

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

Article

Accepted Version

Smith, N. ORCID: <https://orcid.org/0009-0003-9217-1164>, Shiney-Ajay, A. ORCID: <https://orcid.org/0009-0006-3090-5101>, Fleurantin, E. ORCID: <https://orcid.org/0000-0002-8640-6305> and Pasmans, I. ORCID: <https://orcid.org/0000-0001-5076-5421> (2024) Investigating ocean circulation dynamics through data assimilation: a mathematical study using the Stommel box model with rapid oscillatory forcings. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34 (10). 103131. ISSN 1089-7682 doi: <https://doi.org/10.1063/5.0215236> Available at <https://centaur.reading.ac.uk/119142/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1063/5.0215236>

Publisher: American Institute of Physics

including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Investigating Ocean Circulation Dynamics Through Data Assimilation: A Mathematical Study Using the Stommel Box Model with Rapid Oscillatory Forcings

Nathaniel Smith<sup>1</sup>, Anvaya Shiney-Ajay<sup>2</sup>, Emmanuel Fleurantin<sup>3</sup>, and Ivo Pasmans<sup>4</sup>

<sup>1</sup>Miami University, Oxford, OH 45056

<sup>2</sup>University of Minnesota, Minneapolis, MN 55455

<sup>3</sup>George Mason University, Fairfax, VA 22030

<sup>4</sup>University of Reading, Reading RG6 6AH, United Kingdom

\*email: smithni@miamioh.edu, ajayx006@umn.edu, efleuran@unc.edu, 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.

**Keywords:** Conceptual climate models, bistability, regime transition, data assimilation, box-averaged Argo observations, non-autonomous systems

**MSC Classification:** 86A08, 37N10, 65K99, 65P99

## 1 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] 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 a 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. [3, 4, 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 are 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, 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, 7, 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 [9, 10, 11, 12, 13, 14, 15, 16], has established thresholds for tipping occurrence. In [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 criticised [20]. The aim of this study is twofold. First, 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. Second, 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 like the strength of the forcing.

## **1.1 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 like the AMOC is motivated by several important considerations. Firstly, 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, 22, 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].

Lastly, 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 [26, 27, 12, 28, 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.

## 1.2 Outline of the Paper

The outline of the paper is as follows. First, we introduce the autonomous and non-autonomous Stommel model in Section 2 and study the qualitative behavior of the solutions as time goes to infinity. In Section 3, we explain our data assimilation method and the assimilated observations. In Section 4, we conduct a twin experiment to investigate the potential of data assimilation to estimate model parameters. In Section 5, 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 Section 6, 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 Section 7, we discuss our findings.

## 2 Mathematical Analysis of the Stommel box Model: Autonomous and Non-Autonomous Perspectives

### 2.1 The Stommel Box Model

The Stommel model comprises two fully mixed water containers. 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.

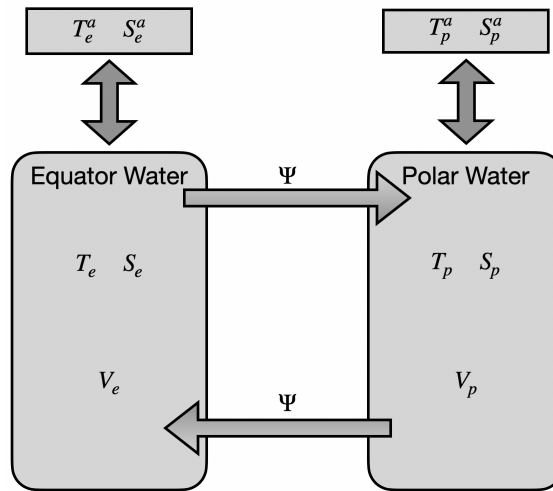


Figure 1: A conceptual depiction of how Henry Stommel modeled thermohaline circulation. The circulation between the two connected bodies of well-mixed water in the Stommel model is driven by density gradients between the two boxes.

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

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

with  $\Delta z$  and  $\Delta x$  being the depth and the zonal width of the boxes,  $\Delta y_p$  and  $\Delta y_e$  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 accomodates 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:

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

where  $\gamma$  is the advection coefficient and  $\rho_0$  is a reference density. Since those are both constants, the flow rate is linearly related to the difference in 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 reference value,  $\alpha_T$  is the thermal expansion coefficient and  $\alpha_S$  the haline coefficient.

But 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

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

Here,  $\eta_1$ ,  $\eta_2$ , and  $\eta_3$  are parameters independent of the ocean state. Their full values are as follow:

$$\eta_1 = \frac{\Delta z \gamma \alpha_T}{\Delta y k_T} (T_e^a - T_p^a) \quad (5)$$

$$\eta_2 = \frac{k_S}{k_T} \frac{\Delta z \gamma \alpha_S}{\Delta y k_T} (S_e^a - S_p^a) \quad (6)$$

$$\eta_3 = \frac{k_S}{k_T} \quad (7)$$

Setting  $\frac{dT}{dt} = 0$ ,  $\frac{dS}{dt} = 0$  in Equation (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

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

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 temperature difference. We find that

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

The solutions from Equation (8) and Equation (9) for fixed  $\eta_1$ ,  $\eta_3$  can be shown together as a function of the dimensionless transport. From this, it is clear that for a range of  $\eta_2$ , 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 the TH regime provides two solutions for certain values of  $\eta_2$  and the SA regime provides one.

## 2.2 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  $\eta_1$  and  $\eta_2$  are no longer constant in time but oscillate, i.e.

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

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

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

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

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 climate change, i.e., a system in which  $m^* = 0$ .

In order to get a full picture of the dynamics of Equation (12), it is thus fitting to analyze its  $\Psi$ -vs.- $\eta_{2,auto}$  bifurcation diagram. It's worth noting that the discontinuity surface  $\Sigma_0 = \{(T, S) \in \mathbb{R}_+^2 | T = S\}$  give rise to a Non-Smooth Fold (NSF) bifurcation (marked by a black star in Figure 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 insight into the dynamical impact of NSF bifurcations.

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 (Equation (12)), under high frequency forcing, we observe amplification or attenuation of patterns in  $\psi$  which depends linearly on the amplitude-to-frequency ratio ( $A/\Omega$ ). That is to say, solutions starting near the NSF bifurcation point may naturally shift away towards higher values of  $\eta_2$ , or be driven towards lower values under the effect of data assimilation. The foundational work for this analysis can be found in a paper by Budd, Griffith, and Kuske [30]. Their study provides valuable insights into the expected behaviors of the parameters  $\eta_1$  and  $\eta_2$ , 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.

### 3 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  $\gamma$ . 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 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].

#### 3.1 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. 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



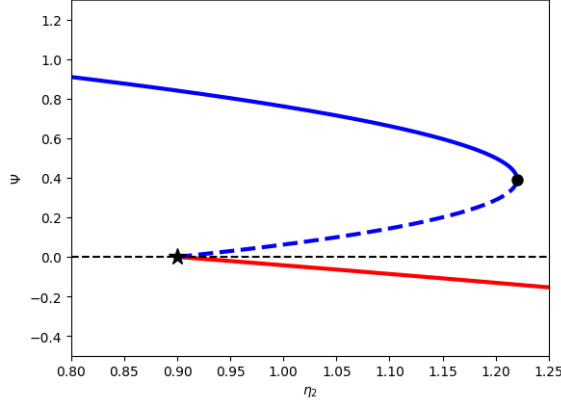


Figure 2: A  $\Psi$ -vs.- $\eta_{2,auto}$  bifurcation diagram for the system in equation 12 with  $\eta_{1,auto} = 3$ ,  $\eta_3 = 0.1$ , and  $A=B=0$ . The solid red and blue lines are stable equilibria, while the dotted blue line represents an unstable equilibrium. The solid black circle indicates the system's saddle-node bifurcation, while the black star indicates the NSF bifurcation. The dotted black line emphasizes the discontinuity set  $\Sigma_0$ .

imperfect observations (see Section 3.2) to produce an 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 a priori 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.

### 3.2 Observations and Model Parameters

Geometry of the two boxes in the model, box-average temperature and salinity, 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 1x1 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 such 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 to be 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 a mechanism to release heat from the ocean is present. This choice is not entirely unproblematic (see section 7) and deviates from the one used by [41, 42, 43]. Their boxes extend to  $23^\circ S$ , cover the whole Arctic ocean, not only the Nordic seas. Furthermore, they simply define the equatorial box to be the part of the Atlantic Ocean between  $30^\circ S$  and  $40^\circ N$  as they base deep-ocean temperatures and salinities on fit a to 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 [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.

	Polar	Equator
$\Delta x$ [km]	7274	7274
$\Delta y$ [km]	324	2262
$\Delta z$ [m]	3148	3148

Table 1: The zonal width, meridional width and depth of the two boxes used in the Stommel model based the EN4 data.

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 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 Figure 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 1.

	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 2: Time-average of the assimilated temperature and salinity observations together with the observational standard deviation, i.e. the square-roots of the diagonal of  $\Sigma_d$ .

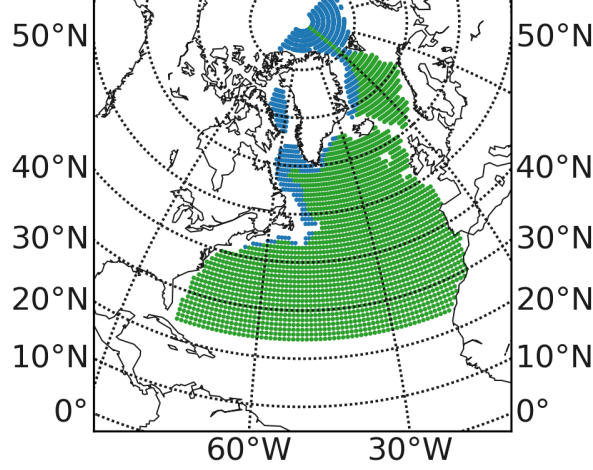


Figure 3: An example of EN4 profile assignment in the year 2010. Green points signify profiles assigned to the equatorial box, and blue points signify profiles assigned to the polar box.

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

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

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

$$\sigma_{T_*}^2(t) = \left( \sum_i V_i \sigma_{T_{i,*}}^2(t) \right) \left( \sum_i V_i \right)^{-1}. \quad (14)$$

These calculations are repeated for salinity  $S$ . The values after January 2004 are used to create the observations to be assimilated in  $\mathbf{d}(t) = [T_{p*}(t), T_{e*}(t), S_{p*}(t), S_{e*}(t)]^T$ . The  $4 \times 4$  observational error covariance matrix  $\Sigma_d$  is assumed to be diagonal. The values on the diagonal are chosen to be the maximal values for  $\sigma_{T_*}^2(t)$  and  $\sigma_{S_*}^2(t)$  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 2.

Field	$\cdot_0$	$\cdot_{\cos}$	$\cdot_{\sin}$	seasonal amplitude
$T_p^a$ [°C]	1.5	-1.5	-1.1	1.9
$T_e^a$ [°C]	16.7	-2.4	-2.3	3.4
$S_p^a$ [ppt]	33.05	0.22	0.32	0.39
$S_e^a$ [ppt]	35.77	0.04	0.05	0.07

Table 3: Regression coefficients found by fitting model Equation 15 to the surface observations for temperature  $^a$  and salinity  $^a$  in polar  $_p$  and  $_e$  boxes. Also shown is the seasonal amplitude  $\sqrt{.2_{\cos} + .2_{\sin}}$

The calculations in Equation 13 and Equation 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

$$T^a(t) \approx T_0^a + T_{\sin}^a \sin\left(\frac{2\pi t}{\tau}\right) + T_{\cos}^a \cos\left(\frac{2\pi t}{\tau}\right), \quad (15)$$

with  $\tau = 1$  year and  $T_0, T_{\cos}, T_{\sin}$  determined by a weighted linear regression against the box-averaged surface values using  $\sigma_{T_*}^{-2}(t)$  as weights. The regression coefficient for surface temperature and their equivalents for surface salinity are given in Table 3.

	dimensional	scale	nondimensional
$\Omega$	$2\pi \text{ rad yr}^{-1}$	$\frac{k_T}{\Delta z}$	3564
$B$	$1.51 \text{ }^\circ\text{C}$	$\frac{k_T}{\alpha_T \gamma \Delta z} \frac{\Delta y_p \Delta y_e}{\Delta y_p + \Delta y_e}$	1.18
$\hat{B}$	$0.32 \text{ ppt}$	$\frac{k_T}{k_S} \frac{k_T}{\alpha_S \gamma \Delta z} \frac{\Delta y_p \Delta y_e}{\Delta y_p + \Delta y_e}$	0.66
$A$	-	-	0.52

Table 4: Relation between the coefficients  $\Omega, A, B, \hat{B}$  in [30] and the surface forcing obtained from the EN4 observation data.

The regression coefficients in Table 3 can be related to the perturbations added to  $\eta_1$  and  $\eta_2$  in Equation (10) and Equation (11). These relations can be found in Table 4.

Table 5: Parameters, forcings, initial conditions and their sources.

Data	Values	Source	Details
Temperature (ocean temperature)	Pole: $T_p = 1.23 \text{ }^\circ\text{C}$ , Equator: $T_e = 5.4 \text{ }^\circ\text{C}$		Initial values from EN4 data
Salt (ocean salinity)	Pole: $S_p^* = 34.82 \text{ ppt}$ Equator: $S_e^* = 35.15 \text{ ppt}$		Initial values from EN4 data
temp_diff (surface heat mixing coefficient)	$k_T = 3.7 \cdot 10^{-6} \text{ ms}^{-1}$		equation 16
salt_diff (surface salinity mixing coefficient)	$k_S = 1.2 \cdot 10^{-6} \text{ ms}^{-1}$		equation 16
$\rho_0$ (reference density)	$1027 \text{ kgm}^{-3}$	[45]	
$S_0$ (reference salinity)	$35 \text{ ppt}$	[45]	
$T_0$ (reference temperature)	$10 \text{ }^\circ\text{C}$	[45]	
$\alpha_T$ (thermal expansion coefficient)	$\frac{0.15 \text{ }^\circ\text{C}^{-1}}{\text{ref}}$	[45]	
$\alpha_S$ (haline expansion coefficient)	$\frac{0.78 \text{ ppt}^{-1}}{\text{ref}}$	[45]	
Initial estimate $Q_{\text{overturning}}$ (meridional overturning flux)	$18 \text{ Sv}$	[46]	
$\gamma$ (advective transport coefficient)	$2.0 \text{ ms}^{-1}$		equation 16
$T^a(t)$	Oscillates over time		Calculated from EN4 dataset; see table 3
$S^a(t)$	Oscillates over time		Calculated from EN4 dataset; see table 3
$V_{\text{ice}}$	$2.9 \times 10^6 \text{ km}^3$		
Greenland ice sheet melt period $t_{\text{melt}}$	$10000 \text{ y}$	[47, 48, 49]	
Equatorial warming rate	$0.03 \text{ }^\circ\text{Cyr}^{-1}$	[50, 51, 52]	
Polar warming rate	$0.06 \text{ }^\circ\text{Cyr}^{-1}$	[53, 52]	
Model time step	$1 \text{ month}$		

The numerical scheme for Stommel model uses a finite volume approach with an explicit 4th-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  $\Sigma_d$  as covariance. Initial values for the logarithm of the diffusion and advection parameters are estimated by minimizing, using Scipy's Nelder-Mead method,

$$\frac{(\Delta T_* - \Delta T_{eq})^2}{\sigma_{T,e}^2 + \sigma_{T,p}^2} + \frac{(\Delta S_* - \Delta S_{eq})^2}{\sigma_{S,e}^2 + \sigma_{S,p}^2} + \frac{(Q_* - Q_{eq})^2}{(2.5 \text{ Sv})^2} \quad (16)$$

with  $\Delta T_*, \Delta S_*$  the difference between equatorial and polar box's initial ocean temperature and salinity,  $\sigma_{T,e}, \sigma_{S,e}$  the observational standard deviations in the 3rd column of table 2,  $Q = 18 \text{ Sv}$  the observed meridional transport and  $\Delta T_{eq}, \Delta S_{eq}, Q_{eq}$  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 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 5. 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 the poles warm around twice the global average [55, 56], so we used this ratio to inform our model.

## 4 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 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  $\Sigma_d$  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 ( $0.06 \frac{\text{°C}}{\text{year}}$  for polar box) than at lower latitudes ( $0.03 \frac{\text{°C}}{\text{year}}$  for equatorial box). Second, that the Greenland ice sheets melts add a constant rate  $m^* = \frac{V_{ice}}{t_{ice}}$  with  $V_{ice}$  and  $t_{ice}$  given in Table 5. 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 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 3rd column are non-dimensionalized using the scaling constants in Section 2.1. 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 function of the difference in temperature and salinity between the equatorial and polar box. Therefore the results in Figure 4 are depicted as function of these differences only.

As seen in Figure 4, none of the ensemble members predict a change in 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

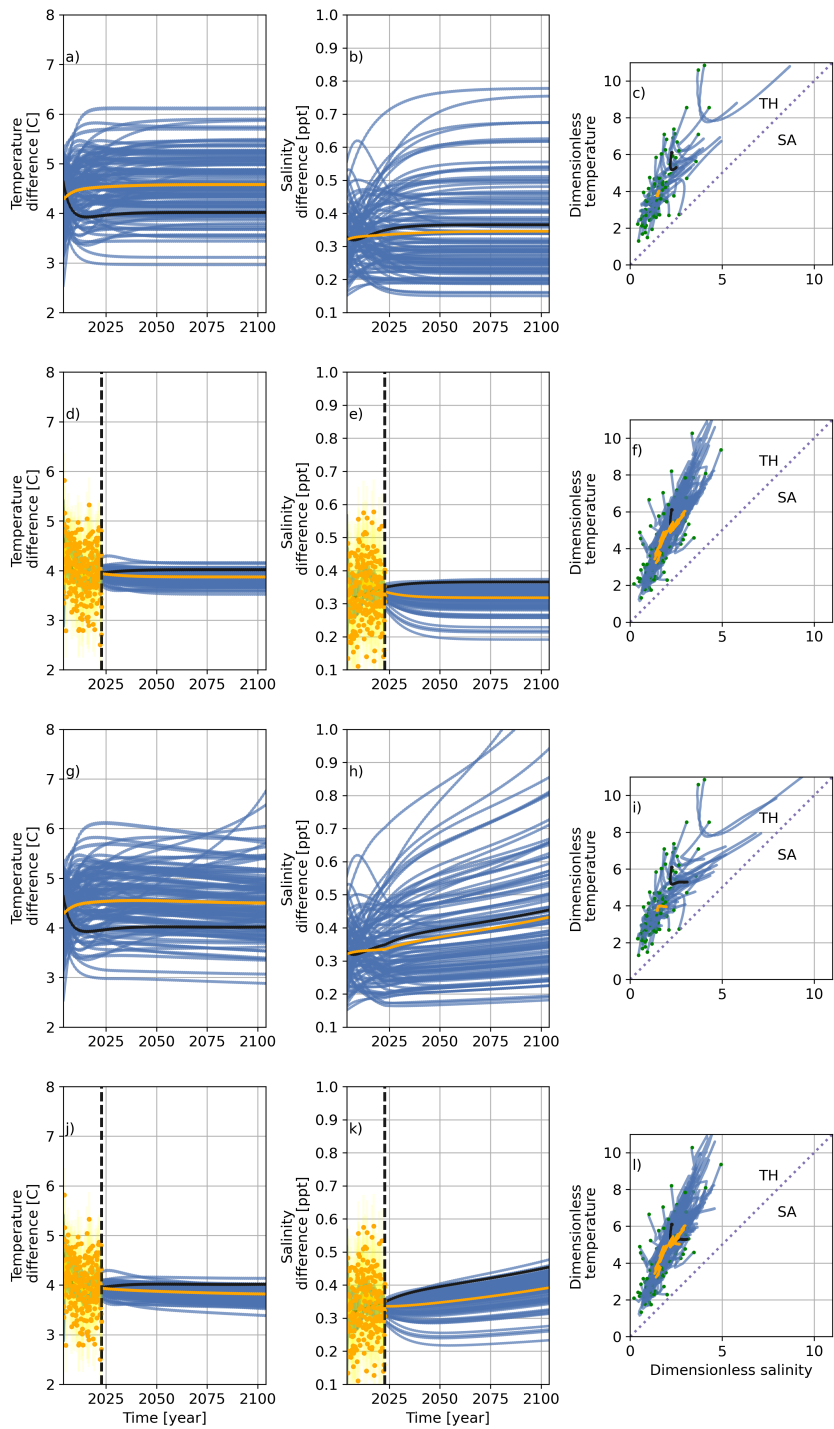


Figure 4: Forecast time series of temperature and salinity differences between equatorial and pole boxes for 100 ensemble members, with corresponding dimensionless salinity-vs-temperature phase portraits. Green dots: initial conditions; dashed blue line: TH-SA regime boundary; black paths: synthetic truth; orange paths: ensemble mode. Rows show different experiments: a)-c) nCC-nDA, d)-f) nCC-yDA, g)-i) yCC-nDA, 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.

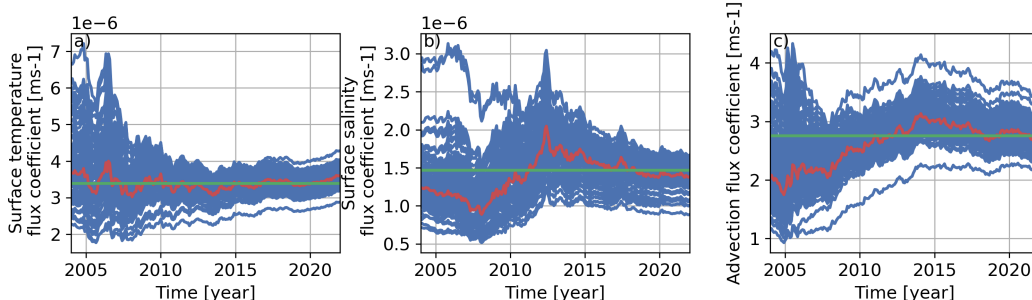


Figure 5: Read left to right: a) temperature ( $\kappa_T$ ), b) salinity ( $\kappa_S$ ), and c) advective flux coefficients ( $\gamma$ ) of the ensemble plotted over the assimilation period for an experiment assimilating 18 years of synthetic data without any global warming effects (the experiment with global warming is effectively identical). The red curve represents the ensemble mode, while the green line represents the true coefficient value.

not form different clusters. This is a necessary condition for the EnKF to function properly. Comparison of Figure 4a,b,g,h with Figure 4d,e,j,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 Figures 4g,h,j,k the solutions no longer converges towards 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 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, temperature diffusivity parameter, salinity diffusivity parameter and advection parameter are shown in Figure 5 as 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 Figure 4 in more detail, the ensemble standard deviation, or ensemble spread, and RMSE for the different experiments are shown in Figure 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 (Figure 6g) and temperature diffusion parameter (Figure 6e) are decimated while that of the salinity parameter is reduced but not to the same extent (Figure 6f). This reflects that salinity diffusion parameters is smaller. Consequently, it takes the system more time to equilibriate with surface salinity forcing and consequently the DA is less sensitive to changes in the salinity diffusion coefficient than in the advection and temperature diffusion coefficient. The fact that by the end of the DA period the spread in the salinity diffusion parameter is still decreasing (Figure 6l), whilst that in the temperature diffusion parameter (Figure 6j) and advection parameter ((Figure 6n) 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 root mean-square error (RMSE) should be equal. Comparison of the

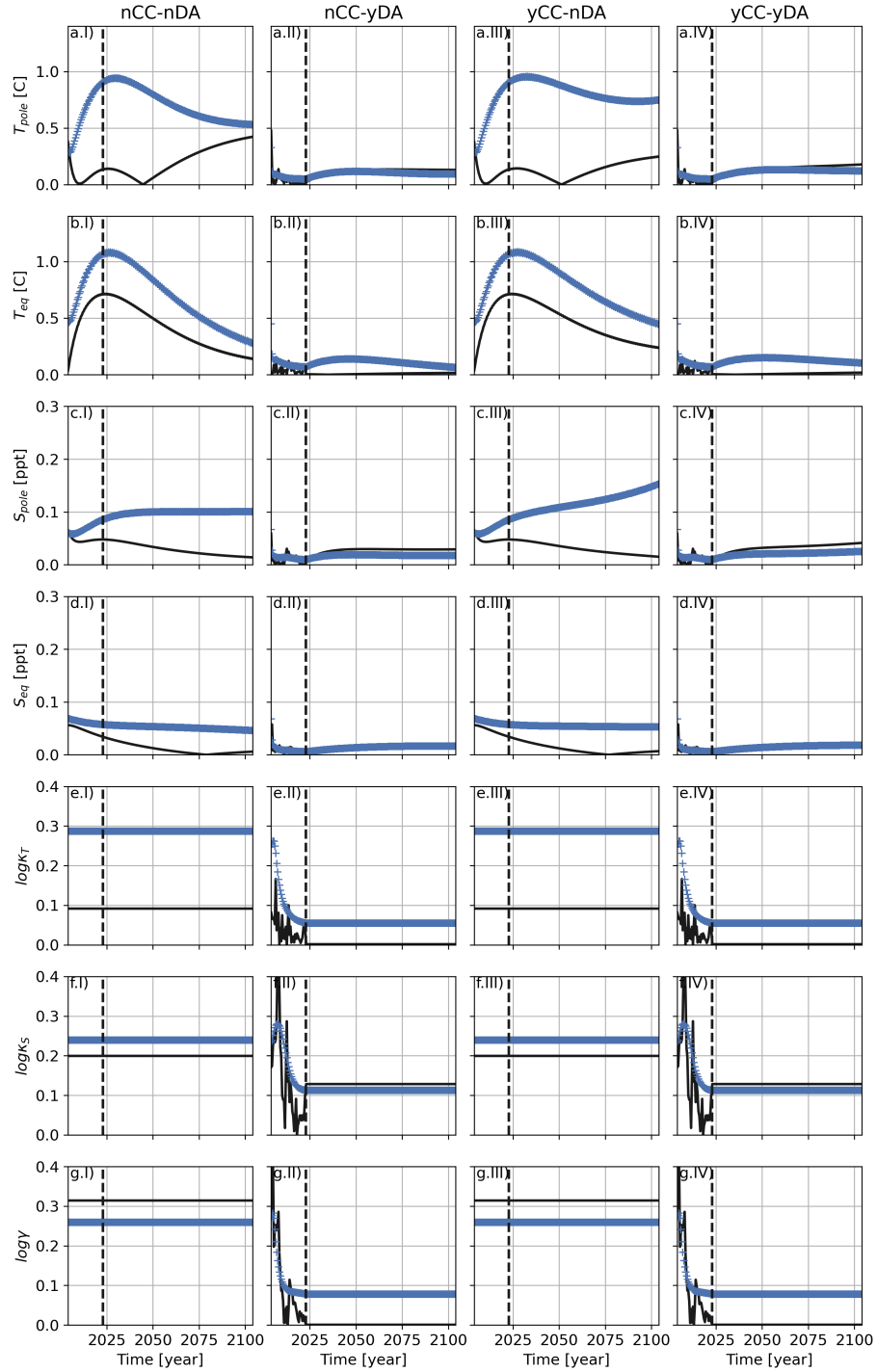


Figure 6: Read left to right, top to bottom. (Black solid lines) RMSE and (blue pluses) ensemble standard deviation of (rows) different variables as function of time for (columns) the four different experiments in Figure 4. For visibility only every 6th month is shown.



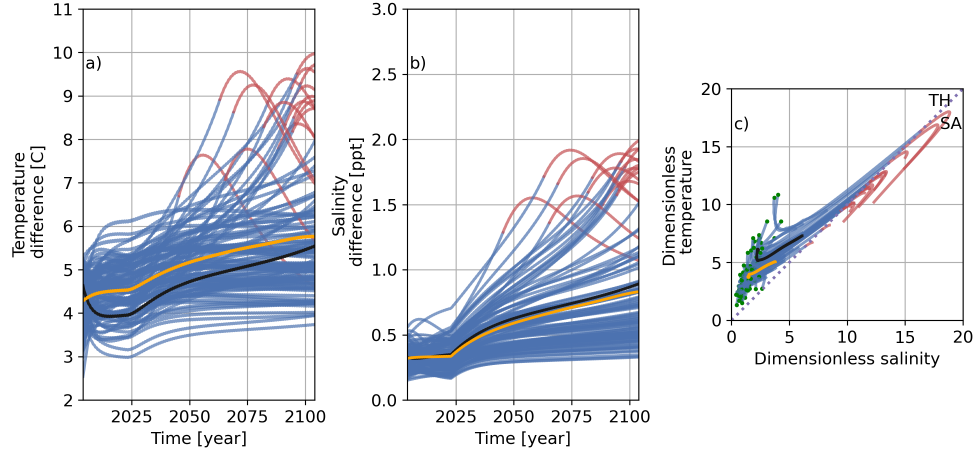


Figure 7: Read left to right: a) temperature and b) salinity time series, as well as c) a phase portrait for an experiment with yearly warming of  $0.03^{\circ}\text{C}$  per year in the equator ( $0.06^{\circ}\text{C}$  per year at the pole) and a melt period of 1000 years without DA. Trajectories turning red indicate a predicted flipping of flow direction. The black trajectories represent the synthetic truth values, and the orange represents the ensemble mode.

RMSEs and ensemble standard deviations at the end of the DA period shows they are of similar magnitude with the spread slightly overestimating the uncertainty. Thus confirming that DA is performing as expected.

All ensemble members in Figure 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 equatorial box is increased to  $0.07^{\circ}\text{C}$  per year (the polar box warming rate is raised to  $0.14^{\circ}\text{C}$  per year) and the melting period for the Greenland ice sheet is reduced from 10,000 to 1,000 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 Figure 7b salinity between-box difference in all ensemble members increases due to the effects of climate change. This reduces the circulation and temperature difference start to increase as consequence of the temperature difference present in the surface. Eventually, Figure 7a,b show that depending on initial conditions and the model parameters of the ensemble member, there is a rapid increase in the temperature and salinity between-box difference resulting in a reversal of the AMOC circulation direction. So, 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 advection parameter. The 18 year DA period used is sufficient to reduce the uncertainty in temperature diffusivity and 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 Section 5.

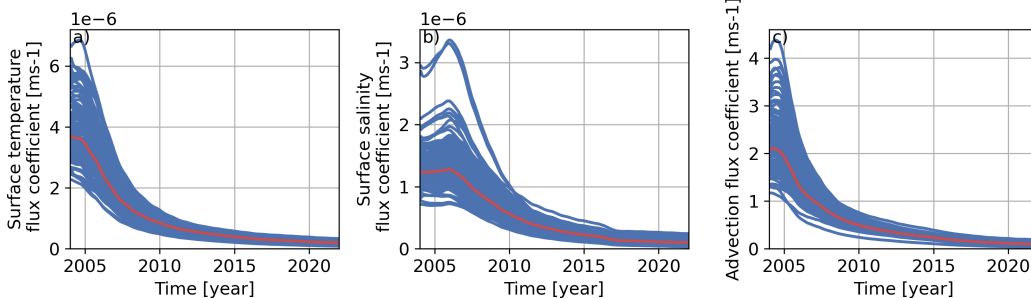


Figure 8: Read left to right: a) surface temperature ( $\kappa_T$ ), b) surface salinity diffusivity ( $\kappa_S$ ) and c) the advection flux coefficients ( $\gamma$ ) for the different ensemble members as functions of time. The red curves represents the ensemble mode.

## 5 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 Section 3.2. In this section we pursue a two-fold aim. First, we attempt to find realistic values for the salinity diffusivity  $k_S$ , temperature diffusivity  $k_T$  and advection parameter  $\gamma$  as well as the non-dimensional values  $\eta_1$ ,  $\eta_2$  and  $\eta_3$  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.

### 5.1 Interfered Model Parameters

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

Table 6: 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.

The three different model parameters are shown in Figure 8 as function of time. As the effects of climate change are never applied during the DA period (see Section 4), 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 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 Section 4 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 6.

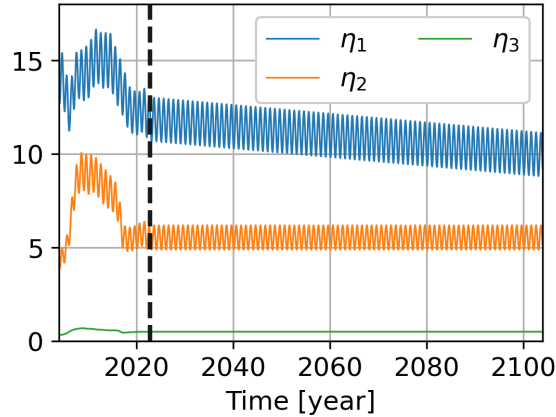


Figure 9: The values of parameters  $\eta_1$  (Equation (10)),  $\eta_2$  (Equation (11))  $\eta_3$  (Equation (7)) as function of time for the experiment with climate change.

The values for the non-dimensional parameters  $\eta_1$ ,  $\eta_2$  and  $\eta_3$  as appearing in Equation 4 are shown in Figure 9. The annual oscillations clearly visible in  $\eta_1$  and  $\eta_2$  are the result of seasonal cycle imposed on the surface forcing as Equation (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  $\eta_1$  after the DA period. The annually averaged values for the ensemble mode of  $\eta_1$ ,  $\eta_2$  and  $\eta_3$  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 [59, 30]. The absence of certain physical relevant processes for the AMOC and might explain difference. See also the discussion of transport at the end of this section.

## 6 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 Figure 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 Section 2.1, the autonomous solutions converge to one of at most two possible stable equilibria each associated with a different circulation regime. 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 largersmaller at 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 Section 4 that the spread in salinity diffusion parameter is larger than that in the temperature diffusivity parameter. The values

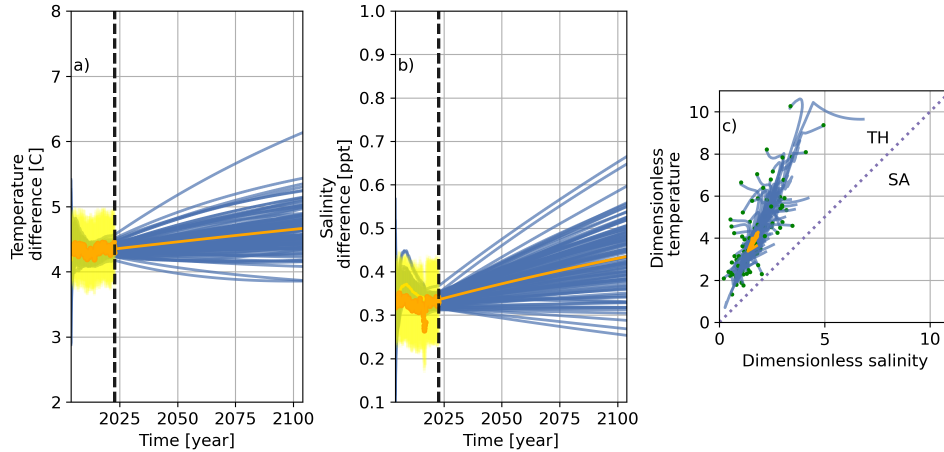


Figure 10: Read left to right. Forecast time series starting in 2004 representing the equator-pole difference in a) temperature and b) salinity for the different ensemble members as well as c) the ensemble trajectories in a dimensionless temperature vs salinity phase portrait. The orange dots with bright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

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  $0.06\text{ }^{\circ}\text{C}\text{y}^{-1}$ , the equatorial surface temperature with  $0.03\text{ }^{\circ}\text{C}\text{y}^{-1}$  and in which the Green land ice sheet melts with a constant rate over 10 thousand years are shown in Figure 11. Compared to Figure 10b the salinity differences increase after the DA period (Figure 11b). 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 (Figure 11a). Apart from these points, there are no qualitative differences between the figures. In particular, none of the ensemble members changes circulation regime.

The actual future rate of surface warming and ice melt will depend on, among other things, future carbon emission and are consequently uncertain. Therefore the DA experiment has been repeated several time for a range of plausible ice melting and between-box surface warming rate differences [47, 48, 49, 52, 60]. The percentage of ensemble members in which the circulation regime changes from TH to SA are shown in Figure 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  $286$  to  $487\text{ Gty}^{-1}$  which corresponds to a melt period of 7-12 thousand years [47, 48, 49]. This would place the future climate firmly within the zone of Figure 12 in which circulation the AMOC does not reverse direction.

The transport between the two boxes is shown in Figure 13. Transports are an order of magnitude smaller than the typical observational transect estimates over this period that lie in the range  $10 - -25\text{ Sv}$  [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

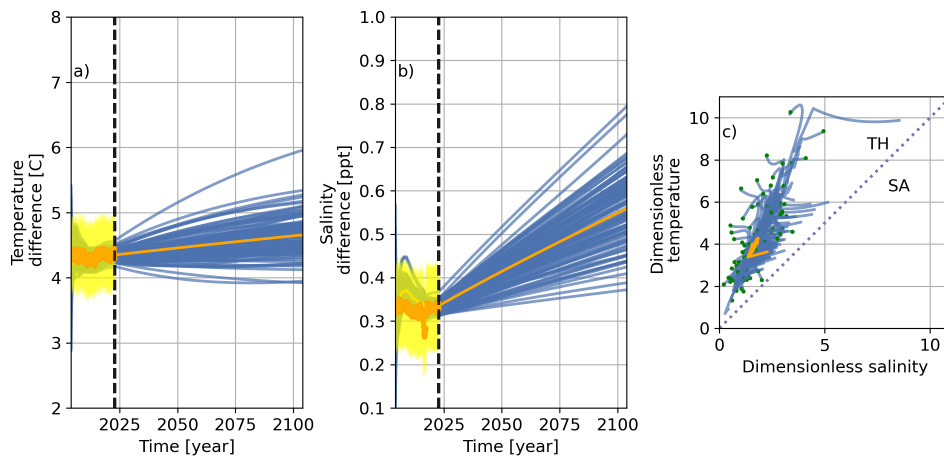


Figure 11: As Figure 10, read left to right. Forecast time series starting in 2004 representing the equator-pole difference in a) temperature and b) salinity for the different ensemble members as well as c) the ensemble trajectories in a dimensionless temperature vs salinity phase portrait, but now climate 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 a TH circulation are shown in blue, those with a SA circulation (if any) in red. End of the DA period is marked by a vertical dashed line

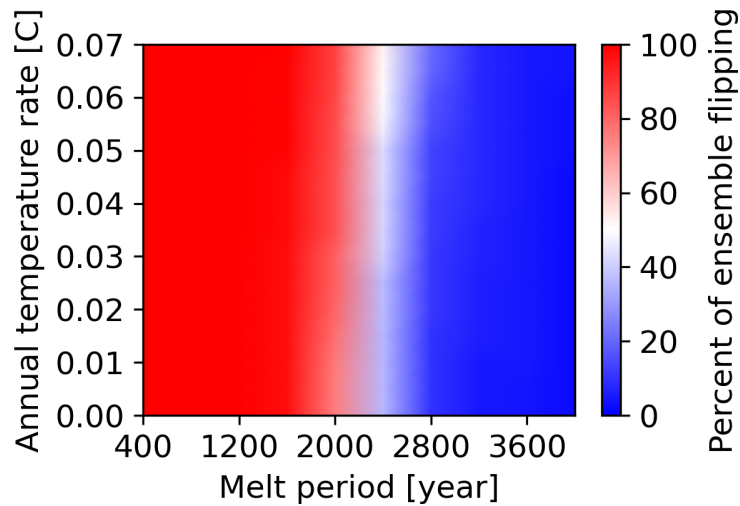


Figure 12: Percentage of ensemble members that undergo circulation regime change. All runs use a similar setup to the experiment with climate change and real data assimilation, but using different ice melt periods (x-axis) and different warmings rates for the equator box (y-axis). The warming rate in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

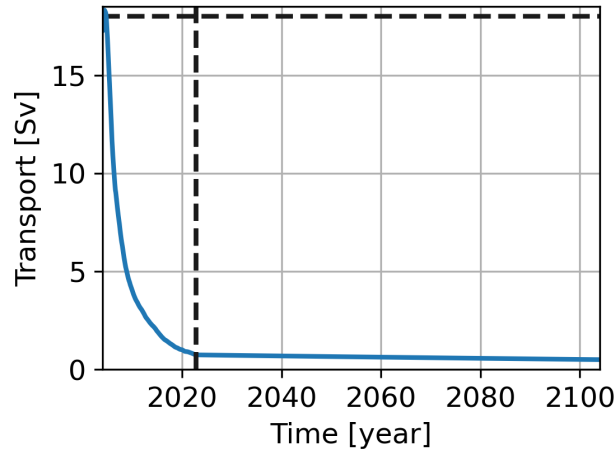


Figure 13: The value of the between-box transport as function of time for the experiment with climate change. The vertical dashed line marks the end of the assimilation period. The horizontal dashed line indicates observed current transport [46].

transports. The source of this discrepancy might lie in our box definition. This will be discussed in more detail in Section 7. 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 Figure 12 at which the circulation flips are in all likelihood an overestimation.

## 7 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 advection parameter by assimilating ocean temperature and salinity parameters and to forecast the likelihood of 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 rate 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 bring 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 is 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 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 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 [62, 41, 42, 43] 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 [63, 39] it has been argued that not the buoyancy forcing, but wind stresses are the primary driver behind the AMOC. And indeed [12] has 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 towards the Antarctic Circumpolar Current nor for the northward flow of Antarctic Bottom Water and waters from the Angulhas current [1, sec. 14.2.14] which can increase transports by creating diapycnal mixing [64]. Accommodating these concepts requires more complex conceptual models like 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.

## **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 Equation (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

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

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

$$\frac{\partial(\phi f)}{\partial T} + \frac{\partial(\phi g)}{\partial S} = \frac{\partial f}{\partial T} + \frac{\partial g}{\partial S} = -1 - \eta_3 - 3|T - S| \quad (17)$$

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 which crosses between the two areas, over the line  $T = S$ .

Now, assume we did have such a trajectory, a periodic solution  $C$  which enclosed a region  $D$  within  $T, S \geq 0$ , and let us call the part 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 Equation 17 allows for a continuous extension from  $D_1$  ( $D_2$ ) to its compact closure  $\bar{D}_1$  ( $\bar{D}_2$ ),  $\int_{\bar{D}_1} (\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}) dT dS$  ( $\int_{\bar{D}_2} (\frac{\partial f}{\partial T} + \frac{\partial g}{\partial S}) dT dS$ ) exists and is, as shown above, strictly smaller than zero. Applying Green's Theorem now gives:

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

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

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

But earlier we showed this was less than zero, so we have a contradiction, so no such loop  $C$  can exist.

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



## B Ensemble Kalman filter

Let

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

represent a possible state of the system at time  $t$  with  $\kappa_{ref} = 1 \text{ m}^2\text{s}^{-1}$  a reference values to nondimensionlize 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  $\kappa_T, \kappa_S$  and  $\gamma$ , it is referred to as the augmented state. Let  $\int_{\mathcal{X}} p(\mathbf{x}_0) d\mathbf{x}_0$  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

$$\underbrace{p(\mathbf{x}_t | \mathbf{d}_{0:t})}_{\text{posteriori}} \sim p(\mathbf{d}_t | \mathbf{x}_t) \underbrace{\int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{d}_{0:t-1}) d\mathbf{x}_{t-1}}_{\text{apriori}} \quad (18)$$

where  $\sim$  indicates equivalence up to a normalisation constant,  $d_t$  a vector with observations at timestep  $t$  and  $d_{0:t}$  a vector of all observations upto and including timestep  $t$ . Once the apriori is specified for the initial timestep  $t = 0$ , equation 18 posteriori and apriori distributions for  $t > 0$  can be constructed by induction using Equation 18. This occurs in two steps. The *prediction step* in which the posteriori distribution at time  $t - 1$  is transformed into the apriori distribution at time  $t$ , and the *update step* in which the DA correction is applied to transform the apriori distribution at time  $t$  into the posteriori distribution at the same time.

In an ensemble Kalman filter it is assumed that the apriori and 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  $\Sigma_x = \frac{1}{M-1} \sum_{m=1}^M (\mathbf{x}_t^{(m)} - \mu_{x,t})(\mathbf{x}_t^{(m)} - \mu_{x,t})^\top$  estimated from an  $M$ -member ensemble of model runs  $\{\mathbf{x}_t^{(m)} : 1 \leq m \leq M\}$ . Each ensemble member represents an equally likely realisation of these probability distributions. The integral in the apriori distribution in Equation 18 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 model error and hence  $p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \delta(\mathbf{x}_t - \mathcal{M}_{t-1 \rightarrow t} \mathbf{x}_{t-1})$  with  $\delta$  the delta function and  $\mathcal{M}_{t-1 \rightarrow t}$  the model operator that propagates the ocean temperatures and salinities from time  $t - 1$  to  $t$  according to equations 1-3.  $\mathcal{M}_{t-1 \rightarrow t}$  whilst 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  $\Sigma_d$ . In our case the observation operator  $H$  is such that  $H\mathbf{x}_t = [T_{p^*}(t), T_{e^*}(t), S_{p^*}(t), S_{e^*}(t)]^\top$  and  $\Sigma_d$  is assumed to be diagonal with its diagonal elements (variances) given by Equation 14.

It can be shown that under these assumptions, the 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  $\Sigma_{\hat{x},t} = \frac{1}{M-1} (\hat{\mathbf{x}}_t^{(m)} - \mu_{\hat{x},t})(\hat{\mathbf{x}}_t^{(m)} - \mu_{\hat{x},t})^\top$  and

$$\hat{\mathbf{x}}_t^{(n)} = \mathbf{x}_t^{(n)} + K(\mathbf{d}_t - H\mu_{t,x}) + AQA^\frac{1}{2}Q^\top \hat{\mathbf{e}}^{(n)}, \quad (19)$$

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

The most-likely state, or mode, of the apriori and 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  $e^{\mu_p}$  as naively might be expected, but by  $e^{\mu_p - \Sigma_p^{-1}}$  with  $\mu_p$  the 3-dimensional vector containing the ensemble mean of the logarithms of the model parameters and  $\Sigma_p$  the  $3 \times 3$  matrix with their ensemble covariance [67]. I.e.  $\mu_p$  and  $\Sigma_p$  are the last 3-elements or  $3 \times 3$ -block of the  $\mu_x$  and  $\Sigma_x$  respectively.

## 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 is 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 [68] GIT repository. 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.

## Acknowledgement

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 NSF grant DMS-1722578. Ivo Pasmans would like to acknowledge the support of the Scale-Aware Sea Ice Project (SASIP) funded by Schmidt Sciences – a philanthropic initiative that seeks to improve societal outcomes through the development of emerging science and technologies - grant G-24-66154. Emmanuel Fleurantin was supported by NSF grant 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.

## References

- [1] Lynne Talley, George Pickard, William Emery, and James Swift. *Descriptive Physical Oceanography*. Elsevier, London, 6th edition, 2011.
- [2] Henry Stommel. Thermohaline convection with two stable regimes of flow. *Tellus*, 13(2):224–230, 1961.
- [3] Frank Bryan. High-latitude salinity effects and interhemispheric thermohaline circulations. *Nature*, 323(6086):301–304, September 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: Dynamic Meteorology and Oceanography*, 40(2):162, January 1988.
- [5] E. Maier-Reimer and U. Mikolajewicz. Experiments with an OGCM on the cause of the Younger Dryas. Technical Report 39, Max-Planck-Institut fuer Meteorologie, Hamburg, Germany, October 1989.

- [6] Vasilis Dakos, Marten Scheffer, Egbert H. van Nes, Victor Brovkin, Vladimir Petoukhov, and Hermann Held. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, 105(38):14308–14312, 2008.
- [7] Peter D. Ditlevsen and Sigfus J. Johnsen. Tipping points: Early warning and wishful thinking. *Geophysical Research Letters*, 37(19), 2010.
- [8] Marten Scheffer, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, September 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. *Journal of Geophysical Research: Oceans*, 124(8):5336–5375, 2019.
- [10] Bas J. H. Van de Wiel, Etienne Vignon, Peter Baas, Ivo G. S. van Hooijdonk, Steven J. A. van der Linden, J. Antoon van Hooft, Fred C. Bosveld, Stefan R. de Roode, Arnold F. Moene, and Christophe Genthon. Regime transitions in near-surface temperature inversions: A conceptual model. *Journal of the Atmospheric Sciences*, 74(4):1057 – 1073, 2017.
- [11] Ilona M. Otto, Jonathan F. Donges, Roger Cremades, Avit Bhowmik, Richard J. Hewitt, Wolfgang Lucht, Johan Rockström, Franziska Allerberger, Mark McCaffrey, Sylvanus S. P. Doe, Alex Lenferna, Nerea Morán, Detlef P. van Vuuren, and Hans Joachim Schellnhuber. Social tipping dynamics for stabilizing earth’s climate by 2050. *Proceedings of the National Academy of Sciences*, 117(5):2354–2365, 2020.
- [12] Andrea A. Cimadoribus, Sybren S. Drijfhout, and Henk A. Dijkstra. Meridional overturning circulation: Stability and ocean feedbacks in a box model. *Climate Dynamics*, 42(1):311–328, January 2014.
- [13] Daniele Castellana and Henk A. Dijkstra. Noise-induced transitions of the atlantic meridional overturning circulation in cmip5 models. *Scientific Reports*, 10(1):20040, 2020.
- [14] D. Ryvkine and M. I. Dykman. Pathways of activated escape in periodically modulated systems. *Phys. Rev. E (3)*, 73(6):061109, 11, 2006.
- [15] Paul E. O’Keeffe and Sebastian Wicczorek. Tipping phenomena and points of no return in ecosystems: beyond classical bifurcations. *SIAM J. Appl. Dyn. Syst.*, 19(4):2371–2402, 2020.
- [16] Xiang Zhou and Weinan E. Study of noise-induced transitions in the Lorenz system using the minimum action method. *Commun. Math. Sci.*, 8(2):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 System Dynamics*, 12(3):819–835, 2021.
- [18] Peter Ditlevsen and Susanne Ditlevsen. Warning of a forthcoming collapse of the Atlantic meridional overturning circulation. *Nature Communications*, 14(1):4254, July 2023.

- [19] René M. van Westen, Michael Kliphuis, and Henk A. Dijkstra. Physics-based early warning signal shows that AMOC is on tipping course. *Science Advances*, 10(6):eadk1189, February 2024.
- [20] Maya Ben-Yami, Andreas Morr, Sebastian Bathiany, and Niklas Boers. Uncertainties too large to predict tipping times of major Earth system components, September 2023.
- [21] Julie Sherman, Christian Sampson, Emmanuel Fleurantin, Zhimin Wu, and Christopher KRT Jones. A data-driven study of the drivers of stratospheric circulation via reduced order modeling and data assimilation. *Meteorology*, 3(1):1–35, 2023.
- [22] PJ Smith, GD Thornhill, SL Dance, AS Lawless, DC Mason, and NK Nichols. Data assimilation for state and parameter estimation: application to morphodynamic modelling. *Quarterly Journal of the Royal Meteorological Society*, 139(671):314–327, 2013.
- [23] Florian Sévellec and Alexey V Fedorov. Unstable amoc during glacial intervals and millennial variability: The role of mean sea ice extent. *Earth and Planetary Science Letters*, 429:60–68, 2015.
- [24] Ed Hawkins, Robin S Smith, Lesley C Allison, Jonathan M Gregory, Tim J Woollings, Holger Pohlmann, and B De Cuevas. Bistability of the atlantic overturning circulation in a global climate model and links to ocean freshwater transport. *Geophysical Research Letters*, 38(10), 2011.
- [25] LC Jackson, E Alastrué de Asenjo, K Bellomo, G Danabasoglu, H Haak, A Hu, J Jungclaus, W Lee, VL Meccia, O Saenko, et al. Understanding amoc stability: the north atlantic hosing model intercomparison project, geosci. model dev., 16, 1975–1995, 2023.
- [26] Henk A. Dijkstra and Michael Ghil. Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Reviews of Geophysics*, 43(3), 2005.
- [27] Ruth Chapman, Sacha Sinet, and Paul D. L. Ritchie. Tipping mechanisms in a conceptual model of the Atlantic Meridional Overturning Circulation. *Weather*, n/a(n/a), August 2024.
- [28] Niklas Boers. Observation-based early-warning signals for a collapse of the atlantic meridional overturning circulation. *Nature Climate Change*, 11(8):680–688, 2021.
- [29] Mark M Dekker, Anna S Von Der Heydt, and Henk A Dijkstra. Cascading transitions in the climate system. *Earth System Dynamics*, 9(4):1243–1260, 2018.
- [30] Chris Budd, Cody Griffith, and Rachel Kuske. Dynamic tipping in the non-smooth Stommel-box model, with fast oscillatory forcing. *Physica D: Nonlinear Phenomena*, 432:132948, April 2022.
- [31] Patrick N. Raanes, Yumeng Chen, Colin Grudzien, Maxime Tondeur, and Remy Dubois. DAPPER. Nansen Environmental and Remote Sensing Center, September 2023.
- [32] Craig H. Bishop, Brian J. Etherton, and Sharanya J. Majumdar. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*,

129(3):420–436, March 2001.

- [33] Simon A. Good, Matthew J. Martin, and Nick A. Rayner. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716, 2013.
- [34] Viktor Gouretski and Franco 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 Research Part I: Oceanographic Research Papers*, 57(6):812–833, June 2010.
- [35] Claus W. Böning, Frank O. Bryan, William R. Holland, and Ralf Döscher. Deep-Water Formation and Meridional Overturning in a High-Resolution Model of the North Atlantic. *Journal of Physical Oceanography*, 26(7):1142–1164, July 1996.
- [36] Argo. Argo float data and metadata from global data assembly centre (Argo GDAC). <https://doi.org/10.17882/42182>.
- [37] Tillys Petit, M. Susan Lozier, Simon A. Josey, and Stuart A. Cunningham. Atlantic Deep Water Formation Occurs Primarily in the Iceland Basin and Irminger Sea by Local Buoyancy Forcing. *Geophysical Research Letters*, 47(22):e2020GL091028, 2020.
- [38] David A. Bailey, Peter B. Rhines, and Sirpa Häkkinen. Formation and pathways of North Atlantic Deep Water in a coupled ice–ocean model of the Arctic–North Atlantic Oceans. *Climate Dynamics*, 25(5):497–516, October 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. *Reviews of Geophysics*, 45(2), 2007.
- [40] M. Susan Lozier. Deconstructing the Conveyor Belt. *Science*, 328(5985):1507–1511, June 2010.
- [41] Sybren S. Drijfhout, Susanne L. Weber, and Eric van der Swaluw. The stability of the MOC as diagnosed from model projections for pre-industrial, present and future climates. *Climate Dynamics*, 37(7):1575–1586, October 2011.
- [42] Richard A. Wood, José M. Rodríguez, Robin S. Smith, Laura C. Jackson, and Ed Hawkins. Observable, low-order dynamical controls on thresholds of the Atlantic meridional overturning circulation. *Climate Dynamics*, 53(11):6815–6834, December 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, May 2024.
- [44] Adam Hugh Monahan. Stabilization of climate regimes by noise in a simple model of the Thermohaline Circulation. *Journal of Physical Oceanography*, 32(7):2072–2085, July 2002.
- [45] Jonas Nycander, Magnus Hieronymus, and Fabien Roquet. The nonlinear equation of state of sea water and the global water mass distribution. *Geophysical Research Letters*, 42(18):7714–7721, 2015.

- [46] G.D. McCarthy, D.A. Smeed, W.E. Johns, E. Frajka-Williams, B.I. Moat, D. Rayner, M.O. Baringer, C.S. Meinen, J. Collins, and H.L. Bryden. Measuring the atlantic meridional overturning circulation at 26°n. *Progress in Oceanography*, 130:91–111, 2015.
- [47] Shfaqat A. Khan, Andy Aschwanden, Anders A. Bjørk, John Wahr, Kristian K. Kjeldsen, and Kurt H. Kjær. Greenland ice sheet mass balance: A review. *Reports on Progress in Physics*, 78(4):046801, March 2015.
- [48] Jérémie Mouginot, Eric Rignot, Anders A. Bjørk, Michiel van den Broeke, Romain Millan, Mathieu Morlighem, Brice Noël, Bernd Scheuchl, and Michael Wood. Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018. *Proceedings of the National Academy of Sciences*, 116(19):9239–9244, May 2019.
- [49] Kenneth D. Mankoff, Anne Solgaard, William Colgan, Andreas P. Ahlstrøm, Shfaqat Abbas Khan, and Robert S. Fausto. Greenland Ice Sheet solid ice discharge from 1986 through March 2020. *Earth System Science Data*, 12(2):1367–1383, June 2020.
- [50] Climate change: Global temperature projections. URL <https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature-projections>. Last accessed 2022-11-21.
- [51] Climate change 2007: Synthesis report. URL [https://www.ipcc.ch/site/assets/uploads/2018/02/ar4\\_syr\\_full\\_report.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/ar4_syr_full_report.pdf). Last accessed 2022-11-21.
- [52] Qinglong You, Ziyi Cai, Nick Pepin, Deliang Chen, Bodo Ahrens, Zhihong Jiang, Fangying Wu, Shichang Kang, Ruonan Zhang, Tonghua Wu, Pengling Wang, Mingcai Li, Zhiyan Zuo, Yanhong Gao, Panmao Zhai, and Yuqing Zhang. Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences. *Earth-Science Reviews*, 217:103625, June 2021.
- [53] Jinlun Zhang. Warming of the arctic ice-ocean system is faster than the global average since the 1960s. *Geophysical Research Letters*, 32(19), 2005.
- [54] John E Walsh. Intensified warming of the arctic: Causes and impacts on middle latitudes. *Global and Planetary Change*, 117:52–63, 2014.
- [55] J Richter-Menge and M Druckenmiller. State of the climate in 2019. *Arctic*, 101(8):S239–S286, 2020.
- [56] Eystein Jansen, Jens Hesselbjerg Christensen, Trond Dokken, Kerim H Nisancioglu, Bo M Vinther, Emilie Capron, Chuncheng Guo, Mari F Jensen, Peter L Langen, Rasmus A Pedersen, et al. Past perspectives on the present era of abrupt arctic climate change. *Nature Climate Change*, 10(8):714–721, 2020.
- [57] G. R. Halliwell, A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. Le Hénaff, and R. Atlas. Rigorous Evaluation of a Fraternal Twin Ocean OSSE System for the Open Gulf of Mexico. *Journal of Atmospheric and Oceanic Technology*, 31(1):105–130, January 2014.

- [58] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450, 1965.
- [59] Henk A. Dijkstra. *Dynamical Oceanography*. Springer Berlin,, Heidelberg, 1st edition edition, 2011.
- [60] Aihong Xie, Jiangping Zhu, Shichang Kang, Xiang Qin, Bing Xu, and Yicheng Wang. Polar amplification comparison among Earth’s three poles under different socioeconomic scenarios from CMIP6 surface air temperature. *Scientific Reports*, 12(1):16548, October 2022.
- [61] Gokhan Danabasoglu, Frederic S. Castruccio, R. Justin Small, Robert Tomas, Eleanor Frajka-Williams, and Matthias Lankhorst. Revisiting AMOC transport estimates from observations and models. *Geophysical Research Letters*, 48(10):e2021GL093045, 2021.
- [62] S. Rahmstorf. On the freshwater forcing and transport of the Atlantic thermohaline circulation. *Climate Dynamics*, 12(12):799–811, November 1996.
- [63] Anand Gnanadesikan. A Simple Predictive Model for the Structure of the Oceanic Pycnocline. *Science*, 283(5410):2077–2079, March 1999.
- [64] Laura Cimoli, Ali Mashayek, Helen L. Johnson, David P. Marshall, Alberto C. Naveira Garabato, Caitlin B. Whalen, Clément Vic, Casimir de Lavergne, Matthew H. Alford, Jennifer A. MacKinnon, and Lynne D. Talley. Significance of diapycnal mixing within the Atlantic Meridional Overturning Circulation. *AGU Advances*, 4(2):e2022AV000800, 2023.
- [65] Claes Rooth. Hydrology and ocean circulation. *Progress in Oceanography*, 11(2):131–149, January 1982.
- [66] Geir Evensen, Femke C. Vossepoel, and Peter Jan van Leeuwen. *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer Nature, 2022.
- [67] S.J. Fletcher. Mixed Gaussian-lognormal four-dimensional data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 62(3):266–287, January 2010.
- [68] Ivo Pasmans. DAPPER/inversion, November 2023.