**DSCS**

OXFORD

Review Article

# A numerical framework to understand transitions in high-dimensional stochastic dynamical systems

**Henk A Dijkstra[a,*], Alexis Tantet[a], Jan Viebahn[a],Erik Mulder[a], Mariët Hebbink[a], Daniele Castellana[a], Henri van den Pol[b], Jason Frank[b], Sven Baars[c], Fred Wubs[c], Mickaël Chekroun[d] and Christian Kuehn[e]**

[a]Department of Physics and Astronomy, Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, The Netherlands, [b]Department of Mathematics, Utrecht University, Utrecht, The Netherlands, [c]Department of Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands, [d]Department of Oceanic and Atmospheric Sciences, University of California at Los Angeles, Los Angeles, CA, USA and [e]Faculty of Mathematics, Technical University of Munich, München, Germany.

*Correspondence: Henk A Dijkstra, Department of Physics and Astronomy, Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands; E-mail: h.a.dijkstra@uu.nl

## Abstract

Dynamical systems methodology is a mature complementary approach to forward simulation which can be used to investigate many aspects of climate dynamics. With this paper, a review is given on the methods to analyse deterministic and stochastic climate models and show that these are not restricted to low-dimensional toy models, but that they can be applied to models formulated by stochastic partial differential equations. We sketch the numerical implementation of these methods and illustrate these by showing results for two canonical problems in climate dynamics.

**Key words:** Climate dynamics, ergodic theory, flow transitions, numerical bifurcation analysis, stochastic dynamical systems.

## 1. Introduction

Climate research starts with 'reality' which we humans experience incompletely and inaccurately (Peixoto and Oort, 1992). With the instrumental record, a very rough view of the climate system has been obtained over the last 150 years, with up to the 1950s, mainly through sea surface temperature and sea surface pressure observations. Global coverage of several surface quantities, for example, of sea surface height is monitored since the satellite era. Proxy data such as oxygen isotope records probe a very incomplete development of the climate system in the (geological) past (Zachos et al., 2001). Based on this data and our imagination, we want to build a picture how particular phenomena arise through the interaction of more elementary processes, i.e. we want an understanding (Ghil, 2001). For some

phenomena, such as the flows in the ocean and atmosphere, we can make use of more general formulated 'laws of nature', such as mass, momentum and energy balances. For other phenomena, such as phytoplankton blooms and in general the behaviour of the marine biosphere, it is difficult to derive precise laws from first principles.

Combining data and imagination leads to models of phenomena. Actually, a whole hierarchy of models usually exists, depending on the detail of the phenomenon, which is aimed to be described (Dijkstra, 2013). Models on the macroscopic level of the hierarchy only describe very gross characteristics of the behaviour, for example only a particular timescale of variability. Other models are aimed to describe the spatial structure of the phenomenon and its context, e.g. its coupling to other phenomena. Which model to develop or pick from the zoo of models available at each (vague) level of the hierarchy depends on the particular question asked. Development of new models can be an art and may depend on the imagination of the scientist. What is important is that the model is 'fit for purpose', i.e. it captures all processes which are relevant for answering the question posed with the 'right' of amount of detail (Katzav et al., 2012).

Eventually, every climate model (whether discrete or continuous) can be formulated as a dynamical system. This dynamical system may be autonomous, non-autonomous or stochastic. Each such system has parameters and may contain delay terms. After a model has been specified, efforts are directed to determine the structure of its solutions for different initial conditions and varying parameters. In climate models, non-autonomous effects easily arise through diurnal, seasonal and/or through Milankovitch astronomical forcing, with known frequencies (Berger and Loutre, 1991). Based on the behaviour of linear systems, one might anticipate that solutions of the model will display these 'forced' timescales of variability. However, in many models, the solution behaviour is contrary to what one expects. Variability can appear with frequencies very different from those in the forcing. A prominent example is the El Niño/Southern Oscillation interannual variability (Philander, 1990) in the equatorial Pacific, which arises due to coupled ocean–atmosphere processes. Also, transitions in observables can occur on much shorter timescales than those of the forcing, for example the changes in the Atlantic Ocean circulation (Rahmstorf, 2000). Non-linear elementary processes, for example advection and radiation, are responsible for this 'unexpected' behaviour.

One may find changes in the behaviour of a model by changing its parameters. Interesting behaviour appears when qualitative changes occur once a parameter is changed only gradually. For example, for $\mu = \mu_0$, all initial conditions may converge in time to a stationary solution, whereas for the $\mu = \mu_1$, suddenly all initial conditions converge to a periodic orbit. Critical transitions, also called 'tipping points' (Gladwell, 2000), appear when sufficiently large or drastic changes in the dynamics are observed as parameters are changed (Kuehn, 2013). Climate phenomena in which such tipping behaviour can occur, such as the decay of the Greenland Ice Sheet, are referred to as 'tipping elements' (Lenton et al., 2008).

For autonomous finite-dimensional dynamical systems—where the dimension or the number d.f. refers to the number of dependent variables in the state vector—there is a beautiful mathematical theory of transition behaviour, i.e. bifurcation theory (Guckenheimer and Holmes, 1990). The concept of a bifurcation of a fixed point (or steady state solution of the dynamical system) is directly defined through the behaviour of eigenvalues of the Jacobian matrix. The theory of normal forms provides a classification of types of bifurcations which is sufficiently detailed for most practical applications. For non-autonomous finite dimensional systems, bifurcation theory is much less advanced and is still in development (Ghil et al., 2008). An exception is the theory of fast–slow systems, where a decoupling can be made between the autonomous behaviour and the slow transient changes in the system (Berglund and Gentz, 2006; Kuehn, 2011). For random dynamical systems, bifurcation theory development is a highly active area of research and currently a complete normal form theory is not available (Arnold, 1998).

Bifurcation theory is widely applied to study the behaviour of models with a relatively few degrees of freedom (d.f.) and is considered as a useful approach to analyse low-dimensional conceptual models in the climate community. However, for answering the most relevant scientific questions posed on climate variability and climate change, only high-dimensional models are considered fit for purpose. The main aim of this review paper is to demonstrate that a stochastic dynamical systems approach can also be applied to high-dimensional climate models. The setup of the paper is to first introduce two relevant climate variability phenomena (section 2), and then show the methods of the numerical approach to determine the behaviour in high-dimensional models of these phenomena (section 3). These methods will then be applied to the phenomena in the sections 4 and 5 and a summary and discussion concludes the paper (section 6).

## 2. Two canonical problems

The ocean circulation plays an important role in the climate system, in particular regarding climate variability on interannual-to-millennial timescales. Ocean flows are forced by the surface wind stress and affected by density

differences, the latter mainly caused by the surface buoyancy flux. The wind-stress forcing leads to strong western boundary currents in the ocean, such as the Gulf Stream in the North Atlantic and the Kuroshio in the North Pacific. The effects of density lead to a meridional overturning circulation (MOC) which is mainly responsible for the meridional heat transport in the oceans. The two problems here deal with variations in the path of the Kuroshio and transitions in the Atlantic MOC (AMOC).

## 2.1. Path transitions of the Kuroshio

The Kuroshio, the western boundary current in the North Pacific, flows south of the islands of Japan. After separating from the coast, the Kuroshio flows almost directly eastwards into the open ocean and is known as the Kuroshio Extension. The Kuroshio Current system as a whole has been observed to have multiple paths (Taft, 1972). The path taken by the Kuroshio south of Japan is divided into two categories, the Large Meander (LM, or meandering path), where the current meanders offshore and then returns to the coast before its separation point (Fig. 1a), and the Non-Large Meander (NLM, or alongshore path), where the current stays close to the coast (Fig. 1b) until its separation point (Kawabe, 1995). Every few years there are transitions between the LM and NLM Kuroshio path states and several mechanisms have been proposed. In Qiu and Miao (2000), it has been proposed that a self-sustained internal variability mechanism is responsible, whereas in Schmeits and Dijkstra (2001), it has been suggested that they are caused by switches between multiple steady Kuroshio states.

The Kuroshio Extension also displays interannual-to-decadal variability between so-called 'elongated' and 'contracted' states. In the elongated state, the current has larger eastward surface transport, a more northerly path and a more intense southern recirculation gyre. The contracted state has smaller eastward surface transport, a more southerly path and a weaker southern recirculation gyre (Qiu and Chen, 2005). The eddy kinetic energy of the contracted state is also generally higher than that of the elongated state. By analysing altimetry data over the period 1993–2009, Qiu and Chen (2010) clearly demonstrate that the transitions between the two Kuroshio Extension states coincide with the arrival of sea surface height anomalies associated with long baroclinic Rossby waves. These waves have been generated by wind variations such as these related to the Pacific Decadal Oscillation (PDO) (Mantua et al., 1997). Using a 1950–2003 hind cast simulation with a high-resolution ocean model, Taguchi et al. (2007) showed that the broad scale response in the Kuroshio Extension could indeed be attributed to the Rossby wave mechanism. However, also strong internal variability is found in the same high-resolution simulation in the absence of any PDO wind variations.

Given the fact that the low-frequency changes of the Kuroshio Current play an important role in air–sea interaction and fisheries, it is important to determine the predictability of these changes and improve model prediction skill (Nonaka et al., 2012). A sound theoretical framework, where the role of the forced Rossby waves, the internal variability, the mesoscale eddy field and the interaction with the bathymetry is clarified, is highly desired. This theoretical framework should not only clarify what causes the variability in the Kuroshio Extension but also explain the processes causing the transitions between the NLM and LM states of the Kuroshio.
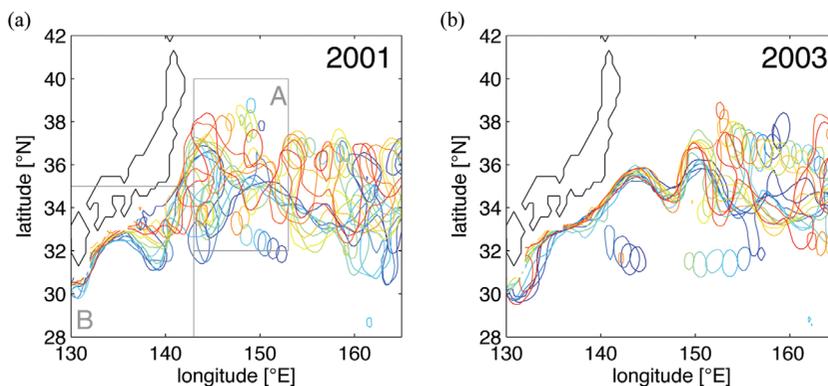


**Figure 1**. (a and b) Current paths (shown by the 100 cm sea surface height contour) for every fourth week for 2 years as observed by satellite altimetry (left: 2001; right: 2003). Regions A and B indicate the Kuroshio Extension and the Kuroshio Current region, respectively.

## 2.2. Transitions of the AMOC

In the North Atlantic, the Gulf Stream transports relatively warm and saline waters northwards. Part of the heat is taken up by the atmosphere, making the water denser. In certain areas (e.g. the Labrador Sea), when there is strong cooling in winter, the water column becomes unstably stratified resulting in strong convection. The interaction of this convection with boundary currents (Spall, 2003) eventually leads to the formation of deepwater, which overflows the various ridges that are present in the topography and enters the Atlantic basin (Fig. 2). This deepwater is transported southwards at a depth of about 2 km, where it enters the Southern Ocean. Through upwelling, the water is slowly brought back to the surface and eventually transported back in the upper ocean to the sinking areas in the North Atlantic. In the Southern Ocean, also bottom water is formed which has a higher density than that from the northern North Atlantic and therefore appears in the abyssal Atlantic and Pacific; in the North Pacific, no deep water is formed (Kuhlbrodt et al., 2007).

The AMOC is the zonally integrated volume transport at each latitude and depth, and its strength is usually taken as the maximum of the meridional overturning stream function. This transport is mainly responsible for the meridional heat transport in the Atlantic. There are no observations available to reconstruct the pattern of the AMOC but its strength at 26°N is now routinely monitored (Cunningham et al., 2007). The currently available time series of the AMOC strength indicates a mean of about 19 Sv (1 Sv = $10^6$ m$^3$ s$^{-1}$), a SD of 5 Sv and a decreasing trend of about 1.6 Sv over the last decade (Smeed et al., 2014). At 26°N, the heat transport associated with the AMOC is estimated (Johns et al., 2011) to be 1.2 PW (1 PW = $10^{15}$ W). This heat is transferred to higher latitudes leading to a relatively mild climate over Western Europe, compared to similar latitudes on the eastern Pacific coast.

The AMOC is sensitive to freshwater anomalies (Stommel, 1961). Freshening of the surface waters in the Nordic and Labrador Seas diminishes the production of deepwater that feeds the deep southward branch of the AMOC. The weakening of the MOC leads to reduced northward salt transport freshening the northern North Atlantic and amplifying the original freshwater perturbation. An AMOC collapse, where a transition to a different equilibrium (weak AMOC) state occurs within a few decades, may occur due to the existence of a saddle-node bifurcation associated with this salt-advection feedback. If the AMOC is in a multiple equilibrium regime, it may undergo transitions due to the impact of noise (Cessi, 1994). The stochastic nature of the forcing (wind-stress and/or buoyancy flux) may even lead to transitions before the actual saddle-node bifurcation has been reached (Ditlevsen and Johnsen, 2010). Hence, it is of central importance to determine what perturbations can cause a
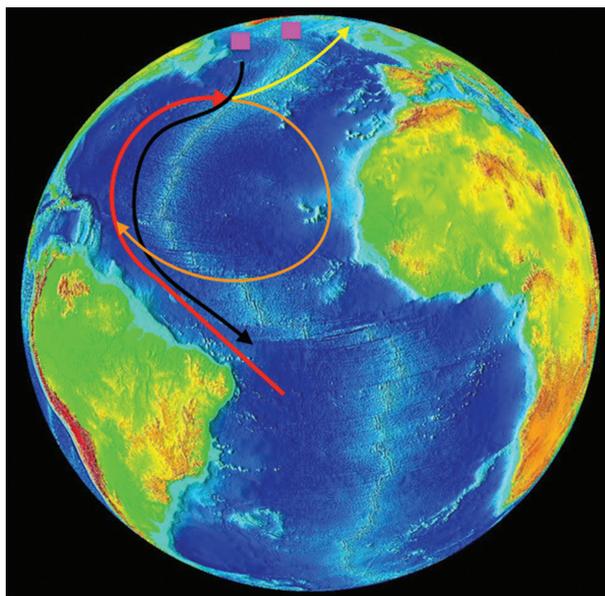


**Figure 2.** Sketch of the North Atlantic Circulation with the Gulf Stream (red), the subtropical gyre (orange), the deep western boundary current (black) and the North Atlantic Current (yellow). The two pink squares indicate regions of deep water formation in the North Atlantic.

transition to the weak AMOC state, in particular, which spatial and temporal correlations in the noise are most effective.

## 3. Numerical stochastic dynamical systems approach

Most climate models are formulated as a set of partial differential equations (which may be deterministic or stochastic) with boundary and initial conditions. To determine solutions of these equations, they are discretized, e.g., with a finite difference, finite element or spectral method. After discretization, a (large) set of ordinary differential equations with algebraic constraints will result which can be written as

$$M(\boldsymbol{p})\frac{d\boldsymbol{x}}{dt} = L(\boldsymbol{p})\boldsymbol{x} + N(\boldsymbol{x},\,\boldsymbol{p}) + F(\boldsymbol{p}),$$

(3.1)

where $\boldsymbol{x} \in \mathbb{R}^d$ is the state vector and $\boldsymbol{p} \in \mathbb{R}^p$ the parameter vector. In addition, $L$ represents the discretized linear part of the equations, $N$ its non-linear part, $F$ represents the forcing and $M$ is in most models a singular diagonal matrix of which every zero on the diagonal is associated with an algebraic constraint (e.g. an incompressibility condition).

To answer questions on flow transitions (such as discussed in section 2 for the Kuroshio and the AMOC) with these models, we follow the systematic approach provided by dynamical systems theory. The first step is to determine the fixed points of the model versus specific control parameters and analyse their linear stability. Methods to perform this numerically, and to detect simple bifurcation points (i.e. pitchfork, transcritical, saddle-node and Hopf), are described in section 3.1. To understand the physics of the occurrence of these bifurcation points, it is helpful to have locally reduced models (with fewer d.f.) and methods to determine these models numerically are given in section 3.2. In regimes where complicated behaviour occurs (even in deterministic models), one is interested in the probability density function of specific observables and techniques to obtain information on these distributions through transfer operator methods are described in section 3.3. In the sections 3.4 to 3.6, we turn to stochastic dynamical systems and provide methods to determine local (Gaussian) probability density functions near deterministic fixed points (section 3.4), methods to numerically determine reduced stochastic models (section 3.5) and methods to determine approximate probability density functions by directly solving the full stochastic equations numerically (section 3.6).

### 3.1. Continuation methods

Often the first step in the analysis of the dynamical system (3.1) is the determination of its fixed points, or steady states. One can simply integrate the model in time to get to an equilibrium for different values of parameters, but there are more efficient techniques to determine the dependence of fixed points upon parameters, i.e. continuation methods.

Finding steady states of the system (3.1) versus a single parameter $p$ amounts to solving the algebraic system

$$\Phi(\boldsymbol{x}, p) = L(p)\boldsymbol{x} + N(\boldsymbol{x}, p) + F(p) = 0.$$

(3.2)

The idea of pseudo-arclength continuation (Keller, 1977; Seydel, 1994) is to parametrize branches of solutions $\Gamma(s) \equiv (\boldsymbol{x}(s), p(s))$ with an arclength parameter $s$ (or an approximation of it, thus the term 'pseudo') and choose $s$ as the continuation parameter instead of $p$. Because of the extra d.f., an extra equation is added by approximating the normalization condition of the tangent $\dot{\Gamma}(s) = (\dot{\boldsymbol{x}}(s), \dot{p}(s))$ to the branch $\Gamma(s)$, where the dot refers to the derivative with respect to $s$, with $|\dot{\Gamma}|^2 = 1$. More precisely, for a given solution $(\boldsymbol{x}_0, p_0)$, the next solution $(\boldsymbol{x}, p)$ is required to satisfy the constraint

$$\dot{\boldsymbol{x}}_0^\top(\boldsymbol{x} - \boldsymbol{x}_0) + \dot{p}_0(p - p_0) - \Delta s = 0,$$

(3.3)

where $\dot{\Gamma}_0 = (\dot{\boldsymbol{x}}_0, \dot{p}_0)$ is the normalized direction vector of the solution family $\Gamma(s)$ at $(\boldsymbol{x}_0, p_0)$ and $\Delta s$ is the step size. There are many approaches to finding a starting point and initial tangent on a branch of solutions (Dijkstra, 2005; Kuehn, 2015).

The solution of the system (3.2) and (3.3) is obtained through a Newton-Raphson iteration, with iteration index $k$. Starting from $\boldsymbol{x}^0 = \boldsymbol{x}_0 + \Delta s\, \dot{\boldsymbol{x}}_0$, $p^0 = p_0 + \Delta s\, \dot{p}_0$, at each iteration a linear system of the form

$$\begin{pmatrix} \mathcal{J}(\boldsymbol{x}^k, p^k) & \frac{\partial}{\partial p}\Phi(\boldsymbol{x}^k, p^k) \\ \dot{\boldsymbol{x}}_0^\top & \dot{p}_0 \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x}^{k+1} \\ \Delta p^{k+1} \end{pmatrix} = \begin{pmatrix} -\Phi(\boldsymbol{x}^k, p^k) \\ r^k \end{pmatrix}.$$

(3.4)

has to be solved, with the Jacobian matrix $\mathcal{J} = D_x\Phi$. Here $r^k = \Delta s - \dot{\boldsymbol{x}}_0^T(\boldsymbol{x}^k - \boldsymbol{x}_0) - \dot{p}_0(p^k - p_0)$. Once $(\Delta\boldsymbol{x}^{k+1}, \Delta p^{k+1})$ is found, one sets $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \Delta\boldsymbol{x}^{k+1}$ and $p^{k+1} = p^k + \Delta p^{k+1}$.

When a steady state has been determined, the linear stability of the solution is considered and transitions that mark qualitative changes such as transitions to multiple equilibria (pitchfork bifurcations or limit points) or periodic behaviour (Hopf bifurcations) can be detected. The linear stability analysis amounts to solving a generalized eigenvalue problem of the form

$$\mathcal{J}\boldsymbol{v} = \sigma M \boldsymbol{v}$$

(3.5)

The matrix $\mathcal{J}$ is in general non-symmetric, $M$ is an, often singular, diagonal matrix and $\sigma$ is a complex number $\sigma = \sigma_r + i\sigma_i$. Simple bifurcations can be detected by monitoring the sign of $\sigma_r$ and they can be distinguished by computing additional properties (Seydel, 1994).

The advantage of pseudo-arclength methods over traditional continuation methods is that the Jacobian of the extended system $\left[\mathcal{J} \dfrac{\partial}{\partial p}\Phi\right]$ has rank $d + 1$, even at saddle-node bifurcations where the Jacobian $\mathcal{J}$ becomes singular. Hence, one can easily continue around these bifurcations (Keller, 1977). Another advantage of pseudo-arclength continuation is that it can compute branches of unstable as well as stable solutions (Seydel, 1994). The unstable solutions are often useful to understand the dynamical behaviour of the system. Continuation methods can also be used to determine periodic solutions versus parameters and to analyse their stability (Dijkstra et al., 2014; Doedel, 1980; Doedel and Tuckermann, 2000).

## 3.2. Local amplitude equations

The discretized system of equations (3.1) will be referred to be in this subsection as the 'full model'. Suppose that at some value of the control parameter, say at $p = p_B$, a bifurcation has been detected within this full model using the methods described above. The idea of reduced models is to construct a low-order model locally near the bifurcation which is capable of representing the dynamics of the full model. The main advantage of such reduced models is that the elementary processes leading to the bifurcation can be studied in detail.

The Jacobian matrix $\mathcal{J}$ near a fixed point has an eigenvector decomposition,

$$L^H\mathcal{J}R = \Sigma \; ; \; L^H M R = I$$

(3.6)

Here $R$ and $L$ denote the right- and left-hand eigenspaces of the problem (3.5), and $L^H$ is the conjugate transpose of $L$. $I$ is the identity and the diagonal matrix $\Sigma$ contains the corresponding eigenvalues, i.e. $R = (r_1 \quad r_2 \quad \cdots \quad r_r), L = (l_1 \quad l_2 \quad \cdots \quad l_r)$ and $\Sigma = \text{diag}(\sigma_1 \quad \sigma_2 \quad \cdots \quad \sigma_r)$, where $r = rank(\Sigma) < d$; the latter inequality is due to the singular nature of $M$. Relation (3.6) states that $L$ and $R$ are a bi-orthogonal set of (generalized) eigenvectors.

The equations governing the perturbations $\boldsymbol{v}$ from the steady state can be written as

$$M\frac{d\boldsymbol{v}}{dt} = \mathcal{J}\boldsymbol{v} + \mathcal{N}(\boldsymbol{v}, \boldsymbol{v})$$

(3.7)

where $\mathcal{N}$ represents the non-linear terms. With the use of the eigenbasis, such a perturbation is expanded in $n$ right-hand eigenvectors, according to

$$\boldsymbol{v}_n = R_n \boldsymbol{a} = \sum_{j=1}^{n} r_j a_j(t)$$

(3.8)

where $\boldsymbol{a} = L_n^H \boldsymbol{v}$. The matrix $R_n$ denotes the $n$-dimensional subspace of $R$ of suitably chosen right-hand vectors, and $L_n$ is its adjoint subspace.

Substitution of (3.8) into (3.7), using (3.2), subsequent projection onto the left-hand eigenbasis $L_n$ and the use of the bi-orthogonality relation (3.6), results in the set of coupled equations for the $a_j, j = 1, ..., n$, i.e.

$$\frac{da_j}{dt} - \sum_{k=1}^{n} b_{jk} \, a_k - \sum_{k=1}^{n} \sum_{l=1}^{n} c_{jkl} \, a_k a_l = 0, \quad j = 1, ..., n \tag{3.9}$$

where the coefficients in the projected system are defined as $b_{jk} = l_j^H \mathcal{J} r_k$ and $c_{jkl} = l_j^H \mathcal{N}(r_k, r_l)$.

While this approach is fairly standard, an issue appears regarding the domain in parameter space for which a close correspondence exists between the dynamical behaviour contained in (3.9), for a chosen value of $n$, and the full model (3.1) exists. In addition, the selection of the appropriate eigenmodes is also highly problem dependent (Van der Vaart et al., 2002).

### 3.3. Transfer operator methods

In the previous sections, focus was on individual trajectories and their asymptotic behaviour (e.g. to fixed points or periodic orbits). However, the climate system displays chaotic behaviour in many of its subsystems, e.g. in its midlatitude atmospheric flow. One is then interested in statistical properties of trajectories, or ergodic properties, of a model. For example, a common approach to improve the skill of weather forecasts and quantify their uncertainty is to use an ensemble of trajectories with slightly different initial conditions (Slingo and Palmer, 2011).

Mathematically, let the model satisfy a system of well-posed ordinary differential equations

$$\dot{x} = \Phi(x), \quad x \in X \ ; \ x(0) = x_0, \tag{3.10}$$

where $X = \mathbb{R}^d$ is the Euclidean state space of dimension $d$ and $\Phi: X \to X$ a smooth vector field with associated flow $S_t$, which associates to an initial state $x_0$ the solution of (3.10) at some time $t \geq 0$. An initial ensemble is then represented by the initial distribution $\rho_0 : X \to \mathbb{R}$ with

$$\rho_0(x) = \frac{1}{m} \sum_{i=1}^{m} \delta(x - x_0^{(i)}) \tag{3.11}$$

where the $x_0^{(i)}$ in $X$ are the initial states of the members of the ensemble and $\delta$ is the Dirac distribution. The average of any smooth observable $g : X \to \mathbb{R}$, such as the sea surface temperature at some location, over the initial ensemble $\rho_0$ is then given by

$$\int_X \rho_0(x)g(x)\mathrm{d}x = \frac{1}{m} \sum_{i=1}^{m} g(x_0^{(i)}), \tag{3.12}$$

where the right-hand side is the usual formula for numerical applications. Similarly, the ensemble average of $g$ after a given time $t \geq 0$, i.e. after propagating the ensemble members by the flow $S_t : x_0^{(i)} \to x^{(i)}(t)$, is

$$\int_X \rho_0 g(S_t x)\mathrm{d}x = \frac{1}{m} \sum_{i=1}^{m} g(x^{(i)}(t)) \, . \tag{3.13}$$

The latter can then be used to define an ensemble $\rho(x, t)$ at time $t$ such that

$$\int_X \rho(x, t)g(x)\mathrm{d}x := \int_X \rho_0 g(S_t x)\mathrm{d}x, \tag{3.14}$$

which, for $\rho_0$ defined by (3.11), gives

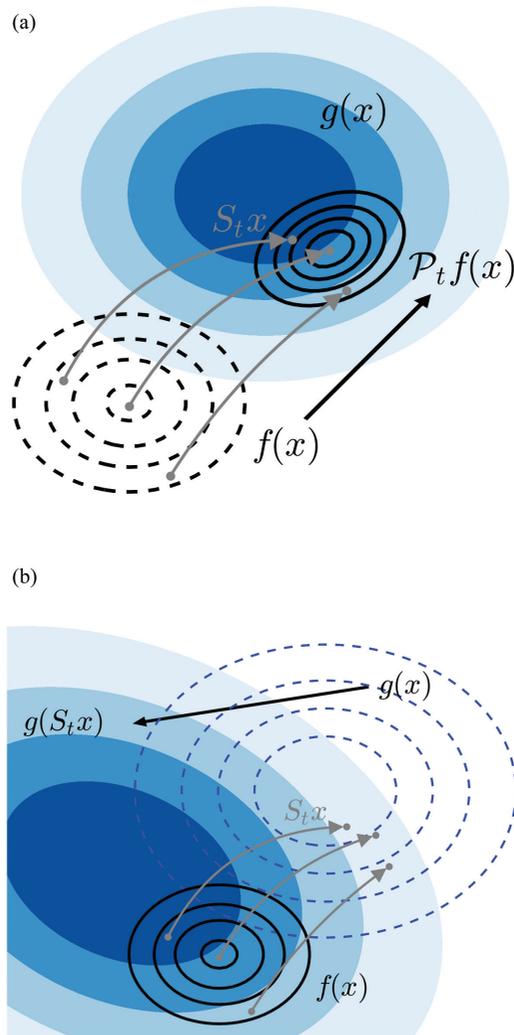$$\rho(x, t) = \frac{1}{m} \sum_{i=1}^{m} \delta(x - x^{(i)}(t)) \, . \tag{3.15}$$

**Figure 3.** Illustration of the definition (3.14) of the transferred density $\rho(\boldsymbol{x}, t) = \mathcal{P}_t\rho_0(\boldsymbol{x})$. The average with respect to the initial ensemble $\rho_0$ of the observable $g$ applied to the propagated states $S_t\boldsymbol{x}$ coincides with the average of $g$ with respect to the transferred density $\mathcal{P}_t\rho_0(\boldsymbol{x})$.

More generally, there exists (Lasota and Mackey, 1994) a linear operator $\mathcal{P}_t : L^1(X) \to L^1(X)$ such that (3.14) holds with $\rho(\boldsymbol{x}, t) = \mathcal{P}_t\rho_0(\boldsymbol{x})$, for any density $\rho_0$ (our initial ensemble) in $L^1(X)$ and observable $g$ in $L^\infty(X)$ (Fig. 3). In particular, taking $g$ in (3.14) to the characteristic function $\chi_A$ of any Borel set $A$, i.e. $\chi_A(\boldsymbol{x}) = 1$ if $\boldsymbol{x}$ is in $A$, 0 otherwise, one has that

$$\int_A \mathcal{P}_t\rho_0(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{S_t^{-1}A} \rho_0(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x},$$

$$(3.16)$$

which expresses the fact that the probability to find a member $\boldsymbol{x}$ sampled from an initial ensemble $\rho_0$ in a set $A$ after some time $t$ is none other than the probability of this member to be initially in the preimage of this set by the flow.

The family $\mathcal{P}_t, t \geq 0$ inherits from the semigroup property of the flow, i.e.

$$\mathcal{P}_{s+t} = \mathcal{P}_t \mathcal{P}_s, \quad t \geq s \geq 0 \tag{3.17a}$$

$$\mathcal{P}_0 = I, \tag{3.17b}$$

and satisfies the Liouville equation

$$\frac{d}{dt}\mathcal{P}_t \rho_0 = -\nabla \cdot (\Phi \mathcal{P}_t \rho_0), \quad \rho_0 \in L^1(X), \tag{3.18}$$

which expresses the fact that probabilities are conserved in phase space, thus playing the role of the continuity equation for hydrodynamic flows in physical space.

A key concept in ergodic theory is that of an invariant measure for the non-singular flow $S_t$, i.e. a probability measure $\mu$ such that $\mu(S_t^{-1}A) = \mu(A)$, for any Borel set $A$. In other words, a measure is invariant if the probability of a state to be in some set does not change as this state is propagated by the flow. It follows then that the average $\overline{g}$ with respect to the invariant measure $\mu$ of any integrable observable $g$ is also invariant with time, i.e.

$$\int_X g(S_t \boldsymbol{x})\mu(d\boldsymbol{x}) = \int_X g(\boldsymbol{x})\mu(d\boldsymbol{x}). \tag{3.19}$$

A flow $S_t$ with an invariant measure $\mu$ has interesting statistical properties when $\mu$ is ergodic, i.e. when the sets $A$ which are invariant, i.e. $S_t^{-1}A = A$, are either of measure 0 or 1. Then, by the celebrated individual ergodic theorem of Birkhoff, the average of any integrable observable $g$ is such that

$$\int_X g(\boldsymbol{x})\mu(d\boldsymbol{x}) = \lim_{T\to\infty}\frac{1}{T}\int_0^T g(S_t\boldsymbol{x})dt, \quad \text{for almost every } \boldsymbol{x}, \tag{3.20}$$

where the right-hand side is none other than the infinitely long version of the time mean used in numerical applications, for which the ergodic theorem is in fact implicitly invoked. When $\mu$ is ergodic, the time mean thus depends on the initial state $\boldsymbol{x}$ only for a set of measure 0.

Unfortunately, there may exist many ergodic measures. For example, multi-stable dissipative systems, i.e. with more than one attractor, such as climate models with both a warm and a snow-ball steady state, may have several ergodic measures supported by different attractors. In this case, the equality (3.20) between ensemble averages and time averages only holds for initial states belonging to the attractor supporting the measure. On the other hand, time averages for two initial conditions in different basins of attractions would not coincide in general. This leads to the definition of a physical measure for which the equality (3.20) between ensemble averages and time averages holds almost everywhere with respect to the Lebesgue measure $d\boldsymbol{x}$. When several physical measures exist, as may be the case when several attractors coexist, one may choose to restrict the phase space to the basin of attraction of only one of the attractors, e.g. by discarding all initial data converging to the snow-ball attractor, in the case of multi-stable climate models. Let us then assume for convenience that $\mu$ is the unique physical measure of the $S_t$ and denote the time mean of $g$ in the right-hand side of (3.20) by $\overline{g}$.

In climate studies, one is not only interested in the mean but also in the variability of observables. This can be done by looking at the correlation function $C_{f,g}$ between two observable $f$ and $g$ in $L_\mu^2(X)$, defined as

$$C_{f,g}(t) = \int_X f(\boldsymbol{x})g(S_t\boldsymbol{x})\mu(d\boldsymbol{x}) - \int_X g(\boldsymbol{x})\mu(d\boldsymbol{x})\int_X f(\boldsymbol{x})\mu(d\boldsymbol{x}) \tag{3.21}$$

For the physical measure $\mu$, the correlation function can be evaluated as the cross-correlation function

$$\widehat{C}_{f,g}(t) = \lim_{T\to\infty}\frac{1}{T}\int_0^T (f(\boldsymbol{x}) - \overline{f})(g(S_t\boldsymbol{x}) - \overline{g})\,dt \tag{3.22}$$

which is much used in practice (Von Storch and Zwiers, 1999). Furthermore, in the definition of the correlation function, only information about the dynamics on the support of the measure $\mu$ is required. A family of transfer operators $\mathcal{L}_t, t \geq 0$ can be defined on $L^2_\mu(X)$ by

$$\int_X \mathcal{L}_t f(\boldsymbol{x}) g(\boldsymbol{x}) \mu(\mathrm{d}\boldsymbol{x}) = \int_X f(\boldsymbol{x}) g(S_t \boldsymbol{x}) \mu(\mathrm{d}\boldsymbol{x}), \quad t \geq 0, \quad f, g \in L^2_\mu(X) \tag{3.23}$$

$$= C_{f,g}(t) \quad \text{when } \mu(f) = \mu(g) = 0. \tag{3.24}$$

Note that compared to the semigroup $\mathcal{P}_t, t \geq 0$ defined by (3.14) with respect to the Lebesgue measure $\mathrm{d}\boldsymbol{x}$, the semigroup of transfer operators $\mathcal{L}_t, t \geq 0$ is defined with respect to $\mu$. As will become clear, this comes as a restriction of the method presented here. In particular, when $\mu$ is supported by an attractor, no information on the dynamics away from the attractor is carried by $\mathcal{L}_t$. On the other hand, when $\mu$ has a density $\rho_\infty(\boldsymbol{x}, t)$ with respect to the Lebesgue measure, then both semigroups are related by

$$\rho_\infty \mathcal{L}_t f = \mathcal{P}_t(\rho_\infty f). \tag{3.25}$$

Particularly important are the eigenfunctions of the transfer operators $\mathcal{L}_t$ and their adjoints, the Koopman operators $\mathcal{U}_t : g \to g \circ S_t$. A function $\psi_k$ is an eigenfunction of $\mathcal{L}_t$ if

$$\mathcal{L}_t \psi_k = \zeta_k(t) \psi_k, \tag{3.26}$$

where the complex number $\zeta_k(t)$ is the associated eigenvalue. To each $\psi_k$ corresponds an eigenfunction $\psi^*_k$ of $\mathcal{U}_t$ associated with the complex conjugate of the eigenvalue $\zeta_k(t)$ and such that

$$\int_X \psi_k \psi^*_l \mu(\mathrm{d}\boldsymbol{x}) = \delta_{kl}. \tag{3.27}$$

It follows from the semigroup property that the eigenfunctions $\psi^*_k$ and $\psi_k$ are independent of time. Moreover, under sufficient continuity conditions for the spectral mapping theorem to hold (Engel and Nagel, 2001), the eigenvalues $\zeta_k(t)$ of the semigroup $\mathcal{L}_t, t \geq 0$ are also related to the eigenvalues $\lambda_k$ of its generator $K$ by (for all $k$)

$$\zeta_k(t) = \mathrm{e}^{\lambda_k t}, \quad t \geq 0 \tag{3.28}$$

and $\mathrm{Re}(\lambda_k) \leq 0$, since $|\zeta_k(t)| \leq 1$. In particular, one can see from (3.19) that the invariance of $\mu$ implies that constant functions on the support of $\mu$ are eigenvectors of $\mathcal{L}_t$ and $\mathcal{U}_t$ associated with the eigenvalue 1.

The mixing spectrum, denoting the set eigenvalues $\lambda_k$ and eigenvectors $\psi_k$ and $\psi^*_k$ of the generator and its adjoint, provides deep insight into the statistical properties of the trajectories of the dynamical system. In particular, when the eigenvectors form a bi-orthonormal family, i.e. when the eigenvalues are all semi-simple, the correlation function $C_{f,g}$ can be decomposed as

$$C_{f,g}(t) = \sum_{k=2}^{\infty} w_k(f, g) e^{\lambda_k t} \tag{3.29a}$$

$$w_k(f, g) = \left( \int f \psi^*_k d\mu \right) \left( \int \psi_k g d\mu \right), \tag{3.29b}$$

so that mixing eigenvalues close to the imaginary axis are responsible for a slow decay of correlations, at a rate given by the real part $-\mathrm{Re}(\lambda_k)$ and with oscillations at an angular frequency given by the imaginary part $\mathrm{Im}(\lambda_k)$.

In addition, the correlation spectrum, given by the Fourier transform of the correlation function, decomposes as

$$S_{f,g}(\omega) = -\frac{1}{\pi} \sum_{k=2}^{\infty} w_k(f, g) \frac{Re(\lambda_k)}{(\omega - \mathrm{Im}(\lambda_k))^2 + (\mathrm{Re}(\lambda_k))^2} \tag{3.30}$$

The largest peaks in the spectrum are thus related to the subset of eigenvalues $\lambda_k$ closest to the imaginary axis. The position of each peak is located at $\omega = \text{Im}(\lambda_k)$ and the half width of the amplitude is given by $-\text{Re}(\lambda_k)$.

In order to get information on the mixing spectrum for high-dimensional systems, projections of the transfer operators on a reduced space $Y$ subset of the phase space $X$ will be approximated from time series. After defining an observation operator $h : Y \to X$, Ulam's method is applied which relies on a Galerkin approximation of the infinite-dimensional transfer operators by finite-dimensional transition matrices (Dellnitz and Junge, 1999; Froyland, 1998; Ulam, 1961). For this purpose, the reduced space is discretized into a family $G = \{B_i\}_{1 \le i \le m_b}$ of $m_b$ grid boxes corresponding to the orthogonal basis $\left\{ \chi_{B_1}, ..., \chi_{B_{m_b}} \right\}$ of characteristic functions. The transfer operator $\mathcal{L}_\tau$ is then approximated by a time-homogenous Markov chain with transition matrix $P_\tau$ whose elements are the transition probabilities

$$(P_\tau)_{ij} = \mathbb{P}(h(\boldsymbol{x}_{t+\tau}) \in B_j | h(\boldsymbol{x}_t) \in B_i). \tag{3.31}$$

These probabilities can be estimated from a long time series $\{\boldsymbol{x}_t\}_{1 \le t \le N}$ of $\boldsymbol{x}$ using the Maximum Likelihood Estimator

$$(\widehat{P}_\tau)_{ij} = \widehat{\mathbb{P}}(h\boldsymbol{x}_{t+\tau}) \in B_j | h(\boldsymbol{x}_t) \in B_i) = \frac{\#\left\{ (h(\boldsymbol{x}_t) \in B_i) \wedge (h(\boldsymbol{x}_{t+\tau}) \in B_j) \right\}}{\#\{h(\boldsymbol{x}_t) \in B_i\}}, \tag{3.32}$$

where $\#\left\{ (h(\boldsymbol{x}_t) \in B_i) \wedge (h(\boldsymbol{x}_{t+\tau}) \in B_j) \right\}$ is the number of observations $h(\boldsymbol{x})$ in box $B_i$ such that the observation $h(\boldsymbol{x}_{t+\tau})$ (a time $\tau$ later) is in box $B_j$ and $\#\{h(\boldsymbol{x}_t) \in B_i\}$ is the total number of observations $h(\boldsymbol{x})$ in $B_i$.

From the eigenvalues $\widehat{\zeta}_k(\tau)$ of the matrix $\widehat{P}_\tau$, for a chosen delay time $\tau$, eventually the $\widehat{\lambda}_k$ are determined through

$$\widehat{\lambda}_k = \frac{1}{\tau} \ln \widehat{\zeta}_k(\tau) \tag{3.33}$$

There are all kind of technical issues in this approximation of the transfer operator and the ergodicity eigenvalues. One important issue is that the semigroup property of $\mathcal{L}_t$ may no longer hold in the reduced space because memory enters the reduced system due to projection into the low-dimensional subspace. If the projection does not introduce any non-Markovian effects, the family of reduced transfer operators should inherit from the semigroup property and the $\widehat{\lambda}_k$ should be constant. If this is not the case, however, one can choose the lag $\tau$ in a range in which at least the leading eigenvalues $\widehat{\lambda}_k$ vary as little as possible, so that the latter behave as Markovian. Another issue is the choice of the delay time $\tau$. Apart from this, the reduced subspace should be quite small as otherwise the computation of the transition probabilities becomes intractable. Hence, in each application, the robustness of the mixing spectrum has to be addressed carefully (Tantet et al., 2015).

### 3.4. Covariance matrix determination through Lyapunov solvers

When noise is added to the forcing, the evolution of the flow (described by the stochastic process $\boldsymbol{X}_t$) can generally be formulated (after discretization of the continuous equations) by a stochastic differential-algebraic equation (SDAE) of the form

$$M \mathrm{d}\boldsymbol{X}_t = \Phi(\boldsymbol{X}_t)\,\mathrm{d}t + \Theta(\boldsymbol{X}_t)\mathrm{d}\boldsymbol{W}_t, \tag{3.34}$$

where $\Phi(\boldsymbol{x}) = L\boldsymbol{x} + N(\boldsymbol{x}) + F$ is the deterministic right-hand side of (3.1) and $\boldsymbol{W}_t$ represents standard $d$-dimensional multivariate Brownian motion (Gardiner, 2002) and $\Theta(\boldsymbol{X}_t)$ the state-dependency of the noise. Memory effects in the driving noise can be accounted for by increasing the dimension of the state vector $\boldsymbol{X}$.

Suppose that the deterministic part of (3.34) has a stable fixed point $\boldsymbol{x}^*$ for a given range of parameter values. Then linearization around the deterministic equilibrium, and assuming the multiplicative noise part is higher order in the linearization of $\Theta(\boldsymbol{X}_t)$, yields (Kuehn, 2012)

$$M \mathrm{d}\boldsymbol{v} = J\boldsymbol{v}\,\mathrm{d}t + \mathscr{B}\,\mathrm{d}\boldsymbol{W}_t, \tag{3.35}$$

where $\mathcal{J}$ is again the Jacobian matrix and $\mathscr{B} = \Theta(\boldsymbol{x}^*)$. Hence, in this section, we consider only the additive noise case.

When $M$ is a singular diagonal matrix, the stochastic part is assumed non-zero only on the respective components where $M$ is non-singular. In this case, (3.35) can be written as

$$\begin{pmatrix} 0 & 0 \\ 0 & M_{22} \end{pmatrix}\begin{pmatrix} d\boldsymbol{v}_1 \\ d\boldsymbol{v}_2 \end{pmatrix} = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \mathcal{J}_{21} & \mathcal{J}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix}dt + \begin{pmatrix} 0 \\ \mathcal{B}_2 \end{pmatrix}d\mathbf{W}_t, \tag{3.36}$$

where $M_{22}$ represents the non-singular part of $M$. Consequently, for non-singular $\mathcal{J}_{11}$ (3.36) can be separated into an algebraic part and an explicitly time-dependent part according to

$$\mathcal{J}_{11}\boldsymbol{v}_1 + \mathcal{J}_{12}\boldsymbol{v}_2 = 0, \tag{3.37a}$$

$$M_{22}\, d\boldsymbol{v}_2 = S\boldsymbol{v}_2\, dt + \mathcal{B}_2\, d\mathbf{W}_t, \tag{3.37b}$$

where $S = \mathcal{J}_{22} - \mathcal{J}_{21}\mathcal{J}_{11}^{-1}\mathcal{J}_{12}$ is the Schur complement of $\mathcal{J}$. Since $M_{22}$ is non-singular, the stationary covariance matrix $C_{22}$ can be determined by solving the corresponding generalized Lyapunov equation (Kuehn, 2012)

$$SC_{22}M_{22} + M_{22}C_{22}S^\top + \mathcal{B}_2\mathcal{B}_2^\top = 0\,. \tag{3.38}$$

In order to find the full covariance matrix one can use $C_{ij} = \mathbb{E}[\boldsymbol{v}_i\boldsymbol{v}_j^\top]$ together with (3.37a) which gives $C_{12} = -\mathcal{J}_{11}^{-1}\mathcal{J}_{12}C_{22}$, $C_{21} = -C_{22}\mathcal{J}_{12}^\top\mathcal{J}_{11}^{-\top}$ and $C_{11} = -\mathcal{J}_{11}^{-1}\mathcal{J}_{12}C_{21}$.

An estimate of the covariance matrix $C$ of the stochastic dynamical system (3.34) can hence be obtained by providing the matrices $\mathcal{J}$, $\mathcal{B}$ and $M$ and solving the corresponding generalized Lyapunov equation (3.38). Once the covariance matrix $C$ is computed, the stationary probability density function, indicated by q($\boldsymbol{x}$), of the approximating Ornstein-Uhlenbeck process follows from (Gardiner, 2002; Kuehn, 2011) as

$$q(\boldsymbol{x};\boldsymbol{x}^*) = \frac{1}{(2\pi)^{\frac{d}{2}}}|C|^{-1/2}e^{-\frac{1}{2}(x-x^*)^\top C^{-1}(x-x^*)}, \tag{3.39}$$

The limitations of this approach are, firstly, that only a probability density function estimate is provided valid on a subexponential timescale before large deviations occur and, secondly, that only the local (i.e. near the steady state) and Gaussian stochastic behaviour of the system are obtained. When large deviations away from the fixed point are of interest, other approaches must be considered (Freidlin and Wentzell, 2012).

In Baars et al. (2017), a new efficient Lyapunov equation solver was presented which solely relies on matrix-vector products based on a Galerkin projection. The covariance matrix $C$ is approximated by a low-rank approximation $\widetilde{C} = VTV^\top$ and the projected system solved is of the form

$$V^\top\mathcal{J}VTV^\top MV + V^\top MVTV^\top\mathcal{J}^\top V + V^\top\mathcal{B}\mathcal{B}^\top V = 0,$$

With $\widetilde{\mathcal{J}} = V^\top\mathcal{J}V, \widetilde{M} = V^\top MV, \widetilde{\mathcal{B}} = V^\top\mathcal{B}$ is taken, a smaller Lyapunov equation results, i.e.

$$\widetilde{\mathcal{J}}T\widetilde{M} + \widetilde{M}T\widetilde{\mathcal{J}}^\top + \widetilde{\mathcal{B}}\widetilde{\mathcal{B}}^\top = 0, \tag{3.40}$$

is obtained which can be solved by a dense solution routine (Bartels and Stewart, 1972). Details on the computation of the space $V$ are presented in Baars et al. (2017).

## 3.5. Non-Markovian reduced stochastic models

The Galerkin method in section 3.2 was used to numerically reduce a high-dimensional system locally to a low-dimensional system to study the physics of the local bifurcations. Such methodology was recently extended to stochastic case by Chekroun et al. (2015) and improved by the consideration of non-Markovian terms. We here only consider the full multiplicative noise case, where the noise is only applied to components of the state vector

for which the diagonal element of $M$ is non-zero. To the evolution equations of the perturbation state vector $v$ (with respect to the deterministic fixed point), a specific state-dependent noise term is consider here. This gives the SDAE

$$M dv = (\mathcal{J}v + N(v, v)) \, dt + \varepsilon M v \circ d W_t, \tag{3.41}$$

where $W_t$ denotes the scalar Wiener process. Here, $v \circ d W_t$ is understood in the Stratonovich sense, and $\varepsilon \geq 0$ quantifies the intensity of the noise term.

The eigenvalues of the linearized problem (as in section 3.2) are ordered by their real part, with $\mathrm{Re}(\sigma_1) \geq \mathrm{Re}(\sigma_2) \geq \ldots$ and again $r_i$ and $l_i$ are the right and left eigenvector belonging to eigenvalue $\sigma_i$. Next the projection of the perturbation $v$ on the space spanned by the $m$ first eigenvectors is denoted by $v_c$; these are the 'resolved variables'. The projection of $v$ on the space spanned by the $m + 1$-th until $r$-th eigenvector is indicated by $v_s$; these are the 'unresolved variables'.

Expanding $v$ in $r$ eigenvectors gives $v = v_c + v_s = \sum_{i=1}^{r} a_i r_i = R_r a_r$. Substituting this in equation (3.41) and multiplying by $L_r^H$ gives:

$$da_j = \left[ \sigma_j a_j + l_j^H N(R_r a_r, R_r a_r) \right] dt + \varepsilon a_j \circ d W_t, \qquad j = 1, \ldots, r. \tag{3.42}$$

In one version of the parameterized manifold approach developed in Chekroun et al. (2015), the unresolved part of the solution $v_s$ is represented by

$$v_s = h(\xi) = \sum_{n=m+1}^{r} h^n(\xi) r_n, \tag{3.43}$$

where $\xi$ is the vector variable in the parameterized manifold and for $n \geq m + 1$:

$$h^n(\xi) = \sum_{i_1, i_2 = 1}^{m} N_{i_1, i_2}^{n} Q_n^{i_1, i_2} \xi_{i_1} \xi_{i_2}.$$

Here the $Q_n^{i_1, i_2}$ are determined by solving the stochastic differential equation

$$dQ = (1 - (\sigma_{i_1} + \sigma_{i_2} - \sigma_n)Q) dt - \varepsilon Q \circ d W_t. \tag{3.44}$$

and the coefficients $N_{pq}^{k}$ are determined from $N_{pq}^{k} = \, <l_k, N(r_p, r_q)> \, = l_k^H N(r_p, r_q)$.

Substituting this approximation for $v_s$ into (3.42) and projecting again eventually gives the stochastic reduced model

$$
\begin{aligned}
da_j = \Bigg( & \sigma_j a_j + \sum_{i_1, i_2 = 1}^{m} N_{i_1 i_2}^{j} a_{i_1} a_{i_2} \\
& + \sum_{i, i_1, i_2 = 1}^{m} \sum_{n=m+1}^{r} \left( N_{in}^{j} + N_{ni}^{j} \right) N_{i_1 i_2}^{n} Q_n^{i_1 i_2} a_{i_1} a_{i_2} a_i \\
& + \sum_{n_1, n_2 = m+1}^{r} \sum_{i_1, i_2 = 1}^{m} \sum_{i_3, i_4 = 1}^{m} N_{n_1 n_2}^{j} N_{i_1 i_2}^{n_1} N_{i_3 i_4}^{n_2} Q_{n_1}^{i_1 i_2} Q_{n_2}^{i_3 i_4} a_{i_1} a_{i_2} a_{i_3} a_{i_4} \Bigg) dt \\
& + \varepsilon a_j \circ d W_t, \quad 1 \leq j \leq m,
\end{aligned}
\tag{3.45}
$$

These random coefficients $Q_n^{i_1, i_2}$ convey exogenous memory effects resulting from the interactions of the noise with the bilinear term in (3.41) (these coefficients depending on the past of the noise) as explained in Chekroun et al. (2015). For $\varepsilon = 0$, the deterministic case is obtained and compared to the result in section 3.2. One observes that, different from the Galerkin projection, now also the higher order interactions of the modes are taken into account in the reduced model (3.45).
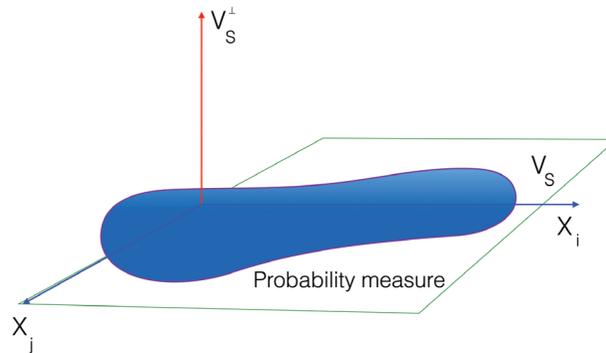
**Figure 4.** Sketch of the orthogonal basis spanning the space $V_S$ and the orthogonal space $V_S^\perp$. A probability measure is defined on $V_S$.

## 3.6. Dynamical Orthogonal (DO) field methods

If a general random state vector is denoted by $X_t$, the starting point of the dynamical orthogonal (DO) method is to use a generalized, time-dependent, Karhoenen-Loeve expansion

$$X_t = \overline{X}(t) + \sum_{i=1}^{s} Y_i(t) X_i(t),$$

(3.46)

where the coefficients $Y_i(t)$ are scalar stochastic processes with arbitrary probability distribution functions and the $X_i$ are deterministic vectors. The employed representation follows from the assumption that the stochastic part of the solution 'lives' in a finite-dimensional space, the stochastic subspace $V_S$ (Fig. 4).

   The finite-dimensionality of the representation is crucial towards the derivation of closed field equations although it is not sufficient. The choice to have all unknown terms varying with time leads to redundancy issues and therefore additional constraints are necessary in order to obtain independent equations for all the quantities involved. In (Sapsis and Lermusiaux, 2009), it is shown that a suitable constraint, which arises naturally to overcome this redundancy issue, is the dynamical orthogonality condition. It requires that the time-dependent basis should vary orthogonally to the space $V_S$ and hence on every time instant:

$$\left\langle \frac{dX_i(t)}{dt}, X_j(t) \right\rangle = 0 , \quad i = 1, ..., s , \quad j = 1, ..., s.$$

(3.47)

and $< , >$ is the standard inner product. The use of time-dependence in the basis elements (the so-called DO modes) allows the representation of the transient character of the solution using much fewer modes. From a physical point of view, the constraint (3.47) imposes the natural need for the computational algorithm to evolve the shape of the perturbations only if there is an important new direction, which is not already covered by the existing modes. In terms of computational performance, it is a trade-off between, on the one hand, fewer modes to consider, but on the other hand, more equations (including interactions between all the considered modes) to solve.

   The orthogonality condition leads to a closed set of equations that allows for the evolution of the mean field $\overline{X}(t)$, the DO modes $X_i(t)$ and the stochastic coefficients $Y_i(t)$. In this way the solution is obtained in the form of random realizations for the scalar stochastic coefficients $Y_i(t)$ (having low-computational and storage cost since they are low-dimensional) and in terms of $s + 1$ deterministic fields containing the full deterministic information (mean field) and the $s$ time-dependent directions (the DO modes) where stochasticity is important. From the mean, the time-dependent patterns of the DO modes plus the distribution of the stochastic coefficients $Y$ (at a certain time $t$), an approximation to the probability density function of the state vector can be obtained.

   The dimension $s$ can be estimated during the time-development by monitoring the eigenvalues of the covariance matrix $C_{Y_i(t)Y_j(t)}$ (Sapsis and Lermusiaux, 2012). The criteria for evolving the subspace size are based on stability arguments which follow directly from the system differential equations. When an eigenvalue of the covariance

matrix $C_{Y_i(t)Y_j(t)}$ exceeds a critical value, $s$ is increased by one; when it drops below a critical value, $s$ is decreased by one. The algorithm also provides the instantaneously most unstable perturbation, which is not included into the stochastic subspace. More details on the methodology can be found in Sapsis (2010).

Having the random coefficients expressed in the form of random realizations has important advantages especially for climate applications. It allows for the simple and efficient representation of any interesting statistical quantity of the solution such as single or joint moments of any order and probability density functions of the velocity field in specific points of the domain but also spatiotemporal correlation functions and length-scales. Based on these realizations, one may compute statistical moments or the associated probability density functions.

## 4. Application: the path variations of the Kuroshio

In this section, we apply part of the methodology to an idealized model of the wind-driven ocean circulation. After presenting the model and its bifurcation diagram in the sections 4.1 and 4.2, respectively, we consider results for the DO method in section 4.3 and of a stochastic reduced model in section 4.4. From the bifurcation diagram in section 4.2 we learn that symmetry breaking can lead to multiple asymmetric Kuroshio jets, even under symmetric wind forcing, and that intrinsic variability (meandering of the jet) can occur through Hopf bifurcations. With the DO analysis in section 4.3 the effects of noise in the wind stress on the transitions between the different Kuroshio states are studied. With the local stochastic model near the first symmetry breaking pitchfork the local probability density function is calculated (section 4.4).

### 4.1. Model of the Kuroshio

We consider ocean flows with constant density $(\rho_o)$ in a square basin of horizontal dimension $L$ and with constant depth $H$, which is situated on a midlatitude $\beta$-plane with Coriolis parameter $f = f_0 + \beta_0 y$. The ocean circulation in the basin is forced at the surface through a wind-stress vector $\tau = \tau_0 \, (\tau^x(x, y), \tau^y(x, y))$.

In a quasi-geostrophic approximation, the flow can be modelled by the well-known barotropic vorticity equation (Pedlosky, 1987) for the geostrophic stream function $\psi$. With introduction of the horizontal velocities $u$ and $v$ and the relative vorticity $\zeta$, the model can be written as

$$\left[\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right](\zeta + \beta_0 y) = A_H \nabla^2 \zeta + \frac{\tau_0}{\rho_o H}\left(\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y}\right) \tag{4.1a}$$

$$\zeta = \nabla^2 \psi \tag{4.1b}$$

$$u = -\frac{\partial \psi}{\partial y} \, , \, v = \frac{\partial \psi}{\partial x} \tag{4.1c}$$

where $A_H$ is the lateral friction coefficient. On the lateral zonal boundaries $x = 0, L$, no-slip conditions are prescribed, whereas at the meridional boundaries ($y = 0, L$), we apply slip conditions

$$x = 0, \, L : \ \psi = 0, \frac{\partial \psi}{\partial x} = 0 \tag{4.2a}$$

$$y = 0, \, L : \ \psi = 0, \zeta = 0 \tag{4.2b}$$

The deterministic (so-called double-gyre) wind stress considered in this section is

$$\tau_d^x(y) = -\frac{1}{2\pi}\cos\left(\frac{2\pi y}{L}\right); \ \tau_d^y = 0 \tag{4.3}$$

where the parameter $a$ controls the asymmetry of the wind-stress field with respect to the mid-axis of the basin. The standard values of the parameters are provided in Table 1 and we use $A_H$ (or the Reynolds number $Re = UL/A_H$) as a control parameter.

Table 1. Fixed model parameters of the two-dimensional ocean model of the wind-driven ocean circulation

| | |
|---|---|
| $H = 6.0 \cdot 10^2$ m | $\rho = 1.0 \cdot 10^3$ kg m$^{-3}$ |
| $L = 1.0 \cdot 10^6$ m | $\beta_0 = 2.0 \cdot 10^{-11}$ (ms)$^{-1}$ |
| $U = 1.6 \cdot 10^{-2}$ m s$^{-1}$ | $\tau_0 = 1.0 \cdot 10^{-1}$ Pa |

### 4.2. Deterministic bifurcation diagram

The equations (4.1) were discretized using a 128 × 128 equidistant grid and the steady states are computed using a pseudo-arclength continuation method as described in section 3.1. The dimension of the state vector $d = 128 \times 128 \times 2$ as both $\psi$ and $\zeta$ are considered in the state vector. The matrix $M$ in (3.1) has zeroes on the diagonal for each $\psi$ equation, i.e. the discretized version of (4.1b).

For the wind-stress forcing (4.3), the structure of the steady solutions is shown through the bifurcation diagram in Figure 5a, where the value of the stream function (scaled with $LHU$) at a point in the southwest part of the domain ($\psi_R = \psi(x = L/4, y = L/4)/(LHU)$) is plotted versus $Re$. At small $Re$, the anti-symmetric double-gyre flow (Fig. 5b) is a unique state. When lateral friction is decreased, this flow becomes unstable at the pitchfork bifurcation $P_1$ and two branches of stable asymmetric states appear for smaller values of $A_H$ (larger $Re$). The solutions on these branches have the jet displaced either southward or northward (Fig. 5c) and are exactly symmetrically related for the same value of $Re$. The asymmetric solutions become unstable at a Hopf bifurcation $H$ located at $Re = 52$.

The branch of unstable steady solutions at $P_1$ is the continuation of the anti-symmetric solution branch. For even smaller friction, this anti-symmetric flow becomes inertially dominated and $\psi_R$ increases rapidly while going through the saddle-node bifurcation $L$. A pitchfork bifurcation $P_2$ occurs on the anti-symmetric branch where an additional pair of asymmetric solution branches appears; all these solutions are unstable.

### 4.3. Wind-stress noise: DO analysis

The wind-stress field over the North Atlantic and North Pacific contains a very strong variability on many timescales. Hence, to understand the variability of wind-driven flows, a stochastic component of the wind-stress needs to be represented. A stochastic wind-stress product $\tau_s$ was suggested in Sura et al. (2001) as

$$\tau_s = \rho_a C_D |u'| u', \tag{4.4}$$

where $\rho_a$ is the air density and $C_D$ the drag coefficient, with

$$u' = \varepsilon \eta_x(t) f(x, y); \quad v' = \varepsilon \eta_y(t) f(x, y) \tag{4.5a}$$

$$f(x, y; \lambda) = \frac{\alpha}{\sqrt{\pi}} \left( \lambda \ Erf\left(\frac{1}{2\lambda}\right) \right)^{-1} e^{-\frac{\left(\frac{x}{L} - \frac{1}{2}\right)^2 + \left(\frac{y}{L} - \frac{1}{2}\right)^2}{2\lambda^2}} \tag{4.5b}$$

where $\lambda$ is a scale factor controlling the spatial extent of the stochastic forcing and $\alpha$ controls its amplitude. The quantities $\eta_x$ and $\eta_y$ represent Gaussian white noise as a generalized (distribution valued) stochastic process [see section 3 in Doering (1990)]. Numerically, at each time value $t_i = i\Delta t, i = 1, ..., n$, the quantities $\eta_x(t_i), i = 1, ..., n$ are independent and normally distributed, and the stochastic process has a non-zero autocorrelation time that is shorter than the time step $\Delta t$; the same holds for $\eta_y(t)$.

Similarly to the choice in Sura et al. (2001), the amplitude $\alpha$ is normalized so that it is the same for every value of $\lambda$, i.e. $\alpha$ is computed from $\alpha = 1/f(0, 0; 1/4)$. The reference amplitude for $\lambda = 1/4$ is chosen such that the variance in the noise $\varepsilon^2 = 28$ m$^2$ s$^{-2}$, similar to that in Sura et al. (2001), with values of $\rho_a = 1.2$ kg m$^{-3}$ and $C_D = 2.0 \cdot 10^{-3}$. The DO results for $Re = 35$ are shown in Figure 6 for $\lambda = 0.25$. The amplitudes of all DO modes approach steady values in time and the first DO mode has the largest amplitude. This mode has the spatial pattern of the $P$-mode in Simonnet and Dijkstra (2002), the normal mode destabilizing the steady state at the pitchfork bifurcation $P_1$. It was shown in Sapsis and Dijkstra (2013) that the spatial structure of the DO modes is the same for each value of $\lambda$. Also, the geometry of the probability density remains the same for different noise forcing although the probability measure is more spread for smaller values of $\lambda$.
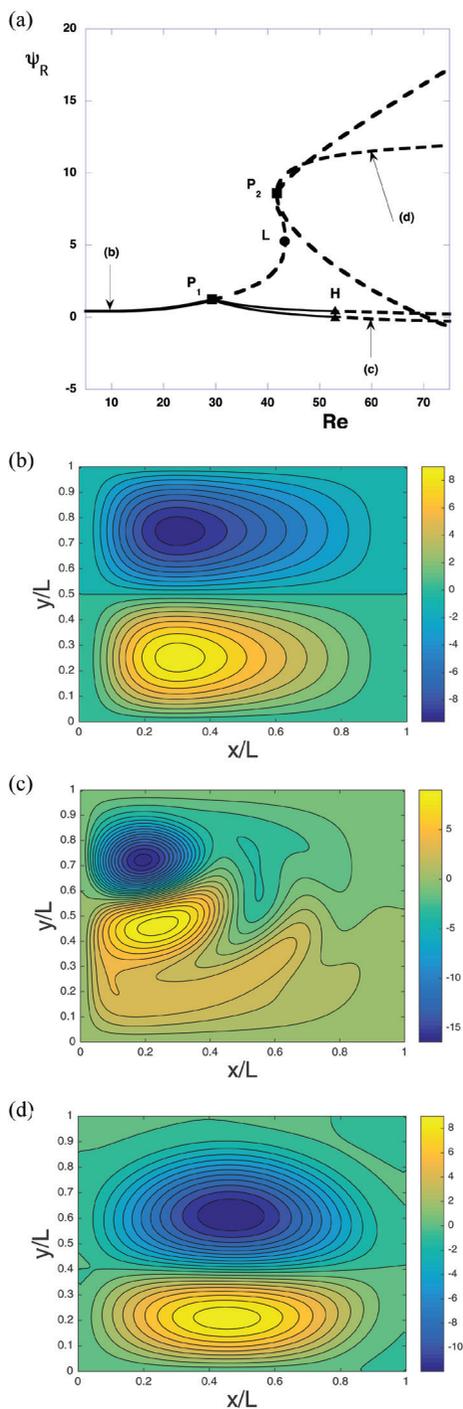
**Figure 5.** (a) Bifurcation diagram for the double-gyre barotropic quasi-geostrophic model for a square basin with $Re = UL/A_H$ as the control parameter. On the y-axis, the dimensionless stream function value $\psi_R = \psi(x = L/4, y = L/4)/(ULH)$ is plotted. The Hopf bifurcation $H$ on both branches $A_1$ (Dijkstra and Katsman, 1997) is located at $Re = 52$. Drawn (dashed) branches indicate linearly stable (unstable) steady states. (b) Pattern of $\psi$ near $Re = 10$ on the lower stable branch in (a); all contour values are with respect to the maximum value of $\psi$ and the contour spacing is 0.1. The x- and y-coordinate are scaled with the basin size $L$. (c) Same for $Re = 60$ along the branch $A_{1u}$; the pattern on the branch $A_{1d}$ at $Re = 60$ is the mirror image of (c) with respect to reflection through the mid-axis of the basin. (d) The pattern at $Re = 60$ on the branch $A_{2d}$ [slightly modified from Sapsis and Dijkstra (2013)].
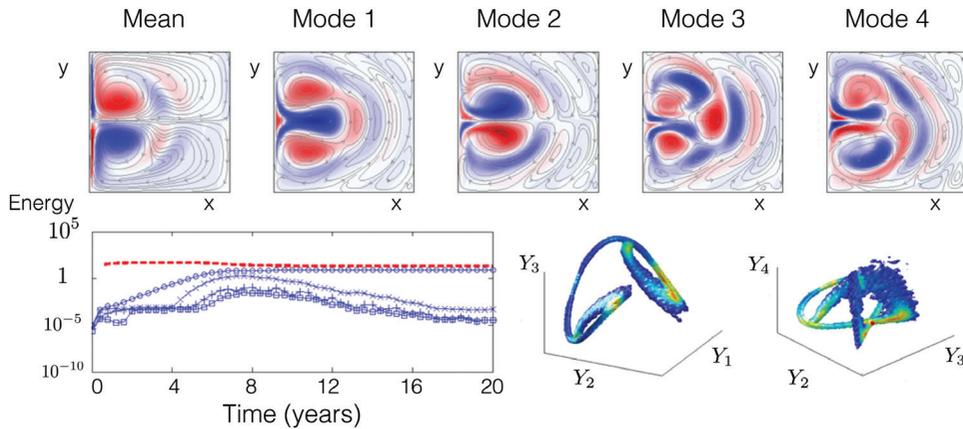
**Figure 6.** Results from the DO field analysis for $Re = 35$ as determined on a $64 \times 64$ grid, with $s = 4$. Here an initial variance $10^{-5}$, a scaled noise amplitude $\alpha = 1/f(0, 0; 1/4)$ was taken and $\lambda = 0.25$ in (4.5). Patterns of the stream function $\psi$ are plotted on top for the mean and four DO modes (at $t = 12.8$ years). On the lower right panel, the marginal distribution functions of the stochastic coefficients $Y_i$ are shown (Sapsis and Dijkstra 2013).

### 4.4. (Stochastic) reduced models

We will show results for the reduced models just above the value of $Re$ at the first pitchfork bifurcation $P_1$ in Figure 5. Thereto the equations (4.1) are first non-dimensionalized using a spatial scale $L$, a velocity scale $U$ and a timescale $L/U$. For $Re = 37$, there are two (dimensionless) real eigenvalues $\sigma_1 = 6.98 \times 10^{-2}$ and $\sigma_2 = -4.196$. In the deterministic case, using the methodology of section 3.2, we obtain the equations (3.9)

$$\frac{da_1}{dt} = \sigma_1 a_1 + (c_{112} + c_{121}) a_1 a_2 \tag{4.6a}$$

$$\frac{da_2}{dt} = \sigma_2 a_2 + c_{211} a_1^2 + c_{222} a_2^2 \tag{4.6b}$$

with $c_{112} = 1.49 \times 10^{-2}$, $c_{121} = -1.53 \times 10^{-2}$, $c_{211} = 1.37 \times 10^{-3}$, $c_{222} = -7.36 \times 10^{-4}$. Apart from $(0, 0)$, there are two additional steady states and the total solution for the dimensionless stream function based on $a_1$ and $a_2$ gives $\psi_{max} = 20.93$, whereas from the numerical solution of the full model, we find $\psi_{max} = 21.16$ indicating that the deterministic reduced model is well capable of approximating the pitchfork bifurcation quantitatively (Van der Vaart et al., 2002).

For the stochastic case the amplitude $a_1$, as determined from (3.45) using $m = 1$ and $r = 2$, is shown versus time for a noise amplitude $\varepsilon = 0.1$ in Figure 7a; the initial amplitude $a_1(0) = 1$. The function $Q_2^{1,1}$ (not shown) varies around 0.23, $a_1$ around 710.8 and the probability density function $p^+$ is plotted in Figure 7b. In this multiplicative noise case considered, there are no stochastically induced transitions between fixed points (note that $Re = 37$ is in a multiple equilibrium regime) and the probability density function depends on the sign of the initial amplitude $a_1(0)$.

## 5. Application: the transitions of the MOC

In this subsection, we apply part of the methodology to an idealized model of the meridional overturning ocean circulation. After presenting the model and its bifurcation diagram of the deterministic version in the sections 5.1 and 5.2, respectively, we consider the Lyapunov solver approach in section 5.3 and the transfer operator approach in section 5.4.

### 5.1. Model of the MOC

The spatially quasi two-dimensional model of the Atlantic circulation as described in den Toom et al. (2011) is fit for purpose to consider transitions of the AMOC (see section 2.2). In the model, there are two active tracers: temperature $T$ and salinity $S$, which are related to the density $\rho_o$ by a linear equation of state

$$\rho = \rho_0 (1 - \alpha_T (T - T_0) + \alpha_S (S - S_0)), \tag{5.1}$$
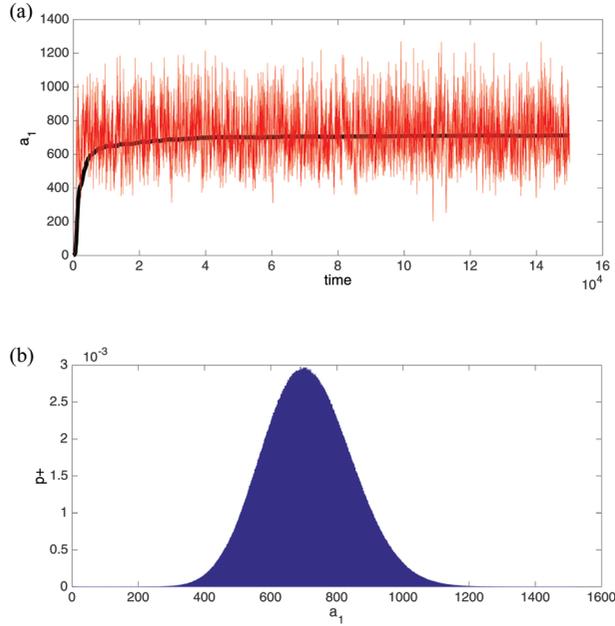
**Figure 7.** Results for the reduced stochastic model (3.45) for $Re = 37$, $m = 1$, $q = 2$ and $\varepsilon = 0.1$. (a) The dimensionless amplitude $a_1$ with the black curve being the (running) mean. (b) The probability density function $p^+$ for an initial perturbation $a_1(0) > 0$; the mean is 710.8.

where $\alpha_T$ and $\alpha_S$ are the thermal expansion and haline contraction coefficients, respectively, and $\rho_0$, $T_0$ and $S_0$ are reference quantities. The mixing of momentum and tracers due to eddies is parameterized by simple anisotropic diffusion. In this case, the zonal velocity $u = 0$, the longitudinal derivatives are zero and the equations for the meridional velocity $v$, vertical velocity $w$, pressure $p$, and the tracers $T$ and $S$ are given by

$$0 = -\frac{1}{\rho_0 r_0}\frac{\partial p}{\partial \theta} + A_V\frac{\partial^2 v}{\partial z^2} + \frac{A_H}{r_0^2}\left(\frac{1}{\cos\theta}\frac{\partial}{\partial\theta}\left(\cos\theta\frac{\partial v}{\partial\theta}\right) + (1 - \tan^2\theta)v\right),$$

(5.2a)

$$0 = -\frac{1}{\rho_0}\frac{\partial p}{\partial z} + g(\alpha_T T - \alpha_S S),$$

(5.2b)

$$0 = \frac{1}{r_0\cos\theta}\frac{\partial v\,\cos\theta}{\partial\theta} + \frac{\partial w}{\partial z},$$

(5.2c)

$$\frac{\partial T}{\partial t} + \frac{v}{r_0}\frac{\partial T}{\partial\theta} + w\frac{\partial T}{\partial z} = \frac{K_H}{r_0^2\cos\theta}\frac{\partial}{\partial\theta}\left(\cos\theta\frac{\partial T}{\partial\theta}\right) + K_V\frac{\partial^2 T}{\partial z^2} + CA(T),$$

(5.2d)

$$\frac{\partial S}{\partial t} + \frac{v}{r_0}\frac{\partial S}{\partial\theta} + w\frac{\partial S}{\partial z} = \frac{K_H}{r_0^2\cos\theta}\frac{\partial}{\partial\theta}\left(\cos\theta\frac{\partial S}{\partial\theta}\right) + K_V\frac{\partial^2 S}{\partial z^2} + CA(S).$$

(5.2e)

Here $t$ is time latitude, $z$ the vertical coordinate, $r_0$ the radius of Earth, $g$ the acceleration due to gravity, $A_H$ ($A_V$) the horizontal (vertical) eddy viscosity and $K_H$ ($K_V$) the horizontal (vertical) eddy diffusivity. The terms with CA represent convective adjustment contributions as specified in den Toom et al. (2011).

The equations are solved on an equatorially symmetric, flat-bottomed domain. The basin is bounded by latitudes $\theta = -\theta_N$ and $\theta = \theta_N = 60\,°$ and has depth $H$. Free-slip conditions apply at the lateral walls and at the bottom. Rigid lid conditions are assumed at the surface and atmospheric pressure is neglected. The wind stress is zero everywhere, and 'mixed' boundary conditions apply for temperature and salinity,

$$K_V\frac{\partial T}{\partial z} = \frac{H_m}{\tau}(\overline{T}(\theta) - T),$$

(5.3a)

$$K_V\frac{\partial S}{\partial z} = S_0 F_s(\theta).$$

(5.3b)

This formulation implies that temperatures in the upper model layer (of depth $H_m$) are relaxed to a prescribed temperature profile $\overline{T}$ at a rate $\tau^{-1}$, while salinity is forced by a net freshwater flux $Q$, which is converted to an equivalent virtual salinity flux by multiplication with $S_0$. The numerical values of the fixed model parameters are given in Table 2.

## 5.2. Deterministic bifurcation diagram

For the deterministic case, the equatorially symmetric boundary forcing is fixed as

$$\overline{T}(\theta) = 10.0\cos\,(\pi\theta/\theta_N),$$

(5.4a)

$$F_s(\theta) = \overline{F}_s(\theta) = \mu F_0\frac{\cos(\pi\theta/\theta_N)}{\cos(\theta)},$$

(5.4b)

where $\mu$ is the strength of the mean freshwater forcing (which is taken as bifurcation parameter) and $F_0 = 0.1$ m yr$^{-1}$ is a reference freshwater flux.

The equations are discretized on a latitude-depth equidistant $M_o \times L_o$ grid using a second-order conservative central difference scheme. An integral condition expressing the overall conservation of salt is also imposed, as the salinity equation is only determined up to an additive constant. The total number of d.f. $d = 6M_oL_o$, as there are six unknowns per point and the standard spatial resolution used is $M_o = 32, L_o = 16$.

The bifurcation diagram of the deterministic model for parameters as in Table 1 is shown in Figure 8a. On the $y$-axis, the sum of the maximum ($\Psi^+$) and minimum $\Psi^-$ values of the meridional stream function $\Psi$ is plotted, where $\Psi$ is defined through

$$\frac{\partial\Psi}{\partial z} = v\cos\,\theta, \; -\frac{1}{r_0\cos\theta}\frac{\partial\Psi}{\partial\theta} = w\,.$$

(5.5)

For the calculation of the transports, the basin is assumed to have a zonal width of $64\,°$. The value of $\Psi^+ + \Psi^-$ is zero when the MOC is symmetric with respect to the equator.

For small values of $\mu$, a unique equatorially anti-symmetric steady MOC exists of which a pattern at location $a$ is shown in Figure 8b. This pattern destabilizes at a supercritical pitchfork bifurcation and two asymmetric pole-to-pole solutions appear. An example of the MOC at location $b$ in Figure 8b shows a stronger asymmetric overturning with sinking in the northern part of the basin. The pole-to-pole solutions cease to exist beyond a saddle-node bifurcation near $\mu = 0.13$ and both branches connect again with the anti-symmetric solution at a second subcritical pitchfork bifurcation. At this

Table 2. Fixed model parameters of the two-dimensional model of the AMOC

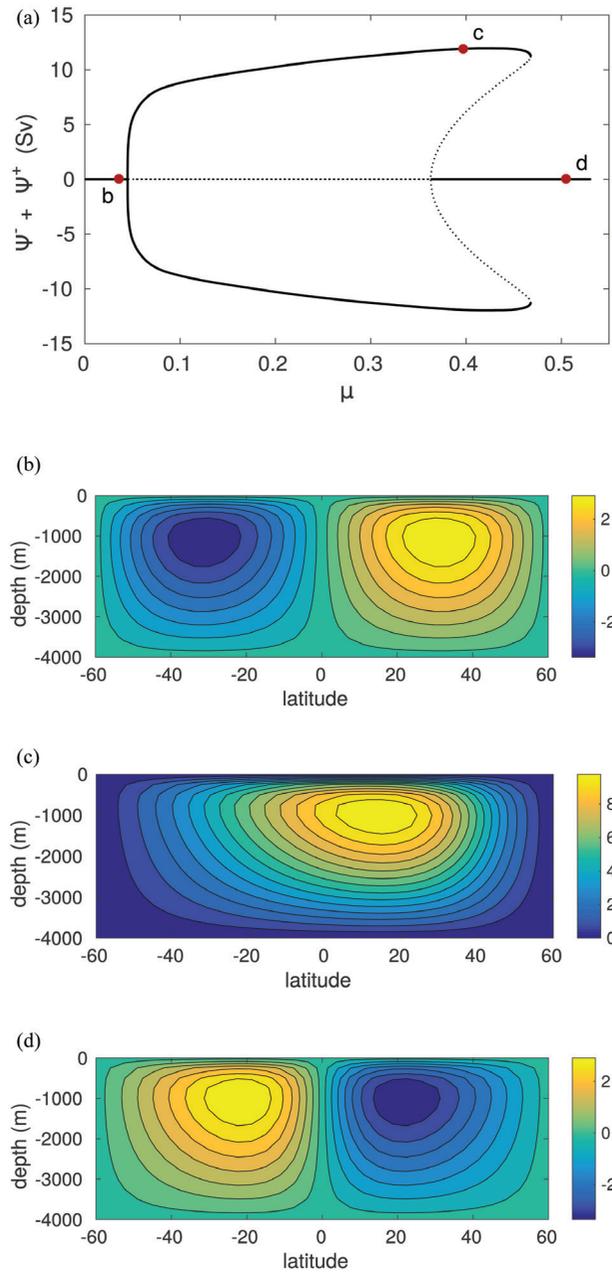| | |
|---|---|
| $H = 4.0 \cdot 10^3$ m | $H_m = 2.5 \cdot 10^2$ m |
| $r_0 = 6.371 \cdot 10^6$ m | $T_0 = 15.0$ °C |
| $g = 9.8$ m s$^{-2}$ | $S_0 = 35.0$ psu |
| $A_H = 2.2 \cdot 10^{12}$ m$^2$ s$^{-1}$ | $\alpha_T = 1.0 \cdot 10^{-4}$ K$^{-1}$ |
| $A_V = 1.0 \cdot 10^{-3}$ m$^2$ s$^{-1}$ | $\alpha_S = 7.6 \cdot 10^{-4}$ psu$^{-1}$ |
| $K_H = 1.0 \cdot 10^3$ m$^2$ s$^{-1}$ | $\rho_0 = 1.0 \cdot 10^3$ kg m$^{-3}$ |
| $K_V = 1.0 \cdot 10^{-4}$ m$^2$ s$^{-1}$ | $\tau = 75.0$ days |

**Figure 8**. (a) Bifurcation diagram of the deterministic equatorially symmetric two-dimensional MOC model, with the forcing as in (5.4). (b–d) Patterns of the meridional overturning stream function at several values of $\mu$ indicated in (a).

bifurcation, the anti-symmetric solution with equatorial sinking (see MOC at location $c$ in Fig. 8b) appears which is stable for larger values of $\mu$. The value of $\mu$ at the point $b$, $\mu_b = 0.110$, is here considered as the reference freshwater forcing.

## 5.3. Lyapunov solver approach

Next, the symmetric freshwater forcing is chosen as

$$F_s(\theta, t) = (1 + \varepsilon\ \eta(t))\overline{F}_s(\theta) \tag{5.6}$$

where $\eta(t)$ represents Gaussian white noise [again in the generalized distribution sense (Doering, 1990)], i.e. with $\mathbb{E}[\eta(t)] = 0$ and $\mathbb{E}[\eta(t)\eta(s)] = \delta(t - s)$, where $\delta$ is the Dirac distribution. For the numerical implementation and the interpretation of the stochastic process, the same holds for $\eta(t)$ as for $\eta_x(t)$ in (4.5). In this case, the noise matrix $\mathcal{B}$ in (3.35) simply represents additive noise, which is (i) only active in the freshwater component, (ii) only present at the surface, (iii) uncorrelated and (iv) at each latitude $\theta$ the magnitude $\sigma$ is 10% of the deterministic freshwater forcing amplitude [see (5.6)]. Note that the noise in (5.6) is assumed to be highly correlated spatially which is not very realistic considering the spatial irregularity of the atmospheric precipitation field.

Using the available Jacobian $\mathcal{J}$ of the deterministic continuation, the mass matrix $M$, which is a diagonal matrix which is only non-zero on the $T$ and $S$ parts, and the forcing $\mathcal{B}$ as described above, we can determine the local probability distribution of a steady state using the generalized Lyapunov equation (3.38). The patterns of the stream function, temperature and salinity of the first eigenvector of the covariance matrix for $\mu = \mu_b$ using the method as described in section 3.4 are shown in Figure 9 and are the same as the first empirical orthogonal function (EOF) (Baars et al., 2017) determined from a long time series. The patterns of the first eigenvector of the covariance matrix are here closely related to the first normal mode [the first eigenvector of the generalized stability problem (3.5)]. Perturbing the system along this direction with a small amplitude will induce the longest decay time back to the steady state in the presence of the salt-advection feedback. This is not a generic result; there may be a sum of eigenmodes which may end up in the first EOF due to the existence of non-normal dynamics (Alexander and Monahan, 2009; Sévellec et al., 2007; Tziperman et al., 2008).

## 5.4. Transfer operator approach

In the bifurcation diagram Figure 8a, a saddle-node bifurcation occurs when $\mu$ is increased slightly from the value of $\mu_b$, say at $\mu_S$. When the freshwater forcing is such that $\mu > \mu_S$, then the MOC will collapse into a flow with sinking at the equator (Fig. 8d), with severe changes in the meridional heat transport. Hence, when $\mu$ slowly increasing from
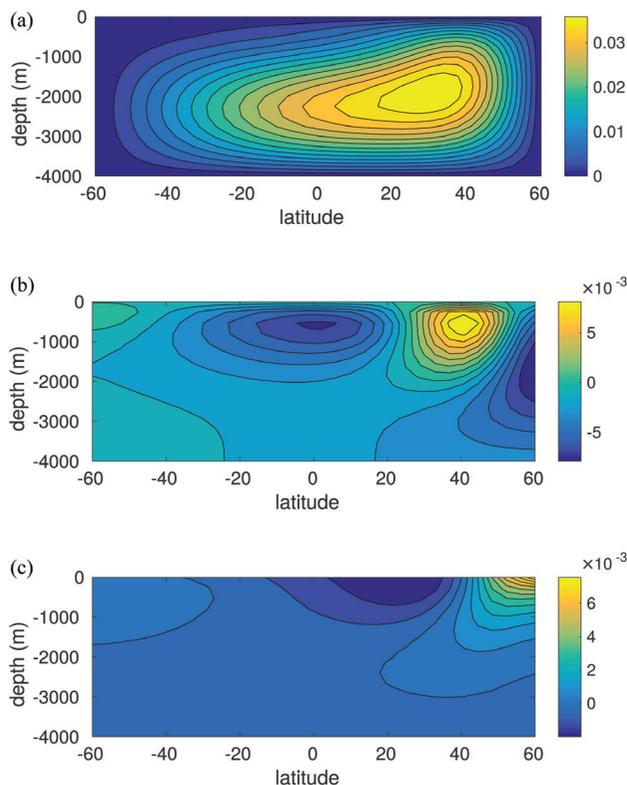


**Figure 9.** Patterns of the stream function $\Psi$ (a), the isothermals (b) and the isohalines (c), respectively, for the first eigenvector of the covariance matrix $C$ obtained by solving the Lyapunov equation for the steady state at $\mu = \mu_b$.

$\mu_b$, one wants to design an early warning indicator of the approaching collapse. Traditional methods make use of the increase in lag-1 autocorrelation or the increase in variance (Lenton, 2011), which are expected near a saddle-node bifurcation as one eigenvalue of the Jacobian matrix approaches zero. However, in practical applications this method may lead to false alarms and too early warning. It is therefore interesting to consider the behaviour of the correlation function (and hence the ergodicity spectrum) near the saddle-node bifurcation.

From (3.29), it is seen that the decay in the correlation function is controlled by the real part of the eigenvalue of the generator $\text{Re}(\lambda_2)$—note that $\lambda_1 = 0$—near the imaginary axis. To approximate the transfer operator, time series of 50 000 years of the stream function $\Psi$ are used with a resolution of 1 year. First, the eigenfunctions of the covariance matrix are calculated and next the time series are projected into the (EOF1, EOF2) space using the principle component time series (PC1, PC2). Second, the transition matrix is calculated as described in section 3.3 using a grid of $40 \times 40$ and the time lag $\tau$ of 51 years.

In the spectrum of the generator at $\mu_b$ (Fig. 10a), there is large spectral gap with a strong decay of correlations. When $\mu$ is increased, $\lambda_2$ becomes real and approaches the imaginary axis (Fig. 10b) leading to long decay rates and explaining the critical slowdown near the saddle-node bifurcation (Gaspard et al., 1995). The results are robust for different grid sizes in the (EOF1, EOF2) space and are also robust for time series with a length equal to or bigger than 40 000 years. While this slowing down is here visible from the decay of the correlation function alone, the spectrum of appropriately reduced transfer operators can reveal a more complete
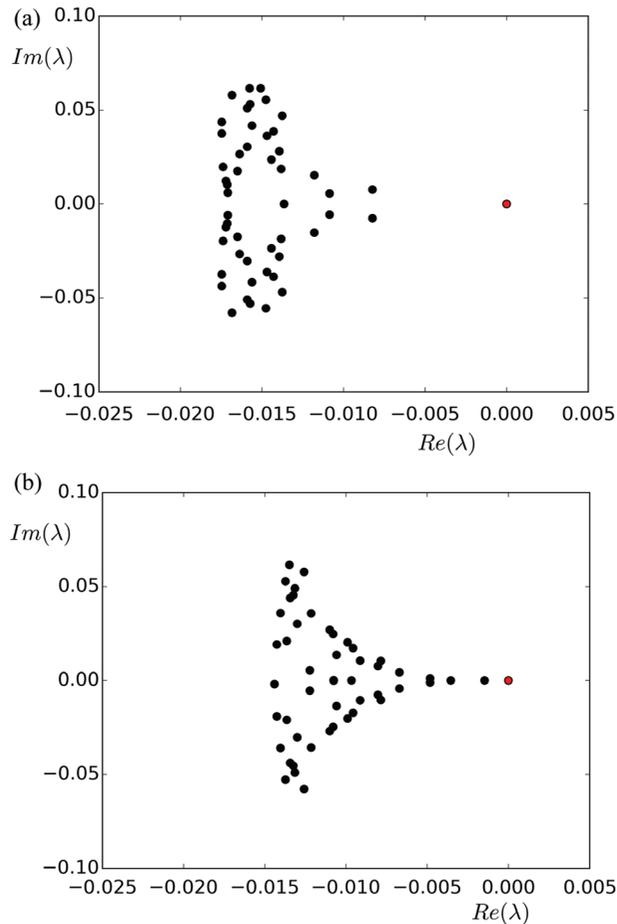


**Figure 10.** Eigenvalues of the generator at the different points before the saddle-node. The transfer operator is calculated using the methods in section 3.3 with a stream function data, projected on the first two EOFs and using a $40 \times 40$ grid in the (EOF1, EOF2) space with a lag of 51 years. (a) $\mu = \mu_b$, (b) at a value of $\mu = (\mu_b + \mu_S)/2$, where $\mu_S$ is the value of $\mu$ at the saddle-node bifurcation.

information about the system, such as the presence of resonance for different observables. However, as this method is 'data hungry', its applicability to determine early warning signals for transitions in the AMOC is very limited.

## 6. Summary and discussion

In this paper, we have given a review of a number of numerical techniques to study transition behaviour in high-dimensional stochastic dynamical systems, such as those arising from the discretization of stochastic partial differential equations. This allows the application of tools from dynamical systems theory (both bifurcation theory and ergodic theory) to problems involving spatially extended systems in many domains.

At the basis of all dynamical systems analysis is a determination of the fixed points of the autonomous 'backbone' system in parameter space (i.e. for which the forcing and parameters are assumed stationary). While these fixed points may be less relevant to explain complicated spatial-temporal behaviour of practical interest, the bifurcation diagram indicates possible intervals in parameter space where multiple states can occur and where transitions to time-dependence take place. Continuation methods, combined with sophisticated solvers for linear systems and generalized eigenvalue problems, are now capable of handling these computations. Practical difficulties in evaluating Jacobian matrices and choices for preconditioners may be application dependent but general strategies exist (Dijkstra et al., 2014).

Further analyses depend on the specific questions asked and whether the mathematical model system is autonomous, non-autonomous or stochastic. Interest in specific mechanisms of transitions in deterministic systems may benefit from the deterministic reduction techniques as presented in section 3.2, where a system of local ordinary differential equations is generated numerically by Galerkin projection. Analysis of this reduced model in the form of amplitude equations enables a very detailed analysis of instability mechanisms (Dijkstra and Katsman, 1997), such as the symmetry breaking at the pitchfork $P_1$ in Figure 5.

When a non-autonomous system is considered, e.g., by adding a time-dependent forcing, the behaviour can often be interpreted with help of the autonomous bifurcation diagram. This includes so-called fast–slow systems, where the timescale of the forcing/parameters is much slower than the internal response timescales of the autonomous system. This timescale is determined by the eigenvalues of the linear stability problem of the fixed points and hence is available from the bifurcation analysis. In many cases, rate-dependent tipping behaviour (Ashwin et al., 2012) can also be explained in terms of classical bifurcation theory although one may also need other theoretical approaches or full trajectory computation.

Under stochastic noise, one can either look at the probability density function under linearized dynamics by solving a Lyapunov equation or directly tackle the full equations by DO field methods. The matrix formulation of the latter method allows a relatively easy application as the generation of the equations for the stochastic coefficients and the DO modes is carried out numerically. Methods to obtain stochastic reduced (non-Markovian) equations, where the interaction of the noise and the non-linearities can be investigated near any background flow, can also be very useful to study the interaction of noise and non-linearities near a (deterministic) bifurcation.

We have demonstrated the potential of the methods by applying them to two canonical problems in physical oceanography, the midlatitude Atlantic wind-driven ocean circulation and the AMOC. Which method to use depends on the eventual question and the type of noise included in a particular model. Each method has its limitations in terms of applicability (e.g. only additive noise) and in the information eventually provided (e.g. only local dynamics). In addition, the methods presented here can be technically demanding for a particular model chosen as Jacobian matrices have to be computed, or transfer operators have to be reconstructed.

The techniques in principle also allow to tackle problems in control and predictability. The availability of the Jacobian matrix is crucial to determine transfer functions in optimal control problems. For predictability studies, an estimate of the evolution of the probability density function, even in a reduced space, is valuable. Furthermore, the adjoint of the tangent linear model is easily constructed from the Jacobian matrix as it is its transpose. Hence, optimal modes, in any norm, can be computed by solving the corresponding optimization problem (van Scheltinga and Dijkstra, 2008).

To conclude, efficient numerical methods are available to study non-linear behaviour in a broad class of climate models which are described by stochastic partial differential equations and we hope that this paper will stimulate their usage, the improvement of current methods and the development of new tools.

# References

Alexander J, Monahan AH. Nonnormal perturbation growth of pure thermohaline circulation using a 2d zonally averaged model. *J Phys Oceanogr* 2009; 39: 369–86.

Arnold L. *Random Dynamical Systems*. Berlin-Heidelberg: Springer Verlag, 1998.

Ashwin P, Wieczorek S, Vitolo R, Cox P. Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philos Trans A Math Phys Eng Sci* 2012; 370: 1166–84.

Baars S, Mulder E, Viebahn J et al. Continuation of probability density functions using a generalized Lyapunov approach. *J Comput Phys* 2017 (in press).

Bartels RH, Stewart G. Solution of the matrix equation AX+XB=C [F4]. *Commun ACM* 1972; 15: 820–6.

Berger A, Loutre MF. Insolation values for the climate of the last 10 million years. *Quat Sci Rev* 1991; 10: 297–317.

Berglund N, Gentz B. *Noise-Induced Phenomena in Slow-Fast Dynamical Systems. A Sample-Paths Approach*. New York: Springer-Verlag, 2006.

Cessi P. A simple box model of stochastically forced thermohaline flow. *J Phys Oceanogr* 1994; 24: 1911–20.

Chekroun MD, Liu H, Wang S. *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II*. Heidelberg: Springer, 2015.

Cunningham SA, Kanzow T, Rayner D et al. Temporal variability of the Atlantic meridional overturning circulation at 26.5 N. *Science* 2007; 317: 935–8.

Dellnitz M, Junge O. On the approximation of complicated dynamical behavior. *SIAM J Numer Anal* 1999; 36: 491–515.

den Toom M, Dijkstra HA, Wubs FW. Spurious multiple equilibria introduced by convective adjustment. *Ocean Modelling* 2011; 38: 126–37.

Dijkstra HA. *Nonlinear Climate Dynamics*. Cambridge, UK: Cambridge University Press, 2013.

Dijkstra HA. *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño*. 2nd Revised and Enlarged Edition. New York: Springer, 2005.

Dijkstra HA, Katsman CA. Temporal variability of the wind-driven quasi-geostrophic double gyre ocean circulation: basic bifurcation diagrams. *Geophys Astrophys Fluid Dyn* 1997; 85: 195–232.

Dijkstra H, Wubs F, Cliffe A et al. Numerical bifurcation methods and their application to fluiddynamics: analysis beyond simulation. *Commun Comput Phys* 2014; 15: 1–45.

Ditlevsen PD, Johnsen SJ. Tipping points: early warning and wishful thinking. *Geophys Res Lett* 2010; 37: L19703.

Doedel EJ. AUTO: a program for the automatic bifurcation analysis of autonomous systems. In: *Proceedings 10th Manitoba Conference on Numerical Mathematics and Computing*. 1980, pp. 265–74.

Doedel EJ, Tuckermann LS. *Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems*. New York: Springer-Verlag, 2000.

Doering C. Modelling stochastic systems: stochastic processes, stochastic differential equations, and Fokker-Planck equations. In Nagel L, Stein D (eds). *Lecture Notes in Complex Systems*. Santa Fe: Santa Fe Institute,1990, pp. 3–51.

Engel K-J, Nagel R. One-parameter semigroups for linear evolution equations. *Semigroup Forum* 2001; 63: 278–80.

Freidlin M, Wentzell A. *Random Perturbations of Dynamical Systems*. Cambridge, UK: Cambridge University Press, 2012.

Froyland G. Approximating physical invariant measures of mixing dynamical systems in higher dimensions. *Nonlinear Anal Theory Methods Appl* 1998; 32: 831.

Gardiner CW. *Handbook of Stochastic Methods*. 2nd edn. Heidelberg-Berlin: Springer, 2002.

Gaspard P, Nicolas G, Provata A. Spectral signature of the pitchfork bifurcation: Liouville equation approach. *Phys Rev E* 1995; 51 : 74–94.

Ghil M. Hilbert problems for the geosciences in the 21st century. *Nonlin Proc Geophys* 2001; 8: 211–22.

Ghil M, Chekroun M, Simonnet E. Climate dynamics and fluid mechanics: natural variability and related uncertainties. *Physica D* 2008; 237: 2111–26.

Gladwell M. *The Tipping Point*. New York-Boston: Little Brown, 2000.

Guckenheimer J, Holmes P. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. 2nd edn. Berlin/Heidelberg: Springer-Verlag, 1990.

Johns WE, Baringer MO, Beal LM. Continuous, array-based estimates of Atlantic Ocean heat transport at 26.5 N. *J Climate* 2011; 24: 2429–49.

Katzav J, Dijkstra HA, de Laat ATJJ. Assessing climate model projections State of the art and philosophical reflections. *Stud Hist Philos Mod Phys* 2012; 43: 258–76.

Kawabe M. Variations of current path, velocity, and volume transport of the Kuroshio in relation with the large meander. *J Phys Oceanogr* 1995; 25: 3103–17.

Keller HB. Numerical solution of bifurcation and nonlinear eigenvalue problems. In: Rabinowitz PH (ed). *Applications of Bifurcation Theory*. New York: Academic Press, 1977, pp. 359–84.

Kuehn C. A mathematical framework for critical transitions: bifurcations, fast–slow systems and stochastic dynamics. *Physica D* 2011; 240: 1020–35.

Kuehn C. A mathematical framework for critical transitions: normal forms, variance and applications. *J Nonlinear Sci* 2013; 23: 457–510.

Kuehn C. Deterministic continuation of stochastic metastable equilibria via Lyapunov equations and ellipsoids. *SIAM J Sci Comput* 2012; 34: A1635–58.

Kuehn C. Efficient gluing of numerical continuation and a multiple solution method for elliptic PDEs. *Appl Math Comput* 2015; 266: 656–74.

Kuhlbrodt T, Griesel A, Montoya M et al. On the driving processes of the Atlantic meridional overturning circulation. *Rev Geophys* 2007; 45: 1–32.

Lasota A, Mackey MC. *Chaos, Fractals and Noise*. 2nd edn. Berlin/Heidelberg: Springer-Verlag, 1994.

Lenton TM. Early warning of climate tipping points. *Nat Clim Change* 2011; 1: 201–9.

Lenton TM, Held H, Kriegler E et al. Tipping elements in the Earth's climate system. *Proc Natl Acad Sci USA* 2008; 105: 1786–93.

Mantua NJ, Hare S, Zhang Y, Wallace JM, Francis RC. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull Amer Meteor Soc* 1997; 78: 1069–79.

Nonaka M, Sasaki H, Taguchi B, Nakamura H. Potential predictability of interannual variability in the Kuroshio extension jet speed in an eddy-resolving ogcm. *J Climate* 2012; 25: 3645–52.

Pedlosky J. *Geophysical Fluid Dynamics*. 2nd edn. New York: Springer-Verlag, 1987.

Peixoto JP, Oort AH. *Physics of Climate*. New York: AIP Press, 1992.

Philander SGH. *El Niño and the Southern Oscillation*. New York: Academic Press, 1990.

Qiu B, Chen S. Variability of the Kuroshio Extension jet, recirculation gyre and mesoscale eddies on decadal time scales. *J Phys Oceanogr* 2005; 35: 2090–103.

Qiu B, Chen S. Eddy-mean flow interaction in the decadally modulating Kuroshio extension system. *Deep Sea Res II* 2010; 57: 1098–110.

Qiu B, Miao W. Kuroshio path variations south of Japan: bimodality as a self-sustained internal oscillation. *J Phys Oceanogr* 2000; 30: 2124–37.

Rahmstorf S. The thermohaline circulation: a system with dangerous thresholds? *Clim Change* 2000; 46: 247–56.

Sapsis TP. Dynamically Orthogonal Field Equations for Stochastic Fluid Flows and Particle Dynamics. PhD Thesis, MIT. 2010.

Sapsis TP, Dijkstra HA. Interaction of additive noise and nonlinear dynamics in the double-gyre wind-driven ocean circulation. *J Phys Oceanogr* 2013; 43: 366–81.

Sapsis TP, Lermusiaux PFJ. Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D* 2009; 238: 2347–60.

Sapsis TP, Lermusiaux PFJ. Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty. *Physica D* 2012; 241: 60–76.

Schmeits MJ, Dijkstra HA. Bimodality of the Kuroshio and the Gulf Stream. *J Phys Oceanogr* 2001; 31: 2971–85.

Sévellec F, Jelloul MB, Huck T. Optimal surface salinity perturbations influencing the thermohaline circulation. *J Phys Oceanogr* 2007; 37: 2789–808.

Seydel R. *Practical Bifurcation and Stability Analysis: From Equilibrium to Chaos*. New York: Springer-Verlag, 1994.

Simonnet E, Dijkstra HA. Spontaneous generation of low-frequency modes of variability in the wind-driven ocean circulation. *J Phys Oceanogr* 2002; 32: 1747–62.

Slingo J, Palmer T. Uncertainty in weather and climate prediction. *Philos Trans A Math Phys Eng Sci* 2011; 369: 4751–67.

Smeed DA, McCarthy GD, Cunningham SA, et al. Observed decline of the Atlantic meridional overturning circulation 2004–2012. *Ocean Sci* 2014; 10: 29–38.

Spall MA. On the thermohaline circulation in flat bottom marginal seas. *J Mar Res* 2003; 61: 1–25.

Stommel H. Thermohaline convection with two stable regimes of flow. *Tellus* 1961; 2: 244–30.

Sura P, Fraedrich K, Lunkeit F. Regime transitions in a stochastically forced double-gyre model. *J Phys Oceanogr* 2001; 31: 411–26.

Taft B. Characteristics of the flow of the Kuroshio south of Japan. In: Stommel H, Yoshida K (eds). *Kuroshio, Physical Aspects of the Japan Current*. Seattle, WA: University of Washington Press, 1972, pp. 17–43.

Taguchi B, Xie SP, Schneider N et al. Decadal variability of the Kuroshio Extension: observations and an eddy-resolving model hincast. *J Climate* 2007; 20: 2357–77.

Tantet A, van der Burgt FR, Dijkstra HA. An early warning indicator for atmospheric blocking events using transfer operators. *Chaos* 2015; 25: 036406.

Tziperman E, Zanna L, Penland C. Nonnormal thermohaline circulation dynamics in a coupled oceanatmosphere gcm. *J Phys Oceanogr* 2008; 38: 588–604.

Ulam SM. *A Collection of Mathematical Problems*. New York: Interscience Publ., 1961. Problems in Modern Mathematics, 1964.

Van der Vaart PCF, Schuttelaars HM, Calvete D, Dijkstra HA. Instability of time-dependent wind-driven ocean gyres. *Phys Fluids* 2002; 14: 3601–15.

van Scheltinga ADT, Dijkstra HA. Conditional nonlinear optimal perturbations of the double-gyre ocean circulation. *Nonlinear Process Geophys* 2008; 15: 727–34.

Von Storch H, Zwiers FW. *Statistical Analysis in Climate Research*. Cambridge, UK: Cambridge University Press, 1999.

Zachos J, Pagani M, Sloan L, Thomas KBE. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 2001; 292: 686–93.