Stability of the thermohaline circulation examined with a one-dimensional fluid loop

Article

Published Version

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

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online
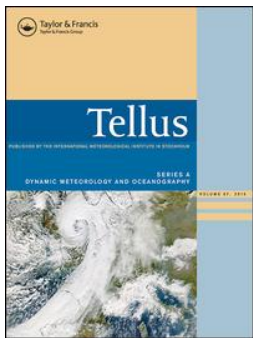
Stability of the thermohaline circulation examined with a one-dimensional fluid loop

Fabien Roquet, Rickard Lindqvist, Friederike Pollmann, David Ferreira & Gurvan Madec

Stability of the thermohaline circulation examined with a one-dimensional fluid loop

By FABIEN ROQUET[1]*, RICKARD LINDQVIST[1], FRIEDERIKE POLLMANN[2], DAVID FERREIRA[3] and GURVAN MADEC[4], [1]*Department of Meteorology (MISU), Stockholm University, Stockholm, Sweden;* [2]*Institut für Meereskunde, Universität Hamburg, Hamburg, Germany;* [3]*Department of Meteorology, University of Reading, Reading, UK;* [4]*LOCEAN Sorbonne University (UPMC, Paris 6)/CNRS/IRD/MNHN, Paris, France*

ABSTRACT

The Stommel box model elegantly demonstrates that the oceanic response to mixed boundary conditions, combining a temperature relaxation with a fixed salt flux forcing, is non-linear owing to the so-called salt advection feedback. This non-linearity produces a parameter range of bi-stability associated with hysteresis effects characterised by a fast thermally driven mode and a slow salinity-driven mode. Here, we investigate whether a similar dynamical behaviour can be found in the thermohaline loop model, a one-dimensional analogue of the box model. A semi-analytical method to compute possible steady states of the loop model is presented, followed by a linear stability analysis carried out for a large range of loop configurations. While the salt advection feedback is found as in the box model, a major difference is obtained for the fast mode: an oscillatory instability is observed near the turning point of the fast mode branch, such that the range of bi-stability is systematically reduced, or even removed, in some cases. The oscillatory instability originates from a salinity anomaly that grows exponentially as it turns around the loop, a situation that may occur only when the salinity torque is directed against the loop flow. Factors such as mixing intensity, the relative strength of thermal and haline forcings, the non-linearity of the equation of state or the loop geometry can strongly affect the stability properties of the loop.

Keywords: thermohaline circulation, box model, linear stability, thermohaline loop, Meridional Overturning Circulation

1. Introduction

The global ocean circulation connects water masses of the different ocean basins, inducing a large-scale redistribution of heat, carbon and other passive tracers of importance for Earth's climate. The largest scale component of this circulation is generally referred to as the 'thermohaline circulation' to acknowledge the central importance of surface fluxes of heat and freshwater in its dynamics, although it has been noted that the wind forcing, eddy stirring and internal mixing also play a critical role in providing the mechanical forcing needed to maintain a substantial thermohaline circulation (Wunsch and Ferrari, 2004).[1]

At the core of the thermohaline circulation lies the Atlantic Meridional Overturning Circulation (AMOC). The AMOC carries relatively warm and saline upper waters northwards across the entire Atlantic basin, which are then transformed into colder and fresher North Atlantic Deep Water in the Nordic Seas before

returning southwards at intermediate depth. It is estimated that without the AMOC, the North Atlantic climate would be about 5 °C colder than at present (Vellinga and Wood, 2002; Stouffer et al., 2006) due to a decrease in meridional heat transport (direct effect) and a subsequent increase in sea ice extent in the North Atlantic (indirect ice–albedo effect). The AMOC is also responsible for the mean northward shift of the intertropical convergence zone with important consequences for tropical climate variability (Marshall et al., 2013). Past abrupt climate events such as the Heinrich events have been linked to rapid changes in the AMOC strength and a link between the AMOC and North Atlantic decadal climate variability is debated (Rahmstorf, 2002).

Despite its importance for climate, the AMOC remains poorly understood and such fundamental questions as 'Why is there an AMOC and no equivalent in the Pacific Basin?', 'Has there always been an AMOC?' or 'What is the relative importance of wind and thermohaline forcing in setting the variability of the AMOC?' remain largely open. The very nature of the AMOC makes it difficult to study, as it is characterised by a nearly

---

*Corresponding author. e-mail: roquet@misu.su.se

1

planetary spatial scale and a centennial to millennial time scale while its dynamics relies on processes with scales several orders of magnitude smaller (Wunsch and Heimbach, 2013). Direct observations are available for the last two decades only and realistic general circulation models (GCM) remain very expensive to run at high resolution or rely on poorly constrained parameterisations at coarse resolution.

Simple idealised models of the AMOC are very useful in this context to develop hypotheses and concepts on the real complex system (Held, 2005). Stommel's two-box model has been by far the most influential conceptual model of the AMOC (Stommel, 1961). This model is described by a single scalar equation for the overturning transport. Assuming a linear equation of state (EOS), the steady-state transport q, which is proportional to the density difference between the two boxes, satisfies the (non-dimensionalised) equation,

$$|q|(1-q) - H_S = 0 \qquad (1)$$

where H_S represents an imposed flux of salt (positive in the warm box, negative in the cold box). Note that we are here using the simplified, elegant version of the two-box model proposed by Marotzke (2000).

The key feature of this model is that despite its simplicity it features a non-linear response to the varying salt forcing because of the so-called salt advection feedback (Fig. 1). Depending on the strength of the salt flux, the transport can be either temperature-controlled ($q > 0$) or salinity-controlled ($q < 0$), and there generally exists a range of salt forcing for which the two states are possible (bi-stability for $0 < H_S < 0.25$). This means that when the salt forcing is increased, the transport is prone to sudden collapse associated with a hysteresis behaviour preventing the transport to jump back to the temperature-controlled state until a substantial reduction of the salt forcing has occurred (Marotzke, 2000).

The Stommel two-box model is the simplest of a hierarchy of models that can be used to study overturning circulations such as the AMOC (Welander, 1986). A natural way to classify the large variety of existing conceptual models is by way of the number of spatial dimensions used to describe the tracer fields. The Stommel's model is a zero-dimensional model in this sense, while GCMs are three-dimensional models. Between them, we find one-dimensional loop models and two-dimensional section (depth, latitude) models. Interestingly, models with a given dimensionality can feature the entire spectrum of dynamical behaviours of a lower dimensionality, but they also display new types of behaviours. Here, we will focus on the thermohaline loop model, viewed as an intermediate case between the simplistic box model and more realistic GCMs.

The thermohaline loop model consists of a fluid loop with infinitesimal section area along which a circulation is induced by applying fluxes of temperature and salt or by applying a torque
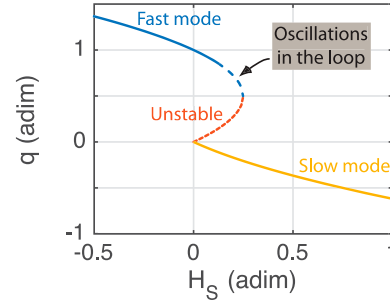

Fig. 1. Steady-state solutions of a Stommel two-box model, showing a range of bi-stability between a temperature-controlled fast mode (blue) and a salinity-controlled slow mode (yellow). The intermediate state (red) is unstable. In this study, we will show that in the thermohaline loop model, the fast mode becomes subject to an oscillatory instability before the turning point, effectively reducing the range of bi-stability.

(Keller, 1966; Welander, 1967; Malkus, 1972; Huang, 1999; Wunsch, 2005; Sévellec et al., 2006; Pollmann et al., 2015). The greatest innovation of the thermohaline loop model compared to box models is the explicit resolution of the advective–diffusive balance, allowing for a more realistic tracer distribution and delayed oscillations due to the advection of anomalies around the loop. Yet, the thermohaline loop model remains rather simple and thus easy to handle because the velocity must be constant around the loop by continuity and thus follows a single scalar equation that can be made diagnostic.

In this paper, the dynamical behaviour of the loop model will be investigated. In analogy with the AMOC, fluxes will be imposed at the same vertical level (horizontal convection case) and the top of the loop will be 'folded' as in Pollmann et al. (2015). Mixed boundary conditions will be used, here by applying a temperature relaxation together with fixed salt fluxes. The effect of imposing a torque will also be investigated. The thermohaline loop model is simple enough to readily explore the parameter range. A numerical version of the loop model has also been implemented in Fortran[2] and is used to investigate transitory responses of the loop model.

The loop model is derived in Section 2. A semi-analytic method to compute possible steady states in the thermohaline loop are presented in Section 3. Section 4 presents the linear stability analysis of the loop system. Conclusions as well as a discussion on how the box and loop models compare with each other will be provided in Section 5.

2. The loop model

The loop model used here is a modified version of the model of Pollmann et al. (2015), which is itself a development of the Wunsch (2005) model. It consists of a one-dimensional circular loop of constant cross section with one source and one sink of
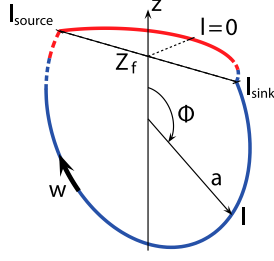
Fig. 2. Schematics of the folded loop configuration. Adapted from Pollmann et al. (2015).

temperature and salinity situated at the same height Z_f (Fig. 2). The loop is assumed sufficiently thin that there is no cross-sectional temperature, salinity or velocity gradient and the velocity vector is tangent to the loop direction.

The loop model of Pollmann et al. (2015) introduced two important innovations compared to Wunsch (2005). It allows for the use of a non-linear EOS, including a representation of cabbeling and thermobaricity effects. It also introduced a so-called folded geometry, wherein the upper part of the circular loop situated above the level of applied forcings (Z_f) has been folded down in the horizontal direction (Fig. 2). The non-folded version of the loop is comparable to previously published versions (e.g. Wunsch, 2005; Welander, 1967; Huang, 1999), but it is unrealistic in the sense that its 'surface' thermohaline forcings are applied below the top of the loop. In the folded version, the upper part of the circulation is at the same height as the applied buoyancy source and sink. It can thus capture the observed asymmetry between the fundamentally horizontal surface flow and the interior return flow.

Compared to Pollmann et al. (2015), the loop model used here can employ both relaxation and fixed-flux forcing types, i.e. mixed boundary conditions. At the point source and sink, the temperature will be relaxed towards prescribed values, while a fixed salinity flux will be imposed.

2.1. *Governing equations*

Two coordinates play an important role in the loop model: the loop (curvilinear) coordinate l and the loop angle $\phi = l/a$, where a is the radius of the loop. The loop coordinate l varies clockwise between zero at the top of the loop and the circumference $L = 2\pi a$.

Making the seawater Boussinesq approximation (Young, 2010; Roquet et al., 2015), the continuity equation reduces to the condition that the loop tangential velocity w must be constant around the loop, $\partial w / \partial l = 0$. This means there is no momentum advection and the pressure term can be removed by integrating along the loop. Furthermore, as in Pollmann et al. (2015), the flow is assumed inertialess (Stokes

flow type) so that the time-dependent term in equation (2) can be neglected, leading to a simple scalar equation for velocity,

$$w = \frac{1}{\varepsilon}\left(-\frac{g}{L}\int_0^L \sigma \frac{\mathrm{d}z}{\mathrm{d}l}\,\mathrm{d}l + \tau\right) \qquad (2)$$

where ε is a Rayleigh friction coefficient, g the gravitational constant, $\sigma(l)$ the density, $z(l)$ the height as a function of the loop coordinate and τ an applied torque. In the circular case, the height varies along the loop as $z = a\cos(\phi)$, so that its variation is sinusoidal $\mathrm{d}z/\mathrm{d}l = -\sin\phi$. In the case of a folding at level Z_f, the definition of the height z must be modified as follows,

$$z = \min(a\cos(\phi), Z_f). \qquad (3)$$

A general definition for the local change of height as a function of the loop coordinate, valid for both circular and folded configurations, is given as,

$$\frac{\mathrm{d}z}{\mathrm{d}l} = -M(l)\sin\phi \qquad (4)$$

where the mask $M(l) = 1 - b_{\mathrm{fold}}[1 - \mathcal{H}(l - l_{\mathrm{sink}}) + \mathcal{H}(l - l_{\mathrm{source}})]$ is everywhere equal to 1 in the circular case, while in the folded case, it is equal to 0 above the height Z_f and 1 elsewhere. b_{fold} is a boolean equal to 1 if the loop is folded and 0 otherwise. $\mathcal{H}$ is the Heaviside step function. l_{sink} and l_{source} are the loop coordinates of the sink and source, respectively. As we restrict ourselves to the case of horizontal convection, $l_{\mathrm{source}} = L - l_{\mathrm{sink}}$, with $0 < l_{\mathrm{sink}} < L/2$.

In the momentum balance, we have left the possibility to apply an externally forced torque, using the parameter τ. Although previous studies (e.g. Wunsch, 2005; Hazewinkel et al., 2012) have considered the effect of imposing a surface wind stress, here the parameter τ is not necessarily meant to represent a surface wind stress. Rather, it is a mean to drive a 'non-thermohaline' contribution to the overturning circulation, i.e. not forced by buoyancy fluxes. τ may represent a combination of processes such as Ekman pumping, eddies in the Southern Ocean, or zonal pressure gradient in the Northern Basins. No attempt is made here to relate explicitly τ to these processes (as e.g. in Gnanadesikan, 1999). We will instead focus on the consequences of applying a torque in the loop model. In doing so, we will see how the analysis of loop properties can be greatly simplified by introducing this free parameter.

The simplified non-linear EOS of Roquet et al. (2015a) will be used:

$$\sigma = -\frac{C_b}{2}(\theta - \theta_0)^2 + T_h z\theta + \beta_0 S \qquad (5)$$

where θ is the conservative temperature and S the absolute salinity. Parameters C_b and T_h control the strength of cabbeling and

thermobaricity, respectively. β_0 is a constant haline contraction coefficient and θ_0 sets the temperature at which surface thermal expansion vanishes. Realistic values for this EOS are proposed in Roquet et al. (2015a) who demonstrate that this simplified EOS yields a reasonable global circulation when implemented in an ocean GCM.

Finally, two equations are needed for the evolution of temperature and salinity in the loop,

$$\frac{\partial \theta}{\partial t} = -w\frac{\partial \theta}{\partial l} + \kappa\frac{\partial^2 \theta}{\partial l^2} + \mathcal{F}_\theta \qquad (6)$$

$$\frac{\partial S}{\partial t} = -w\frac{\partial S}{\partial l} + \kappa\frac{\partial^2 S}{\partial l^2} + \mathcal{F}_S \qquad (7)$$

where κ is the eddy diffusivity (same for temperature and salinity), and $\mathcal{F}_\theta$ and $\mathcal{F}_S$ are the forcing distributions applied to temperature and salinity, respectively.

A relaxation form is used for temperature, while a fixed-flux form is used for salinity,

$$\mathcal{F}_\theta = 2\pi \xi_\theta \left[(\theta_{\text{source}}^0 - \theta)\delta(l - l_{\text{source}}) \right.$$
$$\left. + (\theta_{\text{sink}}^0 - \theta)\delta(l - l_{\text{sink}}) \right] \qquad (8)$$

$$\mathcal{F}_S = 2\pi \eta \left[\delta(l - l_{\text{source}}) - \delta(l - l_{\text{sink}}) \right] \qquad (9)$$

where δ is the classical Dirac delta function, ξ_θ the temperature relaxation parameter, θ_{source}^0 (θ_{sink}^0) the reference temperature to which temperature is relaxed at the source (sink), and η the strength of the salt forcing. Note that in the real ocean, salinity variations are forced by freshwater fluxes. The approximation of using equivalent salt fluxes rather than freshwater fluxes may have an impact on the solution, especially in cases where diffusion is negligible (Dewar and Huang, 1996). However, it is generally a reasonable approximation that allows for a substantial simplification of the model equations.

2.2. Non-dimensionalisation

For the non-dimensionalisation, variables are decomposed as $T = T^0 + T^\star T^s$, where the star symbol indicates the non-dimensionalised variable, the superscript s the corresponding scaling factor and the superscript 0 the reference value.

We use the radius of the loop as the length scale $l^s = a$ (the non-dimensional loop coordinate $l^\star$ is merely the loop angle ϕ so that non-dimensional equations may be equally expressed as a function of ϕ). The velocity scale is set to the frictional scale $w^s = g\sigma^s/\varepsilon$. The inverse Rayleigh number $R = \kappa\varepsilon/ag\sigma^s$ represents the ratio between diffusive and advective time scales, and the Grasshof number $F = g\sigma^s/a\varepsilon^2$ represents the ratio between frictional and advective time scales.

Reference values are set to zero for all the variables except for temperature, salinity, height and density. The temperature reference value is chosen to be $\theta^0 = (\theta_{\text{source}}^0 + \theta_{\text{sink}}^0)/2$, so that we may define a unique parameter $\theta_r^\star \equiv \theta_{\text{source}}^{0\star} = -\theta_{\text{sink}}^{0\star}$.

The salinity reference value is taken as the loop average salinity, which cannot change through time owing to the (balanced) fixed-flux forcing type. The height reference value is the mean height of the loop.

The non-dimensional system of equations then becomes:

$$w^\star = \frac{1}{2\pi}\int_0^{2\pi} \sigma^\star M \sin(\phi)\,\mathrm{d}\phi + \tau^\star \qquad (10)$$

$$\sigma^\star = -\left(1 + \frac{\lambda^\star}{2}\theta^\star - \mu^\star z^\star\right)\theta^\star + S^\star \qquad (11)$$

$$\frac{\partial \theta^\star}{\partial t^\star} = -w^\star\frac{\partial \theta^\star}{\partial \phi} + R\frac{\partial^2 \theta^\star}{\partial \phi^2} + \mathcal{F}_\theta^\star \qquad (12)$$

$$\frac{\partial S^\star}{\partial t^\star} = -w^\star\frac{\partial S^\star}{\partial \phi} + R\frac{\partial^2 S^\star}{\partial \phi^2} + \mathcal{F}_S^\star \qquad (13)$$

where the following periodic boundary conditions are applied,

$$\mathcal{F}_\theta^\star = 2\pi \xi_\theta^\star \left[(\theta_r^\star - \theta^\star)\delta(\phi - \phi_{\text{source}}) \right.$$
$$\left. - (\theta_r^\star + \theta^\star)\delta(\phi - \phi_{\text{sink}}) \right] \qquad (14)$$

$$\mathcal{F}_S^\star = 2\pi \eta^\star \left[\delta(\phi - \phi_{\text{source}}) - \delta(\phi - \phi_{\text{sink}}) \right] \qquad (15)$$

The non-dimensional EOS has been obtained by developing Equation (5) noting that a Boussinesq fluid is insensitive to the addition or subtraction of any height only-dependent function in the EOS. Moreover, the density scaling factor is used to scale both temperature and salinity fields, $\sigma^s = \alpha_0\theta^s = \beta_0 S^s$. Simple algebraic manipulations then show that $\alpha_0 = C_b(\theta^o - \theta_o) - T_h z^o$, $\sigma^o(z) = -C_b(\theta^o - \theta_o)^2/2 + \beta_o S^o + T_h\theta^o z$, and the nonlinear parameters must have the following values, $\lambda = C_b\theta^s/\alpha_o$ and $\mu = T_h a/\alpha_o$. Note that this scaling is only possible if $\alpha_o \neq 0$ which should always be the case in the real ocean, because $\theta_0 \simeq -4.5^oC$ and $z^o < 0$ (see Roquet et al., 2015a). Setting one of the three scaling factors σ^s, θ^s or S^s is then sufficient to set the two others.

In the following, only non-dimensionalised equations are considered so the star symbols will be dropped for simplicity.

2.3. The numerical loop model

The loop model (Equations 10–15) is integrated numerically using a finite-difference discretisation. In the numerical model, the loop is discretised into N grid points (defined by the angle ϕ), each associated with a value of θ and S. There is only one scalar velocity defined diagnostically by Equation (10). There are thus 2N independent variables in the numerical model. The time discretisation uses a leapfrog algorithm, combined with a Robert–

Asselin–Williams filter to prevent divergence of adjacent time steps. The diffusive term is solved explicitly using a forward-time scheme. The application of a temperature relaxation also induces a possibility for numerical instability. Standard stability criteria constrain the time step that can be used to numerically solve the loop model. In practice, a time step of 10^{-5} time units is found suitable for the standard range of loop parameters used in this study. This approximately corresponds to a 1-day time step for a revolution period of 1000 years. More details on the discretisation can be found in Pollmann et al. (2015).

3. Steady states and the torque diagram

3.1. Determination of possible steady states

A semi-analytical method (first described by Welander (1967) but applied here to a more general case) is used to find the steady states of the system of Equations (10)–(13). The method consists in:

(1) determining analytically the steady-state distribution of temperature θ_e and salinity S_e as a function of the loop velocity,
(2) computing the equilibrium buoyancy torque $\mathcal{B}_e(w)$ from steady-state temperature and salinity distributions using the EOS,
(3) finding the velocity values for which the scalar momentum balance (10) is satisfied.

A Matlab code is provided as Supplementary Material for computing the steady-state tracer distribution and the equilibrium buoyancy torque.

For any given value of velocity w, steady-state distributions of temperature and salinity can be determined analytically by solving the advective–diffusive balance equation:

$$-w \frac{\partial X_e}{\partial \phi} + R \frac{\partial^2 X_e}{\partial \phi^2} + \mathcal{F}_X = 0 \qquad (16)$$

where $X_e(\phi)$ is the steady-state distribution of either temperature or salinity. Equation (16) is a one-dimensional differential equation; it is homogeneous everywhere except at the pointwise source and sink where a discontinuity in the gradient of X is introduced. On each side of the loop, the solution has thus to be exponentially decaying, with a boundary layer thickness $\delta_{BL} = |R/w|$ (except for $w = 0$ for which the solution is piecewise linear). The boundary layers are found upstream of the source and sink locations. Considerations of continuity, local budgets of tracer gradients at the source and sink positions, and tracer conservation for the fixed-flux case allow to determine the two remaining unknowns, i.e. the tracer values at source and sink (details in Appendix 1).

Table 1. Parameter list of study cases. Case 1 is used as the reference case throughout this study. Other cases are mainly discussed in section 4c (see Fig. 6).

Case	1	2	3	4	5	6
folded	yes	–	–	–	–	–
Z_f	0.5	–	–	–	–	0
R	0.1	0.05	–	–	–	–
λ	0	–	–	–	0.5	–
μ	0	–	–	–	–	–
θ_r	5	–	–	10	–	–
ζ	1	–	0	–	–	–

Once the temperature and salinity distributions are determined analytically as a function of the loop velocity, the equilibrium buoyancy torque $\mathcal{B}_e(w) = \int_0^{2\pi} \sigma_e M \sin(\phi) \, d\phi / (2\pi)$ can be derived using the EOS (11). Vertical sections of the loop contribute most to the buoyancy torque, and the folded sector does not contribute at all. A steady state solution must then satisfy the momentum balance (10):

$$\mathcal{B}_e(w) = w - \tau \qquad (17)$$

Intersections between the buoyancy torque curve, $\mathcal{B}_e(w)$, and the (linear) friction minus applied torque curve, $w - \tau$, can be determined numerically. Once the steady-state velocity solutions are obtained semi-analytically, it is straightforward to determine the corresponding temperature and salinity distribution using the analytical solution of the tracer equations.

Examples of torque curves are presented in Fig. 3a. Contributions of the temperature relaxation (blue curve), the fixed salinity flux (red curve) and their combination (yellow curve) to the buoyancy torque are shown. The buoyancy torque for the combined case is a linear superposition of the salinity and temperature contributions (salinity and temperature act independently in the simplified EOS). For the shown settings, salinity has a negative effect and thus partially compensates for the positive temperature contribution. The sum of the two curves gives a 'M'-shaped curve which is intersected three times by the frictional torque line (for zero applied torque), yielding three possible steady states (denoted A, B and C). The possibility of multiple states critically depends on the temperature and salinity contributions having similar magnitudes. The salinity curve tends to be steeper than the temperature curve because the salinity flux is constant, while the temperature flux tends towards zero for small velocities. By nature, a fixed salinity flux will always tend to dominate over a temperature relaxation flux for small velocities (and vice versa).

Varying the applied torque can modify the number of possible states. For $\tau = 1$ (dashed black line), the combined case features only one steady state, but in the salinity-only scenario
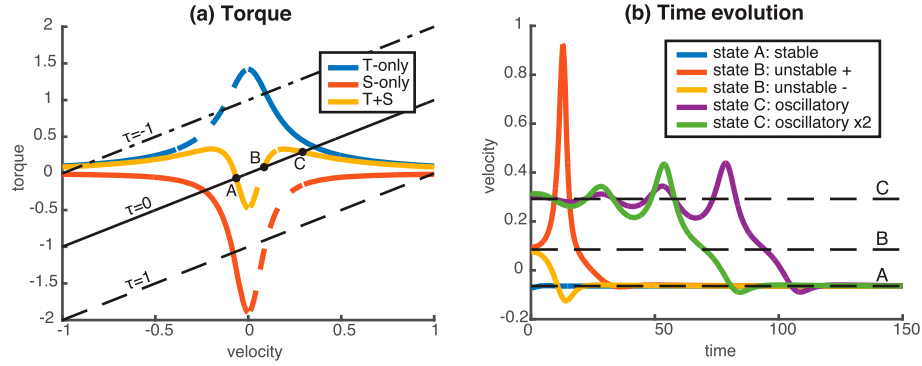
Fig. 3. (a) Torque curves for Case 1 (see Table 1). The buoyancy torque curve $\mathcal{B}_e$ is plotted for a temperature-only (blue curve, T-only), salinity-only (red curve, S-only) or both forcings together (yellow curve, T+S). Possible steady states are determined as the intersection between the buoyancy torque curve and the frictional minus applied torque line. Dashed lines denote unstable regions, where $d\mathcal{B}_e/dw > 1$. Multiple equilibria can be found for the three cases when a suitable torque is applied. For example, for the mixed case with no applied torque (yellow curve intersecting the solid black line), three steady states, denoted A, B and C, are found. (b) Time evolution of the loop velocity for the three steady states A, B and C of the no-applied-torque mixed case T+S, following a small perturbation. A perturbation of 0.01 in velocity is applied around steady states, except for the 'oscillatory x2' case where a 0.02 velocity perturbation is applied instead.

three steady states are possible. For $\tau = -1$ (dashed-dotted line), it is the temperature-only case that produces three possible states. In all cases, multiple equilibria are allowed for only when two different forces tend to oppose each other with comparable strength, either temperature versus salinity, or salinity versus applied torque, or temperature versus applied torque. Cases with three equally important forces acting together at the same time can lead to as many as five possible steady states for a given set of parameters (not shown).

3.2. Transitory regimes and stability of steady states

The stability of the steady states can be partially determined from the shape of the buoyancy torque curve. If the slope of the buoyancy torque is steeper than that of the frictional torque, $d\mathcal{B}_e/dw > 1$, the steady state is unstable because buoyancy increases faster than friction as velocity increases, or decreases faster than friction as velocity decreases (i.e. friction does not provide a sufficient restoring force if velocity is changed). In this case, the flow in the loop is subject to an instability, that we will refer to as the *torque instability*. On the contrary, if the slope of the frictional curve is greater than that of the buoyancy curve, the relation is reversed and any change in velocity results in a restoring force that attempts to undo the change (yet the steady state is not necessarily stable as will be shown later).

Stability inferred from the buoyancy torque curve is incomplete as this curve only captures steady-state properties, ignoring transitory regimes. While it is impossible to obtain oscillations in a two-box Stommel model (Ruddick and Zhang, 1996), it has been long known that such behaviour can emerge in the loop model (Welander, 1967). The stability of the steady states for Case 1 (mixed flux, no applied torque; see Table 1 and Fig. 3a,

yellow line) is further analysed empirically by numerical integration of the discretised equations of the loop model. The model is initialised with the semi-analytical steady-state solutions of temperature and salinity perturbed by a small velocity increment. Time evolutions for five different initialisations are shown in Fig. 3b. These simulations illustrate the range of behaviours found in the loop circulation.

The steady state A (negative velocity) is very stable. Following a small perturbation, the velocity quickly returns to its predicted stable value (blue, Fig. 3b). The middle state B is unstable, consistent with the simple stability criterion based on the buoyancy torque curve. A very different behaviour is observed when the state B is perturbed positively or negatively. For a small positive perturbation (red curve), the velocity grows fast until it reaches a maximum value of about $w = 0.9$ at time $t = 12$, and then decays before stabilising around stable state A at $t = 30$. In the other case (yellow), the velocity directly converges towards the stable state A at $t = 20$.

For the fast steady state C (a stable state according to the stability criterion of the buoyancy torque curve), we observe oscillations that grow with time. The period of oscillations is rather constant ($t \simeq 27$). For a $\delta w = 0.01$ perturbation (purple curve), four oscillations are observed until the loop circulation collapses towards the negative stable state, while it takes only three oscillations for the $\delta w = 0.02$ perturbation (green curve). A tipping point is clearly crossed when the loop velocity gets smaller than the velocity of the middle unstable state B ($w \simeq 0.1$ in this case). Contrary to expectations based on the *simple instability criterion* (based on the buoyancy torque curve), only one steady state is truly stable although it can take long before the oscillatory instability is sufficiently strong to induce a collapse of the circulation.
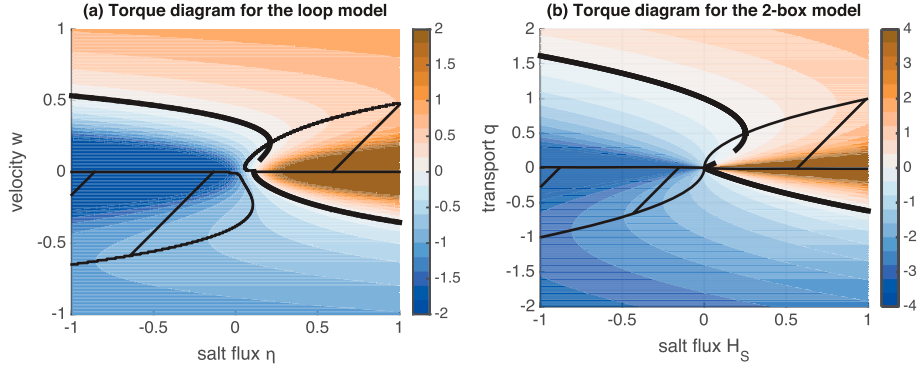
Fig. 4. (a) Torque diagram for Case 1. Colour contours correspond to the torque that must be applied to achieve a steady state, shown as a function of the salt flux intensity η and the loop velocity w. (b) Equivalent torque diagram for the two-box Stommel model, showing the torque τ required to achieve a steady state as a function of the imposed salt flux H_S and the overturning transport q. In both panels, the thick black line indicates possible steady states for $\tau = 0$, which are either stable (solid) or unstable (dashed) based on the simple instability criterion $d\tau/dw < 0$ ($d\tau/dq < 0$ for the two-box model). The corresponding domains of torque instability are hatched. Note the striking similarity between the two torque diagrams.

3.3. *The torque diagram*

The torque balance (17) implies that for any set of loop parameters and loop velocity, a steady state can be achieved by applying a torque compensating for the imbalance between the velocity and the buoyancy torque. This effectively means that the applied torque can be used as an adjustment variable to achieve a unique steady state for any set of loop parameters. This powerful result holds because the equilibrium buoyancy torque is independent of the applied torque.

Instead of using the torque curves to find possible steady states, it is then possible to use a more general tool that will be called the torque diagram. The torque diagram gives the torque that must be applied to achieve a steady state for the chosen parameter values. An example of a torque diagram is shown in Fig. 4a for the standard set of parameters (Case 1, Table 1), except that the salt flux parameter η is now varied between -0.5 and 0.5. The simplicity of the loop model allows for a systematic exploration of the parameter range on a regular η-w mesh grid. By continuity it is then possible to determine possible steady states for a given value of the applied torque, avoiding complications of the continuation methods (Dijkstra and Weijer, 2005).

In Fig. 4a, for example, a typical 'S'-shaped curve is found for the $\tau = 0$ case, including a range of salt flux values (around $0.1 < \eta < 0.2$), for which three steady-state velocities can be found. The multiple steady state at $\eta = 0.15$ found as intersections of the torque curves (Fig. 3) are naturally detected with the torque diagram. The simple instability criterion becomes $d\tau/dw < 0$ in the torque diagram. The corresponding domain of instability is hatched in Fig. 4a, generally structured in two sectors, both at low velocity when salinity tends to compensate the flow (i.e. in areas where the product $\eta w > 0$).

The torque diagram for the loop model can be compared with the Stommel two-box model's one. For an easier comparison, we include the effect of an applied torque τ in the two-box model, which slightly modifies Equation (1) as follows:

$$|q|(1 - q + \tau) - H_S = 0 \qquad (18)$$

The applied torque required to achieve a steady state is then $\tau = q - 1 + H_S/|q|$ (Fig. 4b). The 'S'-shaped curve seen in the torque diagram of the loop model is similarly observed in the two-box model although the transition between negative and positive velocity values has a singularity in the latter case. The difference is partly due to the effective infinite value of the temperature relaxation parameter in the two-box model. Another important difference is that the transport in the box model is proportional to the density difference, while in the loop circulation it depends on the density difference *and* the thickness of the diffusive boundary layer (a function of velocity and diffusion).

4. Linear stability analysis

Oscillatory instabilities have been observed in the loop model that are not possible in a two-box model. A more quantitative approach is needed to analyse their occurrence and properties. To this end, a linear stability analysis of the loop circulation is carried out. This analysis relies on the computation of the temperature and salinity distribution using the semi-analytical method presented in the previous section.

4.1. *Methodology*

Linear stability is computed using a standard methodology (Dijkstra, 2005). Tracer equations are first discretised spatially using centred finite difference techniques, resulting in an

autonomous system of $2N$ ordinary differential equations: N for discrete temperature variables and another N for salinity variables,

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) \qquad (19)$$

where $\mathbf{x}$ is the state vector,

$$\mathbf{x} = [\theta_1 \ldots \theta_N, S_1 \ldots S_N]' \qquad (20)$$

and $\mathbf{f}$ is a smooth vector-valued function. In this formalism, steady-states are simply the states that satisfy $\mathbf{f}(\mathbf{x}_o) = 0$. In practice, the value $N = 360$ was found to be a good trade-off between precision and numerical efficiency.

This set of equations is then linearised about a given steady-state $\mathbf{x}_o$, neglecting quadratic terms, yielding a linear set of equations for the perturbation vector, $\mathbf{x}' = \mathbf{x} - \mathbf{x}_o$,

$$\frac{d\mathbf{x}'}{dt} = \mathbf{J}(\mathbf{x}_o)\mathbf{x}' \qquad (21)$$

where $\mathbf{J} = d\mathbf{f}/d\mathbf{x}$ is the Jacobian matrix.

The next step consists in computing the eigenvalues and eigen-vectors of the Jacobian,

$$\mathbf{J}\hat{\mathbf{X}} = \hat{\mathbf{X}}\mathbf{D} \qquad (22)$$

The matrix $\mathbf{X}' = [\mathbf{x}'_1 \ldots \mathbf{x}'_{2N}]$ contains the 2N eigenvectors (or modes) which form an orthogonal basis of decomposition for any perturbation vector $\mathbf{x}'$. Each eigenvector $\mathbf{x}'_k$ is associated with a complex eigenvalue $\sigma_k = \sigma_k^{Re} + i\sigma_k^{Im}$, found in the diagonal matrix $\mathbf{D}$. Several algorithms exist to decompose the Jacobian, such as the QZ algorithm. In this study, we used the function *eig* of Matlab which automatically selects an optimal algorithm.

For each eigenvector, the solution to equation (21) is of the form,

$$\mathbf{x}'_k = \mathbf{c}_k e^{(\sigma_k^{Re} + i\sigma_k^{Im})t} \qquad (23)$$

where $\mathbf{c}_k$ is the projection of the initial perturbation on the considered eigenvector.

Dynamical system theory shows that 1) the system is linearly stable if none of the Jacobian eigenvalues has a positive real part and 2) the system is at a bifurcation point if at least one of the eigenvalues has a null real part. To analyse stability, it is thus useful to focus on the eigenvalue(s) with the largest real parts.

Two cases must be distinguished when the system is un-stable. Either the eigenvalue with the largest real part has a zero imaginary part, $\sigma_{ti} = \sigma_{ti}^{Re}$. This type of instability, which corresponds to the torque instability, follows an exponentially growing time evolution with a time scale set by the inverse eigenvalue $T_{ti} = 1/\sigma_{ti}$,

$$\mathbf{x}'_{ti} = \mathbf{c}_{ti} e^{t/T_{ti}} \qquad (24)$$

In the second case, there is not one but two eigenvalues with the largest positive real part, that are the conjugate of each other, $\sigma_{oi} = \sigma_{oi}^{Re} \pm i\sigma_{oi}^{Im}$ (the corresponding eigenvectors are also conjugate). This situation induces an oscillatory instability with perturbations of the form:

$$\mathbf{x}'_{oi} = \mathbf{c}_{oi} e^{t/T_{oi}} \cos\left(\frac{2\pi t}{T_{osc}} + \Phi_{oi}\right) \qquad (25)$$

where $\mathbf{c}_{oi}$ and Φ_{oi} are again constants related to the initial perturbation. Two scales are important here: the e-folding time-scale $T_{oi} = 1/\sigma_{oi}^{Re}$ and the oscillation period $T_{osc} = 2\pi/\sigma_{oi}^{Im}$.

Note that it is possible to have several eigenvalues with posi-tive real parts at the same time, for instance, two unstable non-conjugate eigenvalues, or three unstable eigenvalues including two conjugate ones. In these cases, the eigenvalue with greatest real part corresponds to the unstable mode that grows fastest, so it is expected to dominate the observed instability. This is why we focus here only on the eigenvalue with the largest positive real part.

As the perturbation grows, nonlinear effects eventually take over and the trajectory of the loop departs from this linear predic-tion. Typically, after a finite time, the loop trajectory is attracted towards a stable state. In some rare cases where there are no stable states at all, the loop velocity might reach a limit cycle or oscillate between different unstable states (not shown).

4.2. Development of an oscillatory instability

A necessary condition for the development of an oscillatory instability is to have an active compensation between two forces, such as temperature-driven or salinity-driven contributions to the buoyancy torque or the applied torque. The case of a compensat-ing wind stress has been discussed in the literature (Hazewinkel et al., 2012), however, it is generally considered less relevant to the case of the AMOC where temperature gradient and mo-mentum stress are believed to act in the same direction. Here, we are mostly interested in the case of salinity acting against temperature.

Owing to mixed boundary conditions, temperature adjusts faster than salinity. The heat flux is state-dependent, restoring more strongly the further the temperature drifts from the refer-ence values. In contrast, the salt flux is constant and the salinity change close to the forcing locations is inversely proportional to the velocity. That is, a large salinity anomaly can be created whenever the velocity is reduced. The induced salinity anomaly
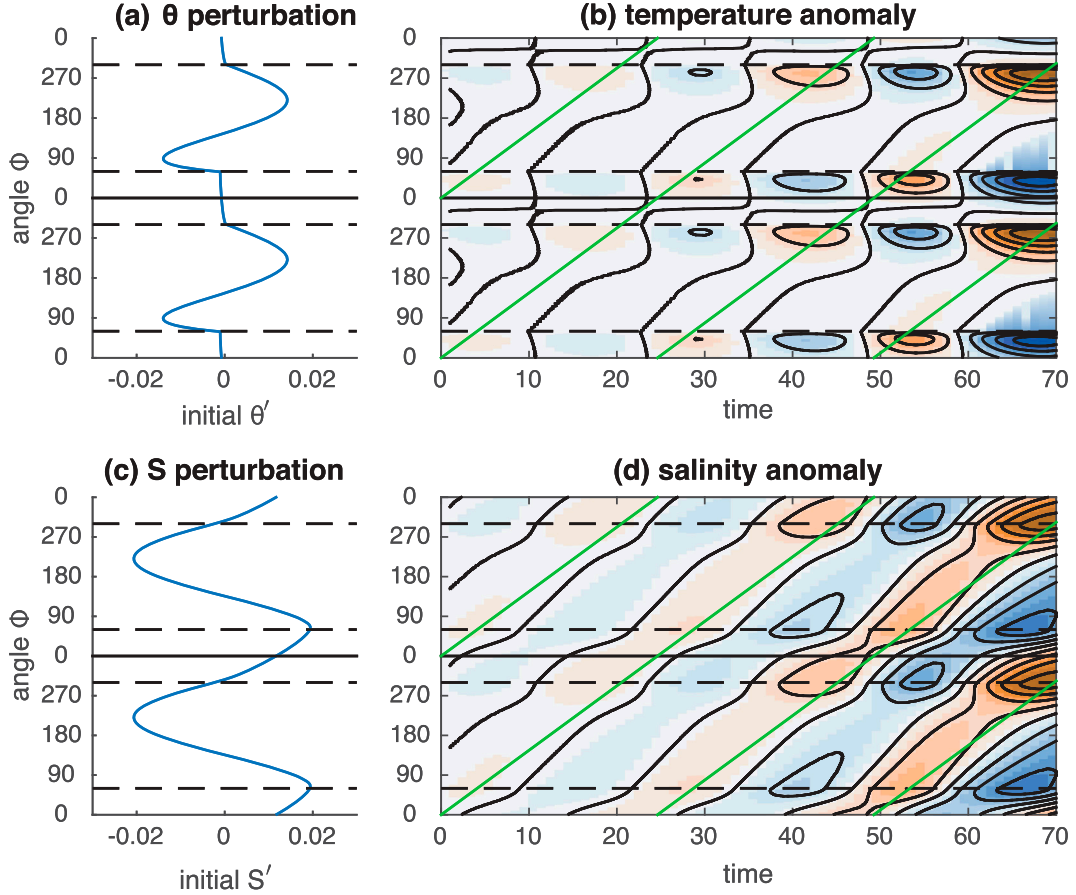
Fig. 5. Time evolution of an oscillatory instability. The initial temperature and salinity perturbations from the state C of Case 1 (no applied torque, see Table 1 and Fig. 3) are shown in panels (a) and (c), respectively. Initial perturbations correspond to the most unstable mode, scaled so that the velocity perturbation equals 0.01. Temperature and salinity anomalies as a function of time are shown in panels (b) and (d), respectively. Anomalies are defined as the difference with the distributions of steady state C. As the anomaly signal propagates around the loop, it amplifies exponentially. The phase of the anomaly as predicted by linear theory is shown in green.

can in turn generate variations in velocity that may promote further amplification of the initial anomaly.

The key to growing oscillations is thus to have a slow loop velocity whenever the positive salinity anomaly crosses the salt source (maximising its growth), and a fast loop velocity when it reaches the salt sink (minimising its decay). This configuration can only happen when salinity is not driving the circulation, otherwise positive salinity anomalies at the salt source would correspond to a minimum (in magnitude) in velocity and the anomaly could not grow.

To develop an understanding of the mechanisms that lead to an oscillatory instability, the fast steady state C of Case 1 (affected by an oscillatory instability, Fig. 3b) is now studied in detail to test if the instability develops in a manner consistent with the linear stability analysis (Fig. 5).

Linear theory predicts two conjugate eigenvalues (i.e. an oscillatory instability) with values $0.038 \pm 0.255\,i$ for stable state C. As the steady-state velocity is equal to $w = 0.291$,

$T_{oi}/T_{\rm rot} = 1.22$ and $T_{\rm osc}/T_{\rm rot} = 1.14$ are both close to unity. This means that the amplitude of the oscillation increases by a factor $e^{1.22} \simeq 3.4$ during each period, in good agreement with the amplification of the successive velocity oscillations in Fig. 3b. Also, since the revolution period is $T_{\rm rot} \simeq 21.6$, the predicted period of oscillation should be $T_{\rm osc} \simeq 24.6$, which is in good agreement with the numerical value of ~ 27.

The most unstable anomaly pattern is described by the real part of the unstable conjugate eigenvectors (Fig. 5, left panels), while the time evolution of this initial anomaly is plotted in the right panels. The asymmetry in forcing types is well reflected in the contrasting distributions of temperature and salinity. Temperature anomalies are induced upstream of the source and sink (at $\phi_{\rm source} = 300°$ and $\phi_{\rm sink} = 60°$, respectively) by velocity fluctuations: a velocity increase induces an anomalous advection of cold water upstream of the thermal source (cold anomaly) while it simultaneously produces a warm anomaly upstream of the thermal sink (the opposite is observed for a velocity decrease).

However, temperature relaxation efficiently removes anomalies as they cross the thermal source and sink: the temperature signal carries almost no memory of past anomalies.

Rather, the memory of the system resides in the salinity field: the salinity anomalies have a quasi-sinusoidal pattern propagating and amplifying over time (Fig. 5d). The growth is especially marked whenever the positive salinity anomaly crosses the salt source. To understand the observed pattern, one must keep in mind that a positive anomaly of salt (and negative anomaly of temperature) near the source (at $\phi \simeq 300°$) produces a slowing down of the loop circulation. This results in a larger salt flux per unit length at the salt source, and thus further growth of the positive salt anomaly. In contrast, when the positive salt anomaly reaches the position of the salt sink, the velocity is increased reducing the ability of the sink to cancel the anomaly. For the same reasons, the negative salt anomaly also tends to grow with time, which is why the oscillation is able to amplify.

Finally, note that the phase speed of anomalies is well predicted by the linear theory (green lines in the right panels of Fig. 5). However, the linear theory is unable to predict the moment when the loop velocity is attracted by a different steady state. In practice, it happens when the velocity decreases below a critical threshold of $w_{critic} \simeq 0.2$, below which it becomes attracted by the unstable steady state before collapsing towards the negative stable steady state.

4.3. Bifurcation diagram

Using the same mesh grid as for the torque diagram, it is possible to map the linear stability properties by computing the Jacobian eigenvalues. Three possible states are defined:

(1) stable state: no eigenvalue with positive real part.
(2) unstable state: the eigenvalue with the largest real part is real and positive, which produces a torque instability.
(3) oscillatory state: the two eigenvalues with largest positive real part are complex conjugates, which produces an oscillatory instability.

The transition between a stable state and an oscillatory state is called a Hopf bifurcation, while the transition between a stable state and an unstable state is a saddle-node bifurcation. Contrary to the 2-box model, which only allows for saddle-node bifurcations, Hopf bifurcations are common in the loop model. Note that an oscillatory state is linearly unstable, yet it can converge towards a small-amplitude limit cycle if the associated Hopf bifurcation is supercritical (Dijkstra, 2005). In this case, the limit cycle is said to be orbitally stable. We left the analysis of limit cycle stability properties for a future study.

The bifurcation diagram of Case 1 is presented in Fig. 6a. Regions of torque instability (dark grey) are well approximated by the simple instability criterion ($d\tau/dw < 0$, hatched domain), although the domain of torque instability is wider showing that
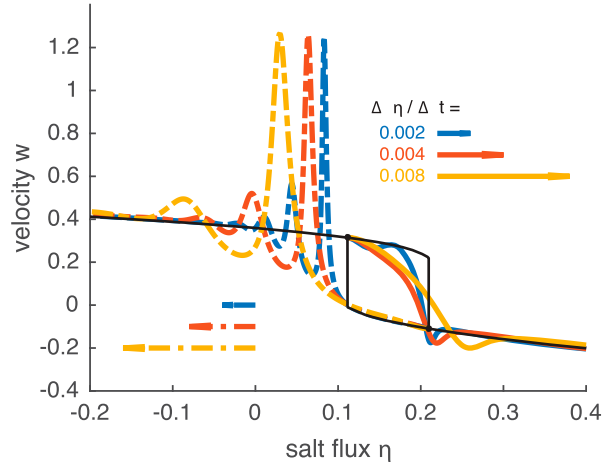

Fig. 8. Transitory response when the salt flux parameter is varied linearly at different time rates ($\Delta\eta/\Delta t = 0.002$ in blue, 0.004 in red and 0.008 in yellow), starting from the fast mode and increasing the salt flux (solid lines), or starting from the slow mode and decreasing the salt flux (dashed lines). The theoretical fast and slow steady-state velocities are superimposed (black solid lines). When the salt forcing is increased starting from the fast mode, the velocity starts to decrease before it reaches the turning point due to the development of an oscillatory instability. By contrast, when the salt is decreased starting from the slow mode, the recovery begins only after the turning point has been passed.

this simple criterion provides a sufficient but not necessary condition for torque instability. What the linear stability analysis also reveals is the presence of large sectors of oscillatory instability (light grey) substantially reducing the domain of stability (white).

Oscillatory instabilities develop almost exclusively in the upper right and lower left quadrants, i.e. where the salinity forcing induces a torque opposing the loop velocity. This is consistent with the idea that such oscillation can only exist in the presence of antagonistic forces.

The transition between torque instability and stability usually happens where the velocity changes sign. The position of the transition between stability and oscillatory instability is found along a slanted curve (slope $\simeq 7$), showing that the salinity forcing must be sufficiently strong to generate growing oscillations. The transition happens at the point where the growth of salt anomalies due to velocity oscillations is exactly cancelled by the diffusive erosion of anomalies.

The linear stability analysis predicts the e-folding timescale of instability. The ratio between this timescale and the revolution time is shown in Fig. 7a using a logarithmic scale. This ratio is much smaller than unity in the case of torque instability. For regions of oscillatory instability, it varies substantially from $\sim$2-3 near the Hopf bifurcation (i.e. the instability develops very slowly there) to values smaller than unity near the transition to torque instability. The period of oscillation can also be predicted in the case of oscillatory instability (see Fig. 7b). Except very
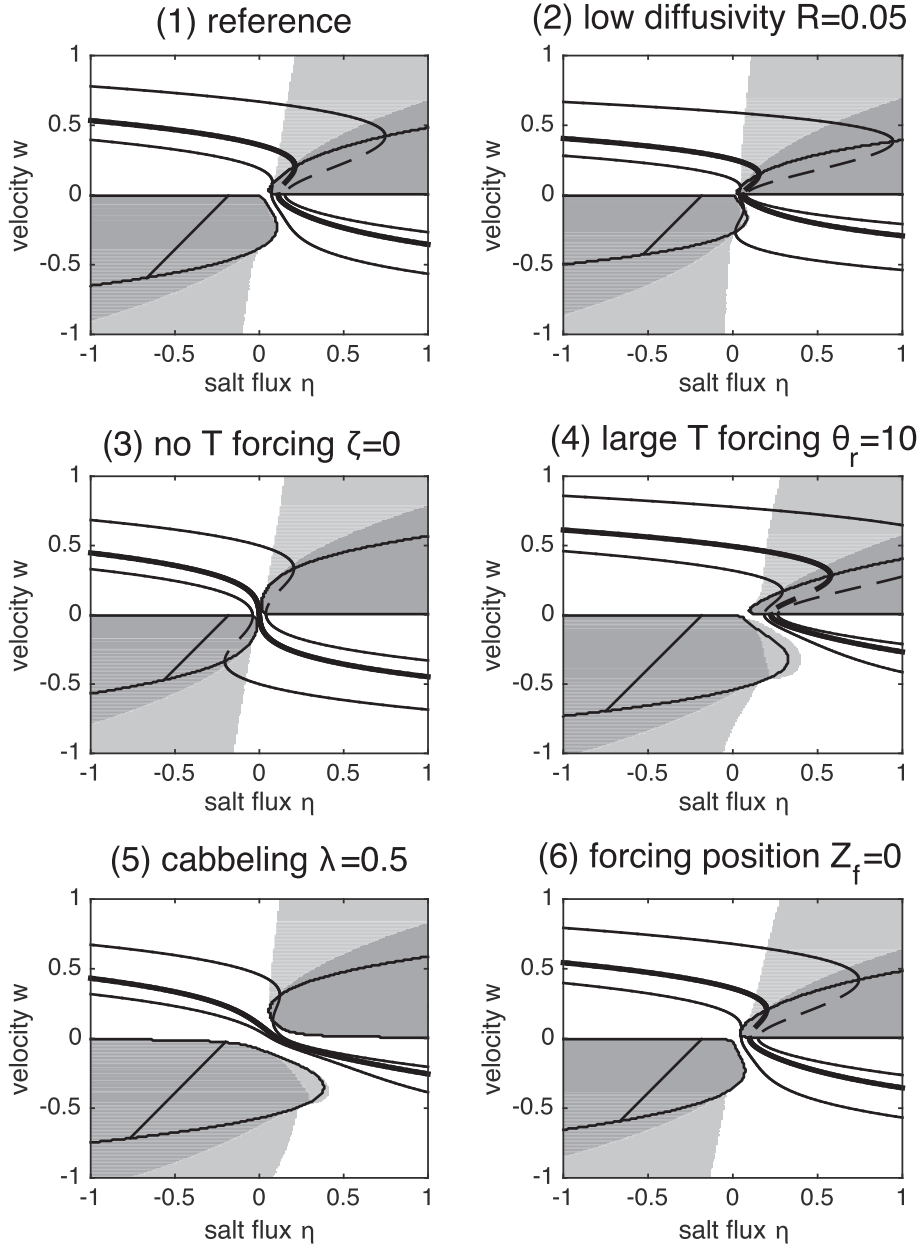
(1) reference

(2) low diffusivity R=0.05

(3) no T forcing ζ=0

(4) large T forcing θ_r=10

(5) cabbeling λ=0.5

(6) forcing position Z_f=0

Fig. 6.   Bifurcation diagrams for a variety of configurations (listed in Table 1). The bifurcation diagram indicates, based on the linear stability analysis, ranges of [white] stability, [light grey] oscillatory instability and [dark grey] torque instability. Regions of torque instability based on the simple instability criterion are hatched, showing that this criterion tends to underestimate the actual domain of torque instability. Each case differs from the reference case by only one modified parameter. The steady-state velocity for a zero applied torque $\tau = 0$ follows the thick black contour. Thin contours associated with $\tau = -0.5$ and $\tau = 0.5$ are also superimposed, found above and below the zero contour, respectively.

close to the torque instability transition, it is everywhere near unity: instability has a period similar to the revolution period. This is understandable as the instability is primarily generated by the loop flow's advecting of a growing salt anomaly.

Following the zero-torque curve in the bifurcation diagram (thick black line in Fig. 6a), starting from a negative value of the salt forcing where the states are stable, it is instructive to see how the stability properties evolve as one gradually increases the salt
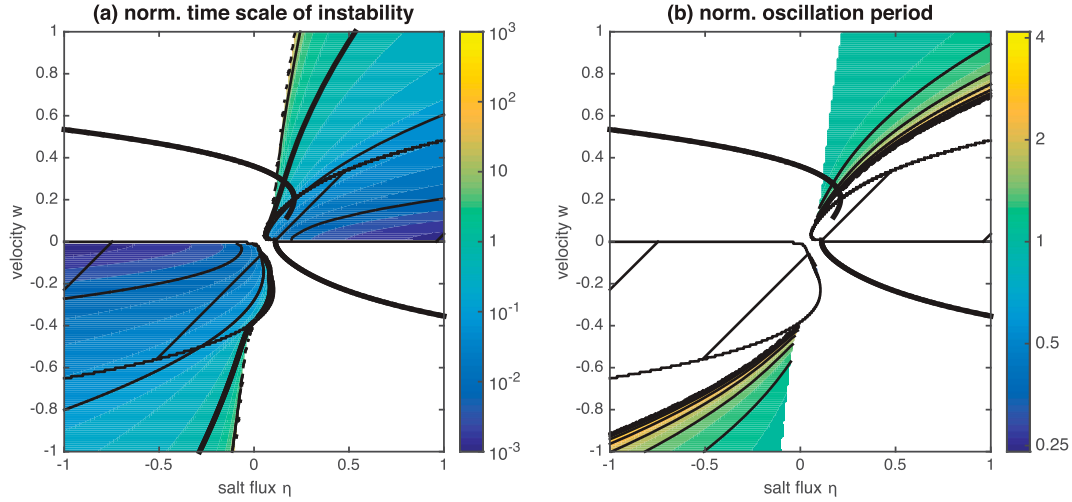
Fig. 7.   (a) Ratio between the e-folding scale of instability and the revolution time (in logarithmic scale) for Case 1, showing that instabilities develop much faster in the domain of torque instability than for oscillatory instabilities. (b) Ratio between the oscillation period and the revolution time (in logarithmic scale) for Case 1, defined only for oscillatory states. The oscillation period is generally of same order of magnitude as the revolution timescale.

forcing. For a negative salt forcing ($\eta < 0$), the loop circulation is temperature-controlled, slowly decaying with increasing salt forcing values as the contribution of salt on the torque rises. Two events happen almost simultaneously (which may be fortuitous) when the salt forcing value reaches $\eta \simeq 0.1$: a Hopf bifurcation is crossed, so that the steady state becomes subject to an oscillatory instability, while two new steady states appear at low velocity following a saddle-node bifurcation, one unstable ($w = 0^+$) and the other stable ($w = 0^-$). A second saddle-node bifurcation is observed at $\eta \simeq 0.22$, beyond which the fast temperature-controlled state vanishes and only the slow salinity-controlled state remains.

The presence of the oscillatory instability before the saddle-node bifurcation along the fast branch means that, in contrast to the simple two-box model, the region of effective bi-stability is largely reduced if not entirely removed in the loop model. Importantly, this remains true when a positive torque is applied to the loop. The dynamical behaviour can be more complex when a negative torque is applied since, as mentioned above, it also allows for interactive regimes between temperature and torque forcings independent of the salt forcing.

The sensitivity of the bifurcation diagram to a few parameters of the loop is also shown in Fig. 6. The general topology of bifurcation diagrams seems robust to parameter changes. Decreasing diffusivity increases the area of oscillatory instability (case 2). Removing the temperature forcing (case 3) produces a diagram exactly symmetrical about the origin, with a monostable no-torque case for any value of η. Raising the reference temperature (case 4) modifies the diagram quite substantially, increasing the area of instability and curving the transition lines. This is expected as a larger salt forcing will be needed to balance the larger temperature forcing.

Adding a large cabbeling effect in the EOS (case 5) can also have a profound effect on stability properties. Here, it removes occurrences of multiple equilibria for the no-torque case, but not for the $\tau = 0.5$ case. In the latter case, it is able to generate a small range of η-values for which there is no stable state at all. Adding a thermobaricity effect (not shown) has a smaller impact on the bifurcation diagram. Modifying the height at which the source and sink of heat and salt are applied (case 6) can also modify the bifurcation diagram, but does not change its overall shape.

4.4. *Transitory response in the presence of an oscillatory instability*

Being in the range of oscillatory instability does not mean that the loop would jump to another steady state right away. If the initial perturbation is small enough, it can take several revolutions around the loop before the instability grows sufficiently large to make the system reach another steady state. In the example shown in Fig. 3b, the oscillatory instability leads to a transition towards the slow stable state after four revolutions in the case with a $w = 0.01$ velocity perturbation but after only three oscillations for a doubled perturbation.

Figure 8 shows examples of the transitory response to a gradual change of the salt forcing, using the Case 1 configuration with no applied torque. In the first set of experiments (solid coloured lines), the salt forcing is increased linearly starting from the first bifurcation point at $\eta = 0.11$ at three different rates. In the three cases, the collapse begins before the turning point of the fast mode (at $\eta = 0.21$) because of the presence of the oscillatory instability.

In the second set of experiments (dashed coloured lines), the salt forcing is reduced linearly, again at three different time rates starting from the value $\eta = 0.21$. When the salt forcing is reduced and the turning point of the slow branch is crossed, the loop velocity starts a rapid increase towards the fast state. Before reaching its new equilibrium, the loop velocity undergoes a large overshoot followed by several oscillations of decreasing amplitude. The maximum amplitude of the velocity following the recovery is about four times larger than the steady-state value.

Note that a similar transitory response would be observed in the presence of a positive torque. Applying a torque stabilises the loop insofar as a larger salt forcing is required to produce a collapse of the velocity. Indeed, the salt forcing must now compensate for both the thermal forcing and the applied torque. However, once this threshold is passed, the transitory response and stability properties will be qualitatively similar to the no-torque case. Applying a negative torque can partially compensate the temperature forcing and can therefore lower the instability threshold. This would happen only if the applied torque and temperature forcing produce torques of similar magnitudes.

5. Summary and discussion

In this paper, we have investigated the dynamical response of a thermohaline loop model to mixed boundary conditions. In particular, we have analysed the conditions of existence of multiple equilibria and the stability properties of the loop model.

This study required the development of a set of tools, both analytical and numerical. A methodology has first been described to derive semi-analytical steady-state solutions of the loop model (section 3). The term 'semi-analytical' refers to the fact that distributions of temperature and salinity are obtained analytically, while the steady-state velocity is obtained using a root-finding numerical algorithm such as Newton's method. It has been shown that the applied torque can be used as an adjustment variable so that a unique steady-state solution exists for any set of loop parameters.

A linear stability analysis has then been carried out on a finite difference approximation of the loop model (section 4). It has been shown that two types of instabilities can be found: a torque instability where the velocity increases exponentially away from the equilibrium and an oscillatory instability where velocity oscillations of growing amplitude are generated until some critical threshold is crossed and the loop velocity is attracted towards a different steady state. Note that the stability analysis of the loop model does not require to deploy adjoint techniques or the continuation method as required for GCMs (Dijkstra and Weijer, 2005).

Welander (1967) described a similar oscillatory instability when heating happens below cooling (Rayleigh–Benard convection). The existence of growing instabilities has also been described in the horizontal convection case (Yuan and Wunsch, 2005; Sévellec et al., 2006). Here, we provide a more systematic

demonstration of its ubiquity in loop models, treating cases with a non-linear EOS, a non-circular geometry or an applied torque. Note that, although we only focused on the horizontal convection case in this paper, the semi-analytical and numerical tools that we presented can be used wherever the sources and sinks of heat and salt are placed around the loop.

The loop model can be viewed as a one-dimensional analogue of the Stommel two-box model in the sense that the active tracers, namely temperature and salinity, are described along the one-dimensional loop instead of 0D boxes. The two-dimensional analogue corresponds to section models, while GCMs are the three-dimensional equivalent. It is of interest to study and compare this hierarchy of models to better grasp their inherent limitations, but also to develop the most adequate concepts to understand the real system.

As for the two-box model, the loop model features a non-linear response to changing salt fluxes, eventually leading to a collapse of the circulation from a temperature-controlled fast mode to a salinity-controlled slow mode (Fig. 4). Most GCMs do agree with the prediction that if the fresh water forcing in the North Atlantic increased sufficiently, a collapse of the current temperature-dominated flow would occur (Stouffer et al., 2006). The occurrence of hysteresis is also observed in several coupled models, although not systematically (Hawkins et al., 2011; Yin et al., 2006; Rahmstorf et al., 2005; Gregory et al., 2003).

However, several important differences are also found compared to the two-box model:

(1) the loop model has no singularity in the vicinity of the zero velocity state. This is because, in our version of the loop model, diffusion is explicitly accounted for and salinity gradients remain finite everywhere,

(2) the loop model has a more realistic treatment of the buoyancy torque contribution, which varies nonlinearly with the variable thickness of the advective–diffusive boundary layer (e.g. Park and Bryan, 2000; de Boer et al., 2010),

(3) finally, the loop model can produce velocity oscillations that may be growing exponentially with time. The period of such oscillations is independent of the forcing, but instead depends on the intrinsic properties of the loop system. Such oscillatory instability is simply impossible in a two-box model (Ruddick and Zhang, 1996).

One major implication of the existence of such oscillatory instability is that the range of bi-stability can be significantly reduced in the loop model compared to the box model. In some cases, the range of bi-stability is even replaced by an oscillatory regime with no stable state at all. This means that the two-box model might not be a suitable model to predict the extent of the bi-stable range, as it is not accounting for possible oscillatory instability. A similar instability through growing oscillations has not been observed in a GCM to our knowledge. However, this is

not surprising owing to the timescale of this type of instability, scaling as the overturning time scale itself. A GCM should be integrated several hundreds of years with a suitable E-P forcing to get a chance to observe the entire cycle of development of such instability. Our study suggests that oscillatory instabilities may play a role in the real climate system on paleo-climate timescales, imprinting a tendency to oscillate at a very low frequency dictated by the MOC transport value.

Here, we have chosen to focus on a comparison of the loop model with the two-box model only, however other box models have been proposed as conceptual models of the MOC in the literature. Of particular relevance, the three-box version proposed by Rooth (1982) better accounts for the inter-hemispheric nature of the MOC – see (Marotzke, 2000) for a discussion of this model. A comparison between the three-box model and the loop model could be attempted by applying three point sources/sinks of temperature and salinity instead of two in the upper branch of the folded loop model; this task, however, is left for a future study. Also, Tziperman et al. (1994) proposed a four-box model, in which the two upper boxes receive the forcing and the other two are located below. Interestingly, this model produces an oscillatory instability just like the loop model. This is perhaps not surprising as the four-box model of Tziperman et al. (1994) can be seen as a loop model in its most extreme discretised form ($N = 4$).

Several aspects of the loop model play a critical role in setting its equilibrium and stability properties, most significantly the mixing intensity, the relative strength of thermal and haline forcings, the applied torque, the non-linear EOS and the loop geometry. Mixing in the loop is represented by a Fickian diffusion with a constant diffusivity value. There would be several ways to improve the representation of mixing in the loop, and it would be interesting as a future study to analyse the sensitivity of the loop to the mixing representation. Following Huang (1999) and Nilsson and Walin (2001), an energy-conserving mixing representation would capture more realistically the variations of mixing intensity with global stratification. Static instability in weakly stratified regions can cause large, localised and sudden variations of the effective mixing intensity. Welander (1982) described heat–salt oscillations that can arise from such instabilities due to a so-called convective feedback. This was shown in a simple two box model with heat and salt fluxes, where boxes are placed one over the other as opposed to the Stommel two-box model where the boxes are placed side by side. It would be useful to analyse whether similar oscillations could be produced in the thermohaline loop model.

The task of comparing the two-box model and the loop model has proved instructive. On one hand, many similarities have been found, including the salt advective feedback producing a fast and a slow state and the possibility of abrupt shifts between the two. On the other, a fundamental innovation is introduced in the loop model with the possibility for oscillatory instabilities. The analysis of these simple models is very useful to develop

appropriate concepts to study the large-scale ocean circulation, yet one must keep in mind that the real ocean is far more complex than zero-dimensional two-box models or one-dimensional loop models. The gap already seen between these two conceptual models gives a hint of how far short both models might still be of representing the time-varying three-dimensional ocean circulation, and how careful one should be when using these conceptual models to interpret model simulations and observations such as paleo-proxy.

Acknowledgements

Disclosure statement

Notes

1. Similarly, the 'wind-driven' circulation critically depends on the stratification that would not exist without the thermohaline forcing. This terminology is useful insofar as it highlights the characteristic differences between the adiabatic (wind-driven) and diabatic (thermohaline) components of the circulation.
2. Publicly available at http://github.com/fabien-roquet/loop.

References

de Boer, A. M., Gnanadesikan, A., Edwards, N. R. and Watson, A. J. 2010. Meridional density gradients do not control the atlantic overturning circulation. *J. Phys. Oceanogr.* **40**(2), 368–380. DOI: 10.1175/2009JPO4200.1.

Dewar, W. K. and Huang, R. X. 1996. On the forced flow of salty water in a loop. *Phys. Fluids* **8**, 954–970. DOI: 10.1063/1.868874.

Dijkstra, H. A. 2005. *Nonlinear physical oceanography: a dynamical systems approach to the large scale ocean circulation and El Nino*, Vol. 28, Springer Science, New York, 523 pp.

Dijkstra, H. A. and Weijer, W. 2005. Stability of the global ocean circulation: Basic bifurcation diagrams. *J. Phys. Oceanogr.* **35**(6), 933–948. DOI: 10.1175/JPO2726.1.

Gnanadesikan, A. 1999. A simple predictive model for the structure of the oceanic pycnocline. *Science* **283**(5410), 2077–2079.

Gregory, J. M., Saenko, O. A. and Weaver, A. J. 2003. The role of the atlantic freshwater balance in the hysteresis of the meridional overturning circulation. *Climate Dyn.* **21**(7), 707–717. DOI: 10.1007/s00382-003-0359-8.

Hawkins, E., Smith, R. S., Allison, L. C., Gregory, J. M., Woollings, T. J., and co-authors. 2011. Bistability of the atlantic overturning circulation in a global climate model and links to ocean freshwater transport. *Geophys. Res. Lett.* **38**(10), DOI: 10.1029/2011GL047208.

Hazewinkel, J., Paparella, F. and Young, W. R. 2012. Stressed horizontal convection. *J. Fluid Mech.* **692**, 317–331. DOI: 10.1017/jfm.2011.514.

Held, I. M. 2005. The gap between simulation and understanding in climate modeling. *Bull. Am. Meteorol. Soc.* **86**(11), 1609–1614. DOI: 10.1175/BAMS-86-11-1609.

Huang, R. X. 1999. Mixing and energetics of the oceanic thermohaline circulation. *J. Phys. Oceanogr.* **29**(4), 727–746. DOI: 10.1175/1520-0485(1999)029<0727:MAEOTO>2.0.CO;2.

Keller, J. B. 1966. Periodic oscillations in a model of thermal convection. *J. Fluid Mech.* **26**(3), 599–606. DOI: 10.1017/S0022112066001423.

Malkus, W. V. 1972. Non-periodic convection at high and low prandtl number. *Mem. Soc. R. Sci. Liege* **4**(6), 125–128.

Marotzke, J. 2000. Abrupt climate change and thermohaline circulation: Mechanisms and predictability. *Proc. Nat. Acad. Sci.* **97**(4), 1347–1350.

Marshall, J., Donohoe, A., Ferreira, D. and McGee, D. 2013. The ocean's role in setting the mean position of the inter-tropical convergence zone. *Climate Dyn.* **42**(7), 1967–1979. DOI: 10.1007/s00382-013-1767-z.

Nilsson, J. and Walin, G. 2001. Freshwater forcing as a booster of thermohaline circulation. *Tellus A* **53**(5), 629–641. DOI: 10.1034/j.1600-0870.2001.00263.x.

Park, Y.-G. and Bryan, K. 2000. Comparison of thermally driven circulations from a depth-coordinate model and an isopycnal-layer model. Part i: Scaling-law sensitivity to vertical diffusivity. *J. Phys. Oceanogr.* **30**(3), 590–605. DOI: 10.1175/1520-0485(2000)030<0590:COTDCF>2.0.CO;2.

Pollmann, F., Roquet, F. and Madec, G. 2015. Effects of the asymmetry between surface and interior flow on the dynamics of a thermohaline loop. *J. Phys. Oceanogr.* **45**, 2544–2563. DOI: 10.1175/JPO-D-15-0022.1.

Rahmstorf, S. 2002. Ocean circulation and climate during the past 120,000 years. *Nature* **419**(6903), 207–214. DOI: 10.1038/nature01090.

Rahmstorf, S., Crucifix, M., Ganopolski, A., Goosse, H., Kamenkovich, I, and co-authors. 2005. Thermohaline circulation hysteresis: a model intercomparison. *Geophys. Res. Lett.* **32**(23), DOI: 10.1029/2005GL023655.

Rooth, C. 1982. Hydrology and ocean circulation. *Prog. Oceanogr.* **11**(2), 131–149. DOI: 10.1016/0079-6611(82)90006-4.

Roquet, F., Madec, G., Brodeau, L. and Nycander, J. 2015a. Defining a simplified yet 'realistic' equation of state for seawater. *J. Phys. Oceanogr.* **45**, 2564–2579. DOI: 10.1175/JPO-D-15-0080.1.

Roquet, F., Madec, G., McDougall, T. J. and Barker, P. M. 2015. Accurate polynomial expressions for the density and specific volume of seawater using the teos-10 standard. *Ocean Model.* **90**, 29–43. DOI: 10.1016/j.ocemod.2015.04.002.

Ruddick, B. and Zhang, L. 1996. Qualitative behavior and nonoscillation of stommel's box model. *J. Climate* **9**(11), 2768–2777. DOI: 10.1175/1520-0442(1996)009<2768:QBANOS>2.0.CO;2.

Sévellec, F., Huck, T. and Ben Jelloul, M. 2006. On the mechanism of centennial thermohaline oscillations. *J. Marine Res.* **64**(3), 355–392. DOI: 10.1357/002224006778189608.

Stommel, H. 1961. Thermohaline convection with two stable regimes of flow. *Tellus* **13**(2), 224–230. DOI: 10.1111/j.2153-3490.1961.tb00079.x.

Stouffer, R. J., Yin, J., Gregory, J. M., Dixon, K. W., Spelman, M. J. and co-authors. 2006. Investigating the causes of the response of the thermohaline circulation to past and future climate changes. *J. Climate* **19**(8), 1365–1387. DOI: 10.1175/JCLI3689.1.

Tziperman, E., Toggweiler, J. R., Bryan, K. and Feliks, Y. 1994. Instability of the thermohaline circulation with respect to mixed boundary conditions: is it really a problem for realistic models? *J. Phys. Oceanogr.* **24**(2), 217–232. DOI: 10.1175/1520-0485(1994)024<0217:IOTTCW>2.0.CO;2.

Vellinga, M. and Wood, R. A. 2002. Global climatic impacts of a collapse of the atlantic thermohaline circulation. *Climatic Change* **54**(3), 251–267. DOI: 10.1023/A:1016168827653.

Welander, P. 1967. On the oscillatory instability of a differentially heated fluid loop. *J. Fluid Mech.* **29**, 17–30.

Welander, P. 1982. A simple heat-salt oscillator. *Dyn. Atmos. Oceans* **6**(4), 233–242. DOI: 10.1016/0377-0265(82)90030-6.

Welander, P. 1986. Thermohaline effects in the ocean circulation and related simple models. In: *Large-Scale Transport Processes in Oceans and Atmosphere* (eds. J. Willebrand and D. L. T. Anderson). D. Reidel, Hingham, Mass., pp. 163–200. DOI: 10.1007/978-94-009-4768-9_4.

Wunsch, C. 2005. Thermohaline loops, stommel box models, and the sandström theorem. *Tellus A* **57**(1), 84–99. DOI: 10.1111/j.1600-0870.2005.00093.x.

Wunsch, C. and Ferrari, R. 2004. Vertical mixing energy and the general circulation of the oceans. *Ann. Rev. Fluid Mech.* **36**, 281–314. DOI: 10.1146/annurev.fluid.36.050802.122121.

Wunsch, C. and Heimbach, P. 2013. Two decades of the atlantic meridional overturning circulation: Anatomy, variations, extremes, prediction, and overcoming its limitations. *J. Climate* **26**, 7167–7186. DOI: 10.1175/JCLI-D-12-00478.1.

Yin, J., Schlesinger, M. E., Andronova, N. G., Malyshev, S. and Li, B. 2006. Is a shutdown of the thermohaline circulation irreversible? *J. Geophys. Res. Atmos.* **111**(D12). DOI: 10.1029/2005JD006562, d12104.

Young, W. R. 2010. Dynamic enthalpy, conservative temperature, and the seawater Boussinesq approximation. *J. Phys. Oceanogr.* **40**, 394–400.

Yuan, S. and Wunsch, C. 2005. Stress-driven thermohaline loops. *Phys. Fluids (1994-present)* **17**(6), 066601. DOI: 10.1063/1.1927887.

Appendix 1. Solution to diffusive–advective equations

In this section, we provide the solution to the advection–diffusion tracer equation for a given tracer X defined along a circular path, both for the case of a fixed-flux type of forcing and a relaxation one. In both cases, forcings are applied on a point source and a point sink that are separated by an angle of $\Delta = \phi_+ - \phi_-$,

$$d\frac{\partial^2 X}{\partial \phi^2} - \frac{\partial X}{\partial \phi} + F_+\delta(\phi - \phi_+) + F_-\delta(\phi - \phi_-) = 0 \quad \text{(A1)}$$

where $d = R/w$ scales the boundary layer thickness.

Except at the source and sink, this equation behaves as a simple homogeneous advection–diffusion equation, with an exponential decaying solution. Taking advantage of the symmetries of the equation, we can only treat the case where the prescribed velocity is negative $w < 0$, and chose to place the origin at the

sink position (i.e. $\phi_- = 0$ and $\Delta = \phi_+$). Equation (16) can be related to the prototypical equation (A1) for any non-zero value of w, using the coordinate transformation $\phi = -s(e - e_-)$ where $s = w/|w|$ corresponds to the sign of w.

Owing to the discontinuity introduced in the tracer derivative by the forcings, the tracer distribution is defined piecewise,

$$X = \begin{cases} X^\infty - \frac{C\mathrm{d}F_+}{1+Ce^{-\Delta/d}}e^{-\phi/d}, & \phi < \Delta \\ X^\infty - \mathrm{d}F_+ + \frac{\mathrm{d}F_+}{1+Ce^{-\Delta/d}}e^{-(\phi-\Delta)/d}, & \phi > \Delta \end{cases} \quad \text{(A2)}$$

where $C = (1 - e^{(2\pi - \Delta)/d})/(1 - e^{\Delta/d})$. To obtain this template of the tracer distribution, we used the fact that the distribution must remain continuous across the source and sink positions, that the divergence of the diffusive flux on either side of the sink must be compensated by the forcing, and finally, that to achieve a steady state, the tracer flux at the source must be equal and opposite to the tracer flux at the sink, $F_- = -F_+$. There is one last constant X^∞ that needs to be determined, which will depend on the type of forcing that is applied.

A.1. Analytical solution of the tracer equation: fixed-flux forcing

The forcing functions are constants of equal but opposite amplitude, $F_+ = -F_- = F_o$. In this case, the mean amount of tracer $\overline{X} = \int_0^{2\pi} X(\phi)\,\mathrm{d}\phi/(2\pi)$ has to be conserved along the loop at any time. From this, it is found that,

$$X^\infty = \overline{X} + dF_o(2\pi - \Delta) \quad \text{(A3)}$$

A.2. Analytical solution of the tracer equation: relaxation forcing

The relaxation forcing are as follows,

$$\begin{cases} F_+ = \zeta(X_r - X) \\ F_- = \zeta(-X_r - X) \end{cases} \quad \text{(A4)}$$

where ζ is the restoring constant and X_r the reference value towards which the tracer is relaxed at the source position. A balance can be achieved only if the net flux is zero, i.e. if $X(\phi_+) = -X(\phi_-)$. This gives the necessary constraint to determine X^∞, following some tedious derivations that are not reproduced here. A Matlab code is provided as Supplementary Material to compute analytically the tracer distribution.

Note that both forcing types can easily be applied together on the same tracer using the fact that the equation is linear so that each forcing can be computed separately. In that case, the solution of the relaxation forcing must be computed first and then the steady-state global mean tracer can be determined, a prerequisite to obtain the contribution of fixed-flux forcing on the tracer distribution.