixed in time. The time dependence reflects that the probability distribution for the truth is changing as more and more observations get assimilated. Had we used the Rauch–Tung–Striebel smoother,[58] we would

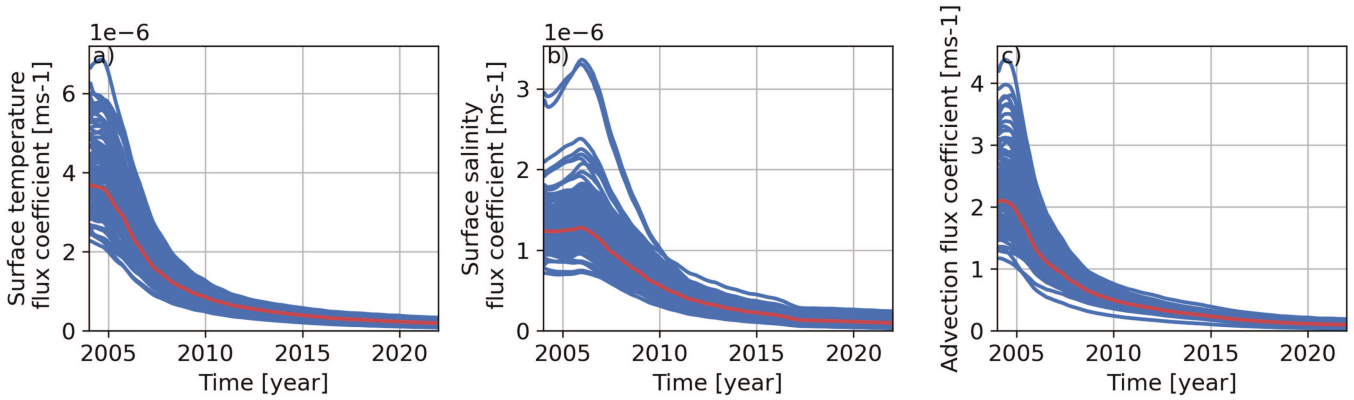

FIG. 8. (a) Surface temperature (κ_T), (b) surface salinity diffusivity (κ_S), and (c) the advection flux coefficients (γ) for the different ensemble members as functions of time. The red curves represent the ensemble mode.

have been able to correct the parameters in the beginning of the DA window using observations at the end of the window. However, as this would have not changed the final estimate, i.e., the value for the model parameters at the end of the DA window used to propagate the ensemble members forward, we have abstained from the additional effort. The figure shows an initial overestimation of the value of all three coefficients: the mode of the ensemble and all ensemble members exhibit a decreasing trend that starts to level off near the end of the DA period. Similar to the twin experiment in Sec. IV, the salinity diffusion parameter is slower to converge to its final value. The ensembles for all three parameters show a steadily decreasing spread indicating that the uncertainty of the parameter's true values decreases. The final estimates for the parameters, i.e., their values at the end of the DA period, together with their 90% confidence interval based on the ensemble spread, are given in Table VI.

The values for the non-dimensional parameters η_1, η_2, and η_3 as appearing in Eq. (4) are shown in Fig. 9. The annual oscillations clearly visible in η_1 and η_2 are the result of a seasonal cycle imposed on the surface forcing as Eq. (15). As climate change warms the polar box faster than the equatorial one, the temperature difference between the two decreases over time resulting in a decreasing slope in η_1 after the DA period. The annually averaged values for the ensemble mode of η_1, η_2, and η_3 at the end of the DA period are 9.2, 4.2, and 0.5, respectively. These are about double the values explored in other Stommel box model literature.[30,59] The absence of certain physical relevant processes for the AMOC and might explain a difference. See also the discussion of transport at the end of this section.

TABLE VI. Most-likely estimates of model parameters after assimilating monthly boxed-averaged temperatures and salinities for the period 2004–2022 together with their 90%-confidence interval.

Parameter	Most-likely	90%-confidence interval
k_T (10^{-7} ms^{-1})	1.8	[1.2, 2.9]
k_S (10^{-7} ms^{-1})	0.8	[0.5, 2.1]
γ (10^{-2} ms^{-1})	8.5	[6.1, 13.9]

VI. FUTURE CIRCULATION

The differences between the forecast temperature (forecast salinity) in the equatorial box and the polar box for the 100 ensemble members for the experiment without climate change are shown in Fig. 10. It can be seen that during the DA period (first 18 years), both temperature and salinity differences follow the trend in the observations. However, they fail to reproduce the multiyear oscillation that is visible in the salinity and especially in the temperature observations. This is because such multiyear oscillations are missing in the surface forcing. This results in the model overestimating the salinity difference between equatorial and polar box, thus underestimating the transport between the boxes. As explained in Sec. II A, the autonomous solutions converge to one of at most two possible stable equilibria each associated with a different circulation regime.

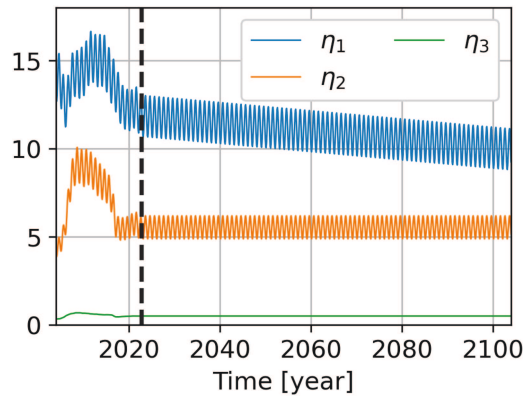

FIG. 9. The values of parameters η_1 [Eq. (10)], η_2 [Eq. (11)], and η_3 [Eq. (7)] as a function of time for the experiment with climate change.

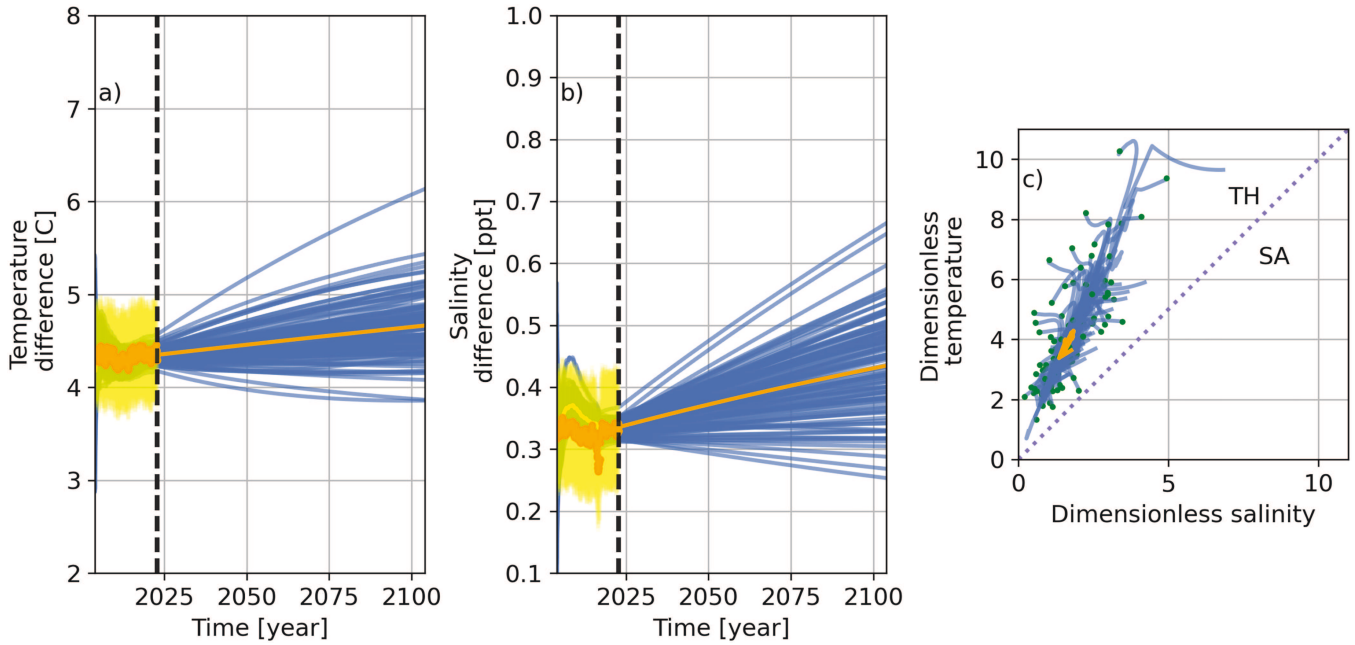

FIG. 10. Forecast time series starting in 2004 representing the equator-pole difference in the (a) temperature and (b) salinity for the different ensemble members as well as (c) the ensemble trajectories in a dimensionless temperature vs a salinity phase portrait. The orange dots with bright yellow errorbars represent the observed values and their standard deviation. Solutions with TH circulation are shown in blue and those with SA circulation (if any) in red.

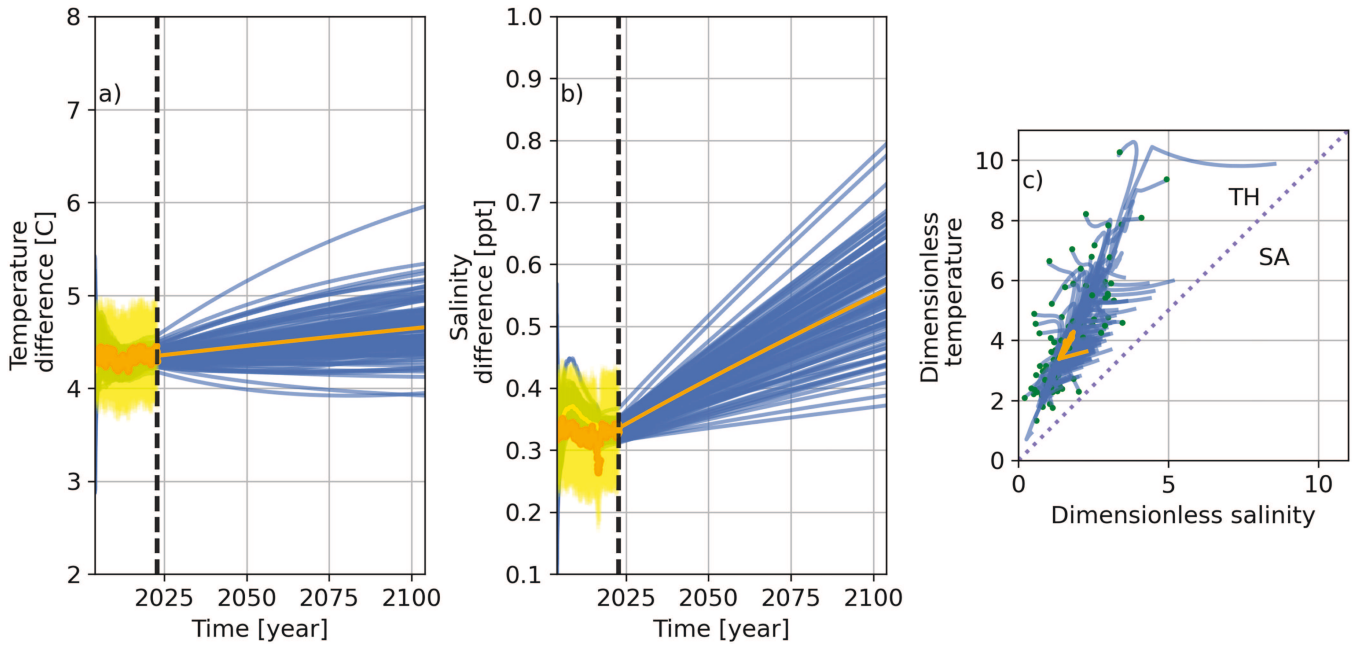

FIG. 11. The same as Fig. 10. Forecast time series starting in 2004 representing the equator-pole difference in (a) temperature and (b) salinity for the different ensemble members as well as (c) the ensemble trajectories in a dimensionless temperature vs a salinity phase portrait, but now climate change forcing is applied outside the assimilation period. The orange dots with bright yellow errorbars represent the observed values and their standard deviation. Solutions with TH circulation are shown in blue and those with SA circulation (if any) in red. The end of the DA period is marked by a vertical dashed line.

In the non-autonomous system, these equilibrium values will oscillate in time and the solution always "chases" one of them or switches from chasing one to chasing the other. The figure shows that in the absence of climate change, the circulation remains in the TH regime for all ensemble members until at least 2104. Furthermore, the solutions tend to values that are close to current conditions: in the Stommel box model, the solutions for the temperature differences lie above the difference at the end of the DA period. However, the differences are still within the range of contemporary temperature differences. The solutions for salinity differences are larger or smaller at the end of the DA period than in 2104, partially offsetting the decreasing trend over the DA period. The relative spread at the end of the century is also larger for the salinity differences than for the temperature differences. This can be explained by our observation in Sec. IV that the spread in the salinity diffusion parameter is larger than that in the temperature diffusivity parameter. The values depend on the model parameters, and consequently, the uncertainty for future ocean salinities is relatively larger than that in ocean temperatures.

The solutions in which after the DA period, the polar surface temperature increases with $0.06\,^{\circ}\mathrm{C\,y^{-1}}$, the equatorial surface temperature with $0.03\,^{\circ}\mathrm{C\,y^{-1}}$, and in which the Greenland ice sheet melts with a constant rate over 10 000 years are shown in Fig. 11. Compared to Fig. 10(b), the salinity differences increase after the DA period [Fig. 11(b)]. This is due to the ice melt. Next to this, the temperature difference increase is smaller due to the differential increase in the surface temperature between the two boxes [Fig. 11(a)]. Apart from these points, there are no qualitative differences between the figures. In particular, none of the ensemble members changes the circulation regime.

The actual future rate of surface warming and ice melt will depend on, among other things, future carbon emission and is consequently uncertain. Therefore, the DA experiment has been repeated several times for a range of plausible ice melting and between-box surface warming rate differences.[47–49,52,60] The percentage of ensemble members in which the circulation regime changes from TH to SA are shown in Fig. 12. The figure shows that after DA, the uncertainty in the regime is small: either all ensemble members are in the TH regime or in the SA regime. Next to this, the close-to-vertical slope of the boundary between the two regimes indicates that faster ice melting is more influential in pushing the solutions to the SA regime than faster warming over the pole than equator. Measured melt rates over the last decade vary between 286 and 487 $\mathrm{Gt\,y^{-1}}$, which corresponds to a melt period of 7000–12 000 years.[47–49] This would place the future climate firmly within the zone of Fig. 12 in which circulation, the AMOC does not reverse direction.

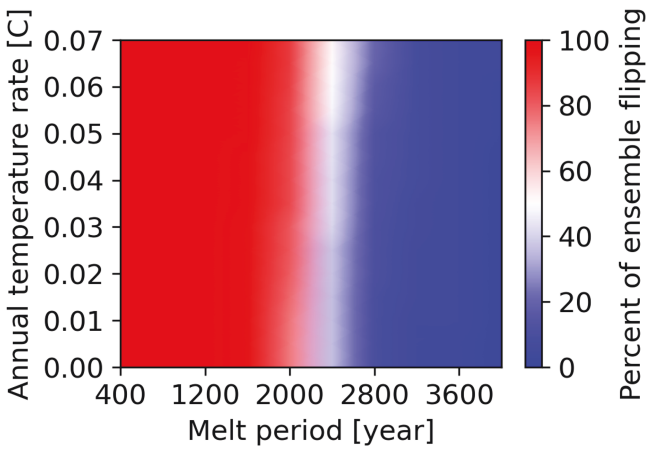

FIG. 12. Percentage of ensemble members that undergo a circulation regime change. All runs use a similar setup to the experiment with climate change and real data assimilation, but using different ice melt periods (x axis) and different warmings rates for the equator box (y axis). The warming rate in the polar box is twice that of the equator box. Bright red denotes mostly flipping, while blue represents mostly remaining in the starting TH regime.

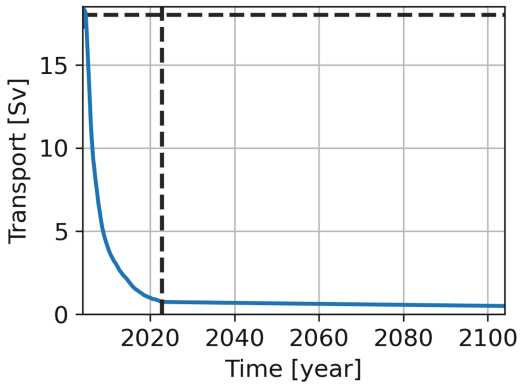

FIG. 13. The value of the between-box transport as a function of time for the experiment with climate change. The vertical dashed line marks the end of the assimilation period. The horizontal dashed line indicates observed current transport.[46]

The transport between the two boxes is shown in Fig. 13. Transports are an order of magnitude smaller than the typical observational transect estimates over this period that lie in the range 10–25 Sv.[61] The decrease is, however, not the result of climate change, but occurs during the DA period; i.e., the DA is not capable of simultaneously correcting temperatures, salinities, and transports. The source of this discrepancy might lie in our box definition. This will be discussed in more detail in Sec. VII. The underestimation of current transports implies that the model is closer to the tipping point than the real AMOC. Consequently, the ice melt and warming rates in Fig. 12 at which the circulation flips are in all likelihood an overestimation.

VII. CONCLUSION AND DISCUSSIONS

We have applied the Stommel 2-box model to replicate the Atlantic meridional overturning circulation and investigated its solutions, the feasibility of retrieving the temperature diffusivity, salinity diffusivity, and the advection parameter by assimilating ocean temperature and salinity parameters and to forecast the likelihood of a circulation regime change in the 21st century.

We introduced seasonal variability and two of the effects of climate change into the model: warming of the ocean surface at different rates for the pole and the equator and the melt of the Greenland ice sheet. Qualitatively, the Stommel box model (4) mirrors certain aspects of the non-autonomous Stommel box model (12) close to the NSF (by setting $A = B = 0$ in the non-autonomous case). The former presents a relatively generic and simple model with a region of bistability of two stable states that lose stability via a smooth saddle-node bifurcation, while the latter undergoes an NSF bifurcation. The non-autonomous Stommel box model presented provides a framework in which we can develop critical (tipping) points by introducing high-frequency oscillatory forcing.

Using the setup, we tested whether it would be possible to reconstruct the true value of the model parameters by assimilating ocean temperature and salinity observation and use those parameters to forecast the future circulation regime of the AMOC. Our twin experiments showed that the DA method is indeed capable of reconstructing temperature diffusivity and the advection parameters from observations. It also brings the salinity diffusivity closer to the truth, but the 18-year DA period used is too short to convergence to the minimal obtainable error. Using observed temperature and salinity from the EN4 dataset, the most-likely values of the model parameters have been determined with the DA system, indicating uncertainties of the same order of magnitude. In the absence of climate change, future temperature and salinity are expected to lie in the same range of temperatures and salinities that have already been observed in the period 2004–2022. With the effects of climate, salinity differences between the equatorial and pole box can move outside the current range because of the freshening caused by the melting Greenland ice sheet. Running the model with different climate change scenarios showed that this freshening is relatively more important to the direction of the AMOC circulation than changes in the surface temperature. In particular, with the currently observed ice sheet melting rates, the model does not expect a reversal of the circulation direction to occur.

Though an interesting proof-of-concept that data assimilation can help with reconstructing Stommel model parameters, the forecasts produced in this study should be interpreted with some care. The simplicity of the Stommel box model means that it cannot represent the wide range of processes occurring in the North Atlantic nor can it represent the North Atlantic geometry. In particular, we have seen that our model underestimates the transports. In other box-based studies,[41–43,62] the question of the transport is avoided by fitting temperature and salinity in boxes to a GCM with realistic levels for the AMOC transport. Consequently, deficiencies in the dynamics of the Stommel model do not impact the parameter fit. In other multibox studies, comparison with observed values is not even attempted.[44] In Kuhlbrodt *et al.*[39] and Gnanadesikan,[63] it has been argued that not the buoyancy forcing, but wind stresses are the primary driver behind the AMOC. Indeed, Cimatoribus *et al.*[12] have shown that the transport in a box-model increases strongly with increasing wind stresses. The fact such wind forcing is entirely absent in the Stommel model might explain why our transport estimates are an order of magnitude smaller than those observed. The omission of the southern Atlantic ocean in our setup might be another contributing factor to the errors in AMOC transports. Our two-box model considers the North Atlantic to be a closed system and does not account for the southward transport of North Atlantic Deep Water past the equator toward the Antarctic Circumpolar Current nor for the northward flow of Antarctic Bottom Water and waters from the Angulhas current (Sec. 14.2.14 in Ref. 1), which can increase transports by creating diapycnal mixing.[64] Accommodating these concepts requires more complex conceptual models, such as the Rooth model[65] or full climate models. Finally, we noticed that the volume of the boxes can have profound impact on the transports, and this dependence deserves more attention in future box-model studies.

As a last note, the simplicity of the chosen reduced-order model makes our analysis possible as we are able to initialize large ensembles at low computational costs. This setup can serve as a baseline in complexity with more detailed versions of the model available for further studies on higher order dynamics.

ACKNOWLEDGMENTS

We would like to thank John Gemmer and Christopher K. R. T. Jones for starting and funding this research project during the 2023 MCRN Summer School and Research Program under the NSF grant (No. DMS-1722578). Ivo Pasmans is supported by Schmidt Sciences LLC, grant G-24-66154, as part of the Scale-Aware Sea Ice Project (SASIP). Emmanuel Fleurantin was supported by the NSF grant (No. DMS-2137947). A part of this project originated from the Mason Experimental Geometry Lab (MEGL) at George Mason University. A special thanks goes out to those MEGL students who started with the twin experiment setup in the Fall 2022 semester.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Nathaniel Smith: Formal analysis (equal); Investigation (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Anvaya Shiney-Ajay:** Formal analysis (equal); Investigation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Emmanuel Fleurantin:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Ivo Pasmans:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Supervision (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The MetOffice EN4 objective analysis temperature and salinity data with XBT and MBT corrections used to create observations for assimilation as well as the surface forcing are available at https://hadleyserver.metoffice.gov.uk/en4/download-en4-2-2.html as the *EN.4.2.2.analyses.g10.XXXX.zip* files. The *hadley_*

obs.py script used to process these observations can be found in the *inversion* branch of the GIT repository.[68] Also included in this fork of DAPPER[31] are the code for the Stommel model (*dapper/mods/stommel/_init_.py*) as well as the scripts used to carry out the data assimilation (*dapper/mods/stommel/experiment.py*) using these observations and the scripts for the twin experiments.

APPENDIX A: LACK OF PERIODIC SOLUTIONS IN THE STOMMEL BOX MODEL

In this Appendix, we go over why we can rule out the existence of periodic solutions in the autonomous Stommel box model. Doing so guarantees that in the non-autonomous model, when we do the twin experiments and begin assimilating real data, the only oscillatory behavior will come from the periodic forcing. The Poincaré–Bendixson theorem asserts that in a continuous dynamical system, any confined solution will inevitably approach either a fixed point or a limit cycle. In the next few paragraphs, we will show that we get the former, not the latter in Eq. (4), where we take $0 < \eta_3 < 1$, $\eta_1 > \frac{\eta_3}{1-\eta_3}$, with $T, S \geq 0$.

To start, we have that

$$\frac{dT}{dt} = f(T,S) = \eta_1 - T - T|T-S|,$$
$$\frac{dS}{dt} = g(T,S) = \eta_2 - \eta_3 S - S|T-S|.$$

We define our Dulac function $\phi(T,S) = 1$. Applying Dulac's criterion, we get

$$\frac{\partial(\phi f)}{\partial T} + \frac{\partial(\phi g)}{\partial S} = \frac{\partial f}{\partial T} + \frac{\partial g}{\partial S} = -1 - \eta_3 - 3|T-S|. \tag{A1}$$

This is always strictly negative, and so by the Bendixson–Dulac theorem, we get that there cannot be any closed trajectories either entirely where $T > S > 0$, as well as in the area where $S > T > 0$. However, there could be a closed trajectory that crosses between the two areas, over the line $T = S$.

Now, assume we did have such a trajectory, a periodic solution C that enclosed a region D within $T, S \geq 0$, and let us call the part of D where $T > S$, D_1, and the part where $S < T$ as D_2 so that $\partial D_1 = C_1 \cup C_3$ and $\partial D_2 = C_2 \cup (-C_3)$.

As Eq. (A1) allows for a continuous extension from D_1 (D_2) to its compact closure $\overline{D}_1$ ($\overline{D}_2$), $\int\int_{\overline{D}_1}\left(\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}\right) dT\,dS$ $\left(\int\int_{\overline{D}_2}\left(\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}\right) dT\,dS\right)$ exists and is, as shown above, strictly smaller than zero. Applying Green's theorem now gives

$$\begin{aligned}
&\iint_D \left(\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}\right) dT\,dS \\
&= \iint_{D_1}\left(\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}\right) dT\,dS + \iint_{D_2}\left(\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}\right) dT\,dS \\
&= \int_{C_1 \cup C_3} (-g\,dT + f\,dS) + \int_{C_2 \cup (-C_3)} (-g\,dT + f\,dS) \\
&= \int_{C_1} (-g\,dT + f\,dS) + \int_{C_3} (-g\,dT + f\,dS) \\
&\quad + \int_{C_2} (-g\,dT + f\,dS) + \int_{(-C_3)} (-g\,dT + f\,dS).
\end{aligned}$$

However, rearranging this, the second and fourth terms simply cancel out. Furthermore, by definition, we have that $\frac{dT}{dt} = f(T,S)$ and $\frac{dS}{dt} = g(T,S)$; therefore, we get

$$\begin{aligned}
\int\int_D \left(\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}\right) dT\,dS &= \int_{C_1} (-g\,dT + f\,dS) + \int_{C_2} (-g\,dT + f\,dS) \\
&= \int_C (-g\,dT + f\,dS) \\
&= \int_C \left(-g\frac{dT}{dt} + f\frac{dS}{dt}\right) dt \\
&= \int_C \left(-\frac{dS}{dt}\frac{dT}{dt} + \frac{dT}{dt}\frac{dS}{dt}\right) dt \\
&= \int_C 0\,dt = 0.
\end{aligned}$$

However, earlier, we showed that this was less than zero; therefore, we have a contradiction. Consequently, no such loop C can exist.

Thus, in the autonomous Stommel box model, it is impossible for a periodic solution to exist for $T, S \geq 0$.

APPENDIX B: ENSEMBLE KALMAN FILTER

Let

$$\mathbf{x}_t = \left[T_{e*}(t), T_{p*}(t), S_{e*}(t), S_{p*}(t), \log\frac{\kappa_T}{\kappa_{ref}}, \log\frac{\kappa_S}{\kappa_{ref}}, \log\frac{\gamma}{\kappa_{ref}}\right]^{\mathsf{T}}$$

represent a possible state of the system at time t with $\kappa_{ref} = 1\ \mathrm{m^2\ s^{-1}}$ being reference values to nondimensionalize the diffusion and advection coefficients. As this state does not only include the state variables $[T_{e*}(t), T_{p*}(t), S_{e*}(t), S_{p*}(t)]$ but also the model parameters κ_T, κ_S, and γ, it is referred to as the augmented state. Let $\int_{\mathcal{X}} p(\mathbf{x}_0)\,\mathrm{d}\mathbf{x}_0$ be the probability that the true state of the system at the beginning of the model run lies in the set $\mathcal{X}$. Using Bayes' theorem and the chain rule for probabilities, it can be shown[66] that

$$\underbrace{p(\mathbf{x}_t|\mathbf{d}_{0:t})}_{posteriori} \sim p(\mathbf{d}_t|\mathbf{x}_t) \underbrace{\int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{d}_{0:t-1})\,\mathrm{d}\mathbf{x}_{t-1}}_{apriori}, \tag{B1}$$

where $\sim$ indicates equivalence up to a normalization constant, d_t a vector with observations at time step t, and $d_{0:t}$ a vector of all observations up to and including time step t. Once the *a priori* is specified for the initial time step $t = 0$, Eq. (B1) for *a posteriori* and *a priori* distributions for $t > 0$ can be constructed by induction using Eq. (B1). This occurs in two steps. The *prediction step* in which *a posteriori* distribution at time $t-1$ is transformed into the *a priori*

distribution at time t, and the *update step* in which the DA correction is applied to transform the *a priori* distribution at time t into *a posteriori* distribution at the same time.

In an ensemble Kalman filter, it is assumed that the *a priori* and *a posteriori* distributions can be approximated by a Gaussian with its mean $\mu_{x,t} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{x}_t^{(m)}$ and covariance $\Sigma_x = \frac{1}{M-1}\sum_{m=1}^{M}\left(\mathbf{x}_t^{(m)} - \mu_{x,t}\right)\left(\mathbf{x}_t^{(m)} - \mu_t\right)^{\mathsf{T}}$ estimated from an M-member ensemble of model runs $\left\{\mathbf{x}_t^{(m)} : 1 \le m \le M\right\}$. Each ensemble member represents an equally likely realization of these probability distributions. The integral in the *a priori* distribution in Eq. (B1) is approximated using a Monte Carlo method. For each ensemble $\mathbf{x}_{t-1}^{(n)}$, an ensemble member $\mathbf{x}_t^{(m)}$ is sampled according to $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. In this work, the model is assumed to be without a model error, and hence, $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \delta\left(\mathbf{x}_t - \mathcal{M}_{t-1\to t}\mathbf{x}_{t-1}\right)$ with δ being the delta function and $\mathcal{M}_{t\to t-1}$ being the model operator that propagates the ocean temperatures and salinities from time $t-1$ to t according to Eqs. (1)–(3). $\mathcal{M}_{t-1\to t}$ while keeping the model parameters constant; i.e., model parameters are only changed during the *update step*.

The probability density distribution for the observations $p(\mathbf{d}_t|\mathbf{x}_t)$ is assumed to be a Gaussian with mean $\mu_d = H\mu_{t,x}$ and covariance Σ_d. In our case, the observation operator H is such that $H\mathbf{x}_t = [T_{p*}(t), T_{e*}(t), S_{p*}(t), S_{e*}(t)]^{\mathsf{T}}$ and Σ_d is assumed to be diagonal with its diagonal elements (variances) given by Eq. (14).

It can be shown that under these assumptions, *a posteriori* distribution is again a Gaussian with mean $\mu_{\hat{x},t} = \frac{1}{M}\sum_{m=1}^{M}\hat{\mathbf{x}}_t^{(m)}$ and covariance $\Sigma_{\hat{x},t} = \frac{1}{M-1}\left(\hat{\mathbf{x}}_t^{(m)} - \mu_{\hat{x},t}\right)\left(\hat{\mathbf{x}}_t^{(m)} - \mu_{\hat{x},t}\right)^{\mathsf{T}}$ and

$$\hat{\mathbf{x}}_t^{(n)} = \mathbf{x}_t^{(n)} + K\left(\mathbf{d}_t - H\mu_{t,x}\right) + AQ\Lambda^{\frac{1}{2}}Q^{\mathsf{T}}\hat{\mathbf{e}}^{(n)}, \tag{B2}$$

with $\hat{\mathbf{e}}^{(n)}$ being the nth unit vector, A a matrix having $\mathbf{x}_t^{(n)} - \mu_{t,x}$ as its nth column, and $Q\Lambda Q^{\mathsf{T}}$ the singular value decomposition of $I - A^{\mathsf{T}}H^{\mathsf{T}}\left((m-1)\Sigma_d + HAA^{\mathsf{T}}H^{\mathsf{T}}\right)^{-1}HA$[66] and $K = AA^{\mathsf{T}}\left(HAA^{\mathsf{T}}H^{\mathsf{T}} + (M-1)\Sigma_d\right)^{-1}$ the Kalman gain.

The most-likely state, or mode, of the *a priori* and *a posteriori* distributions are referred to as the forecast and analysis, respectively. As the ocean temperatures and salinities are assumed to follow Gaussian distributions, their forecasts and analysis coincide with their ensemble mean. For the advection and diffusion parameters, however, we corrected their logarithms instead of the values themselves. This so-called log-transform ensures that the parameters themselves remain strictly positive. For a log-normal distribution, the mode is not given by e^{μ_p} as naïvely might be expected, but by $e^{\mu_p - \Sigma_p\mathbf{1}}$ with μ_p the three-dimensional vector containing the ensemble mean of the logarithms of the model parameters and Σ_p the 3×3 matrix with their ensemble covariance;[67] i.e., μ_p and Σ_p are the last three elements or a 3×3-block of the μ_x and Σ_x, respectively.

REFERENCES

[1] L. Talley, G. Pickard, W. Emery, and J. Swift, *Descriptive Physical Oceanography*, 6th ed. (Elsevier, London, 2011).
[2] H. Stommel, "Thermohaline convection with two stable regimes of flow," Tellus **13**, 224–230 (1961).
[3] F. Bryan, "High-latitude salinity effects and interhemispheric thermohaline circulations," Nature **323**, 301–304 (1986).
[4] J. Marotzke, P. Welander, and J. Willebrand, "Instability and multiple steady states in a meridional-plane model of the thermohaline circulation," Tellus A: Dyn. Meteorol. Oceanogr. **40**, 162 (1988).
[5] E. Maier-Reimer and U. Mikolajewicz, "Experiments with an OGCM on the cause of the younger dryas," Technical Report No. 39, Max-Planck-Institut fuer Meteorologie, Hamburg, Germany, 1989.
[6] V. Dakos, M. Scheffer, E. H. van Nes, V. Brovkin, V. Petoukhov, and H. Held, "Slowing down as an early warning signal for abrupt climate change," Proc. Natl. Acad. Sci. U.S.A. **105**, 14308–14312 (2008).
[7] P. D. Ditlevsen and S. J. Johnsen, "Tipping points: Early warning and wishful thinking," Geophys. Res. Lett. **37**, L19703, https://doi.org/10.1029/2010GL044486 (2010).
[8] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara, "Early-warning signals for critical transitions," Nature **461**, 53–59 (2009).
[9] W. Weijer, W. Cheng, S. S. Drijfhout, A. V. Fedorov, A. Hu, L. C. Jackson, W. Liu, E. L. McDonagh, J. V. Mecking, and J. Zhang, "Stability of the Atlantic meridional overturning circulation: A review and synthesis," J. Geophys. Res.: Oceans **124**, 5336–5375, https://doi.org/10.1029/2019JC015083 (2019).
[10] B. J. H. V. de Wiel, E. Vignon, P. Baas, I. G. S. van Hooijdonk, S. J. A. van der Linden, J. A. van Hooft, F. C. Bosveld, S. R. de Roode, A. F. Moene, and C. Genthon, "Regime transitions in near-surface temperature inversions: A conceptual model," J. Atmos. Sci. **74**, 1057–1073 (2017).
[11] I. M. Otto, J. F. Donges, R. Cremades, A. Bhowmik, R. J. Hewitt, W. Lucht, J. Rockström, F. Allerberger, M. McCaffrey, S. S. P. Doe, A. Lenferna, N. Morán, D. P. van Vuuren, and H. J. Schellnhuber, "Social tipping dynamics for stabilizing Earth's climate by 2050," Proc. Natl. Acad. Sci. U.S.A. **117**, 2354–2365 (2020).
[12] A. A. Cimatoribus, S. S. Drijfhout, and H. A. Dijkstra, "Meridional overturning circulation: Stability and ocean feedbacks in a box model," Clim. Dyn. **42**, 311–328 (2014).
[13] D. Castellana and H. A. Dijkstra, "Noise-induced transitions of the Atlantic meridional overturning circulation in CMIP5 models," Sci. Rep. **10**, 20040 (2020).
[14] D. Ryvkine and M. I. Dykman, "Pathways of activated escape in periodically modulated systems," Phys. Rev. E **73**, 061109 (2006).
[15] P. E. O'Keeffe and S. Wieczorek, "Tipping phenomena and points of no return in ecosystems: Beyond classical bifurcations," SIAM J. Appl. Dyn. Syst. **19**, 2371–2402 (2020).
[16] X. Zhou and W. E, "Study of noise-induced transitions in the Lorenz system using the minimum action method," Commun. Math. Sci. **8**, 341–355 (2010).
[17] J. Lohmann, D. Castellana, P. D. Ditlevsen, and H. A. Dijkstra, "Abrupt climate change as a rate-dependent cascading tipping point," Earth Syst. Dyn. **12**, 819–835 (2021).
[18] P. Ditlevsen and S. Ditlevsen, "Warning of a forthcoming collapse of the Atlantic meridional overturning circulation," Nat. Commun. **14**, 4254 (2023).
[19] R. M. van Westen, M. Kliphuis, and H. A. Dijkstra, "Physics-based early warning signal shows that AMOC is on tipping course," Sci. Adv. **10**, eadk1189 (2024).
[20] M. Ben-Yami, A. Morr, S. Bathiany, and N. Boers, "Uncertainties too large to predict tipping times of major Earth system components," arXiv:2309.08521 (2023).
[21] J. Sherman, C. Sampson, E. Fleurantin, Z. Wu, and C. K. Jones, "A data-driven study of the drivers of stratospheric circulation via reduced order modeling and data assimilation," Meteorology **3**, 1–35 (2023).
[22] P. Smith, G. Thornhill, S. Dance, A. Lawless, D. Mason, and N. Nichols, "Data assimilation for state and parameter estimation: Application to morphodynamic modelling," Q. J. R. Meteorol. Soc. **139**, 314–327 (2013).
[23] F. Sévellec and A. V. Fedorov, "Unstable AMOC during glacial intervals and millennial variability: The role of mean sea ice extent," Earth Planet. Sci. Lett. **429**, 60–68 (2015).
[24] E. Hawkins, R. S. Smith, L. C. Allison, J. M. Gregory, T. J. Woollings, H. Pohlmann, and B. De Cuevas, "Bistability of the Atlantic overturning circulation in a global climate model and links to ocean freshwater transport," Geophys. Res. Lett. **38**, L10605, http://dx.doi.org/10.1029/2011GL047208 (2011).
[25] L. Jackson, E. Alastrué de Asenjo, K. Bellomo, G. Danabasoglu, H. Haak, A. Hu, J. Jungclaus, W. Lee, V. Meccia, O. Saenko *et al.*, "Understanding amoc stability:

The North Atlantic hosing model intercomparison project," Geosci. Model Dev. **16**, 1975–1995 (2023).
[26]H. A. Dijkstra and M. Ghil, "Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach," Rev. Geophys. **43**, RG3002, https://doi.org/10.1029/2002RG000122 (2005).
[27]R. Chapman, S. Sinet, and P. D. L. Ritchie, "Tipping mechanisms in a conceptual model of the Atlantic meridional overturning circulation," Weather **79**, 316 (2024).
[28]N. Boers, "Observation-based early-warning signals for a collapse of the Atlantic meridional overturning circulation," Nat. Clim. Change **11**, 680–688 (2021).
[29]M. M. Dekker, A. S. Von Der Heydt, and H. A. Dijkstra, "Cascading transitions in the climate system," Earth Syst. Dyn. **9**, 1243–1260 (2018).
[30]C. Budd, C. Griffith, and R. Kuske, "Dynamic tipping in the non-smooth Stommel-box model, with fast oscillatory forcing," Physica D **432**, 132948 (2022).
[31]P. N. Raanes, Y. Chen, C. Grudzien, M. Tondeur, and R. Dubois, "DAPPER," Nansen Environmental and Remote Sensing Center, 2023.
[32]C. H. Bishop, B. J. Etherton, and S. J. Majumdar, "Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects," Mon. Weather Rev. **129**, 420–436 (2001).
[33]S. A. Good, M. J. Martin, and N. A. Rayner, "EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates," J. Geophys. Res.: Oceans **118**, 6704–6716, https://doi.org/10.1002/2013JC009067 (2013).
[34]V. Gouretski and F. Reseghetti, "On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database," Deep Sea Res. I: Oceanogr. Res. Pap. **57**, 812–833 (2010).
[35]C. W. Böning, F. O. Bryan, W. R. Holland, and R. Döscher, "Deep-water formation and meridional overturning in a high-resolution model of the North Atlantic," J. Phys. Oceanogr. **26**, 1142–1164 (1996).
[36]Argo, "Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)"; see https://doi.org/10.17882/42182.
[37]T. Petit, M. S. Lozier, S. A. Josey, and S. A. Cunningham, "Atlantic deep water formation occurs primarily in the Iceland basin and Irminger Sea by local buoyancy forcing," Geophys. Res. Lett. **47**, e2020GL091028, https://doi.org/10.1029/2020GL091028 (2020).
[38]D. A. Bailey, P. B. Rhines, and S. Häkkinen, "Formation and pathways of North Atlantic Deep Water in a coupled ice–ocean model of the Arctic–North Atlantic oceans," Clim. Dyn. **25**, 497–516 (2005).
[39]T. Kuhlbrodt, A. Griesel, M. Montoya, A. Levermann, M. Hofmann, and S. Rahmstorf, "On the driving processes of the Atlantic meridional overturning circulation," Rev. Geophys. **45**, RG2001, https://doi.org/10.1029/2004RG000166 (2007).
[40]M. S. Lozier, "Deconstructing the conveyor belt," Science **328**, 1507–1511 (2010).
[41]S. S. Drijfhout, S. L. Weber, and E. van der Swaluw, "The stability of the MOC as diagnosed from model projections for pre-industrial, present and future climates," Clim. Dyn. **37**, 1575–1586 (2011).
[42]R. A. Wood, J. M. Rodríguez, R. S. Smith, L. C. Jackson, and E. Hawkins, "Observable, low-order dynamical controls on thresholds of the Atlantic meridional overturning circulation," Clim. Dyn. **53**, 6815–6834 (2019).
[43]R. Chapman, P. Ashwin, J. Baker, and R. A. Wood, "Quantifying risk of a noise-induced AMOC collapse from northern and tropical Atlantic Ocean variability," arXiv:2405.10929 (2024).
[44]A. H. Monahan, "Stabilization of climate regimes by noise in a simple model of the thermohaline circulation," J. Phys. Oceanogr. **32**, 2072–2085 (2002).
[45]J. Nycander, M. Hieronymus, and F. Roquet, "The nonlinear equation of state of sea water and the global water mass distribution," Geophys. Res. Lett. **42**, 7714–7721, https://doi.org/10.1002/2015GL065525 (2015).
[46]G. McCarthy, D. Smeed, W. Johns, E. Frajka-Williams, B. Moat, D. Rayner, M. Baringer, C. Meinen, J. Collins, and H. Bryden, "Measuring the Atlantic meridional overturning circulation at 26 °N," Prog. Oceanogr. **130**, 91–111 (2015).
[47]S. A. Khan, A. Aschwanden, A. A. Bjørk, J. Wahr, K. K. Kjeldsen, and K. H. Kjær, "Greenland ice sheet mass balance: A review," Rep. Prog. Phys. **78**, 046801 (2015).
[48]J. Mouginot, E. Rignot, A. A. Bjørk, M. van den Broeke, R. Millan, M. Morlighem, B. Noël, B. Scheuchl, and M. Wood, "Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018," Proc. Natl. Acad. Sci. U.S.A. **116**, 9239–9244 (2019).
[49]K. D. Mankoff, A. Solgaard, W. Colgan, A. P. Ahlstrøm, S. A. Khan, and R. S. Fausto, "Greenland Ice Sheet solid ice discharge from 1986 through March 2020," Earth Syst. Sci. Data **12**, 1367–1383 (2020).
[50]See https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature-projections for "Climate Change: Global Temperature Projections" (last accessed November 21, 2022).
[51]See https://www.ipcc.ch/site/assets/uploads/2018/02/ar4_syr_full_report.pdf for "Climate Change 2007: Synthesis Report" (last accessed November 21, 2022).
[52]Q. You, Z. Cai, N. Pepin, D. Chen, B. Ahrens, Z. Jiang, F. Wu, S. Kang, R. Zhang, T. Wu, P. Wang, M. Li, Z. Zuo, Y. Gao, P. Zhai, and Y. Zhang, "Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences," Earth-Sci. Rev. **217**, 103625 (2021).
[53]J. Zhang, "Warming of the arctic ice-ocean system is faster than the global average since the 1960s," Geophys. Res. Lett. **32**, L19602, https://doi.org/10.1029/2005GL024216 (2005).
[54]J. E. Walsh, "Intensified warming of the arctic: Causes and impacts on middle latitudes," Global Planet. Change **117**, 52–63 (2014).
[55]J. Richter-Menge, M. L. Druckenmiller, J. K. Andersen, L. M. Andreassen, E. H. Baker, T. J. Ballinger, L. T. Berner, G. H. Bernhard, B. Wouters *et al.*, "State of the climate in 2019: The Arctic," Bull. Am. Meteorol. Soc. **101**, S239–S286 (2020).
[56]E. Jansen, J. H. Christensen, T. Dokken, K. H. Nisancioglu, B. M. Vinther, E. Capron, C. Guo, M. F. Jensen, P. L. Langen, R. A. Pedersen *et al.*, "Past perspectives on the present era of abrupt Arctic climate change," Nat. Clim. Change **10**, 714–721 (2020).
[57]G. R. Halliwell, A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. L. Hénaff, and R. Atlas, "Rigorous evaluation of a fraternal twin ocean OSSE system for the open Gulf of Mexico," J. Atmos. Oceanic Technol. **31**, 105–130 (2014).
[58]H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," AIAA J. **3**, 1445–1450 (1965).
[59]H. A. Dijkstra, *Dynamical Oceanography*, 1st ed. (Springer, Berlin, 2011).
[60]A. Xie, J. Zhu, S. Kang, X. Qin, B. Xu, and Y. Wang, "Polar amplification comparison among Earth's three poles under different socioeconomic scenarios from CMIP6 surface air temperature," Sci. Rep. **12**, 16548 (2022).
[61]G. Danabasoglu, F. S. Castruccio, R. J. Small, R. Tomas, E. Frajka-Williams, and M. Lankhorst, "Revisiting AMOC transport estimates from observations and models," Geophys. Res. Lett. **48**, e2021GL093045, https://doi.org/10.1029/2021GL093045 (2021).
[62]S. Rahmstorf, "On the freshwater forcing and transport of the Atlantic thermohaline circulation," Clim. Dyn. **12**, 799–811 (1996).
[63]A. Gnanadesikan, "A simple predictive model for the structure of the oceanic pycnocline," Science **283**, 2077–2079 (1999).
[64]L. Cimoli, A. Mashayek, H. L. Johnson, D. P. Marshall, A. C. Naveira Garabato, C. B. Whalen, C. Vic, C. de Lavergne, M. H. Alford, J. A. MacKinnon, and L. D. Talley, "Significance of diapycnal mixing within the Atlantic meridional overturning circulation," AGU Adv. **4**, e2022AV000800 (2023).
[65]C. Rooth, "Hydrology and ocean circulation," Prog. Oceanogr. **11**, 131–149 (1982).
[66]G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen, *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem* (Springer Nature, 2022).
[67]S. Fletcher, "Mixed Gaussian-lognormal four-dimensional data assimilation," Tellus A: Dyn. Meteorol. Oceanogr. **62**, 266–287 (2010).
[68]I. Pasmans, P. N. Raanes, Y. Chen, C. Grudzien, A. Perrin, F. Curic, and S. Alyaev, "ivopasmans-reading/DAPPER: StommelDapper1.2," Github repository, https://doi.org/10.5281/zenodo.13760128 (accessed 13 September 2024).