# Reduction methods in climate dynamics—A brief review

Felix Hummel [a], Peter Ashwin [b],*, Christian Kuehn [a],**

[a] *Department of Mathematics, TU Munich, Germany*
[b] *Department of Mathematics and Statistics, University of Exeter, Exeter EX4 4QF, UK*

## ARTICLE INFO

## ABSTRACT

Currently the number of reduction methods used in practice in climate applications is vast and tends to be difficult to access for researchers searching for an overview of the area. In this work, we review a range of reduction methods that have been, or may be useful for connecting models of the Earth's climate system of differing complexity. We particularly focus on methods where rigorous reduction is possible. We aim to highlight the main mathematical ideas of each reduction method and also provide several benchmark examples for the reduction from climate modelling. In particular, our goal is to provide a broad overview of available methods.

## Contents

---

* Corresponding author.
** Corresponding author.
  *E-mail addresses:* P.Ashwin@exeter.ac.uk (P. Ashwin), ckuehn@ma.tum.de (C. Kuehn).

## 1. Introduction

In order to understand the Earth's climate system, a larger number of models of varying complexity have been developed:

these cover processes at a wide variety of space and timescales and give rise to hierarchy of models [1,2] for various aspects of the climate. The highest resolution and most complex Earth System Models (ESMs), such as [3], underpin the science Climate Modelling Intercomparison Project (CMIP) forming the scientific basis for the Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC) [4]. ESMs try to describe the Earth System as accurately as possible and to make precise prediction of the evolution of the Earth's climate and its different components, and from these results we can make conclusions not just about natural variability of the climate system but also about variability that is due to anthropogenic effects. Models that attempt to understand the atmosphere and ocean are called general circulation models (GCMs) [5]. Since ESMs and GCMs are of high resolution and include complex physics, they are very difficult to analyse.

Climate and weather models[1] are typically nonlinear differential equations that are mathematically challenging (compressible Navier–Stokes equations with highly nontrivial thermodynamics and chemistry, and topographically non-trivial boundary conditions and a variety of forcing terms) [2]. Over very long timescales different processes become important for the dynamics [6]. Hence, they are usually simulated with a high resolution on very powerful computers but typically only short runs can be achieved at high resolution. For some questions that involve longer timescales, one can consider a hierarchy of models [7] where high resolution ESM/GCMs can be approximated using lower resolution ESMs or Earth system models of intermediate complexity (EMICs). These models usually contain several simplifications and a larger grid size compared to the ESMs/GCMs and they therefore need less computing time. However, EMICs are still mostly too complex for analytical study without further simplification.

To have any chance of analytically understanding components of the Earth's climate, one has to go further down the model hierarchy and study conceptual models such as box models or energy balance models that contain not more than a few variables, where hidden variables may be assumed to be in equilibrium or may be represented as noise terms [2,8–10]. It is remarkable that simple models can often reproduce many aspects of the Earth's climate system [1,11] and while the similarity of the behaviour of complex and simple models may in some cases not be rigorously jusitfiable, in other cases it can be formally linked through a rigorous reduction procedure. For example, one of the key problems in climate science is the determination of Equilibrium Climate Sensitivity (ECS) [12]. This is defined to be the equilibrium increase in global mean surface temperature that results from a doubling of $CO_2$ in the atmosphere and as such is used as the key scientific link between anthropogenic emissions and climate change [13,14]. A rigorous justification of ECS would need to span the model hierarchy as it involves reduction of ESM runs to an equation that just considers global radiation balance [8].

The aim of this paper to give an overview of prominent and rigorously justifiable reduction methods that have found use in climate modelling, and to highlight others that may find use in future. To keep the review sufficiently brief, we focus on reduction methods for differential equations and only comment on other (e.g. data-driven) reduction methods for stochastic differential equations in the outlook, though we do discuss an example of diffusion maps in  Appendix. Methods for model reduction have become so large and diverse in recent years that getting an overview of the main approaches for PDEs is a major challenge when entering the field. Here, we hope to alleviate this

by providing a relatively concise, yet sufficiently broad, entry point to the literature. Furthermore, we point out that beyond getting an overview of the available reduction techniques, several additional tasks are crucial. For example, one should aim to understand, how the dynamics of the reduced model reproduces a more complex one, and vice versa, whether a large-scale model correctly captures basic physical effects and avoids numerical over-stabilization. Such types of tasks, albeit highly interesting, are beyond the scope of this work. Here we focus on the mathematical ideas underlying each reduction method to make it easier to distinguish, as well as understanding why a reduced model has a particular form.

### 1.1. Reduction of deterministic climate models

We present a survey of several major reduction methods that can be applied to climate models. For each approach, we start with its motivation, then present the main reduction idea, giving some concrete examples and comments on the method. In the absence of time-dependent forcing, a climate model can described by

$$\dot{z} = F(z) \tag{1.1}$$

where $z$ represents the discretized state. However, there are slowly evolving aspects (climate/ocean) and faster (weather) variables. This splitting into climate and weather variables, proposed by Hasselmann in [8], has proven to be a very fruitful point of view and has been explored in many papers since; see for example [15,16]. Indeed, the wide range of timescales from tectonic to atmospheric processes [17], recently reviewed in [18], highlight there is a multiscale hierarchy inherent in earth system processes. To illustrate the weather-climate scale separation, consider a fast–slow evolution equation of the form

$$\dot{x} = f(x, y),$$
$$\dot{y} = \varepsilon g(x, y), \tag{1.2}$$

where $z = (x, y)$ is split into slow (climate) variables $y$ and the fast (weather) variables $x$. The separation of timescales is indicated by a small parameter $\varepsilon > 0$. Some of the methods, such as slow manifold reductions, averaging or homogenization, try to make use of the special structure such as (1.2) with the aim of reducing to the climate variable $y$ only. Others, such as certain spectral Galerkin approximations or asymptotic expansions, can be used to formally establish (1.2) from (1.1).

In Section 2, we consider an abstract forced evolution equation and start with direct linear projection methods, which are among the most classical – yet quite coarse – reduction tools. There exist many methods which can help in finding lower-dimensional structures in high dimensional data. In the context of climate data, a reduction via empirical orthogonal functions (EOFs), which we discuss in Section 2.2, is among the most standard. In Section 3, we proceed to methods motivated by asymptotic analysis and the principle of dominant balance, where one identifies suitable asymptotic scales directly within evolution equations and formally discards higher-order terms. In Section 4, we introduce methods based upon invariant manifold theory, which can be rigorously justified in many cases. Note that reduction of the coupled system (1.1) to a system with the climate variable only does not necessarily lead to a deterministic equation. Sometimes, the variability of the climate system is better described by stochastic forcing or by including a delay term. In fact, the reduction from a deterministic equation to an equation with random forcing can be described by averaging and homogenization techniques which we discuss in Section 5. We consider classical moment closure schemes in Section 6. If no direct deterministic reduction for a

---

[1] We refer to both types of models simply as climate models in the remainder.

**Table 1**

A summary of the reduction methods discussed in this paper: the relevant Section in the paper is denoted with §X.Y. Note that ODE/PDE/SDE/DDE refer to ordinary/partial/stochastic/delay differential equation models.

| Reduction method | From | To |
|---|---|---|
| Spectral Galerkin approximations (Section 2.1) | PDE | ODE |
| Empirical orthogonal functions (Section 2.2) | High dimensional data set<br>High dimensional ODE | Low dimensional data set<br>Low dimensional ODE |
| Principal interaction patterns (Section 2.3) | High dimensional ODE | Low dimensional ODE |
| Characteristic scales (Section 3) | PDE | PDE of lesser complexity |
| Inertial manifolds (Section 4.1) | PDE | ODE |
| Slow manifolds (Section 4.2) | Fast–slow system of PDEs | PDE with slow variable only<br>ODE with slow variable only |
| | Fast–slow system of ODEs | ODE with slow variable only |
| Centre manifolds (Section 4.3) | ODE<br>PDE | ODE with neutral directions only |
| | PDE | PDE with neutral directions only |
| Averaging (Section 5.1) | Fast–slow system of ODEs | ODE with slow variable only |
| | Fast–slow system of SDEs | ODE with slow variable only<br>SDE with slow variable only |
| | Fast–slow system of SPDEs | PDE with slow variable only<br>SPDE with slow variable only |
| Homogenization (Section 5.2, 5.3) | PDE with strongly<br>varying coefficients | PDE with constant coefficients |
| | Fast–slow system of ODEs<br>with two time scales | SDE with one<br>time scale |
| Mori–Zwanzig (Section 5.4) | PDE<br>ODE | ODE/DDE |
| Moment closure methods (Section 6) | SDE<br>PDE | ODE |

given evolution equation is available, one often needs (preliminary) statistical/data-based tools and we include an Appendix to introduce one instance of statistical manifold learning to point the reader to the complementary view of primarily data-driven approaches.

Although we do not claim to cover all possible reduction techniques, we hope that laying out the general ideas via mathematically concrete examples will help the reader better understand the (dis-)advantages of each reduction principle.

*1.2. Classification of reduction methods*

Climate models can be described by different types of equations. They can appear as differential equations with possible delay, spatial derivatives and/or stochastic terms. Many of the reduction methods we want to discuss in this article start with one of these types of equations and transform it into another. One can use this to structure the myriad of reduction methods available in the literature. For the few methods we treat in this article, Table 1 gives a summary overview of reduction methods explored later in the paper.

A more precise description of the different methods follows in the subsequent sections, together with some examples of applications to aspects of the climate system. We hope that this provides readers, who have concrete models in mind, a structured view, which methods are available and how to adapt the main idea to their concrete context. Due to complexity of the climate system and the sort of questions that need to be answered, it is clearly too optimistic to expect a "one-catch-all method" for climate model reduction.

## 2. Linear truncation methods

If one seeks to reduce the complexity of a model or the size of a data set, it is a common approach to project it to a lower-dimensional linear subspace of the state space, which is believed to contain the most important interactions of the model or the data set. The hope is that the parts that are lost during this truncation procedure are not essential and that they can be neglected without losing any meaningful information. Such techniques can be very useful in many different situations: Galerkin approximations can be used to derive well-posedness of partial differential equations on an abstract level [19], finite element methods are a standard tool in numerical applications [20], and a principal component analysis is a well established procedure to structure a given data set [21]. There are many different ways to construct the linear subspace on which the model or data set should be projected and it would be beyond the scope of this paper to discuss all of them. Instead, we just treat three of these methods: First, we study a Fourier–Galerkin approach that can be used to reduce a partial differential equation to an ordinary differential equation. It can also be used to establish a splitting like (1.2) in a fast and a slow variable. After that, we discuss empirical orthogonal functions (EOFs) which is a standard technique to reduce the dimensionality of a data set. However, when applied to a dynamical system, it usually only preserves the statistical and not the dynamical properties of the system. There are methods which try to address this issue, one of them being the use of principal interaction patterns (PIPs), which is the last method we study in this section.

*2.1. A Galerkin approach*

Galerkin approximation appears in different situations in the context of partial differential equations. It is a versatile tool and can be used for deriving the abstract well-posedness of partial differential equations as well as for the development of numerical algorithms. Galerkin approximations are common, and there are many textbooks available, see for example [22,23].

**Idea:** In certain situations infinite-dimensional systems may come with a natural (Schauder) basis which one can use to

construct finite dimensional subspaces. For example, if one has a partial differential equation with periodic boundary conditions, this can be the canonical Fourier basis, or if there is a self-adjoint operator on a Hilbert space involved, then this may be the set of eigenfunctions of this operator. This is a spectral Galerkin kind of approach, which is often most convenient if periodic boundary conditions in space are imposed. The idea is to expand the equation into a Fourier series in space and to omit high frequencies. Oftentimes, such a basis comes with a natural order. In the case of the Fourier basis, this order is given by the increasing frequencies, and in the case of eigenfunctions of a self-adjoint operator, it may be given by the order of the eigenvalues on the real axis. The finite-dimensional approximations are then obtained by projecting the system to the linear span of a finite number of basis vectors.

**Outline of the Reduction Procedure:** A standard scenario in which Galerkin approximations can be applied is given by the abstract evolution equation

$$\partial_t u(t) = F(t, u(t)), \quad u(0) = u_0 \tag{2.1}$$

in a Hilbert space $H$. Let $(e_k)_{k \in \mathbb{N}}$ be an orthonormal basis of $H$. Then one can define the $n$-dimensional spaces $H_n := \mathrm{span}\{e_1, \ldots, e_n\}$ ($n \in \mathbb{N}$) and the canonical orthogonal projections $P_n : H \to H_n$. If we truncate (2.1) to $H_n$, we obtain the finite-dimensional system

$$\partial_t v_n(t) = P_n F(t, v_n(t)), \quad v_n(0) = P_n u_0, \tag{2.2}$$

where $v_n := P_n u$. Let $\langle \cdot, \cdot \rangle$ be the inner product of $H$. Then we can rewrite (2.2) in terms of

$$\partial_t \sum_{j=1}^{n} \langle v_n(t), e_j \rangle e_j = \sum_{j=1}^{n} \langle P_n F(t, v_n(t)), e_j \rangle e_j,$$

$$\sum_{j=1}^{n} \langle v_n(0), e_j \rangle e_j = \sum_{j=1}^{n} \langle u_0, e_j \rangle e_j. \tag{2.3}$$

Now, we define $x_j^n(t) := \langle v_n(t), e_j \rangle$, $f_j^n(t, x_1^n(t), \ldots, x_n^n(t)) := \langle P_n F(t, v_n(t)), e_j \rangle$ and $x_{0,j}^n := \langle u_0, e_j \rangle$ so that (2.3) can be rewritten as

$$\frac{d}{dt} x_j^n(t) = f_j^n(t, x_1^n(t), \ldots, x_n^n(t)), \quad x_j^n(0) = x_{0,j}^n, \quad (j = 1, \ldots, n) \tag{2.4}$$

by the linear independence of $(e_k)_{k \in \mathbb{N}}$. If we define

$$x^n := (x_1^n, \ldots, x_n^n),$$
$$f^n := (f_1^n, \ldots, f_n^n),$$
$$x_0^n := (x_{0,1}^n, \ldots, x_{0,n}^n),$$

then we obtain the more compact form

$$\dot{x}^n(t) = f^n(t, x^n(t)), \quad x^n(0) = x_0^n. \tag{2.5}$$

This is now an $n$-dimensional system of ordinary differential equations and is called a Galerkin approximation of (2.1). Moreover, under relatively mild conditions it holds that $x^n \to x$ weakly as $n \to \infty$ in $H$. For more details on the precise assumptions and statements we refer for example to [24, Theorem 2.76] or [25, Chapter III.4].

**Example:** We follow [9], where the equations for barotropic flow on a beta plane with topography and mean flow are considered:

$$\frac{\partial q}{\partial t} + \nabla^\perp \psi \cdot \nabla q + U \frac{\partial q}{\partial x_1} + \beta \frac{\partial \psi}{\partial x_1} = 0,$$
$$q = \Delta \psi + h, \tag{2.6}$$
$$\frac{dU}{dt} = \oint h \frac{\partial \psi}{\partial x_1}.$$

Here, $q(x_1, x_2, t)$ denotes the small-scale potential vorticity, $U(t)$ is the mean flow, $\psi(x_1, x_2, t)$ is the small-scale stream function, and $h(x_1, x_2)$ denotes the underlying topography, whereas $\beta$ approximates the variation of the Coriolis parameter. Moreover, we use the notation $\nabla^\perp := (-\partial_{x_2}, \partial_{x_1})$ and the integral denotes the mean integral over the 2-dimensional torus $\mathbb{T}^2$. In [9], the equation is studied with periodic boundary conditions so that one may take the Fourier basis functions $([x \mapsto e^{ikx}])_{k \in \mathbb{Z}^2}$ as an orthonormal basis of $L_2(\mathbb{T}^2)$. As in [9, Section 3] we choose $\Lambda \in \mathbb{N}$ and define $\sigma_\Lambda = \{k \in \mathbb{Z}^2 : 1 \le |k|^2 \le \Lambda\}$, $B_\Lambda = \{e^{ikx} : k \in \sigma_\Lambda\}$ as well as the orthogonal projection $P_\Lambda$ onto $H_\Lambda := \mathrm{span}\, B_\Lambda$, where orthogonality is understood with respect to the $L_2(\mathbb{T}^2)$-scalar product. Moreover, we write

$$\psi_\Lambda(x) = \sum_{1 \le |k|^2 \le \Lambda} \hat{\psi}_k(t) e^{ikx}, \quad h_\Lambda(x) = \sum_{1 \le |k|^2 \le \Lambda} \hat{h}_k e^{ikx},$$

$$q_\Lambda(x, t) = \sum_{1 \le |k|^2 \le \Lambda} \hat{q}_k(t) e^{ikx}$$

for the truncations to $H_\Lambda$ and substitute $u_k(t) := |k| \psi_k(t)$. By the reduction procedure described above, one obtains the system

$$\frac{dU}{dt} = \mathrm{Im} \sum_{k \in \sigma_\Lambda} H_k \overline{u_k},$$

$$\hat{q}_k = -|k| u_k + \hat{h}_k \quad (k \in \sigma_\Lambda),$$

$$\frac{du_k}{dt} = iH_k U - i(k_x U - \Omega_k) u_k + \sum_{l \in \sigma_\Lambda} L_{kl} u_l$$

$$+ \frac{1}{2} \sum_{\substack{l, m \in \sigma_\Lambda \\ k+l+m=0}} B_{klm} \overline{u_l} \overline{u_m} \quad (k \in \sigma_\Lambda),$$

where

$$L_{kl} = \frac{k_x l_y - k_y l_x}{|k||l|} h_{k-l},$$

$$B_{klm} = (l_y m_x - l_x m_y) \frac{|l|^2 - |m|^2}{|k||l||m|},$$

$$\Omega_k = \frac{k_x \beta}{|k|^2}, \quad H_k = \frac{k_x \hat{h}_k}{|k|}.$$

The Fourier basis could also be used to derive a splitting in a weather and a climate variable as in (1.2), since higher Fourier modes have stronger oscillations which oftentimes lead to faster dynamics. In this sense, one can introduce $1 \le \overline{\Lambda} \le \Lambda$ and consider $u_k, q_k$ with $k \in \sigma_{\overline{\Lambda}}$ and $U$ as slow climate variables and $u_k, q_k$ with $k \in \sigma_\Lambda \setminus \sigma_{\overline{\Lambda}}$ as fast weather variables. We refer to [9] for further details and to Section 5 where we continue with this example.

**Comments:**

- Galerkin-type methods have the advantage that they are often easy to implement and that they provide a nice geometric intuition as solutions are expanded in basis functions. Therefore, one hopes to capture most features that can be expressed in terms of the finite number of basis functions left in the approximation. Numerical tests can often indicate how many basis functions are needed in practice.

- One disadvantage is that models can become large very quickly if a sufficient accuracy is required, which somewhat defies the initial point of model reduction as a system of nonlinear ordinary differential equation with very many variables is analytically frequently not much easier than directly analysing the initial partial differential equation, e.g., in the context of existence and stability analysis of patterns. Particularly for fluid dynamics, e.g., Navier–Stokes-type equations, it is well-known that small-scale turbulence

can spread from very high Fourier modes into low modes in certain cases.

- Beyond global basis functions such as Fourier bases, one can also consider compactly supported basis functions in finite element or finite volume schemes. These methods are numerically also extremely efficient but usually not capable of providing a substantial reduction in model complexity as the set of basis functions required is often much larger in comparison to spectral Galerkin approaches.

In fact, we shall see below that the advantages and disadvantages mentioned above, are quite typical not only for Galerkin-type methods but for all methods based upon linear projection/reduction. The next method shows another classical example of this principle.

### 2.2. Empirical orthogonal functions (EOFs)

There are many different names for the method that yields the construction of the so-called empirical orthogonal functions. Closely related methods are also referred to as singular value decomposition [26], principal component analysis [27] or proper orthogonal decomposition [28]. This technique is well established and widely used, not only for the reduction of climate models or climate data. We refer to [29–32].

**Idea:** The aim is to construct a subspace of a given dimension which carries the largest amount of variance of the given data set. The data set is then projected orthogonally to this subspace. If one starts with a dynamical system instead of a data set, then one can obtain a data set by taking a certain number of snapshots of trajectories of the dynamical system.

**Outline of the Reduction Procedure:** One can start with a collection of $N \in \mathbb{N}$ points $x_1, \ldots, x_N \in \mathbb{R}^d$. Alternatively, if a dynamical system in a $d$-dimensional state space is given, then one can take $N$ snapshots of a trajectory $(x_1, \ldots, x_N) = (z(t_1), \ldots, z(t_N))$ at different times $t_1, \ldots, t_N \geq 0$ or a collection of such snapshots of trajectories with different initial values. Either way, one works with a matrix $X = (x_1 - \bar{x}, \ldots, x_N - \bar{x}) \in \mathbb{R}^{d \times N}$ which is given as a concatenation of the data set relative to the mean $\bar{x} := \frac{1}{N} \sum_{j=1}^{N} x_j$. The matrix $XX^T \in \mathbb{R}^{d \times d}$ is symmetric and positive semidefinite. Therefore, it has $d$ eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$ and corresponding orthonormal eigenvectors $v_1, \ldots, v_d \in \mathbb{R}^d$. From a statistical point of view the matrix is the covariance matrix of the data set multiplied $N$ (or $N-1$ depending on the convention). Thus, the eigenvalue $\lambda_i$ can be considered as a measure of variance the direction of $v_i$ carries. Since the initial idea was to determine the directions with the most variance, one projects the data set to the subspace generated by $v_1, \ldots, v_k$ for some $k \leq d$. More precisely, one defines

$$x_l^k := \sum_{j=1}^{k} \langle x_l, v_j \rangle v_j \quad (l = 1, \ldots, N).$$

Here, $\langle x_l, v_j \rangle$ denotes the usual Euclidean scalar product, but note that also other choices of scalar products are possible, as we will explain later in the comments. The points $x_1^k, \ldots, x_N^k$ are now contained in a $k$-dimensional subspace of $\mathbb{R}^d$. The choice of $k$ depends on the application in mind and can also be chosen according to the given data set. Small values of $k$ have the advantage that one ends up with a very low-dimensional data set. However, the approximation gets rougher with smaller values of $k$. Hence, there is a trade-off between the dimension of the reduced data set and the quality of approximation of the original data set.

**Example:** There are many works that use EOFs and its extensions in a climate context, see for example [33–40]. Let us briefly explain using the example in [35]. Therein, the authors compare reductions of the Charney–DeVore model with different basis functions, among them EOFs. The Charney–DeVore model was derived in [41] as a 6-dimensional Galerkin approximation of the vorticity equation for a large-scale atmospheric flow on a $\beta$-plane channel with topography which is similar to Eq. (2.6). The reduced equations which are considered in [35] are of the form

$$\begin{aligned}
\dot{x}_1 &= \widetilde{\gamma}_1 x_3 - C(x_1 - x_1^*), \\
\dot{x}_2 &= -(\alpha_1 x_1 - \beta_1) x_3 - C x_2 - \delta_1 x_4 x_6, \\
\dot{x}_3 &= (\alpha_1 x_1 - \beta_1) x_2 - \gamma_1 x_1 - C x_3 + \delta_1 x_4 x_5, \\
\dot{x}_4 &= \widetilde{\gamma}_2 x_6 - C(x_4 - x_4^*) + \eta(x_2 x_6 - x_3 x_5), \\
\dot{x}_5 &= -(\alpha_2 x_1 - \beta_2) x_6 - C x_5 - \delta_2 x_4 x_3, \\
\dot{x}_6 &= (\alpha_2 x_1 - \beta_2) x_5 - \gamma_2 x_4 - C x_6 + \delta_2 x_4 x_2,
\end{aligned} \tag{2.7}$$

where $(x_1, x_2, x_3, x_4, x_5, x_6)$ are the unknown functions, $x_1^*, x_4^*$ are forcing terms and the other quantities are model parameters taking the form

$$\alpha_j = \frac{8\sqrt{2}}{\pi} \frac{j^2}{4j^2 - 1} \frac{b^2 + j^2 - 1}{b^2 + j^2},$$

$$\beta_j = \frac{\beta b^2}{b^2 + j^2},$$

$$\delta_j = \frac{64\sqrt{2}}{15\pi} \frac{b^2 - j^2 + 1}{b^2 + j^2},$$

$$\widetilde{\gamma}_j = \gamma \frac{4j}{4j^2 - 1} \frac{\sqrt{2}b}{\pi}, \quad \eta = \frac{16\sqrt{2}}{5\pi},$$

$$\gamma_j = \gamma \frac{4j^3}{4j^2 - 1} \times \frac{\sqrt{2}b}{\pi(b^2 + j^2)} \quad (j \in \{1, 2\}).$$

Here, $b$ models the length–width ration of the $\beta$-channel, $\beta$ comes from the Coriolis force, $\gamma$ from the topographic height and the damping parameter $C$ is determined by the friction in the Ekman layer. In the vorticity equation from which (2.7) is derived, it is assumed that the forcing term points in the zonal direction only. During the Galerkin reduction, this leads to the fact that there are only forcing terms in the equations for $x_1$ and $x_4$. This model can show rapid and chaotic transitions between different flow regimes. As was shown in [42], this is the case for the parameter values

$$(x_1^*, x_4^*, C, \beta, \gamma, b) = (0.95, -0.76095, 0.1, 1.25, 0.2, 0.5).$$

In [35] it was investigated, whether these regime transitions would persist under an EOF reduction. As described in the outline of the EOF reduction procedure above, the EOFs were computed from the data set generated by a numerical integration of the full model. The authors also used different energy metrics, i.e., different scalar products for the computation of the EOFs. It was found that none of the reduced models was able to reproduce the chaotic transition behaviour of the full model. Even though some of the reduced models show regime transitions, they are far too regular and of periodic type. This already indicates one of the problems of EOF reductions. Even though reduced models usually show the same or a similar statistical behaviour as the original model, the dynamical behaviour might still be very different.

**Comments:**

- In applications, variance and distances are sometimes not measured in terms of the Euclidean metric, but, for example, with respect to a given energy metric determined by a symmetric positive definite matrix $M \in \mathbb{R}^{d \times d}$. In this case, orthogonality is not understood with respect to the Euclidean scalar product, but with respect to the scalar product

$$(x, y)_M = \langle x, My \rangle.$$

In this case, the covariance matrix is given by $M^{1/2}X$ $(M^{1/2}X)^\top$ and $\lambda_1^{(M)}, \ldots, \lambda_N^{(M)}$ as well as $\widetilde{v}_1^{(M)}, \ldots, \widetilde{v}_N^{(M)}$ are defined by the relation

$$M^{1/2}X(M^{1/2}X)^T \widetilde{v}_i^{(M)} = \lambda_i^{(M)} \widetilde{v}_i^{(M)}.$$

As above, the $\widetilde{v}_1^{(M)}, \ldots, \widetilde{v}_N^{(M)}$ are orthonormal with respect to the Euclidean scalar product. Therefore, the basis defined by $v_i^{(M)} := M^{-\frac{1}{2}} \widetilde{v}_N^{(M)}$ is orthonormal with respect to $M$, i.e.

$$(v_i^{(M)}, v_j^{(M)})_M := \langle v_i^{(M)}, M v_j^{(M)} \rangle = \delta_{i,j}.$$

The projection of the data points is then given by

$$x_l^{k,(M)} := \sum_{j=1}^{k} (x_l, v_j^{(M)})_M v_j^{(M)} = \sum_{j=1}^{k} \langle x_l, M v_j^{(M)} \rangle v_j^{(M)}$$

where $l = 1, \ldots, N$. In the literature, it is often common to define

$$v_1^{(M)}, \ldots, v_N^{(M)}$$

directly as the eigenvectors of the eigenvalue problem

$$XX^\top M v_i^{(M)} = \lambda_i^{(M)} v_i^{(M)},$$

see for example [35, Section 3] or [43, Section 4.3]. But it is easy to see that both definitions are mathematically equivalent. Yet, it is often natural to re-scale via an energy metric to bring different directions onto a common scale. In fact, introducing the matrix $M$ into the inner product corresponds to an an-isotropic or ellipsoidal deformation of phase space. This is a common trick not only appearing from a physical or geometric perspective but also in the context of statistical techniques, where the Mahalanobis distance uses a scaling matrix $M$ based upon the covariance of a stochastic process [44,45].

- In some sense, adding the energy metric $M$ is an extension of the usual EOF approach. However, this is not the only possible extension. We refer the reader to [30,46] for more content in this direction.
- EOFs are designed to reduce the dimension of a data set while giving a good approximation of its statistical features. However, this method does not aim to preserve the dynamics of a dynamical system and indeed it has problems in doing so. This is pointed out in several papers, e.g. [35,47] and references therein. In particular, one cannot expect that dynamical notions such as invariant manifolds, attractors or bifurcations persist under an EOF-reduction.
- Furthermore, following up on the last drawback, EOFs also struggle with providing a transparent method for parameter tracing through the reduction. Therefore, one often has to add back in parameters after the reduction, see e.g. [48], to perform a bifurcation analysis.

As discussed above, EOFs cannot be expected to preserve the dynamics of a dynamical system in general. A reduction method via principal interaction patterns, which we discuss next, is designed to improve on this issue.

### 2.3. Principal interaction patterns (PIPs)

Now we consider another choice of basis functions, namely principal interaction patterns (PIPs). The idea behind PIPs was introduced by Hasselmann in [49]. It was then used and further refined in several papers, e.g. [43,50–54].

**Idea:** The idea is similar to that motivating the EOF-method in the sense that one is looking for a suitable low-dimensional subspace to which one can project a dynamical system. But in contrast to the EOF-method, principal interaction patterns were introduced with the aim of finding a reduction method which focuses more on the dynamics instead of the statistics of a dynamical system. Therefore, PIPs are constructed as (approximate) minimizers of a certain error functional which measures the distance of the projected solutions to the original solutions.

**Outline of the Reduction Procedure:** The aim of this method is to reduce a system of ordinary differential equations

$$\dot{x}(t) = F(t, x(t)), \quad x(0) = x^0 \tag{2.8}$$

in the state space $\mathbb{R}^d$ with $d \in \mathbb{N}$ being large, to a lower-dimensional system. Even though PIPs are constructed for systems as (2.8), they can also be used for partial differential equations of the form

$$\partial_t u = D(u) \tag{2.9}$$

with a suitable nonlinear operator $D$. But in such a case, (2.9) is approximated by (2.8) through a Galerkin approximation and then the PIPs are constructed for (2.8).

For the construction of PIPs one needs a suitable notion of an average or expectation of several quantities over a set of initial values (2.8). We will use the notation $\langle \cdot \rangle$ for this average. In the literature, this average is not always made precise and may slightly differ in different situations. The idea behind the precise definitions is however similar in most cases. In the original paper by Hasselmann [49], in which the idea of PIPs was first introduced, it is just described as expectation value. In [52] it is an ensemble average over an attractor. Similarly, an ensemble average over a finite number of such initial conditions was chosen in [35]. Since PIPs are eventually determined numerically anyway, we also work with a finite number of initial conditions in the following. Therefore, let $I = \{x^{0,1}, \ldots, x^{0,N}\}$ be a set of $N \in \mathbb{N}$ initial conditions for (2.8). Whenever $V$ is a vector space and $f : \mathbb{R}^d \to V$ a mapping, we use the notation

$$\langle f \rangle := \frac{1}{N} \sum_{j=1}^{N} f(x^{0,j}).$$

But one should keep in mind that other notions of average are also possible.

For a fixed number $k \in \mathbb{N}$ with $k < d$ and linearly independent vectors $p_1, \ldots, p_k \in \mathbb{R}^d$ one defines $P := \text{span}\{p_1, \ldots, p_k\}$. Then, one projects (2.8) orthogonally to the reduced system

$$\dot{x}_P(t) = F_P(t, x_P(t)), \quad x_P(0) = x_P^0. \tag{2.10}$$

Here, $F_P = \text{pr}_P \circ F$ where $\text{pr}_P$ denotes the orthogonal projection onto $P$ and where orthogonality may again be understood with respect to the scalar product $\langle x, My \rangle$ given by the Euclidean scalar product and the energy metric $M \in \mathbb{R}^{d \times d}$.

Now, we split

$$x(t) = \langle x(t) \rangle + \widetilde{x}(t), \quad x_P(t) = \langle x(t) \rangle + \widetilde{x}_P(t)$$

and work with the deviations from the mean state $\langle x(t) \rangle$ of the system at time $t$. We define the error functional

$$Q(P; w) := \left\langle \int_0^\infty \|\widetilde{x}_P(\tau) - \widetilde{x}(\tau)\|_M^2 w(\tau) \, d\tau \right\rangle.$$

Here, $\|v\|_M := \sqrt{\langle v, Mv \rangle}$ and $w : [0, \infty) \to [0, \infty)$ is a suitable weight function. As for the definition of the average, there is some freedom in the choice of the error functional that goes beyond different choices of $w$ or $M$. We refer the reader to [35,49,52,54], where different choices have been made. Now, $P$ is chosen as an approximate minimizer of the error functional $Q(P; w)$ and usually determined by numerical methods. Details of the minimization procedure are carried out in the appendix of [52]. The

principal interaction patterns are now the basis $p_1, \ldots, p_k$ of $P$. Obviously, they are not unique, since one can obtain a new basis by a linear transformation on $P$. Thus, one usually imposes constraints on the choice of $p_1, \ldots, p_k$ so that they can be uniquely determined, see for example [52, Section II.C].

**Example:** We consider again the CDV-model (2.7). In the section on EOF reductions we discussed the reductions carried out in [35] and how an EOF reduction could not reproduce the chaotic transitions between different flow regimes. Since PIPs were introduced with the intention in mind, that they should be able to preserve the dynamical behaviour of a system, one might ask whether they are able to do so in the case of the CDV-model. This was also investigated in [35]. The authors carried out various PIP reductions with different weights $w$ and different sets of initial conditions $I$. It was found that most models were able to reproduce the chaotic regime transitions. Moreover, the power spectra of the trajectories integrated from the reduced models were compared to the ones from the full model and it was found that they look very similar. Therefore, one could say that for the CDV-model a PIP reduction would indeed yield a more precise reduced model than an EOF reduction. However, the dependence of the reduction on the input parameters is in some cases counterintuitive. In particular, it was found that the chaotic regime transitions would not persist if the set of initial values $I$ was too large. This shows that even if a PIP reduction can produce very good results, but the results depend crucially on the input parameters and it is not always clear, what choice of parameters would be suitable.

**Comments:**

- The choice of $w$, $I$ and the averaging procedure is crucial for the outcome of the minimization problem. This can be seen as an advantage of PIPs, as it provides some flexibility. But the downside is that it has not yet been systematically investigated which choices would yield good results for a given model. The outcome can also be counterintuitive to some extend. For example, as discussed above it was observed in [35] that for the Charney–DeVore model PIPs have difficulties in preserving the dynamics if the set $I$ contains too many initial values.
- For example [52, Section II.E] it was observed that one can obtain EOFs as a limiting case of PIPs if $w = \frac{1}{\tau_{\max}} \mathbb{1}_{\tau_{\max}}$ and $\tau_{\max} \to 0$.
- The minimization of $Q(P; w)$ resembles problems encountered in optimal control theory as for example presented in [55]. The problem here is slightly different in the sense that one is looking for a subspace such that the projection error is minimized in terms of $Q$, whereas in optimal control one usually looks for a control function or parameter such that a cost functional is minimized subject to some constraints. But nonetheless, techniques from optimal control also help in numerically solving the minimization task for $Q$, see [52, Appendix B].

In the previous section we have seen how partial differential equations can by reduced to (finite-dimensional) ordinary differential equations by projecting them to finite-dimensional subspaces of the phase space. The knowledge of certain patterns or characteristic parameters may be useful for such an approach, but it is not essential. This is different in the next method we are going the describe, as it is based on the utilization of the characteristic length scales, time scales and amplitudes of the underlying system.

## 3. Characteristic scales

Characteristic length or time scales can sometimes be found in evolution equations of the form

$$\partial_t u(t, x) = F(\partial_x, u(t, x), \varepsilon), \tag{3.1}$$

where $F$ denotes a generally nonlinear spatial differential operator and $\varepsilon$ denotes a small parameter. The notion of characteristic scale is not defined in a rigorous mathematical manner but is instead based on certain parameters or observations of the system described by (3.1).

**Idea:** Even very complex dynamical systems may exhibit characteristic length or time scales at which interesting patterns emerge. Such characteristic scales can for example be induced by certain parameters in the system or sometimes they are just observed as the system evolves. If one knows the underlying equations of the system, then one can rescale them according to the characteristic scales and just consider the leading order terms which appear after the transformation. The hope is that by restricting to the leading order terms the equations become simpler to analyse while their solutions might still show the same patterns whose scales were considered as characteristic. Effectively, this idea is directly motivated by classical formal matching techniques for differential equations [56–58], where one also searches for scalings such that several terms within an equation are in a dominant balance, allowing us to discard higher-order terms based on asymptotic analysis.

**Outline of the Reduction Procedure:** We suppose that (3.1) is given in a reference time and reference length scale in which we denote the time variable by $t$ the space variable by $x$. Moreover, there are $M \in \mathbb{N}$ characteristic time and $N \in \mathbb{N}$ characteristic length scales given by the coordinate transformations $\tau_{(j)} = \varepsilon^{\alpha_{t,j}} t$, $(j = 1, \ldots, M)$, and $\xi_{(k)} = \varepsilon^{\alpha_{x,k}} x$ $(k = 1, \ldots, N)$, where the $\alpha_{t,j}$ and $\alpha_{x,j}$ are given real numbers. Let us fix $j_0 \in \{1, \ldots, M\}$ and $k_0 \in \{1, \ldots, N\}$ and suppose that we are interested in patterns that emerge on the scales given by the variables $(\tau_{(j_0)}, \xi_{(k_0)})$. Therefore, we rescale (3.1) according to this change of variables and obtain an ansatz of the form

$$\varepsilon^{\alpha_{t,j_0}} \partial_{\tau_{(j_0)}} u(\tau_{(j_0)}, \xi_{(k_0)}) = F(\varepsilon^{\alpha_{x,k_0}} \partial_{\xi_{(k_0)}}, u(\tau_{(j_0)}, \xi_{(k_0)}), \varepsilon), \tag{3.2}$$

which is crucial for characteristic scales arguments to work. Depending on the exponents $\alpha_{t,j_0}$ and $\varepsilon^{\alpha_{x,k_0}}$ one has different leading order terms, i.e., terms with the lowest power in $\varepsilon$. At this point, it can already be insightful to study the equation one obtains by omitting the higher order terms. But in practice one sometimes also has more information about the amplitudes of certain patterns so that one can write

$$u(\tau_{(j_0)}, \xi_{(k_0)}) = \sum_{l=1}^{P} \varepsilon^{\beta_l} u_l(\tau_{(j_0)}, \xi_{(k_0)})$$

for certain $P \in \mathbb{N}$, $\beta_1, \ldots, \beta_P \in \mathbb{R}$ and potentially mutually dependent functions $u_1, \ldots, u_P$. Substituting this expansion into (3.2) may yield different leading order terms and thus also different reduced equations. Mathematically, including amplitudes is directly motivated by the theory of amplitude/modulation equations [59–61] used near bifurcations. For ODEs one can hope to making the scaling arguments geometrically rigorous, even near bifurcations, via geometric desingularization [62–65].

**Example:** In [66], formal asymptotic analysis ideas regarding characteristic scales are applied to compressible flow equations for an ideal gas with constant specific-heat capacities and including gravity, rotation, and generalized source terms. In addition, a tangent plane approximation is used, i.e. one neglects the curvature of the sphere but still takes the Coriolis force into account.

More precisely, the system

$$\left(\frac{\partial}{\partial t} + v_\parallel \cdot \nabla_\parallel + w \frac{\partial}{\partial z}\right) v_\parallel + \varepsilon(2\Omega \times v)_\parallel + \frac{1}{\varepsilon^3 \rho} \nabla_\parallel p = Q_{v_\parallel},$$

$$\left(\frac{\partial}{\partial t} + v_\parallel \cdot \nabla_\parallel + w \frac{\partial}{\partial z}\right) w + \varepsilon(2\Omega \times v)_\perp + \frac{1}{\varepsilon^3 \rho} \frac{\partial p}{\partial z} = Q_w - \frac{1}{\varepsilon^3},$$

$$\left(\frac{\partial}{\partial t} + v_\parallel \cdot \nabla_\parallel + w \frac{\partial}{\partial z}\right) \rho + \rho \nabla \cdot v = 0,$$

$$\left(\frac{\partial}{\partial t} + v_\parallel \cdot \nabla_\parallel + w \frac{\partial}{\partial z}\right) \Theta = Q_\Theta$$

(3.3)

is considered. In this system, the subscript $\parallel$ indicates quantities in horizontal direction, i.e., $\nabla_\parallel = (\partial_{x_1}, \partial_{x_2})^T$, $v_\parallel$ denotes the horizontal direction of the flow velocity $v = (v_\parallel, w)$ and $(2\Omega \times v)_\parallel$ denotes the horizontal part of the Coriolis force with $\Omega$ being the Earth's rotation vector. $(2\Omega \times v)_\perp$ in turn stands for the vertical part of the Coriolis force and $w$ for the vertical part of the flow. Moreover, $p$ denotes the pressure, $\rho$ the density, $\Theta = p^{1/\gamma}/\rho$ the potential temperature (with $\gamma$ being the "dry isentropic exponent") and $Q_{[\cdot]}$ some source terms. The variable $z$ corresponds to the vertical direction, $x$ to the horizontal direction and $t$ to the time. First, the pressure $p$ is replaced by the "Exner pressure" $\pi$, i.e.

$$\pi = p^\Gamma \quad \text{with} \quad \Gamma = \frac{\gamma - 1}{\gamma}.$$

Now, time and horizontal coordinates are rescaled by $\tau = \varepsilon^{\alpha_t} x$ and $\xi = \varepsilon^{\alpha_x} x$ with certain $\alpha_t, \alpha_x \geq 0$ which depend on the scale one wants to consider. Fluctuations of the potential temperature are expected to be of order $\varepsilon$, while velocities in horizontal directions scale as $\varepsilon^{\alpha_x}$. Therefore, one may introduce the expansions

$$\Theta(\tau, \xi, z) = 1 + \varepsilon\overline{\theta}(z) + \varepsilon^{\alpha_\pi}\widetilde{\theta}(\tau, \xi, z),$$
$$\pi(\tau, \xi, z) = \overline{\pi}(z) + \varepsilon^{\alpha_\pi} \Gamma \widetilde{\pi}(\tau, \xi, z),$$
$$w(\tau, \xi, z) = \varepsilon^{\alpha_x} \widetilde{w}(\tau, \xi, z).$$

Here, $\overline{\pi}$ satisfies

$$\overline{\pi}(z) = 1 - \Gamma \int_0^z \frac{1}{\overline{\Theta}(z')} \, dz'$$

with $\overline{\Theta} = 1 + \overline{\theta}$ being the horizontally averaged mean stratification of the potential temperature. Moreover, the parameter $\alpha_\pi$ has to be chosen later to balance the pressure gradient in the horizontal momentum equation, i.e., the first equation of (3.4). Using these substitutions, the system (3.3) can be rewritten as

$$\left(\frac{\varepsilon^{\alpha_t}}{\varepsilon^{\alpha_x}} \frac{\partial}{\partial \tau} + v_\parallel \cdot \nabla_\xi + \widetilde{w} \frac{\partial}{\partial z}\right) v_\parallel + \frac{(2\Omega \times v)_\parallel}{\varepsilon^{\alpha_x - 1}}$$
$$+ \frac{\varepsilon^{\alpha_\pi}}{\varepsilon^3} \Theta \nabla_\xi \widetilde{\pi} = Q_{v_\parallel}^\varepsilon,$$

$$\left(\frac{\varepsilon^{\alpha_t}}{\varepsilon^{\alpha_x}} \frac{\partial}{\partial \tau} + v_\parallel \cdot \nabla_\xi + \widetilde{w} \frac{\partial}{\partial z}\right) \widetilde{w} + \frac{(2\Omega \times v)_\perp}{\varepsilon^{2\alpha_x - 1}}$$
$$+ \frac{\varepsilon^{\alpha_\pi}}{\varepsilon^{3+2\alpha_x}} \left(\Theta \frac{\partial \widetilde{\pi}}{\partial z} - \frac{\widetilde{\theta}}{\Theta}\right) = Q_w^\varepsilon,$$

(3.4)

$$\left(\frac{\varepsilon^{\alpha_t}}{\varepsilon^{\alpha_x}} \frac{\partial}{\partial \tau} + v_\parallel \cdot \nabla_\xi + \widetilde{w} \frac{\partial}{\partial z}\right) \widetilde{\pi} + \frac{\gamma \pi}{\varepsilon^{\alpha_\pi}}$$
$$\times \left(\nabla_\xi \cdot v_\parallel + \frac{\partial \widetilde{w}}{\partial z} + \frac{\widetilde{w}}{\gamma \Gamma \pi} \frac{d\widetilde{\pi}}{dz}\right) = \frac{\gamma \pi Q_\Theta^\varepsilon}{\Theta},$$

$$\left(\frac{\varepsilon^{\alpha_t}}{\varepsilon^{\alpha_x}} \frac{\partial}{\partial \tau} + v_\parallel \cdot \nabla_\xi + \widetilde{w} \frac{\partial}{\partial z}\right) \widetilde{\theta} + \frac{\varepsilon}{\varepsilon^{\alpha_\pi}} \widetilde{w} \frac{d\overline{\theta}}{dz} = Q_\Theta^\varepsilon.$$

For different choices of $\alpha_x, \alpha_t, \alpha_\pi$ one obtains different leading order equations. In [66], it is assumed that the order of the source terms is high enough, so that they can be neglected.

**Comments:**

- In [66, Figure 1] it is shown which length and time scales correspond to which models in the case of atmospheric flows. Roughly speaking, a formal matching of characteristic scales induces a hierarchy of models, where if possible it is good to cross-check final analysis results regarding their validity at the next lower or higher scale.

- The reduction in this section is a formal approach based on asymptotic expansions. Although these expansions can sometimes be made rigorous, it is often difficult to prove that solutions of the reduced equations relate to solutions of the original equations. Moreover, at least for atmospheric flows there is not complete consensus that there is a natural scale separation. We refer to [66, Section 1.4] where this is briefly discussed. This is quite a natural point of discussion as it is still a mathematically unsolved problem, how far one has to track small-scale energy cascades to large scales in geophysical fluid dynamics.

Using characteristic scales can be an intuitive and straightforward approach for a model reduction, provided that one has the knowledge of the existence of such characteristic scales. However, there are oftentimes no direct estimates for approximation errors, so that it can be unclear whether there is a rigorous connection between the original and the reduced models. Invariant manifold reductions, which we are going to discuss next, are among the techniques for which this connection can indeed be described rigorously.

## 4. Invariant manifolds

Invariant manifolds are a key element in the toolbox of dynamical systems theory. Yet, although different approaches have been used successfully and rigorously in several applications, climate dynamics poses additional intricate problems. This pertains crucially to the level of detail of a reduction that one aims for. If one is just interested in the behaviour near a single equilibrium point, stable/unstable and centre manifold theory applies to a wide variety of differential equations. Centre manifolds [67] are particularly useful as they isolate the slowest/neutral modes and provide an efficient dimension reduction. At the next level of detail, one can ask for larger invariant manifolds, which contain non-trivial slow dynamics. This brings us to the realm of slow manifolds [65,68,69], which are still generically local and focus on finite time dynamics. Of course, one could even be more ambitious and aim for a global reduction to a single effective low-dimensional dynamical system, where manifolds are usually inertial manifolds [61,70,71]. Here we present these three viewpoints in reverse order, going from the most global view of inertial manifold, to the intermediate and flexible compromise of slow manifolds, and finally we end describing classical local centre manifold theory near a non-hyperbolic equilibrium.

### 4.1. Inertial manifold reduction

The concept of an inertial manifold aims for a global reduction of a differential equation to a simple, hopefully low-dimensional, set of ordinary differential equations. The most classical setup for inertial manifolds occurs in the context of dissipative evolution equations of the form

$$\partial_t u + Au = f(u), \quad u(0) = u_0.$$

(4.1)

In this scenario, $A: H \supset D(A) \to H$ is usually assumed to be a self-adjoint operator with compact inverse on a Hilbert space $H$ and with domain $D(A)$. Moreover, $f: H \to H$ is a Lipschitz continuous nonlinearity. Technically, one may allow Lipschitz continuous functions $f: H \to D(A^\beta)$ for $\beta > -1$, see for example [72, Section 2.5], and this case is also important if the nonlinearity involves derivatives of $u$. But for the sake of simplicity, we refrain from giving a definition of $D(A^\beta)$ and work with the simpler case $f: H \to H$ instead. The dissipativity of (4.1) is understood in the sense that there are positive constants $C, \alpha > 0$ and a nondecreasing function $N: [0, \infty) \to [0, \infty)$ such that solutions of (4.1) satisfy the estimate

$$\|u(t)\|_H \le N(\|u_0\|_H)e^{-\alpha t} + C \quad (t \ge 0).$$

An inertial manifold $\mathcal{M}$ is a smooth manifold in $H$ such that

(i) $\mathcal{M}$ is finite-dimensional,
(ii) $\mathcal{M}$ is invariant, i.e., it holds that $T(t)\mathcal{M} \subset \mathcal{M}$, where $(T(t))_{t \ge 0}$ denotes the semiflow generated by (4.1).
(iii) $\mathcal{M}$ possesses the exponential tracking property, i.e., there are constants $C, \alpha > 0$ such that for all $u_0 \in H$ there exists a $v_0 \in \mathcal{M}$ such that the corresponding solutions $u$ and $v$ satisfy

$$\|u(t) - v(t)\|_H \le Ce^{-\alpha t}\|u_0 - v_0\|_H \quad (t \ge 0).$$

**Idea:** The invariance ensures that an inertial manifold is closed under the dynamics and that the semiflow generated by (4.1) can also be considered just on the inertial manifold instead of the whole underlying space as a phase space. Due to the exponential tracking property, one does not lose any long-term information about the dynamics of (4.1) by this restriction. Nonetheless, since an inertial manifold is finite-dimensional, the restriction of the dynamics to it is expected to yield a much simpler dynamical system than the one corresponding to (4.1).

**Outline of the Reduction Procedure:** Generally, inertial manifolds are known to exist if $A$ satisfies a certain spectral gap condition in relation to $f$, see for example [72, Theorem 2.1]. More precisely, one usually assumes that the spectrum of $A$ consists of a non-decreasing sequence of positive eigenvalues $(\lambda_k)_{k \in \mathbb{N}} \subset (0, \infty)$. Let $(e_k)_{k \in \mathbb{N}}$ be the corresponding sequence of orthonormal eigenvectors and let $L_f$ be the Lipschitz constant of $f$. If there is an $N \in \mathbb{N}$ such that

$$\lambda_{N+1} - \lambda_N > 2L_f, \tag{4.2}$$

then there is an $N$-dimensional inertial manifold $\mathcal{M}$ given as a graph of a Lipschitz-continuous mapping $h: H_+ \to H_-$, i.e.,

$$\mathcal{M} = \{u_+ + h(u_+) : u_+ \in H_+\}.$$

Here, $H_+ := \operatorname{span}\{e_1, \ldots, e_N\}$ and $H_-$ is the space generated by the orthonormal system $(e_k)_{k \in \mathbb{N}, k > N}$. Let $\operatorname{pr}_{H_+}: H \to H_+$ be the canonical projection onto $H_-$. Similar to the procedure for slow manifolds, one may then reduce (4.1) to

$$\partial_t u_+ + A u_+ = \operatorname{pr}_{H_+} f(u_+ + h(u_+)), \quad u = u_+ + h(u_+).$$

The dynamics of this system are now only finite-dimensional.

**Example:** After the concept of an inertial manifold was introduced, there were attempts to prove the existence of inertial manifolds for 2-dimensional Navier–Stokes equations in different geometries, see for example [73,74]. However, there is a gap in the proof in these early attempts so that the existence of inertial manifolds for the Navier–Stokes equations is still an open problem. We refer to [75] in which the techniques from [73,74] are discussed. Since proving the existence of an inertial manifold for Navier–Stokes equations is a difficult problem even in the 2-dimensional case, inertial manifolds for several regularizations have been considered in recent years, such

as Leray $\alpha$-models [76–78], hyperviscous Navier–Stokes equations [79,80],[71, Chapter IX.4.3] or combinations of both [81]. Let us give more details on the hyperviscous case. The underlying equation on the torus is given by

$$\partial_t u + \nu(-\Delta)^l u + (u \cdot \nabla)u + \nabla p = g, \quad (t, x) \in \mathbb{R}_+ \times \mathbb{T}^n, \\ \operatorname{div} u = 0, \quad u|_{t=0} = u_0, \tag{4.3}$$

where $u$ and $p$ are the unknown functions, $\nu$ is a given parameter, $g$ a given forcing and $u_0$ a given initial value. The parameter $l$ equals 1 for the classical Navier–Stokes equation and is taken larger than 1 in the hyperviscous case. The existence of inertial manifolds however has only been shown for an even smaller set of exponents. But let us first bring (4.3) into the abstract form (4.1). One may hide the incompressibility condition $\operatorname{div} u = 0$ as well as the pressure $p$ by projecting the equation to divergence free vector fields using the Helmholtz projection

$$P: L_2(\mathbb{T}^n; \mathbb{R}^n) \to L_2^\sigma(\mathbb{T}^n; \mathbb{R}^n),$$

where $L_2^\sigma(\mathbb{T}^n; \mathbb{R}^n)$ denotes the subspace of $L_2(\mathbb{T}^n; \mathbb{R}^n)$ consisting of the divergence free vector fields. Applying this projection to (4.3) yields

$$\partial_t u + \nu(-\Delta)^l u + P(u \cdot \nabla)u = Pg, \quad (t, x) \in \mathbb{R}_+ \times \mathbb{T}^n. \tag{4.4}$$

Here, we used that $u$ is divergence free and that $P(-\Delta)^l u = (-\Delta)^l Pu = (-\Delta)^l u$. If we now choose

$$H = L_2^\sigma(\mathbb{T}^n; \mathbb{R}^n), \quad D(A) := \{u \in H : (-\Delta)^\beta u \in H\},$$

$$A: H \supset D(A) \to H, \ u \mapsto (-\Delta)^l u,$$

$$f(u) = -P(u \cdot \nabla)u + Pg,$$

then (4.4) almost is of the form (4.1). The only additional step is that one has to carry out a cutoff procedure for $f$ to become Lipschitz continuous. After such a cutoff procedure, one can verify the existence of an inertial manifold for (4.4) if $l > \frac{3}{2}$ using a spectral gap condition. Indeed, the eigenvalues of $(-\Delta)^l$ are of the form $(|k|_2^2)^l$ $(k \in \mathbb{Z}^n)$ and therefore, every eigenvalue will be of the form $m^l$ for a natural number $m \in \mathbb{N}_0$. Moreover, each of these eigenvalues has finite multiplicity so that for every $m \in \mathbb{N}_0$ there is an $N \in \mathbb{N}$ such that $\lambda_{N+1} - \lambda_N = (m+1)^l - m^l \ge lm^{l-1}$. Using the spectral gap condition we can obtain $\lambda_{N+1}^{1/2l} + \lambda_N^{1/2l} \le 2(m+1)^{1/2}$ and so

$$\frac{\lambda_{N+1} - \lambda_N}{\lambda_{N+1}^{-\beta/2} + \lambda_N^{-\beta/2}} \gtrsim m^{l-\frac{3}{2}} \to \infty$$

as $m \to \infty$ or equivalently $N \to \infty$ if $l > 3/2$. This shows that there is an $N \in \mathbb{N}$ depending on the spatial dimension, such that there is an inertial manifold for the hyperviscous Navier–Stokes equation with $l > 3/2$ which is given as a graph over the first $N$ eigenfunction of $(-\Delta)^l$ in $L_2^\sigma(\mathbb{T}^n; \mathbb{R}^n)$. At least for $n \in \{2, 3\}$ the existence of an inertial manifold has also been verified in the critical case $l = 3/2$, see [80,81]. Therein a technique called spatial averaging was used instead of the spectral gap approach. We note that the hyperviscous Navier–Stokes example is primarily of mathematical interest. Furthermore, it turns out that in many applications it is not known yet, how to get good bounds on the dimension of the reduced system.

**Comments:**

- To the best of our knowledge, the concept of an inertial manifold was first explicitly introduced in [82]. However, it is mentioned in the introduction of [82] that related ideas have already been used before, see for example [83–85, Chapters 6,9].
- In practice, nonlinearities are oftentimes not globally Lipschitz continuous but only locally. However, the asserted dissipativity of the system allows one to use cut-off techniques to circumvent this issue without changing the global long-term dynamics.

- The main advantage of an inertial manifold reduction is that, if inertial manifolds exist, they yield a global reduction of an infinite-dimensional to a finite-dimensional system without losing any asymptotic information of the dynamics. However, the downside is that their existence can only be shown under very restrictive conditions and for many important examples such as Navier–Stokes equations, their existence is still an open problem. Therefore, the number of scenarios in which an inertial manifold reduction can be rigorously applied at the moment in practice is rather limited.

The main idea behind an inertial manifold reduction is that the dissipativity of a system should in some sense eventually lead to finite-dimensional dynamics. The next reduction method we are going to discuss uses different timescales instead of the dissipativity of the underlying equation. For so-called slow manifold reduction, on usually has a system of a fast and a slow variable and one aims to reduce the dynamics to the ones of the slow variable only. Let us now describe this idea in more detail.

### 4.2. Slow manifold reduction

Slow manifold reductions are used in systems with multiple time scales. In their standard form such systems are given by

$$\begin{aligned} \varepsilon \partial_t u^\varepsilon &= A u^\varepsilon + f(u^\varepsilon, v^\varepsilon), \\ \partial_t v^\varepsilon &= B v^\varepsilon + g(u^\varepsilon, v^\varepsilon). \end{aligned} \tag{4.5}$$

Here, $0 < \varepsilon \ll 1$ is a small parameter indicating the presence of different time scales, $A, B$ are linear operators, $f, g$ are given Lipschitz continuous functions and $u^\varepsilon, v^\varepsilon$ are the unknown functions. Suppose that $u^\varepsilon$ takes values in the Banach space $X$ and $v^\varepsilon$ takes values in the Banach space $Y$. If (4.5) is an ordinary differential equation, then the framework to reduce its dynamics is provided by classical Fenichel–Tikhonov theory [68,86], which turned out to be of great importance in multiple time scale dynamics. For partial differential equations however, the theory is much less complete. A first step as taken in [87], where suitably regular/bounded perturbations were covered; see [88, Sec. 3] for a technical discussion of the systems where the methods apply. Furthermore, in more general cases, slow manifolds have been constructed for a limited number of concrete examples over the years. The next step to towards a rigorous justification of quite general slow manifold reductions for abstract partial differential equations of the form (4.5) was taken in [88,89].

**Idea:** Under appropriate conditions, and for sufficiently small $\varepsilon$ the fast variable $u^\varepsilon$ rapidly approaches a state in which its dynamics are only determined by the slow variable. More precisely, it gets attracted by a slow manifold $S_\varepsilon$, a subset of the phase space which is given as a graph over the slow variable space and on which the dynamics of the full system evolve on the slow time scale. The fact that trajectories are attracted by the slow manifold is taken as a justification that one only has to consider the flow on the slow manifold. And since the slow manifold is given as a graph, one can reduce the system to a self-contained, potentially lower-dimensional equation which only depends on the slow variable.

**Outline of a Reduction Procedure:** The following procedure is more general than the above in that it allows $v$ to have multi-scale dynamics. However, it only works under certain conditions [88,89]. Below, we will elaborate further, but for the moment we just assume that we can carry out the following steps and that all the objects we use do exist.

One starts with a fast–slow system of the form (4.5) under the assumption that the slow system itself can have a timescale separation $\zeta > 0$. One splits the slow variable space $Y = Y_F^\zeta \oplus Y_S^\zeta$ into a $B$-invariant part $Y_F^\zeta$ which decays quickly under the semiflow

$(e^{tB})_{t \geq 0}$ generated by the linear part $B$ in the slow equation, and a $B$-invariant part $Y_S^\zeta$ on which the operators $(e^{tB})_{t \geq 0}$ are invertible and their inverses $(e^{-tB})_{t \geq 0}$ on $Y_S^\zeta$ do not grow quickly. Which rates of decay or growth are considered as 'quick' is controlled by a parameter $\zeta$. If $\zeta$ gets smaller, then $Y_F^\zeta$ gets smaller and $Y_S^\zeta$ gets larger. Now, if $\varepsilon, \zeta > 0$ are small enough and satisfy $\varepsilon < c\zeta$ for a certain $c \in (0, 1)$, then there is a mapping $h^{\varepsilon,\zeta} = (h_X^{\varepsilon,\zeta}, h_{Y_F^\zeta}^{\varepsilon,\zeta}): Y_S^\zeta \to X \times Y_F^\zeta$ with the following properties:

(i) The so-called slow manifold $S_{\varepsilon,\zeta} := \{(h_X^{\varepsilon,\zeta}(v), h_{Y_F^\zeta}^{\varepsilon,\zeta}(v) + v) : v \in Y_S^\zeta \cap D(B)\}$ is invariant under the semiflow generated by (4.5). Here, $D(B)$ denotes the domain of $B$.

(ii) The slow manifold $S_{\varepsilon,\zeta}$ attracts all solutions of (4.5) exponentially quickly.

(iii) The slow manifold is close to the critical manifold $S_0 := \{(x, y) : Ax + f(x, y) = 0\}$, where $(x, y)$ is taken such that $Ax + f(x, y)$ is well-defined.

(iv) The solutions of (4.5) on the slow manifold are close to solutions of

$$\begin{aligned} 0 &= A u^{0,\zeta} + f(u^{0,\zeta}, v^{0,\zeta}), \\ 0 &= \mathrm{pr}_{Y_F^\zeta} v^{0,\zeta}, \\ \partial_t v^{0,\zeta} &= B v^{0,\zeta} + \mathrm{pr}_{Y_S^\zeta} g(u^{0,\zeta}, v^{0,\zeta}). \end{aligned} \tag{4.6}$$

Here, $\mathrm{pr}_{Y_F^\zeta}: Y \to Y_F^\zeta$ and $\mathrm{pr}_{Y_S^\zeta}: Y \to Y_S^\zeta$ denote the canonical projections according to the splitting $Y = Y_F^\zeta \oplus Y_S^\zeta$

The properties (i) and (ii) can be taken as a justification that if $\varepsilon$ and $\zeta$ are sufficiently small then one may reduce (4.5) to

$$\begin{aligned} u^{\varepsilon,\zeta}(t) &= h_X^{\varepsilon,\zeta}(v_S^{\varepsilon,\zeta}(t)), \\ v_F^{\varepsilon,\zeta}(t) &= h_{Y_F^\zeta}^{\varepsilon,\zeta}(v_S^{\varepsilon,\zeta}(t)), \\ \partial_t v_S^{\varepsilon,\zeta}(t) &= B v_S^{\varepsilon,\zeta}(t) + \mathrm{pr}_{Y_S^\zeta} g(h_X^{\varepsilon,\zeta}(v_S^{\varepsilon,\zeta}(t)), h_{Y_F^\zeta}^{\varepsilon,\zeta}(v_S^{\varepsilon,\zeta}(t)) + v_S^{\varepsilon,\zeta}(t)). \end{aligned}$$

The dynamical part of this system is only determined by the last equation and depends only on the "super-slow" part of the slow variable. It may even be finite-dimensional in many cases, even if both $X$ and $Y$ are infinite-dimensional. The other components are then uniquely determined by $v_S^{\varepsilon,\zeta}(t)$. Moreover, the properties (ii) and (iv) give us information about the location of the slow manifold and the flow on it, even if it is difficult to construct it explicitly in certain scenarios.

The proximity of the slow manifold to the critical manifold also motivates a simpler reduction of (4.5), namely the one to the slow subsystem given by

$$\begin{aligned} 0 &= A u^0 + f(u^0, v^0), \\ \partial_t v^0 &= B u^0 + g(u^0, v^0). \end{aligned} \tag{4.7}$$

This reduction is less accurate in the sense that the flow generated by (4.7) is not contained in the original one generated by (4.5). However, the slow subsystem is still a reasonable approximation of (4.5), since solutions (4.5) and (4.7) with the same initial condition $v_0$ for the slow variable equation approach each other as $\varepsilon > 0$ on any compact time interval not containing 0. We refer to [88, Section 4.5] for a more precise formulation of this statement. In addition, this reduction has more advantages. Most importantly, it does not need the existence of the above splitting $Y = Y_F^\zeta \oplus Y_S^\zeta$ of the slow variable space, which is a rather restrictive assumption. The reduction to the slow subsystem is therefore applicable in many more situations, even if slow manifolds do not exist.

**Example:** A well-known example of a slow manifold reduction (where we do not need to consider a splitting of $v$) is the one for

the different variants of the Stommel model. Here, we take the Stommel–Cessi model given by the fast–slow system

$$\varepsilon\dot{u} = -(u-1) - \varepsilon u[1 + 7.5(u-v)^2],$$
$$\dot{v} = \mu - v[1 + 7.5(u-v)^2], \tag{4.8}$$

see [90,91] and [92, (6.2.27)]. Here, $u$ denotes the temperature difference and $v$ the salinity difference between a higher and a lower latitude box in the North Atlantic. The parameter $\mu$ is proportional to the freshwater flux and can mathematically be seen as a bifurcation parameter for the dynamics of the thermohaline circulation in the North Atlantic. By Fenichel theory, there is a slow manifold which is a perturbation of order $\varepsilon$ of the critical manifold given by the constant function $h^0(v) \equiv 1$. One may thus approximate (4.8) by

$$\dot{v} = \mu - v[1 + 7.5(1-v)^2]. \tag{4.9}$$

After this reduction, the bifurcation analysis of this equation boils down to studying the roots of the third order polynomial on the right-hand side. One can verify that it undergoes two saddle–node bifurcations at

$$\mu_\pm = \frac{1}{9}\left(11 \pm 3 \cdot \sqrt{\frac{3}{5}}\right).$$

If $\mu < \mu_-$ or $\mu > \mu_+$ there is only one stable equilibrium which corresponds to a low salinity difference or high salinity difference, respectively. If $\mu \in (\mu_-, \mu_+)$, then there are two stable states the system may be in.

A similar reduction can be carried out in certain situations if we add a spatial component. For example, the system

$$\varepsilon\partial_t u(t,x) = \partial_{xx}u(t,x) - (u(t,x)-1) - \varepsilon u(t,x)$$
$$\times [1 + 7.5(u(t,x)-v(t,x))^2],$$
$$\partial_t v(t,x) = \partial_{xx}v(t,x)\mu - v(t,x)[1 + 7.5(u(t,x)-v(t,x))^2], \tag{4.10}$$

on the 1-dimensional torus, i.e., with $x \in \mathbb{T}$ and periodic boundary conditions, can be approximated by the equation

$$\partial_t v(t,x) = \partial_{xx}v(t,x)\mu - v(t,x)[1 + 7.5(1-v(t,x))^2], \quad u(t,x) \equiv 1. \tag{4.11}$$

One may even reduce this equation further to a system of ordinary differential equations by truncating to a finite set of Fourier modes. We refer to [88, Section 6.1] for more details on the reduction.

**Comments:**

- Let us now elaborate a bit more on the necessary assumptions: If (4.5) is an ordinary differential equation, then it is comparably easy to put into an abstract framework. The functions $f$ and $g$ are naturally given by the application and act on $X \times Y = \mathbb{R}^m \times \mathbb{R}^n$ as a phase space, i.e., $u^\varepsilon$ takes values in $X = \mathbb{R}^m$ and $v^\varepsilon$ takes values in $Y = \mathbb{R}^n$ for some $n, m \in \mathbb{N}$. If the nonlinearities are smooth enough, then the only condition one needs for the reduction procedure is that the linearization of $u \mapsto Au + f(u,v)$ does not have eigenvalues on the imaginary axis for every $v \in \mathbb{R}^n$. The splitting of the slow variable space is trivial for ordinary differential equations: One can just take $Y_F^\zeta = \{0\}$ and $Y_S^\zeta = \mathbb{R}^n$. The slow manifolds $S_{\varepsilon,\zeta}$ will then be independent of $\zeta$ so that they only depend on $\varepsilon$ as in classical Fenichel theory.
- However, if (4.5) is a partial differential equation, it requires more effort to put it into an abstract form which allows one to treat it using the setting of [88]. The main reason is that in infinite dimensions, one has many more possible

choices of phase spaces and corresponding topologies. The assumptions for the theory in [88] might not be satisfied in some natural looking choice of phase spaces while they may be satisfied in other phase spaces. Moreover, the functions $f$ and $g$ will usually not act on the whole phase space but only on a certain subset. If for example $X = C^\alpha(\mathbb{R})$ is the space of $\alpha$-Hölder continuous functions on $\mathbb{R}$ with some $\alpha \in (0,1)$ and $f(u^\varepsilon, v^\varepsilon) = \Delta u^\varepsilon$, then one formally needs that $u^\varepsilon$ is twice differentiable and that its derivatives are $\alpha$-Hölder continuous. One is therefore forced to work with different spaces and different topologies. Aside from the technicalities concerning different topologies and spaces, there are two important conditions one has to verify. Firstly, the semi-group generated by $A$ should have a negative growth bound and the Lipschitz constant of $f$ should be small enough so that solution of the fast equation would decay without the forcing from the slow variable. Secondly, one has to find a splitting $Y = Y_F^\zeta \oplus Y_S^\zeta$ as mentioned above. The space $Y_F^\zeta$ contains the parts which decay quickly under the $C_0$–semigroup $(e^{tB})_{t\geq0}$ generated by $B$. On $Y_S^\zeta$ the operators $(e^{tB})_{t\geq0}$ are invertible with inverses $(e^{-tB})_{t\geq0}$ these inverses do not grow too quickly as $t \to \infty$. Moreover, the rates of decay and growth of $(e^{tB})_{t\geq0}$ and $(e^{-tB})_{t\geq0}$, respectively, have to be separated well enough so that it persists even if the nonlinearity $g$ is taken into account. The separation should even become stronger as $\zeta$ gets smaller for convergence results to hold true. Often this means that $B$ has to have spectral gaps which get arbitrarily large as one moves towards the left in the complex plane. Unfortunately, this assumption is very restrictive. For example, if $B = \Delta$ is the Laplacian on the $n$-dimensional torus $\mathbb{T}^n$, then the distance between the eigenvalues does not get arbitrarily large unless $n \in \{1, 2\}$.

- Nonlinearities in many applications are actually not Lipschitz continuous, but only locally Lipschitz-continuous. Nonetheless, they can be made Lipschitz continuous by a standard cut-off procedure. In such cases however, the slow manifolds also only have a local meaning instead of a global one.
- A slow manifold reduction allows one to remove the fast variable $u^\varepsilon$ from the dynamics. Moreover, this method does not only reduce the dimension of the dynamics, but also provides a nice geometric interpretation of the reduction, since the reduced system is contained in an attracting invariant manifold of the original system. However, the assumptions of a slow manifold reduction are not directly satisfied at bifurcation points of the fast variable. Thus, many interesting pattern forming problems are not accessible through a slow manifold reduction alone and instead, one has to use additional desingularization techniques such as the blow-up method, see for example [93–95].
- Difficulties also arise if the equation for the slow variable is given by a partial differential equations. In this case, the conditions under which a slow manifold reduction can be rigorously justified are as restrictive as the ones for an inertial manifold reduction. However, one may still approximate trajectories of the fast–slow system by the ones of the slow subsystem under less restrictive assumptions also in the infinite-dimensional case, see [88].
- Additional difficulties can arise if the domains become extremely large so that their scaling decreases the spectral gap. This is a core theme in the area of amplitude/modulation equations [96]. Recently, efforts have begun to extend geometric singular perturbation theory also to PDEs formulated on unbounded domains [97].

While inertial and slow manifold reductions aim to describe the dynamics of a system in a larger part of the phase space, we are now going to discuss a reduction method which only aims to reduce the dynamics around a non-hyperbolic equilibrium: the so-called centre manifold reduction.

### 4.3. Centre manifold reduction

Another important classical reduction method using invariant manifolds is the centre manifold reduction. It is used to simplify the dynamics of a system of the form

$$\dot{x} = f(x) \tag{4.12}$$

locally around a non-hyperbolic fixed point $x_0 \in X$ with $X$ being a Banach space. A common extension is to consider parametrized bifurcation problems of the form

$$\dot{x} = f(x, \lambda) \tag{4.13}$$

with $\lambda$ a finite dimensional parameter set [98].

**Idea:** A centre manifold consists of those trajectories around a fixed point, which are neither attracted nor repelled by it at an exponential rate. Roughly speaking, the long-term behaviour of the former trajectories shrinks down to the equilibrium point itself, while the latter ones are pushed away from the fixed point so that they are not important for the local dynamics. The trajectories on the centre manifold in turn persist and stay close to the fixed point. It is therefore sufficient to restrict the system to the centre manifold in order to describe its effective dynamics around the fixed point.

**Outline of the Reduction Procedure:** Assume that $f$ is smooth enough. As already mentioned above, centre manifolds are constructed around a non-hyperbolic steady state, i.e., around an $x_0 \in X$ such that $f(x_0) = 0$ and that the spectrum of the linearization $Df(x_0)$ splits into three parts

$$\sigma(Df(x_0)) = \sigma^s \cup \sigma^c \cup \sigma^s$$

with $\sigma_c \neq \emptyset$, where

$$\sigma^s := \{\lambda \in \sigma(Df(x_0)) : \operatorname{Re} \lambda < 0\},$$
$$\sigma^c := \{\lambda \in \sigma(Df(x_0)) : \operatorname{Re} \lambda = 0\}, \quad \text{and}$$
$$\sigma^u := \{\lambda \in \sigma(Df(x_0)) : \operatorname{Re} \lambda > 0\}$$

denote the stable, neutral (centre), and unstable part of the spectrum, respectively. In infinite dimensions one has to impose additional conditions which partly depend on the desired properties of the centre manifold. One possibility, which is carried out in [99], is to assume that:

(i) The linearization $Df(x_0)$ generates a $C_0$-semigroup $(T(t))_{t \geq 0} \subset \mathcal{B}(X)$.

(ii) The three parts of the spectrum $\sigma^s, \sigma^c$ and $\sigma^u$ are closed and open, and there are corresponding spectral projections $\operatorname{pr}_s : X \to X^s$, $\operatorname{pr}_c : X \to X^c$ and $\operatorname{pr}_u : X \to X^s$, respectively.

(iii) $\dim X^c < \infty$, $\dim X^u < \infty$.

(iv) There are constants $M, c > 0$ such that $\|T(t)|_{X^s}\|_{\mathcal{B}(X^s)} \leq Me^{-ct}$ holds for all $t \geq 0$.

Variations on these assumptions are for example discussed in [100]. A more general approach is to assume there is a spectral gap parameter $p > 1$ and an $\alpha > 0$ such that

$$\sigma^s := \{\lambda \in \sigma(Df(x_0)) : \operatorname{Re} \lambda < -p\alpha\},$$
$$\sigma^c := \{\lambda \in \sigma(Df(x_0)) : |\operatorname{Re} \lambda| < \alpha\}, \quad \text{and}$$
$$\sigma^u := \{\lambda \in \sigma(Df(x_0)) : \operatorname{Re} \lambda > p\alpha\}$$

and this is outlined for example in papers leading up to [101]. Note that one can generalize the setting so that it is not necessary

to assume that the centre space $X^c$ is finite-dimensional, as the conservative case in [99] shows.

One locally constructs a centre manifold $\mathcal{C}$ as the graph of a mapping $h : \{x_0\} + U \to X^s \oplus X^u$ over the centre subspace, i.e.,

$$\mathcal{C} = \{h(x_0 + x) + x_0 + x^c : x \in U\},$$

where $U \subset X^c$ is a certain neighbourhood of 0 in $X^c$. This is usually done by one of the standard approaches such as a Lyapunov–Perron or a graph transform argument. Similar to the slow and inertial manifold reductions, one can now reduce (4.12) to a dynamical system in $U \subset X^c$, more precisely to

$$\dot{x}^c = \operatorname{pr}_c f(x^c + h(x^c)), \quad x = x^c + h(x^c)$$

with initial conditions in $\{x_0\} + U$.

**Example:** A slow manifold reduction can be seen as a specific case of a centre manifold reduction. To this end, we rescale time by $\tau = \varepsilon t$ so that (4.5) turns into

$$
\begin{aligned}
\partial_\tau u^\varepsilon &= Au^\varepsilon + f(u^\varepsilon, v^\varepsilon), \\
\partial_\tau v^\varepsilon &= \varepsilon B v^\varepsilon + \varepsilon g(u^\varepsilon, v^\varepsilon).
\end{aligned}
\tag{4.14}
$$

Now one adds the new equation

$$\partial_\tau \varepsilon = 0$$

to the system. This yields the new system

$$
\begin{aligned}
\partial_\tau u^\varepsilon &= Au^\varepsilon + f(u^\varepsilon, v^\varepsilon), \\
\partial_\tau v^\varepsilon &= \varepsilon B v^\varepsilon + \varepsilon g(u^\varepsilon, v^\varepsilon), \\
\partial_\tau \varepsilon &= 0
\end{aligned}
\tag{4.15}
$$

Setting $\varepsilon = 0$ leads to the fast subsystem

$$
\begin{aligned}
\partial_\tau u^\varepsilon &= Au^\varepsilon + f(u^\varepsilon, v^\varepsilon), \\
\partial_\tau v^\varepsilon &= 0.
\end{aligned}
$$

Let now $(u, v) \in S_0$ be a point on the critical manifold $S_0$ of (4.14). Then there is an equilibrium of (4.15) of the form $(u, v, 0)$ where $Au_0 + f(u_0, v_0) = 0$. The linearization of this at $\varepsilon = 0$ has block-matrix of the form

$$
L := \begin{pmatrix} A + f_u^0 & f_v^0 & 0 \\ 0 & 0 & B_v^0 + g \\ 0 & 0 & 0 \end{pmatrix}
$$

and so if $A + f_u^0$ has no spectrum on the imaginary axis, then $(u, v, 0)$ is a non-hyperbolic equilibrium point of (4.15) with centre subspace $\{0\} \times Y \times \mathbb{R}$. Hence one can use centre manifold theory to obtain a neighbourhood $U \subset Y \times \mathbb{R}$ of $(v, 0)$ and a centre manifold $S$ of (4.14) containing $(u, v, 0)$ which is given as a graph over $U$. For a fixed $\varepsilon > 0$, the set $S_\varepsilon := S \cap \{(u, v, \varepsilon) \in X \times Y \times \mathbb{R}\}$ now defines a slow manifold. Therefore, one could again take the Stommel model (4.8) as an example for a centre manifold reduction, where the slow manifold is the perturbation of the critical manifold that consists of steady states for the fast subsystem.

**Comments:**

- One of the advantages of centre manifold theory is that is it well-developed and widely used, also in the infinite-dimensional or stochastic setting. We refer to [67,100,102–105].

- Compared to an inertial manifold reduction, a centre manifold reduction usually has the drawback that it is not a global but only a local reduction method. Hence, the reduction is only valid in a certain part of the phase space.

The main benefits of invariant manifold reductions are that they can be rigorously described with explicit estimates for the approximation error, and that they give a nice geometric interpretation of the reduction. However, the reduction is not always

rigorously provable, especially in systems with strong oscillations. In such cases, one can try to use other classical reduction techniques called averaging and homogenization, which we are going to discuss in the next section.

## 5. Averaging and homogenization

Averaging and Homogenization are important and similar methods of reducing systems with strong oscillations to averaged/homogenized systems without these oscillations. Both methods are particularly useful for the derivation of stochastic climate models, as they can in certain situations characterize in which sense solutions of (1.2) are approximated by solutions of an equation with random forcing. In other words, the intuitive idea that small-scale nonlinear self-interactions can effectively be replaced by a random forcing can be made rigorous. There are many works treating the mathematical ideas behind this theory, see for example [106–110]. Other works such as [16,111–113] explain them with their importance for climate models in mind. Explicit applications to climate models are mainly given by Majda and co-authors, see for example [9,15,114] and references therein.

We are not aware of a coherent abstract framework that includes both, or a classification of the problems which can be treated by the methods as well as a general reduction procedure which applies to all relevant examples. In fact, the literature is not even consistent about the classification of concrete examples: the reduction of a parabolic partial differential equation with small diffusion and strongly varying coefficients for the advection terms such as the one in [110, Section 14, Section 21] is referred to as averaging by some authors and as homogenization by others. Sometimes, both names seem even to be used interchangeably, see for example [115,116]. The reason is that there are different conceptions of what averaging and homogenization should mean.

One possible definition is given in [110, Section 1.3]. Therein, the authors write that averaging and homogenization can be applied to perturbations of equations of the form

$$L^\varepsilon u^\varepsilon = f$$

or

$$\partial_t u^\varepsilon = L^\varepsilon u^\varepsilon$$

where the linear operator $L^\varepsilon$ has the form

$$L^\varepsilon = \varepsilon^{-1} L_0 + L_1 \tag{5.1}$$

or

$$L^\varepsilon = \varepsilon^{-2} L_0 + \varepsilon^{-1} L_1 + L_2. \tag{5.2}$$

Usually, it is assumed that $L_0$ has a non-trivial null space and it is argued that the interesting behaviour takes place on this null space. The techniques treating problems with (5.1) are referred to as averaging, while those treating problems with (5.2) are called homogenization.

Another possibility to distinguish averaging and homogenization is to distinguish between temporal and spatial oscillations. In this case, averaging is usually referred to techniques which simplify temporal effects, while homogenization is used to simplify spatial effects. This distinction was for example made in [106], even though it was not explicitly stated as a definition. One should note that this is not consistent with the definition above from [110, Section 1.3]. For this work, it is not our aim to decide what the best characterizations of averaging and homogenization should be. Instead, we just formulate the idea for both methods at once and afterwards, we describe the two reduction methods in scenarios in which the classification is widely accepted.

**Idea:** In systems with different spatial or temporal scales it is not essential to study the details of the behaviour of the system on finer scales. Instead, only the average or homogeneous part of the high-resolution/fast effects is important for the macroscopic behaviour. In order to make this precise, one has to derive an averaged or homogenized version of the system and to show that solutions of the original and the averaged or homogenized system are close to each other in a suitable sense.

### 5.1. Averaging

One of the standard situations in which averaging is applied is similar to the one in which a slow manifold reduction is used, but with different conditions. Consider the fast–slow system of stochastic differential equations

$$\varepsilon du^\varepsilon = f(u^\varepsilon, v^\varepsilon) dt + \sqrt{\varepsilon} \sigma_1(u^\varepsilon, v^\varepsilon) dW_1(t),$$
$$dv^\varepsilon = g(u^\varepsilon, v^\varepsilon) dt + \sigma_2(u^\varepsilon, v^\varepsilon) dW_2(t), \tag{5.3}$$
$$u^\varepsilon(0) = u_0, \quad v^\varepsilon(0) = v_0.$$

Here, $W_1$ is a $k$-dimensional and $W_2$ an $l$-dimensional Brownian motion which are independent and the nonlinearities $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma_1 : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{m \times k}$ and $\sigma_2 : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{n \times l}$ are assumed to be Lipschitz continuous. The aim is to reduce (5.3) to an equation of the form

$$dv(t) = \overline{g}(v(t)) dt + \overline{\sigma}(v(t)) dW_3(t), \quad v(0) = v_0, \tag{5.4}$$

where $\overline{g}$ and $\overline{\sigma}$ are certain averaged versions of $g$ and $\sigma$, and $W_3$ is a Brownian motion to be determined by the reduction. The solution $v$ of (5.4) is supposed to approximate $v^\varepsilon$. Formally, the equations look similar to the ones which are suitable for a slow manifold reduction. The main difference concerning the assumptions compared to the setting for the slow manifold reduction is not only the stochastic forcing we allow. Also in the deterministic setting the assumptions are less restrictive: While slow manifold reductions are carried out for systems in which the fast subsystem is normally hyperbolic in a suitable sense, averaging treats the case in which the fast variable is assumed to be sufficiently mixing.

**Outline of the Reduction Procedure:** One starts with the fast–slow system (5.3). There are many slight variants for the main assumption and the precise approximation results, see for example [117, Chapter II.3], [110, Chapters 11,12, 17 and 18] or [106, Chapter 5]. Here, we briefly summarize the approach from [117, Chapter II.3]. Apart from smoothness and growth assumptions on the nonlinearities $f, g, \sigma_1$ and $\sigma_2$ as well as boundedness assumptions on the fast variable $u^\varepsilon$ (we refer to [117, Chapter II.3, Theorem 12] for the precise conditions), it is crucial that (5.3) satisfies the following ergodicity condition: For each fixed $v_0 \in \mathbb{R}^n$ we consider the equation

$$du(t) = f(u(t), v_0) dt + \sigma_1(u(t), v_0) dW_1(t), \quad u(0) = u_0.$$

We assume that for all $v_0 \in \mathbb{R}^n$ this equation has a solution $u_{u_0, v_0}$ and an invariant ergodic distribution $\mu_{v_0}$ such that for all $r > 0$ and all $\varphi \in C(\mathbb{R}^m)$ the uniform ergodicity condition

$$\lim_{T \to \infty} \sup_{|u_0|, |v_0| \le r} \mathbb{E} \left| \frac{1}{T} \int_0^T \varphi(u_{u_0, v_0}(t)) dt - \int \varphi(z) d\mu_{v_0}(z) \right| = 0$$

is satisfied. In this case, one may define

$$\overline{g}(v) := \int_{\mathbb{R}^n} g(u, v) d\mu_v(u)$$

and $\overline{\sigma}(v) \in \mathbb{R}^{n \times n}$ such that

$$\overline{\sigma}(v)^2 = \int_{\mathbb{R}^m} \sigma_2(u, v) \sigma_2(u, v)^\top d\mu_v(u).$$

Moreover, the process $v^\varepsilon$ from (5.3) converges in distribution as $\varepsilon \to 0$ to a process $v$ which is defined as the solution of

$$dv(t) = \overline{g}(v(t)) \, dt + \overline{\sigma}(v(t)) \, dW_1(t), \quad v(0) = v_0.$$

In cases where the averaged system (5.4) is not structurally stable there will be higher order corrections; see for example [118].

**Example:** Similar techniques have already been applied to climate models. Let us continue with the example from Section 2.1 that we took from [9] where the equations of a barotropic flow on a beta plane with topography and mean flow given by

$$0 = \frac{\partial q}{\partial t} + \nabla^\perp \psi \cdot \nabla q + U \frac{\partial q}{\partial x} + \beta \frac{\partial \psi}{\partial x},$$
$$q = \Delta \psi + h, \tag{5.5}$$
$$\frac{dU}{dt} = \oint h \frac{\partial \psi}{\partial x}.$$

have been reduced in several steps. Here, $q(x, y, t)$ denotes the small-scale potential vorticity, $U(t)$ the mean flow, $\psi(x, y, t)$ the small-scale stream function, $h(x, y)$ models the topography and $\beta$ the variation of the Coriolis parameter. A couple of reduction procedures that resemble the averaging method presented here are carried out in [9]. Let us briefly look at one of the examples, namely the one considered in [9, Section 7.4]. After a truncation to finitely many Fourier modes as in Section 2.1 the mean flow $U$ is declared as the only climate variable and all other variables are treated as unresolved variables. The nonlinear self-interactions of the unresolved variables are replaced by a stochastic forcing with drift and since one is looking for real-valued solutions, one may impose $w_k = w_{-k}^*$, $(k \in \mathbb{Z}^2)$, for any of the Fourier coefficients coming from (5.5). This way, the equations

$$dU = \frac{2}{\varepsilon} \sum_{k=(k_x, k_y) \in \overline{\sigma}} \mathrm{Im}\left(\frac{k_x \hat{h}_k^*}{|k|} w_k\right) dt,$$

$$dw_k = \frac{ik_x \hat{h}_k}{\varepsilon |k|} U \, dt - \frac{i}{\varepsilon}\left(k_x U - \frac{k_x \beta}{|k|^2}\right) e_k \, dt$$
$$- \frac{1}{\varepsilon^2} \gamma_k(w_k - \overline{w}_k) dt + \frac{\sigma_k}{\varepsilon} dW_k(t) \quad (k = (k_x, k_y) \in \overline{\sigma})$$
$$\tag{5.6}$$

are derived. Here, $\hat{h}_k^*$ denotes the complex conjugate of the $k$th Fourier coefficients of $h$. $\gamma_k, \sigma_k, \overline{w}_k, \varepsilon$ are real parameters coming from the replacement of the nonlinear self-interactions by the stochastic forcing. The parameter $\gamma_k$ measures the strength of the drift to the mean $\overline{w}_k$, $\sigma_k$ the strength of the noise and $\varepsilon$ denotes the scale separation parameter. Moreover, the $W_k(t)$ are independent Brownian motions and $\overline{\sigma}$ is a subset of $\{k \in \mathbb{Z}^2 : |k|^2 \leq N\}$ $(N \in \mathbb{N})$ that contains exactly one of the terms $k$ and $-k$ $(k \in \sigma)$. Except for the $\mathcal{O}(\varepsilon^{-1})$-terms in the equations for the unresolved variables $w_k$, (5.6) already has the form (5.7). And indeed, as was shown in [9, Theorem 7.9] this system can be reduced to

$$dU = -\gamma_u(U - \overline{U}) \, dt + \sigma_u \, dW(t),$$

where

$$\gamma_u = 2\sum_{k \in \overline{\sigma}} \frac{k_x^2 |\hat{h}_k|^2 (\hat{h}_k - |k|^2 \overline{w}_k)}{|k|^2 \gamma_k^2 \hat{h}_k}, \quad \overline{U} = -\frac{2\beta}{\gamma_u} \sum_{k \in \overline{\sigma}} \frac{k_x^2 \overline{w}_k |\hat{h}_k|^2}{|k|^2 \gamma_k \hat{h}_k},$$

$$\sigma_u = 2\left(\sum_{k \in \overline{\sigma}} \frac{k_x^2 \sigma_k^2 |\hat{h}_k|^2}{\gamma_k^2}\right)^{1/2}.$$

**Comments:**

- Averaging techniques have been further developed in the past few decades and can also be applied to certain classes of

stochastic partial differential equations. We refer the reader to [106, Section 5.2], where stochastic partial differential equations of the form

$$\partial_t u^\varepsilon = \frac{1}{\varepsilon} \Delta u^\varepsilon + \frac{1}{\varepsilon} f(u^\varepsilon, v^\varepsilon) + \frac{\sigma_1}{\sqrt{\varepsilon}} \partial_t W_1,$$
$$\partial_t v^\varepsilon = \Delta v^\varepsilon + g(u^\varepsilon, v^\varepsilon) + \sigma_2 \partial_t W_2$$

are studied. Here $f, g$ satisfy certain Lipschitz conditions and $W_1, W_2$ are two independent trace class Brownian motions in $L_2(\mathcal{O})$. Similar problems have also been treated in [119,120].

- Eq. (5.3) can also be seen as a special case of

$$\frac{du^\varepsilon}{dt} = \frac{1}{\varepsilon^2} f(u^\varepsilon, v^\varepsilon) + \frac{1}{\varepsilon} \sigma_1(u^\varepsilon, v^\varepsilon) \frac{dW_1(t)}{dt},$$
$$\frac{dv^\varepsilon}{dt} = \frac{1}{\varepsilon} g_1(u^\varepsilon, v^\varepsilon) + g_2(u^\varepsilon, v^\varepsilon) + \sigma_2(u^\varepsilon, v^\varepsilon) \frac{dW_2(t)}{dt}, \tag{5.7}$$

with $g_1 = 0$. Solutions of this can be approximated by a reduced system $v^\varepsilon \to v$ in many cases[2] — for example [110, Chapter 11] outline a method that works if the $\sigma_1 \sigma_1^\top$ is uniformly positive definite, if $\mu_v$ has a density and if the centring condition $\int g_1(u, v) \, d\mu_v(u) = 0$ holds for all $v$. In such a case, (5.7) can also be reduced to an equation of the form

$$dv(t) = \overline{g}(v(t)) \, dt + \overline{\sigma}(v(t)) \, dW_1(t),$$

but this time with

$$\overline{g}(v) = \int g_2(u, v) + (\nabla_v \Phi(u, v)) g_1(u, v) \, d\mu_v(u)$$

and $\overline{\sigma}(v)$ given by the relation

$$\overline{\sigma}(v)\overline{\sigma}(v)^T = \int \left[\sigma_2(u, v)\sigma_2(u, v)^T + g_1(u, v) \otimes \Phi(u, v) + [g_1(u, v) \otimes \Phi(u, v)]^T\right] d\mu_v(u).$$

Here, $\Phi$ is defined as the solution of

$$-\langle f, \nabla \rangle \Phi - \frac{1}{2} \mathrm{tr}(\sigma_1^T \sigma_1 D_y^2)\Phi = f_1, \quad \int \Phi(u, v) \, d\mu_v(u) = 0$$

with $D_y^2$ denoting the Hessian matrix and tr denoting the trace of a matrix. Moreover, for two vectors $x, y \in \mathbb{R}^l$ the tensor product $x \otimes y$ is defined to be the matrix $x \otimes y \in \mathbb{R}^{l \times l}$ such that $(x \otimes y)z = \langle y, z \rangle x$ for all $z \in \mathbb{R}^l$. In [110, Chapter 18] the reduction is studied more rigorously, but in a more specific setting. Note however, that the reduction of (5.7) is called homogenization instead of averaging in [110].

- Related methods, which are often referred to as fast wave averaging, have also been applied the geophysical flows. These methods are based on techniques that were developed in [121,122] for singular limits of quasilinear hyperbolic systems. For example, the systems that are treated in [122] are of the form

$$A_0(\varepsilon U^\varepsilon) \partial_t U^\varepsilon + \sum_{i=1}^{N} A_i(U^\varepsilon, \varepsilon) \partial_{x_i} U^\varepsilon +$$
$$\sum_{j=1}^{M} \left(\frac{1}{\varepsilon} K_j + D_j(U^\varepsilon, \varepsilon)\right) \partial_{y_j} U^\varepsilon = F(U^\varepsilon, \varepsilon), \tag{5.8}$$
$$U^\varepsilon(0, x, y) = U_0(x, y)$$

where $N, M \in \mathbb{N}$, $A_0$ takes values in the set of positive definite matrices and $A_i, K_j, D_j$ take values in the set of

---

2 We discuss convergence at the end of Section 5.2.

symmetric matrices. Moreover, it is assumed that $A_i$, $K_j$, $D_j$ are continuous in $\varepsilon$ uniformly over bounded $U$ and that they are $C^s$ in $U$ uniformly in $\varepsilon$ for $s - 1 \geq s_0 := \frac{N+M}{2}$. The initial value $U_0$ is assumed to be an element of $H^s(\mathbb{R}^N \times \mathbb{T}^M)$. For the reduced equation, we introduce the $C_0$-semigroup $(S(\tau))_{\tau \geq 0}$ generated by the system

$$\partial_\tau v + \sum_{j=1}^{M} K_j \partial_{y_j} v = 0.$$

Moreover, we define the inner product

$$\langle u, v \rangle_{L_2([0,\infty), L_2(\mathbb{R}^N \times \mathbb{T}^M))} :=$$
$$\lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau \int_{\mathbb{R}^N \times \mathbb{T}^M} u(s, x, y) v(s, x, y) \, d(x, y) \, ds$$

$L_\infty([0, \infty), L_2(\mathbb{R}^N \times \mathbb{T}^M))$ as well as the orthogonal projection $P$ with respect to this inner product to the closed subspace generated by expressions of the form $S(\tau)g(x, y)$ for $g \in H^{s-1}(\mathbb{R}^N \times \mathbb{T}^M)$. With this notation at hand, we may define the reduced equation given by

$$\partial_t V^0 + P \Bigg( [\partial_U A^0(0) V^0] \partial_\tau V^0 + \sum_{i=1}^{N} A_i(V^0, 0) \partial_{x_i} V^0$$
$$+ \sum_{j=1}^{M} D_j(V^0, 0) \partial_{y_j} V^0 - F(V^0, 0) \Bigg) = 0, \tag{5.9}$$

$$V^0(0, x, \tau, y) = S(\tau) U_0(x, y).$$

This is now a reduced equation in the following sense (see [122, Theorem 2.3]): Let $V^0$ be the solution of (5.9) and suppose that there is a $T > 0$ such that the mapping

$$\widetilde{V}^0 : t \mapsto V^0(t, \cdot_x, t/\varepsilon, \cdot_y)$$

is an element of $L_\infty([0, T]; H^s(\mathbb{R}^N \times \mathbb{T}^M))$ for small $\varepsilon > 0$. Then, there are $\varepsilon_0$, $C > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ and all $t_0 \in [0, T]$ the solution $U^\varepsilon$ of (5.8) exists and it holds that

$$\|U^\varepsilon - \widetilde{V}^0\|_{C([0,t_0], H^{s-1}(\mathbb{R}^n \times \mathbb{T}^M))} \leq C t_0.$$

This result can be improved under further assumptions: The crucial assumption is that the operator that maps $(\tau, x, y) \to S(\tau) U_0(x, y)$ to

$$\Bigg( \partial_\tau + \sum_{j=1}^{M} K_j \partial_{y_j} \Bigg)^{-1} (1 - P) \Bigg( \partial_U A^0(0) V^0 \partial_\tau V^0 + $$

$$\sum_{i=1}^{N} A_i(V^0, 0) \partial_{x_i} V^0 + \sum_{j=1}^{M} D_j(V^0, 0) \partial_{y_j} V^0 - F(V^0, 0) \Bigg) \Bigg|_{\tau=t/\varepsilon}$$

for all $t \in [0, T]$ has to be bounded from $H^s(\mathbb{R}^n \times \mathbb{T}^M)$ to $H^{s-1-p}(\mathbb{R}^n \times \mathbb{T}^M)$ for some $p \in \mathbb{R}$. Now, if the other assumptions are satisfied even for $s - 1 \geq s_0 + \max\{0, p\}$ and if $A_i, D_j, F$ are continuously differentiable in $\varepsilon$, then it even holds that

$$\|U^\varepsilon - \widetilde{V}^0\|_{C([0,t_0], H^{s-1}(\mathbb{R}^n \times \mathbb{T}^M))} \leq C t_0 \varepsilon,$$

see [122, Corollary 2.4]. We refer to [122, Lemma 2.5] for an analysis of these additional assumptions.

Already in [121,122], these techniques for singular limits of hyperbolic systems have been used to derive rigorous results for incompressible limits of fluid equations. They have been further refined in [123]. This version of fast wave averaging has been used in the literature to study the limiting behaviour for geophysical flows with different combinations of low or finite Froude and Rossby numbers, see [124,125].

## 5.2. Homogenization of spatially periodic structures

A standard homogenization problem which is introduced in many textbooks is the one for elliptic partial differential equations of the form

$$-\nabla \cdot (A^\varepsilon \nabla u^\varepsilon) = f \tag{5.10}$$

on a domain $\mathcal{O} \subset \mathbb{R}^n$ with Dirichlet boundary conditions. Here, $A_\varepsilon = A(y/\varepsilon)$ for some 1-periodic, $\mathbb{R}^{n \times n}$-valued function $A$ such that (5.10) is uniformly elliptic. $f : \mathcal{O} \to \mathbb{R}$ is a given forcing term. For small values of $\varepsilon > 0$, the coefficients coming from $A_\varepsilon$ have strong spatial oscillations. The aim is to reduce (5.10) to an equation of similar type, but without the oscillating coefficients. More precisely, the reduction allows one to derive a reduced equation for $u_0$, namely

$$-\sum_{i,j=1}^{n} \bar{a}_{i,j} \partial_{y_i} \partial_{y_j} u_0 = f,$$

where the $\bar{a}_{i,j}$ are just scalars and where $u_0$ nonetheless is a good approximation of $u^\varepsilon$ in (5.10).

**Outline of the Reduction Procedure:** A popular approach is to make the ansatz

$$u^\varepsilon(y) = u_0(\tfrac{y}{\varepsilon}, y) + \varepsilon u_1(\tfrac{y}{\varepsilon}, y) + \varepsilon^2 u_2(\tfrac{y}{\varepsilon}, y) + \cdots$$
$$= u_0(x, y) + \varepsilon u_1(x, y) + \varepsilon^2 u_2(x, y) + \cdots$$

so that one obtains an asymptotic expansion in $\varepsilon$ together with a two-scale structure of the variables. Using this two-scale ansatz one can rewrite

$$-\nabla \cdot (A^\varepsilon \nabla) =: L = \frac{1}{\varepsilon^2} L_1 + \frac{1}{\varepsilon} L_2 + L_3$$

where

$$L_1 w := -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i} a_{i,j}(x) \frac{\partial}{\partial x_j} w,$$

$$L_2 w := -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i} a_{i,j}(x) \frac{\partial}{\partial y_j} w + \frac{\partial}{\partial y_i} a_{i,j}(x) \frac{\partial}{\partial x_j} w,$$

$$L_3 w := -\sum_{i,j=1}^{n} \frac{\partial}{\partial y_i} a_{i,j}(x) \frac{\partial}{\partial y_j} w,$$

with $A(x) = (a_{i,j}(x))_{1 \leq i,j, \leq n}$. Now one plugs the expansion for $u^\varepsilon$ into (5.10) and separates the different orders in $\varepsilon$. This yields the equations

$$0 = L_1 u_0,$$
$$0 = L_1 u_1 + L_2 u_0,$$
$$f = L_1 u_2 + L_2 u_1 + L_3 u_0.$$

One obtains that the solution of $L_1 u_0 = 0$ only depends on $y$ and is constant in $x$. Introducing the 1-periodic solution $\chi_i$ of the cell problem

$$L_1 \chi_i = -\sum_{j=1}^{n} \frac{\partial}{\partial x_j} a_{i,j}(x)$$

allows one to derive a reduced equation for $u_0$, namely

$$-\sum_{i,j=1}^{n} \bar{a}_{i,j} \partial_{y_i} \partial_{y_j} u_0 = f \tag{5.11}$$

with Dirichlet boundary conditions, where

$$\bar{a}_{i,j} := \int_{\mathbb{T}^n} a_{i,j}(x) - \sum_{k=1}^{n} a_{i,k}(x) \frac{\partial \chi_i}{\partial x_k} \, dx,$$

i.e. the $\overline{a}_{i,j}$ are just constants. Finally, one can show that $u^\varepsilon \to u_0$ in the $L_2(\mathcal{O})$-sense as $\varepsilon \to 0$, so that one may reduce (5.10) to (5.11).

**Example:** One can use homogenization to derive Darcy's law, which models the laminar flow of a fluid in porous media, from Navier–Stokes type equations. There exist different variants of the reduction procedure. We refer to [126] and references therein for an overview of them. Here, we briefly describe the reduction in [127]. Therein, the incompressible Navier–Stokes equation

$$
\begin{aligned}
\partial_t v^\varepsilon + (v^\varepsilon \cdot \nabla)v^\varepsilon + \nabla p^\varepsilon - \mu \Delta v^\varepsilon &= f^\varepsilon && \text{in } \Omega_\varepsilon \times (0,T), \\
\operatorname{div} v^\varepsilon &= 0 && \text{in } \Omega_\varepsilon \times (0,T), \\
v^\varepsilon &= h && \text{on } \Gamma \times (0,T), \quad (5.12) \\
v^\varepsilon &= 0 && \text{on } S_\varepsilon \times (0,T), \\
v^\varepsilon(x,0) &= v_0^\varepsilon(x) && \text{in } \Omega_\varepsilon
\end{aligned}
$$

is studied. Here, the periodicity comes from the special shape of the domain $\Omega_\varepsilon$ which is given as a smooth domain with periodically distributed holes. More precisely, one starts with a smooth domain $\Omega \subset \mathbb{R}^n$ with $\Gamma := \partial\Omega$ and a second smooth domain $\mathcal{D} \subset [0,1]^n$. Now, one defines the set $T_\varepsilon := \{k \in \mathbb{Z}^n : \varepsilon k + [0,\varepsilon]^n \subset \Omega\}$ as well as

$$
\mathcal{D}_\varepsilon := \bigcup_{k \in T_\varepsilon} \varepsilon(k + \mathcal{D}), \quad S_\varepsilon := \partial\mathcal{D}_\varepsilon, \quad \Omega_\varepsilon := \Omega \setminus \overline{\mathcal{D}}_\varepsilon,
$$

where $\overline{\mathcal{D}}_\varepsilon$ denotes the closure of $\mathcal{D}_\varepsilon$. The given data is assumed to satisfy

$$
\begin{aligned}
& h \in C([0,T]; W_\infty^1(\Omega; \mathbb{R}^n)), \quad \partial_t h = C([0,T]; W_6^1(\Omega; \mathbb{R}^n)), \\
& \quad \operatorname{div} h = 0, \\
& \quad \varepsilon \|v_0^\varepsilon\|_{L_2(\Omega_\varepsilon; \mathbb{R}^n)} = \mathcal{O}(1), \quad \operatorname{div} v_0^\varepsilon = 0 \quad \text{in } \Omega_\varepsilon, \\
& \quad f^\varepsilon \in L_2((0;T) \times \Omega_\varepsilon; \mathbb{R}^n), \quad \varepsilon^2 f^\varepsilon \to f \quad \text{as } \varepsilon \to 0,
\end{aligned}
$$

for a certain $f \in L_2((0;T) \times \Omega_\varepsilon; \mathbb{R}^n)$. Note that $h$, which defines the flux through the outer boundary $\Gamma$, is chosen such that

$$
\int_\Gamma \langle h(x), \nu(x) \rangle \, dS(x) = 0,
$$

where $\nu$ denotes the unit outer normal on $\Gamma$ and $S$ denotes the surface measure.

For the reduction to Darcy's law, the solutions of a suitable cell problem play again an important role. This time, it takes the form

$$
\begin{aligned}
-\nabla \pi^j + \Delta w^j + e^j &= 0 && \text{in } [0,1]^n \setminus \mathcal{D}, \\
\operatorname{div} w^j &= 0 && \text{in } [0,1]^n \setminus \mathcal{D}, \quad (5.13) \\
w^j &= 0 && \text{on } \partial\mathcal{D},
\end{aligned}
$$

with periodic boundary conditions on $\partial[0,1]^n$, where $e^j$ denotes the $j$th canonical unit normal vector in $\mathbb{R}^n$. From the unique solution $(w^j, \pi^j)$ (up to constants in $\pi$) of (5.13) one can construct the so-called permeability tensor $K = (K_{j,k})_{1 \le j,k \le n}$ by means of

$$
K_{j,k} = \int_{[0,1]^n \setminus \mathcal{O}} (w^j)_k(x) \, dx.
$$

This permeability tensor now appears in the homogenized equation which is given by

$$
\begin{aligned}
\operatorname{div} v &= 0 && \text{in } L_2(\Omega \times (0,T)), \\
v = &= \frac{K}{\mu}(f - \nabla p) && \text{in } L_2(\Omega \times (0,T)), \quad (5.14) \\
v \cdot \nu &= h \cdot \nu && \text{in } L_2((0,T); H^{1/2}(\Gamma)).
\end{aligned}
$$

Solutions $(v^\varepsilon, p^\varepsilon)$ of (5.12) are defined on $\Omega_\varepsilon$. However, the aim of the homogenization procedure is to average out the effect of the holes so that (5.14) and its solutions $(v, p)$ should be defined on

$\Omega$. Therefore, if one wants to compare the approximation $(v, p)$ to the solution of the original problem $(v^\varepsilon, p^\varepsilon)$, one should either restrict $(v, p)$ to $\Omega_\varepsilon$ or extend $(v^\varepsilon, p^\varepsilon)$ in a suitable way to $\Omega$. The latter approach, which has the advantage that the approximation result can be formulated in terms of convergence in a fixed space with a fixed topology, was chosen in [127]. While $v^\varepsilon$ can just be extended by 0, the adjoint of a certain restriction operator is used to define the extension $\widetilde{P}^\varepsilon$ of a time integrated version $P^\varepsilon$ of the pressure $p^\varepsilon$. We refer to [127, Section 3] for the details of the construction. By [127, Theorem 4.1] the extensions $v^\varepsilon$ and $\widetilde{P}^\varepsilon$ now satisfy

$$
\begin{aligned}
& v^\varepsilon \to v \quad \text{weakly in } L_2((0,T) \times \Omega; \mathbb{R}^n), \\
& \varepsilon^2 \partial_t \widetilde{P}^\varepsilon \to p \quad \text{weakly in } W_2^{-1}((0,T); L_\beta^0(\Omega)),
\end{aligned}
$$

where $1 < \beta < \frac{n}{n-1}$ and where $L_\beta^0(\Omega) := \{f \in L_\beta(\Omega) : \int_\Omega f(x) \, dx = 0\}$.

**Comments:**

- An important concept for homogenization techniques is the so-called two-scale convergence [110], as it is used in many proofs related to homogenization. Let $\mathcal{O} \subset \mathbb{R}^n$ be a domain and $\mathbb{T}^n$ the $n$-dimensional torus. Then we say that a family of functions $(u^\varepsilon)_{\varepsilon \in (0,1)} \subset L_2(\mathcal{O})$ two-scale converges to $u^0 \in L_2(\mathcal{O} \times \mathbb{T}^n)$ if for all $\varphi \in L_2(\mathcal{O}; C(\mathbb{T}^n))$ it holds that

$$
\lim_{\varepsilon \to 0} \int_\mathcal{O} u^\varepsilon(x)\varphi(x, \tfrac{x}{\varepsilon}) \, dx = \int_\mathcal{O} \int_{\mathbb{T}^n} u^0(x,y)\varphi(x,y) \, dy \, dx.
$$

- Homogenization techniques also exist for parabolic partial differential equations. They are usually applied to equations of the form

$$
\begin{aligned}
& \partial_t u^\varepsilon(t,x) = \frac{1}{\varepsilon} \langle b(\tfrac{x}{\varepsilon}), \nabla \rangle u^\varepsilon(t,x) + D\Delta u^\varepsilon(t,x) \quad ((t,x) \in \mathbb{R}_+ \times \mathbb{R}^n), \\
& u^\varepsilon(0,x) = u_0(x) \quad (x \in \mathbb{R}^n),
\end{aligned}
$$
(5.15)

where $D > 0$ and $b: \mathbb{R}^n \to \mathbb{R}^n$ is smooth and 1-periodic in all directions. Such equations are for example treated in [110, Chapter 13 & Chapter 20].

- Another approach that does not involve passing to the $\varepsilon = 0$ limit is to embed the problem in a larger system where all phase-shifts are considered; see for example [128, Section 4].

Yet another important variant of homogenization in the context of climate dynamics are techniques reducing fast chaotic degrees of freedom to a lower-dimensional stochastic input, resulting in a reduced model.

## 5.3. Stochastic modelling of deterministic systems via homogenization

It is a common idea in climate dynamics that variability corresponding to chaotic parts of the dynamics which happen on a fast time scale can effectively be approximated by stochastic terms. This idea goes back to Hasselmann [8] and has since been used in many climate models, with significant contributions from Majda and co-workers [9,10,15]. From a mathematical perspective, not all the reductions from multi-scale ordinary differential equations to stochastic differential equations are rigorously justified. Nonetheless, there are mathematical techniques which can make the reductions rigorous. These techniques are usually also called homogenization, but are different from the ones in Section 5.2. The starting point of such homogenization techniques

are fast–slow systems of ordinary differential equations of the form

$$\frac{du^\varepsilon}{dt} = \frac{1}{\varepsilon} f_0(u^\varepsilon, v^\varepsilon) + f_1(u^\varepsilon, v^\varepsilon),$$
$$\frac{dv^\varepsilon}{dt} = \frac{1}{\varepsilon^2} g_0(u^\varepsilon, v^\varepsilon) + \frac{1}{\varepsilon} g_1(u^\varepsilon, v^\varepsilon) + g_2(u^\varepsilon, v^\varepsilon),$$
(5.16)

with nonlinearities $f_0, f_1, g_0, g_1, g_2$ and $u^\varepsilon$ representing slow variables forced by fast variables $v^\varepsilon$. The aim is to find reduced equations that are stochastic differential equations of the form

$$dU(t) = \overline{g}(U(t))\,dt + \overline{\Sigma}(U(t))\,dW(t),$$
(5.17)

with nonlinearities $\overline{g}$ and $\overline{\sigma}$, and a certain noise term $W$ which is a Brownian motion in many but not in all cases. It is important to note that unlike in (5.7) the equations we start with are deterministic, i.e., we do not assume that $\sigma_1\sigma_1^T$ is uniformly positive definite. Rigorous reduction results from (5.16) to (5.17) are much more recent than the ones for (5.7). We refer to [107,109,129,130].

**Outline of the Reduction Procedure:** As already mentioned above, a system with multiple time scales of the form (5.16) is reduced to a stochastic differential equation of the form (5.17) with different choices of $\overline{g}$, $\overline{\sigma}$ and $W$, depending on the assumptions. In fact, one may even have to use different kinds of stochastic integrals.

This reduction is shown in a simpler setting in [109] starting with

$$\frac{du^\varepsilon}{dt} = \frac{1}{\varepsilon} f_0(v^\varepsilon) + f_1(u^\varepsilon, v^\varepsilon),$$
$$\frac{dv^\varepsilon}{dt} = \frac{1}{\varepsilon^2} g_0(v^\varepsilon),$$
(5.18)

for $u \in \mathbb{R}^d$, with initial condition $(u^\varepsilon(0), v^\varepsilon(0)) = (\xi, \eta)$, there is an invariant measure $\mu$ for the fast dynamics $v^\varepsilon$ that is assumed uniformly or nonuniformly hyperbolic with fast decay of correlations and $f_0$ has zero mean with respect to $\mu$. It is then possible to show [109, Theorem 1.1] that under reasonable regularity and dynamical assumptions there is weak convergence of solutions $u^\varepsilon \to_w U$ for $t > 0$ as $\varepsilon \to 0$, where $U$ satisfies an SDE on $\mathbb{R}^d$ given by

$$U(t) = \xi + \int_0^t F(U(s))\,ds + \sqrt{\Sigma}W(t)$$

where $W$ is Brownian motion, $\Sigma$ a covariance matrix and $F$ is obtained by averaging $f$ with respect to $\mu$. This relies on assumptions on the fast dynamics that ensure there is a Weak Invariance Principle (WIP) of the form

$$\frac{1}{\sqrt{n}} \int_0^{nt} f_0(v^1(\tau))\,d\tau \to_w \sqrt{\Sigma}W(t)$$

in $C^0([0, T], \mathbb{R}^d)$ as $n \to \infty$ as well as a large deviation assumption on the fast dynamics.

This is generalized in [107,130] to cases that include not only additive but more general (such as multiplicative) noise, where $f_0$ depends not just on $v^\varepsilon$ but also on $u^\varepsilon$, including cases (with slow decay of correlations) where the limiting SDE is driven by a stable Lévy process rather than a Wiener process. Interesting further work in this direction includes [131] who use Edgeworth expansions to find corrections to asymptotic homogenization results for small $\varepsilon$.

**Comments:**

- Although, since the work of Hasselmann [8], homogenization has arguably been one of the most applied methods in climate dynamics, it is also one of the methods that is most

difficult to justify in a rigorous manner. Results so far tend to assume a skew product structure, quite strong assumptions on the forcing chaos and also a timescale separation close to an asymptotic limit. All of these assumptions cannot be verified except in some fairly limited settings, however the utility of the stochastic modelling approach goes well beyond what can be rigorously justified.

- The methods discussed in Section 4.2 as well as the methods discussed in this section do rely on scale separation. Yet, there are interesting problems, where scale separation breaks down. For example, in the context of time scales, classical non-autonomous dynamical systems theory [132] provides tools to formulate and understand relevant notions of attraction and attractor in the non-autonomous context, in particular pullback notions of attraction. In climate dynamics and other applications, this has gained recent attention under the theme of rate-induced instability effects [133–136] which presents some methods for understanding the location where scale separation breaks, and what happens at such rate-induced tipping points.

Related to averaging and homogenization is another classical idea from statistical physics, where a reduction is performed using the concept of averaging in a slightly different variant.

### 5.4. Mori–Zwanzig reduction

Another reduction method that can be applied to remove fast degrees of freedom from an evolution equations is the projection method originally developed by Mori and Zwanzig [137–139] and for example discussed in [140]. This method has recently found use in several climate applications [141,142]. It is unusual in relation to the other techniques discussed in this review in that it includes a memory term (though there are other methods that also will have a memory term, see for example the stochastic centre manifold approach of [111,118]).

**Outline of Reduction Procedure:** If we consider an evolution equation of the form

$$\frac{d}{dt}\varphi = R(\varphi), \quad \varphi(0) = x$$

that generates a flow $\varphi(x, t)$, for the sake of simplicity we assume $\varphi$ (and therefore $x$) are in $\mathbb{R}^n$. Suppose we would like to predict the evolution of some observable $u(x, t) = g(\varphi(x, t))$. Then by considering the linear PDE

$$\frac{\partial}{\partial t}u(x, t) = \mathcal{L}u(x, t), \quad u(x, 0) = g(x)$$
(5.19)

where $[\mathcal{L}u](x) = \sum_i R_i(x)\partial_{x_i}u(x)$, $\mathcal{L}$ is the Liouville operator, and the solution of (5.19) can be written

$$u(x, t) = [\exp(t\mathcal{L})g](x).$$

We aim to understand the evolution of $g(\varphi(x, t))$ in terms of the evolution under (5.19). Consider now a projection $P$ of $\mathbb{R}^n$ onto an $m$-dimensional subspace, with $m < n$. We wish to understand the dynamics of the observable using projection by $P$ onto a number of resolved variables $\hat{\varphi} := P\varphi$, noting the unresolved variables can be written $\overline{\varphi} = (1 - P)\varphi$ (various choices for $P$ are discussed in [140]).

For example, if we consider $g(x) = x_i$ as projection onto a single component then it is possible to show (for example, [140,142]) that it evolves according to a generalized Langevin-type equation of the form

$$\frac{\partial}{\partial t}\hat{\varphi}_i(x, t) = R_i(\hat{\varphi}(x, t)) + \int_0^t K_i(\hat{\varphi}(x, t - s), s)\,ds + F_i(x, t)$$
(5.20)

where

$$F_i(x, t) = [\exp(t(1 - P)\mathcal{L})(1 - P)\mathcal{L}x]_i, \quad K_i(\hat{x}, t) = [P\mathcal{L}F]_i(x, t).$$

Note that for the resolved variables, the first term on the right hand side of (5.20) corresponds to a Markovian term that only depends on the current values of the resolved variables, the second term embodies a memory kernel while the third and final term encapsulates the influence of the unresolved variables. Depending on the situation, one may end up with an ODE reduction (if the last two terms in (5.20) are small) a DDE reduction (if the last term in (5.20) is small) or an SDE reduction (in cases where the middle term in (5.20) is small and the last term can be understood as a stochastic term).

**Examples:** The Mori–Zwanzig reduction method can be a useful method to derive a delay differential equation from a more complex PDE with hyperbolic terms. For example, [142] use this formalism to justify a rational reduction of a PDE model of the El Niño Southern Oscillation (ENSO) to a delay differential equation, while [143] use the method to justify a reduction of a PDE model of Atlantic Multidecadal Variability (AMV) to a delay differential equation.

## 6. Moment closure methods

The manifold reductions as well as techniques related to averaging and homogenization described in Sections 2–5 directly work with the equations describing a certain system. However, if a system is very complex one may prefer to only work with a set of observables of the original system. In complex nonlinear systems it is common that such a set of observables depends on other observables again. If one wants to describe all the observables through a closed set of differential equations, then one might obtain infinitely many observables with infinitely many equations.

The following reduction method, the so-called moment closure methods, treat exactly such a kind of problems. Moment closure methods have been used in many different scientific disciplines, including climate [144]. They can be applied if one works with an abstract evolution equation of the form

$$\partial_t u = N(u). \tag{6.1}$$

This equation is to be understood in a formal way. For example, we also allow the case in which $N$ contains noise terms. Frequently, the method is applied to stochastic differential equations [145] of the form

$$dx_t = f(x_t)\,dt + F(x_t)\,dW_t. \tag{6.2}$$

But also other very important scenarios are possible such as kinetic PDEs [146] of the form

$$\partial_t \varrho + v \cdot \nabla_x \varrho = Q(\varrho), \tag{6.3}$$

where $\rho = \rho(t, x, v)$ can be interpreted – if normalized – as the probability density of a single particle being located at $x$ and having velocity $v$ at time $t$. The so-called collision operator $Q$ usually only depends on the velocity $v$. Moment closure methods are also frequently used in network dynamics [147]. We also refer to the references in the review [148].

**Idea:** An equation like (6.1) may be too complex to be studied analytically or even numerically. However, one can still try to describe certain quantities extracted from the full system, the so-called moments. Usually, the moments are scalar-valued and have a certain hierarchy in the sense that there is a natural ordering which allows one the speak of higher order and lower order moments. One may use (6.1) to derive new equations describing the dynamics of the moments. These new equations are referred to as moment equations. Technically, this step is not a reduction

procedure in the sense that the moment equations do not approximate the original system. Instead, only certain aspects of the original system are considered, which in exchange are described at full resolution and not only approximated. The approximation procedure is contained in the next step: Often, the moment equations will be a fully coupled system of infinitely many differential equations and such a system may be as complicated as the original equation (6.1). In order to reduce the complexity, the moment closure is performed, i.e., the higher order moments are assumed to depend in a certain way on the lower order moments so that one can derive a closed system with finitely many variables. If the moment closure is performed in the right way, then the reduced system should be much simpler to study while still being a good approximation of the full system of moment equations. While in the abstract framework one might think that it is hard to find a good moment closure, many straightforward choices in applications work surprisingly well.

**Outline of the Reduction Procedure:** The procedure usually consists of four steps

(1) One has to determine a set of moments $\{m_k : k \in \mathbb{N}\}$ one wants to consider. In applications, moments are often of the following form: One starts with a suitable solution concept for (6.1) and a solution $u : \mathbb{R}_+ \to X$ with some state space $X$. Moreover, one has a set of mappings $\{M_k : k \in \mathbb{N}\}$ with $M_k : X \to X$ as well as a mapping $\langle \cdot \rangle : X \to \mathbb{K}$, where $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, which can be interpreted as a notion of average. One may think of $\langle \cdot \rangle$ as being a tool to remove complexity from the system, for example in the form of randomness or spatial dependence. The mappings $M_k$ in turn are used to still preserve the information on certain aspects of the system. Now, one defines the moments $m_k(t) := \langle M_k(u(t)) \rangle$. The precise choice of $\{M_k : k \in \mathbb{N}\}$ and the mapping $\langle \cdot \rangle$ depends a lot on the specific problem. The main requirement is that it should be possible to carry out the second step, the derivation of a system of moment equations, which we briefly explain below.

One of the most common scenarios, which also motivates the terminology in the context of moment closure methods, is the case of a stochastic differential equation (6.2), where one may oftentimes just use the usual moment of a random variable. The solution to (6.2) will be a stochastic process, and if it takes values in $X = \bigcap_{p>0} L_p(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the underlying probability space, then one may define

$$M_k : X \to X, \ u \mapsto u^k \quad \text{and} \quad \langle \cdot \rangle : X \to \mathbb{R}, \ u \mapsto \mathbb{E}[u].$$

Therefore, $m_k$ is just the $k$th moment of the solution process.

(2) One has to derive the moment equations. The details of this step depend a lot on the choice of moments in the first step. But generally, the aim is to derive an infinite system of differential equations for $\{m_k : k \in \mathbb{N}\}$, i.e. a system of the form

$$\begin{aligned} \dot{m}_1 &= h_1(m_1, m_2, m_3, \ldots), \\ \dot{m}_2 &= h_2(m_1, m_2, m_3, \ldots), \\ \dot{m}_3 &= h_3(m_1, m_2, m_3, \ldots), \\ \ldots &= \qquad \ldots, \end{aligned} \tag{6.4}$$

for certain $h_k : \mathbb{K}^{\mathbb{N}} \to \mathbb{K}$, $k \in \mathbb{N}$. Since we defined $m_k(t) = \langle M_k(u(t)) \rangle$, one might already guess that one would have to combine a chain rule together with the fact that $u$ solves (6.1). Indeed, this approach usually leads in the right direction, but in practice, this may cause difficulties. For example, $u$ might not be differentiable as in the case of a stochastic differential equation. Fortunately, one can still

try to use Itô's lemma in such a situation. Further below, we will carry out this step for the example of a stochastic version of the Stommel–Cessi model.

Even though $h_k$ will not actually depend on all the moments in most applications, the system of moment equations can still be an infinite system, since the $h_k$ usually depend on higher order moments $m_{k+n_1}, \ldots, m_{k+n_l}$ for some natural numbers $l, n_1, \ldots, n_l \in \mathbb{N}$. The problem of reducing this system to a closed finite systems of ordinary differential equations will be the main task of the third step.

(3) One has to close the system of moment equations in a suitable way. There are many ways to do this. In general, one looks for a mapping $H: \mathbb{R}^N \to \mathbb{R}^{\mathbb{N}}$ with suitable $N \in \mathbb{N}$ that determines the higher order moments through the lower order moments, i.e.

$$H(m_1, \ldots, m_N) = (m_{N+1}, m_{N+2}, m_{N+3}, \ldots).$$

Naive approaches which are used quite often and which still work very well are $0 = m_{N+1} = m_{N+2} = \cdots$ or more generally constant higher moments. One can also use certain assumptions coming from the application or the expected outcome. In the case of a stochastic differential equation, one can for example assume that one is looking for random variables with a certain distribution such as Gaussian random variables.

(4) Finally, one should justify that at least the lower order moments are approximated well by the closure procedure. However, since there are many different ways to carry out the first three steps, there is also no general rule for the last step.

**Example:** As mentioned above, the reduction of stochastic differential equations is a standard application of moment closure methods. An example in climate dynamics is a stochastic version of the Stommel–Cessi model (see [90,91]) given by

$$dU_t = \mu - U_t[1 + 7.5(1 - U_t)^2] \, dt + \sigma \, dW_t. \tag{6.5}$$

Here, $U_t$ models the salinity difference between a higher and a lower latitude box in the North Atlantic, $\mu$ models the difference in freshwater flux, $(W_t)_{t\geq0}$ denotes a Brownian motion, and $\sigma$ is a small parameter which models the strength of the noise. This equation is derived from the Stommel–Cessi model by a reduction to the critical manifold and by adding the stochastic forcing, see for example [92, Section 6.2.1] or [65, Example 19.9.3]. In order to carry out a moment closure procedure, we define

$$M_k: X \to X, \ u \mapsto u^k \quad \text{and} \quad \langle \cdot \rangle: X \to \mathbb{R}, \ u \mapsto \mathbb{E}[u].$$

with $X = \bigcap_{p>0} L_p(\Omega, \mathcal{F}, \mathbb{P})$ as above. For the moment equations, we use Itô's Lemma and obtain

$$
\begin{aligned}
dU_t^k =& kU_t^{k-1}dU_t + \frac{k(k-1)}{2}U_t^{k-2}d[U_t] \\
=& k\big[-7.5U_t^{k+2} + 15U_t^{k+1} - 8.5U_t^k + k\mu U_t^{k-1} \\
& + \frac{(k-1)\sigma^2}{2}U_t^{k-2} \, dt + \sigma U_t^{k-1} \, dW_t\big]
\end{aligned}
$$

for $k \in \mathbb{N}$, $k \geq 2$. Taking the expectation and differentiating with respect to $t$ yields

$$
\begin{aligned}
\dot{m}_k(t) =& \big[-7.5m_{k+2}(t) + 15m_{k+1}(t) - 8.5m_k(t) + k\mu m_{k-1}(t) \\
& + \frac{(k-1)\sigma^2}{2}m_{k-2}(t)\big]
\end{aligned}
$$

Using the convention $m_0 = 1$ and $m_{-1} = 0$, this equation is valid for all $k \in \mathbb{N}$ and yields the system of moment equations.

**Comments:**

- Formulating abstract results on the validity of moment closure methods still seems to be an open problem. However, moment closure methods often work very well in practice, when they are benchmarked via numerical simulations.
- In some sense, we may view moment closure methods as related to other reduction methods. The idea to use set of basis functions from Galerkin or the idea to use a linear subspace from EOF could also be viewed as one choice of the space of observables. The idea to use averaged observables is connected to previously presented ideas on averaging and homogenization, while selecting a suitable closure is connected to invariant manifolds [148].

The reduction methods discussed so far aim to reduce the complexity or dimension of a dynamical system which is given by an explicit equation. However, nowadays large amounts of data on the Earth's climate system are collected so that compressing and finding structures in large data sets can be considered as least as important for the understanding of the Earth's climate as the formal reduction of dynamical systems. In principle, the EOFs approach described in Section 2.2 fulfils this purpose very well, but as a linear method, it may have difficulties in preserving nonlinear structures such as lower-dimensional manifolds. Diffusion maps, which we discuss in Appendix, are designed for exactly that purpose and are just one of the many, data-driven reduction methods that start to permeate many applications in climate science. They have already found quite widespread applications across numerous areas [149]. Of course, diffusion maps are just one instance of a broader class of data-driven reduction methods that are likely to gain more traction within climate science applications during the next decade. In particular, we have not attempted to cover the full range of data-driven reduction methods [150] that presumably will be of great significance in future, including for climate model reduction; see for example [151–153]. Nonetheless, many of the explainable data-driven methods have their roots within reduction methods such as those discussed here, and indeed the methods discussed here may contribute to such methods in future.

## 7. Discussion

In this review we have focused primarily on reduction methods of use (or potential use) in climate science that can be rigorously justified. This is a very diverse area and we unavoidably have had to limit ourselves to an incomplete set of reduction methods. These principles are likely to remain themes for future research as new techniques emerge and improve on existing reduction methods. Our brief review aims to make it easier to compare the mathematical setup of these methods: we try to highlight essential ideas for each method and provide references to concrete examples that serve as test cases. Of course, a natural next step would be to actually code all the different methods on an even broader set of examples and provide computational comparisons for each ansatz. Although this is certainly worthwhile, it is beyond this work, where we have focused on describing the main idea and mathematical setup for each method.

Probably the most relevant omission are some reduction methods for stochastic differential equations, either SDEs or SPDEs, and more general data-based empirical reduction methods. We have not reviewed this area here, primarily to keep this work sufficiently accessible across disciplines, but there are undoubtedly many important methods also available for the stochastic case. As a basic principle, one can often generalize a reduction method for a deterministic differential equation to the stochastic case, usually at the expense of considerable additional technical work. Examples are path-wise stochastic slow manifold reductions via covariance tubes [45,154], reduction to

Markov chain switching models via large deviation theory [155], asymptotic analysis of stochastic slow manifolds [156], centre/slow manifold reduction using a skew-product flow formalism [157–160], distribution-based approaches simplifying Fokker–Planck equations [161], nonlocal Mori–Zwanzig type reductions [162,163], moment closure [148] and amplitude/modulation equations [164], just to name a few.

One might conjecture that a more coherent and rigorous view of the area will emerge within the coming years as some reduction methods for stochastic systems have been applied successfully in practical climate models already using numerical approaches [165]. It is clear that such reductions may be very informative, especially statistical mechanics and large deviations approaches to weather/climate extreme problems; see for example [166–169] as well as for multistable regimes of the climate system [170]. Another area where rigorous reduction methods are challenged is in justifying the predominantly empirical sub-grid parametrizations used in climate models [171–174].

It is important to acknowledge that many reduction methods of great utility for climate models are being, and need to be, applied in cases where it is impossible to provide rigorous proofs of the reduction along the lines presented here. Nonetheless, rigorous reduction methods are still useful in such contexts because better understanding of when and whether these methods are appropriate. In particular, these can (a) help to find an optimal set of conditions and results such that reduction method can be applied and (b) help to understand the circumstances and details of when such a reduction method fails to work, and what happens in cases where it cannot be applied.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

No data was used for the research described in the article.

### Acknowledgements

### Appendix. Diffusion maps

Diffusion maps are a relatively young and mainly data-driven method of dimensionality reduction, which has first been introduced in [175]. Formally, it can be applied whenever EOFs can be used. But while EOFs are a linear dimensionality reduction method, and as such have problems in finding lower-dimensional but nonlinear structures in the data set, diffusion maps specifically aim at finding low-dimensional structures in high-dimensional data and representations thereof in a low-dimensional Euclidean space.

**Idea:** On a set of data one defines a kernel which is then used to construct a Markov chain with the data set as a state space. The Markov chain can be thought of as a form of diffusion which takes the geometry of the data set into account. The hope is to get some insight on this geometry by studying the structure of the Markov chain. This is achieved by defining the so-called diffusion maps, which are constructed from the eigenvalues and eigenfunctions of the Markov chain. The diffusion maps embed the data set into Euclidean space such that the Euclidean distance of two embedded points is close to the distance of the original points in the geometry of the data set.

**Outline of the Reduction Procedure:** One starts with a measure space $(X, \mathscr{A}, \mu)$ and a map $k : X \times X \rightarrow [0, \infty)$ which is symmetric in the sense that $k(x, y) = k(y, x)$ for $x, y \in X$. This map is thought to measure the distance between two points in the data set $X$: the larger $k(x, y)$ the closer are $x, y$ thought to be. Technically, one has to impose further assumptions on $k$ so that the objects we define in the following are well-defined. But since these assumptions are very mild and satisfied in most situations, we refrain from being precise here. Instead, we refer to [175] for more details on the construction.

Having chosen $k$ one defines

$$d(x) := \int_X k(x, y) \, d\mu(y)$$

which can be interpreted as a local measure of volume. This local measure of volume is used to renormalize $k$ and define

$$p(x, y) := \frac{k(x, y)}{d(x)}.$$

The mapping $p$ is not symmetric anymore, but it satisfies $\int_X p(x, y) \, dy = 1$. It can therefore be seen as the transition kernel of the Markov chain $((P^*)^n)_{n \in \mathbb{N}}$ given as the powers of operator

$$P^* : L_2(X, \mathscr{A}, \mu) \to L_2(X, \mathscr{A}, \mu), \ P^* f(y) := \int_X p(x, y) f(x) \, d\mu(x).$$

This operator is the adjoint of the so-called diffusion operator

$$P : L_2(X, \mathscr{A}, \mu) \to L_2(X, \mathscr{A}, \mu), \ Pf(x) := \int_X p(x, y) f(y) \, d\mu(y).$$

which leaves constant functions invariant due to $\int_X p(x, y) \, dy = 1$. Under the right conditions on $k$ (again we refer to [175]) there is a unique stationary distribution of the Markov chain given by

$$\pi(x) := \frac{d(x)}{\int_X d(z) \, d\mu(z)}.$$

Using this stationary distribution one can define the modified kernel

$$a(x, y) := \frac{\sqrt{\pi(x)}}{\sqrt{\pi(y)}} p(x, y) = \frac{k(x, y)}{\sqrt{\pi(x)} \sqrt{\pi(y)}}.$$

This kernel is symmetric so that the corresponding operator

$$A : L_2(X, \mathscr{A}, \mu) \to L_2(X, \mathscr{A}, \mu), \ Af(x) := \int_X a(x, y) f(y) \, d\mu(y)$$

is self-adjoint. If $\int_X \int_X a(x, y)^2 \, d\mu(x) \, d\mu(y) < \infty$, then $A$ is even compact and there is a sequence of eigenvalues $1 = \lambda_0 \geq |\lambda_1| \geq |\lambda_2| \geq \ldots$ and orthonormal eigenfunctions $(\eta_l)_{l \in \mathbb{N}_0}$ in $L_2(X, \mathcal{A}, \mu)$ such that

$$a(x, y) = \sum_{l \in \mathbb{N}_0} \lambda_l \eta_l(x) \eta_l(y).$$

If we now define

$$\psi_l(x) := \frac{\eta_l(x)}{\sqrt{\pi(x)}}, \quad \varphi_l(y) := \eta_l(y) \sqrt{\pi(y)}$$

then we have the representation

$$p(x, y) = \sum_{l \in \mathbb{N}_0} \lambda_l \psi_l(x) \varphi_l(y).$$

Moreover, the $t$th powers of $P$ are given as an integral operator with kernel

$$p_t(x, y) = \sum_{l \in \mathbb{N}_0} \lambda_l^t \psi_l(x) \varphi_l(y).$$

With all this notation at hand, we can define the diffusion distances $D_t$ by

$$D_t(x, y) := \left( \int_X \frac{(p_t(x, u) - p_t(y, u))^2}{\pi(u)} d\mu(u) \right)^{1/2}$$

$$= \left( \sum_{l \in \mathbb{N}_0} \lambda_j^{2t} (\psi_l(x) - \psi_l(y))^2 \right)^{1/2}.$$

For fixed $t \geq 0$ this defines a metric on the data set $X$. This expression is small if there is a large number of short paths between $x$ and $y$, i.e. if there is a high probability that the Markov chain ends up in $y$ after $t$ steps if it starts in $x$. In this sense, the diffusion distance contains a lot of information on the internal structure of the data set $X$.

The reduction is now carried via the so-called diffusion maps given by

$$\Psi_t : X \to \mathbb{R}^N, \quad \Psi_t(x) = \begin{pmatrix} \lambda_0^t \psi_0(x) \\ \vdots \\ \lambda_N^t \psi_N(x) \end{pmatrix}$$

up to a desired dimension $N$. In [175] the authors include a certain rule of the dimensionality that should be considered. They take $N = s(\delta, t) := \max\{l \in \mathbb{N} : |\lambda_l|^t > \delta|\lambda_1|^t\}$. This way, the diffusion maps preserve the diffusion distance to a certain extent in the sense that

$$\|\Psi_t(x) - \Psi_t(y)\|_2 = \left( \sum_{l=0}^{s(\delta, t)} \lambda_j^{2t} (\psi_l(x) - \psi_l(y))^2 \right)^{1/2} := D_{t, \delta}(x, y),$$

where $\| \cdot \|_2$ denotes the usual Euclidean distance. Therefore, the diffusion maps give a representation of the data set $X$ in an $N$-dimensional Euclidean space, such that the internal notion of distance in $X$ is quantitatively preserved in $\mathbb{R}^N$ with the Euclidean distance up to a small error.

**Examples of Applications:** Since diffusion maps are mainly a data-driven method, the results one obtains from the reduction procedure are probably most interesting when it is applied to real world data. Instead of explicitly carrying out such an example, let us briefly explain a few common scenarios. A standard situation is if $X = \{x_1, \ldots, x_K\} \subset \mathbb{R}^d$ is a finite set of points in $\mathbb{R}^d$. Then the $\sigma$-field would just be the power set $\mathscr{P}(X)$ of $X$ and $\mu$ could for example be the counting measure $\zeta$. In this case, the integrals just turn into sums, i.e.

$$\int_X f(x) \, d\mu(x) = \sum_{i=1}^{K} f(x_i).$$

A common choice for a kernel in this situation would be

$$k_\varepsilon(x, y) = \exp\left( \frac{\|x - y\|^2}{2\varepsilon} \right)$$

for a parameter $\varepsilon > 0$. With these choices one could technically apply a diffusion map reduction whenever an EOF-based reduction is also possible. As for EOFs, one can also use diffusion maps to study dynamical systems by applying them to simulated data from the system. As an example, we mention [176].

An interesting use case is if $X := M$ is a compact submanifold $\mathbb{R}^d$ (possibly with boundary) with its Riemannian measure $\nu$. In practice, $X$ might only consist of a finite number of sample points

with a non-uniform distribution on $M$. Although we focus first on the continuous case, i.e. $X = M$, instead of the discrete case, we adopt the idea of a non-uniform distribution and introduce a density $q : M \to (0, \infty)$ on $M$. Then we take $d\mu = q \, d\nu$. Obviously, the outcome of the diffusion map reduction can depend on the density $q$ as well as the geometry of $M$. By choosing the right kernels, one can control the effect of the density. Depending on what is thought to be more important in a particular application, one can choose a corresponding kernel. More precisely, the kernels are constructed as follows: One starts with rotation-invariant reference kernels

$$k_\varepsilon(x, y) := h(\|x - y\|/\varepsilon),$$

where $h : [0, \infty) \to (0, \infty)$ is an exponentially decaying function, for example $h(x) = e^{-x^2/2}$ as above. An approximation of the density is defined by

$$q_\varepsilon(x) := \int_X k_\varepsilon(x, y) q(y) \, dy.$$

Now, a new family of kernels is constructed by

$$k_\varepsilon^{(\alpha)}(x, y) := \frac{k_\varepsilon(x, y)}{q_\varepsilon^\alpha(x) q_\varepsilon^\alpha(y)} \quad (\alpha \in \mathbb{R}).$$

After carrying out the above procedure with the kernel $k = k_\varepsilon^{(\alpha)}$ and the measure $\mu = q \, d\nu$, we obtain a Markov chain $P_{\varepsilon, \alpha}$. Let $(\varphi_l)_{l \in \mathbb{N}_0} \subset L_2(M, \nu)$ be the sequence of orthonormal eigenfunctions of the Laplace–Beltrami operator $\Delta$ on $M$ with Neumann boundary conditions. For $K \in \mathbb{N}_0$ let further

$$E_K := \text{span}\{\varphi_0, \ldots, \varphi_K\}.$$

It was derived in [175, Theorem 2] that the spaces $E_K$ are contained in the domain of the generator of the Markov chain $P_{\varepsilon, \alpha}$. More precisely, for such $f \in E_K$ it holds that

$$\lim_{\varepsilon \to 0} \frac{P_{\varepsilon, \alpha} - I}{\varepsilon} f = \frac{\Delta(fq^{1-\alpha})}{q^{1-\alpha}} - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}} f,$$

where the limit is taken in $L_2(M)$. Note that the authors of [175] use a different convention concerning the sign of the Laplacian: While we take $\Delta = \sum_{j=1}^{n} \partial_j^2$, the authors of [175] work with $\Delta = -\sum_{j=1}^{n} \partial_j^2$. The function $u(t) := P_{t, \alpha} u_0$ solves the equation

$$\partial_t u(t) = \frac{\Delta(u(t) q^{1-\alpha})}{q^{1-\alpha}} - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}} u(t), \quad u(0) = u_0.$$

Substituting $v(t) := q^{1-\alpha} u(t)$ leads to

$$\partial_t v(t) = \Delta v(t) - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}} v(t), \quad u(0) = q^{1-\alpha} u_0. \tag{A.1}$$

Hence, one can study the operator

$$\psi \mapsto \Delta \psi - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}} \psi \tag{A.2}$$

and the dynamics generated by it in terms of the Markov chain $P_{\varepsilon, \alpha}$. There are three values for $\alpha$ which are particularly interesting:

- $\alpha = 0$: This case is traditionally referred to in the literature as normalized graph Laplacian.
- $\alpha = 1$: In this case (A.2) turns into the Laplace–Beltrami operator $\Delta$. In particular, the influence of the density $q$ is removed and the limit only depends on the geometry of the manifold.
- $\alpha = \frac{1}{2}$: Here we obtain $\Delta - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}$. If the density is of the form $q = e^{-U}$, then (A.1) turns into

$$\partial_t v(t) = \Delta v(t) - \left( \frac{\|\nabla U\|_2^2}{4} - \frac{\Delta U}{2} \right) v(t)$$

for (A.2), where $\|\nabla U\|_2$ denotes the Euclidean norm of the gradient of $U$. If we also substitute $f(t, x) := e^{-U(x)/2} v(t, x)$, then we obtain the equation

$$\partial_t f(t, x) = \nabla \cdot (\nabla f(t, x) + f(t, x) \nabla U(x)),$$

which is the forward Fokker–Planck equation to the stochastic differential equation

$$dX_t = -\nabla U(X_t) \, dt + \sqrt{2} \, dW_t \tag{A.3}$$

with reflecting boundary conditions at $\partial M$, where $(W_t)_{t \geq 0}$ is a Brownian motion on $M$. Hence, the Markov chain $P_{\varepsilon, \frac{1}{2}}$ can be used to study the probability densities of solutions of a stochastic differential equation of the form (A.3).

In the discrete situation, the set $X$ is not given by a whole manifold $M$, but by a finite number of points sampled according to $q$. In this case integrals are replaced by sums, i.e. one works with

$$\bar{q}_\varepsilon(x_i) = \sum_{j=1}^m k_\varepsilon(x_i, x_j), \quad \bar{d}_\varepsilon^{(\alpha)}(x_i) = \sum_{j=1}^m \frac{k_\varepsilon(x_i, x_j)}{q_\varepsilon(x_i)^\alpha q_\varepsilon(x_j)^\alpha},$$

$$\bar{p}_{\varepsilon, \alpha}(x_i, x_j) = \frac{k_\varepsilon(x_i, x_j)}{\bar{d}_\varepsilon^{(\alpha)}(x_i)}, \quad \bar{P}_{\varepsilon, \alpha} f(x_i) = \sum_{j=1}^m \bar{p}_{\varepsilon, \alpha}(x_i, x_j) f(x_j)$$

where $m$ denotes the number of points. It was derived in [177] that

$$|P_{\varepsilon, \alpha} f(x_i) - \bar{P}_{\varepsilon, \alpha} f(x_i)| = O(m^{-\frac{1}{2}} \varepsilon^{\frac{1}{2} - \frac{d_M}{4}})$$

with high probability, where $d_M$ denotes the dimension of the underlying manifold. Therefore, the generators satisfy with high probability that

$$|L_{\varepsilon, \alpha} f(x_i) - \bar{L}_{\varepsilon, \alpha} f(x_i)| = O(m^{-\frac{1}{2}} \varepsilon^{-\frac{1}{2} - \frac{d_M}{4}}).$$

# References

[1] H.A. Dijkstra, Nonlinear Climate Dynamics, Cambridge University Press, 2013.
[2] M. Ghil, M.D. Chekroun, E. Simonnet, Climate dynamics and fluid mechanics: Natural variability and related uncertainties, Physica D 237 (14–17) (2008) 2111–2126.
[3] J.W. Hurrell, M.M. Holland, P.R. Gent, S. Ghan, J.E. Kay, P.J. Kushner, J.-F. Lamarque, W.G. Large, D. Lawrence, K. Lindsay, W.H. Lipscomb, M.C. Long, N. Mahowald, D.R. Marsh, R.B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W.D. Collins, J.J. Hack, J. Kiehl, S. Marshall, The community earth system model: A framework for collaborative research, Bull. Am. Meteorol. Soc. 94 (9) (2013) 1339–1360.
[4] V. Masson-Delmotte, A.P. Zhai, S. Pirani, C. Connors, S. Péan, N. Berger, Y. Caud, L. Chen, M. Goldfarb, M. Gomis, K. Huang, E. Leitzell, J. Lonnoy, T. Matthews, T. Maycock, O. Waterfield, R. Yelekçi, B.Z.e. Yu, IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2021.
[5] G.K. Vallis, Atmospheric and Oceanic Fluid Dynamics, Cambridge University Press, 2017.
[6] M. Ghil, V. Lucarini, The physics of climate variability and climate change, Rev. Modern Phys. 92 (3) (2020) 035002.
[7] S.H. Schneider, R.E. Dickinson, Climate modeling, Rev. Geophys. 12 (3) (1974) 447–493.
[8] K. Hasselmann, Stochastic climate models, Part I, Theory, Tellus 28 (1976) 473.
[9] A.J. Majda, I. Timofeyev, E. Vanden Eijnden, A mathematical framework for stochastic climate models, Comm. Pure Appl. Math. 54 (8) (2001) 891–974.
[10] A.J. Majda, I. Timofeyev, E. Vanden-Eijnden, Systematic strategies for stochastic mode reduction in climate, J. Atmos. Sci. 60 (14) (2003) 1705–1722.
[11] H. Kaper, H. Engler, Mathematics and Climate, SIAM, 2013.
[12] J.G. Charney, Carbon Dioxide and Climate: A Scientific Assessment, National Academy of Science. National Academy of Science, 1979.
[13] R. Knutti, M.A. Rugenstein, G.C. Hegerl, Beyond equilibrium climate sensitivity, Nat. Geosci. 10 (10) (2017) 727–736.
[14] M. Ghil, A mathematical theory of climate sensitivity or, how to deal with both anthropogenic forcing and natural variability? in: Climate Change: Multidecadal and beyond, World Scientific, 2016, pp. 31–51.
[15] A.J. Majda, C. Franzke, D. Crommelin, Normal forms for reduced stochastic climate models, Proc. Natl. Acad. Sci. USA 106 (10) (2009) 3649–3653.
[16] C.L. Franzke, T.J. O'Kane, J. Berner, P.D. Williams, V. Lucarini, Stochastic climate theory and modeling, Wiley Interdiscip. Rev. Clim. Change 6 (1) (2015) 63–78.
[17] J. Mitchell, An overview of climatic variability and its causal mechanisms, Quat. Res. 6 (4) (1976) 481–493.
[18] A.S. von der Heydt, P. Ashwin, C.D. Camp, M. Crucifix, H.A. Dijkstra, P. Ditlevsen, T.M. Lenton, Quantification and interpretation of the climate variability record, Glob. Planet. Change 197 (2021) 103399.
[19] L. Evans, Partial Differential Equations, AMS, 2002.
[20] S. Brenner, L. Scott, The Mathematical Theory of Finite Element Methods, third ed., Springer, 2008.
[21] I. Jollife, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. A Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.
[22] C.A.J. Fletcher, Computational Galerkin Methods, in: Springer Series in Computational Physics, Springer-Verlag, New York, 1984.
[23] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems, second ed., in: Springer Series in Computational Mathematics, 25, Springer-Verlag, Berlin, 2006.
[24] M. Ruzicka, Nichtlineare Funktionalanalysis, Springer Spektrum, 2020.
[25] R.E. Showalter, Monotone Operators in Banach Space and Nonlinear Partial Differential Equations, in: Mathematical Surveys and Monographs, 49, American Mathematical Society, Providence, RI, 1997.
[26] C.S. Bretherton, C. Smith, J.M. Wallace, An intercomparison of methods for finding coupled patterns in climate data, J. Clim. 5 (6) (1992) 541–560.
[27] R.W. Preisendorfer, C.D. Mobley, Principal component analysis in meteorology and oceanography, in: Developments in Atmospheric Science, Vol. 17, 1988.
[28] G. Chen, Z. Luo, Proper orthogonal decomposition methods for partial differential equations, in: Mathematics in Science and Engineering, Academic Press, 2019.
[29] H. Björnsson, S. Venegas, A manual for EOF and SVD analyses of climatic data, CCGCR Rep. 97 (1) (1997) 112–134.
[30] A. Hannachi, I.T. Jolliffe, D.B. Stephenson, Empirical orthogonal functions and related techniques in atmospheric science: A review, Int. J. Climatol. 27 (9) (2007) 1119–1152.
[31] E.N. Lorenz, Empirical Orthogonal Functions and Statistical Weather Prediction, Massachusetts Institute of Technology, Department of Meteorology Cambridge, 1956.
[32] A. Navarra, V. Simoncini, A Guide To Empirical Orthogonal Functions for Climate Data Analysis, Springer Netherlands, 2010.
[33] J.-M. Beckers, M. Rixen, EOF calculations and data filling from incomplete oceanographic datasets, J. Atmos. Ocean. Technol. 20 (12) (2003) 1839–1856.
[34] G. Biau, E. Zorita, H. von Storch, H. Wackernagel, Estimation of precipitation by kriging in the EOF space of the sea level pressure field, J. Clim. 12 (4) (1999) 1070–1085.
[35] D.T. Crommelin, A.J. Majda, Strategies for model reduction: comparing different optimal bases, J. Atmospheric Sci. 61 (17) (2004) 2206–2217.
[36] R. Kawamura, A rotated EOF analysis of global sea surface temperature variability with interannual and interdecadal scales, J. Phys. Oceanogr. 24 (3) (1994) 707–715.
[37] W.S. Kessler, EOF representations of the madden–julian oscillation and its connection with enso, J. Clim. 14 (13) (2001) 3055–3061.
[38] F.M. Selten, Baroclinic empirical orthogonal functions as basis functions in an atmospheric model, J. Atmos. Sci. 54 (16) (1997) 2099–2114.
[39] T.M. Smith, R.W. Reynolds, R.E. Livezey, D.C. Stokes, Reconstruction of historical sea surface temperatures using empirical orthogonal functions, J. Clim. 9 (6) (1996) 1403–1420.
[40] J.M. Wallace, R.L. Panetta, J. Estberg, Representation of the equatorial stratospheric quasi-biennial oscillation in EOF phase space, J. Atmos. Sci. 50 (12) (1993) 1751–1762.
[41] H. De Swart, Analysis of a six-component atmospheric spectral model: Chaos, predictability and vacillation, Physica D 36 (3) (1989) 222–234.
[42] D.T. Crommelin, J. Opsteegh, F. Verhulst, A mechanism for atmospheric regime behavior, J. Atmos. Sci. 61 (12) (2004) 1406–1419.
[43] F. Kwasniok, The reduction of complex dynamical systems using principal interaction patterns, Physica D 92 (1–2) (1996) 28–60.
[44] P. Mahalanobis, On the generalised distance in statistics, Proc. Natl. Inst. Sci. India 2 (1) (1936) 49–55.
[45] C. Kuehn, Deterministic continuation of stochastic metastable equilibria via Lyapunov equations and ellipsoids, SIAM J. Sci. Comput. 34 (3) (2012) A1635–A1658.
[46] R. Preisendorfer, C. Mobley, Principal component analysis in meteorology and oceanography, in: Developments in Atmospheric Science, Elsevier, 1988.

[47] A. Monahan, J. Fyfe, M. Ambaum, D. Stephenson, G. North, Empirical orthogonal functions: The medium is the message, J. Clim. 22 (2009).

[48] D. Crommelin, Homoclinic dynamics: a scenario for atmospheric ultra-low frequency variability, J. Atmos. Sci. 59 (2002) 1533–1549.

[49] K. Hasselmann, PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns, J. Geophys. Res.: Atmos. 93 (D9) (1988) 11015–11021.

[50] U. Achatz, G. Schmitz, On the closure problem in the reduction of complex atmospheric models by PIPs and EOFs: A comparison for the case of a two-layer model with zonally symmetric forcing, J. Atmos. Sci. 54 (20) (1997) 2452–2474.

[51] U. Achatz, G. Schmitz, K.-M. Greisiger, Principal interaction patterns in baroclinic wave life cycles, J. Atmos. Sci. 52 (18) (1995) 3201–3213.

[52] F. Kwasniok, Optimal Galerkin approximations of partial differential equations using principal interaction patterns, Phys. Rev. E (3) 55 (5, part A) (1997) 5365–5375.

[53] F. Kwasniok, Low-dimensional models of the Ginzburg-Landau equation, SIAM J. Appl. Math. 61 (6) (2001) 2063–2079.

[54] F. Kwasniok, Empirical low-order models of barotropic flow, J. Atmos. Sci. 61 (2) (2004) 235–245.

[55] J.-L. Lions, Optimal control of systems governed by partial differential equations, in: Die Grundlehren Der Mathematischen Wissenschaften, Band, Vol. 170, Springer-Verlag, New York-Berlin, 1971, Translated from the French by S. K. Mitter.

[56] C. Bender, S. Orszag, Asymptotic Methods and Perturbation Theory, Springer, 1999.

[57] E.D. Jager, J. Furu, The Theory of Singular Perturbations, North-Holland, 1996.

[58] J. Kevorkian, J. Cole, Multiple Scale and Singular Perturbation Methods, Springer, 1996.

[59] M. Cross, P. Hohenberg, Pattern formation outside of equilibrium, Rev. Modern Phys. 65 (3) (1993) 851–1112.

[60] R. Hoyle, Pattern Formation: An Introduction To Methods, Cambridge University Press, 2006.

[61] C. Kuehn, PDE Dynamics: An Introduction, SIAM, 2019.

[62] F. Dumortier, R. Roussarie, Canard cycles and center manifolds, Mem. Am. Math. Soc. 121 (1996) AMS.

[63] H. Jardon-Kojakhmetov, C. Kuehn, A Survey on the Blow-Up Method for Fast-Slow Systems, AMS, 2019, pp. 1–41, accepted at: Contemporary Mathematics, arXiv:1901.01402.

[64] M. Krupa, P. Szmolyan, Extending geometric singular perturbation theory to nonhyperbolic points - fold and canard points in two dimensions, SIAM J. Math. Anal. 33 (2) (2001) 286–314.

[65] C. Kuehn, Multiple Time Scale Dynamics, in: Applied Mathematical Sciences, 191, Springer, Cham, 2015.

[66] R. Klein, Scale-dependent models for atmospheric flows, in: Annual Review of Fluid Mechanics. Vol. 42, in: Annu. Rev. Fluid Mech, 42, Palo Alto, CA, 2010, pp. 249–274, Annual Reviews.

[67] J. Carr, Applications of Centre Manifold Theory, Springer, 1981.

[68] N. Fenichel, Geometric singular perturbation theory for ordinary differential equations, J. Differential Equations 31 (1) (1979) 53–98.

[69] C. Jones, Geometric singular perturbation theory, in: Dynamical Systems (Montecatini Terme, 1994), in: Lect. Notes Math., 1609, Springer, 1995, pp. 44–118.

[70] J. Robinson, Infinite-Dimensional Dynamical Systems, CUP, 2001.

[71] R. Temam, Infinite-Dimensional Dynamical Systems in Mechanics and Physics, second ed., in: Applied Mathematical Sciences, 68, Springer-Verlag, New York, 1997.

[72] S. Zelik, Inertial manifolds and finite-dimensional reduction for dissipative PDEs, Proc. Roy. Soc. Edinburgh Sect. A 144 (6) (2014) 1245–1327.

[73] M. Kwak, Finite-dimensional inertial forms for the 2D Navier-Stokes equations, Indiana Univ. Math. J. 41 (4) (1992) 927–981.

[74] R. Temam, S.H. Wang, Inertial forms of Navier-Stokes equations on the sphere, J. Funct. Anal. 117 (1) (1993) 215–242.

[75] A. Kostianko, S. Zelik, Kwak transform and inertial manifolds revisited, J. Dynam. Differential Equations (2021).

[76] M. Abu Hamed, Y. Guo, E.S. Titi, Inertial manifolds for certain subgrid-scale $\alpha$-models of turbulence, SIAM J. Appl. Dyn. Syst. 14 (3) (2015) 1308–1325.

[77] A. Kostianko, Inertial manifolds for the 3D modified-Leray-$\alpha$ model with periodic boundary conditions, J. Dynam. Differential Equations 30 (1) (2018) 1–24.

[78] X. Li, C. Sun, Inertial manifolds for the 3D modified-Leray-$\alpha$ model, J. Differential Equations 268 (4) (2020) 1532–1569.

[79] J. Avrin, The asymptotic finite-dimensional character of a spectrally-hyperviscous model of 3D turbulent flow, J. Dynam. Differential Equations 20 (2) (2008) 479–518.

[80] C.G. Gal, Y. Guo, Inertial manifolds for the hyperviscous Navier-Stokes equations, J. Differential Equations 265 (9) (2018) 4335–4374.

[81] A. Kostianko, X. Li, C. Sun, S. Zelik, Inertial manifolds via spatial averaging revisited, 2020, pp. 1–34, arXiv:2006.15663.

[82] C. Foias, G.R. Sell, R. Temam, Inertial manifolds for nonlinear evolutionary equations, J. Differential Equations 73 (2) (1988) 309–353.

[83] D. Henry, Geometric Theory of Semilinear Parabolic Equations, in: Lecture Notes in Mathematics, 840, Springer-Verlag, Berlin-New York, 1981.

[84] R. Mañé, Reduction of semilinear parabolic equations to finite dimensional $C^1$ flows, in: Geometry and Topology (Proc. III Latin Amer. School of Math. Inst. Mat. Pura Aplicada CNPQ, Rio de Janeiro, 1976), in: Lecture Notes in Math., 597, 1977, pp. 361–378.

[85] X. Mora, Finite-dimensional attracting manifolds in reaction–diffusion equations, in: Nonlinear Partial Differential Equations (Durham, N.H. 1982), in: Contemp. Math., 17, Amer. Math. Soc., Providence, R.I., 1983, pp. 353–360.

[86] A. Tikhonov, Systems of differential equations containing small small parameters in the derivatives, Mat. Sbornik N. S. 31 (1952) 575–586.

[87] P.W. Bates, K. Lu, C. Zeng, Existence and persistence of invariant manifolds for semiflows in Banach space, Mem. Amer. Math. Soc. 135 (645) (1998) viii+129.

[88] F. Hummel, C. Kuehn, Slow manifolds for infinite-dimensional evolution equations, Comment. Math. Helv. 97 (1) (2022) 61–132.

[89] F.H.M. Engel, C. Kuehn, Connecting a direct and a Galerkin approach to slow manifolds in infinite dimensions, Proc. Amer. Math. Soc. 8 (2021) 252–266.

[90] P. Cessi, A simple box model of stochastically forced thermohaline flow, J. Phys. Oceanogr. 24 (9) (1994) 1911–1920.

[91] H. Stommel, Thermohaline convection with two stable regimes of flow, Tellus 13 (2) (1961) 224–230.

[92] N. Berglund, B. Gentz, Noise-induced phenomena in slow-fast dynamical systems, in: Probability and Its Applications (New York), Springer-Verlag London, Ltd., London, 2006, A sample-paths approach.

[93] M. Engel, C. Kuehn, Blow-up analysis of fast-slow PDEs with loss of hyperbolicity, 2020, pp. 1–35, arXiv:2007.09973.

[94] M. Krupa, P. Szmolyan, Extending geometric singular perturbation theory to nonhyperbolic points—fold and canard points in two dimensions, SIAM J. Math. Anal. 33 (2) (2001) 286–314.

[95] M. Krupa, P. Szmolyan, Extending slow manifolds near transcritical and pitchfork singularities, Nonlinearity 14 (6) (2001) 1473–1491.

[96] G. Schneider, H. Uecker, Nonlinear PDEs: A Dynamical Systems Approach, AMS, 2017.

[97] F. Hummel, S. Jelbart, C. Kuehn, Geometric blow-up of a dynamic Turing instability in the Swift-Hohenberg equation, 2022, pp. 1–69, arXiv:2207.03967.

[98] Y.A. Kuznetsov, Elements of Applied Bifurcation Theory, Volume 112, Springer Science & Business Media, 2013.

[99] P.W. Bates, C.K.R.T. Jones, Invariant manifolds for semilinear partial differential equations, in: Dynamics Reported, Vol. 2, in: Dynam. Report. Ser. Dynam. Systems Appl., 2, Wiley, Chichester, 1989, pp. 1–38.

[100] A. Vanderbauwhede, G. Iooss, Center manifold theory in infinite dimensions, in: Dynamics Reported: Expositions in Dynamical Systems, in: Dynam. Report. Expositions Dynam. Systems (N.S.), 1, Springer, Berlin, 1992, pp. 125–163.

[101] B. Aulbach, T. Wanner, The hartman—grobman theorem for carathéodory-type differential equations in banach spaces, Nonlinear Anal. TMA 40 (1–8) (2000) 91–104.

[102] J. Escher, G. Simonett, A center manifold analysis for the Mullins-Sekerka model, J. Differential Equations 143 (2) (1998) 267–292.

[103] M. Haragus, G. Iooss, Local Bifurcations, Center Manifolds, and Normal Forms in Infinite-Dimensional Dynamical Systems, in: Universitext, Springer-Verlag London, Ltd., EDP Sciences, London, Les Ulis, 2011.

[104] C. Kuehn, A. Neamtu, Rough center manifolds, 2018, pp. 1–48, arXiv: 1811.10037.

[105] G. Simonett, Center manifolds for quasilinear reaction–diffusion systems, Differential Integral Equations 8 (4) (1995) 753–796.

[106] J. Duan, W. Wang, Effective dynamics of stochastic partial differential equations, in: Elsevier Insights, Elsevier, Amsterdam, 2014.

[107] G.A. Gottwald, I. Melbourne, Homogenization for deterministic maps and multiplicative noise, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 469 (2156) (2013) 20130201.

[108] R.Z. Khas'minskii, Principle of averaging for parabolic and elliptic differential equations and for markov processes with small diffusion, Theory Probab. Appl. 8 (1) (1963) 1–21.

[109] I. Melbourne, A.M. Stuart, A note on diffusion limits of chaotic skew-product flows, Nonlinearity 24 (6) (2011) 1361–1367.

[110] G.A. Pavliotis, A.M. Stuart, Multiscale Methods, in: Texts in Applied Mathematics, 53, Springer, New York, 2008, Averaging and homogenization.

[111] L. Arnold, Hasselmann's program revisited: the analysis of stochasticity in deterministic climate models, in: Stochastic Climate Models (Chorin, 1999), in: Progr. Probab, 49, Birkhäuser, Basel, 2001, pp. 141–157.

[112] Y. Kifer, Averaging and climate models, in: Stochastic Climate Models (Chorin, 1999), in: Progr. Probab., 49, Birkhäuser, Basel, 2001, pp. 171–188.

[113] C. Rödenbeck, C. Beck, H. Kantz, Dynamical systems with time scale separation: averaging, stochastic modelling, and central limit theorems, in: Stochastic Climate Models (Chorin, 1999), in: Progr. Probab., 49, Birkhäuser, Basel, 2001, pp. 189–209.

[114] A.J. Majda, C. Franzke, B. Khouider, An applied mathematics perspective on stochastic modelling for climate, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 366 (1875) (2008) 2429–2455.

[115] F. Campillo, M. Kleptsyna, A. Piatnitski, Homogenization of random parabolic operator with large potential, Stochastic Process. Appl. 93 (1) (2001) 57–85.

[116] M.A. Diop, E. Pardoux, Averaging of a parabolic partial differential equation with random evolution, in: Seminar on Stochastic Analysis, Random Fields and Applications IV, in: Progr. Probab., 58, Birkhäuser, Basel, 2004, pp. 111–128.

[117] A.V. Skorokhod, Asymptotic Methods in the Theory of Stochastic Differential Equations, in: Translations of Mathematical Monographs, 78, American Mathematical Society, Providence, RI, 1989, Translated from the Russian by H. H. McFaden.

[118] A.J. Roberts, Normal form transforms separate slow and fast modes in stochastic dynamical systems, Phys. A 387 (1) (2008) 12–38.

[119] S. Cerrai, Averaging principle for systems of reaction–diffusion equations with polynomial nonlinearities perturbed by multiplicative noise, SIAM J. Math. Anal. 43 (6) (2011) 2482–2518.

[120] W. Wang, A.J. Roberts, Average and deviation for slow-fast stochastic partial differential equations, J. Differential Equations 253 (5) (2012) 1265–1286.

[121] S. Klainerman, A. Majda, Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids, Comm. Pure Appl. Math. 34 (4) (1981) 481–524.

[122] S. Schochet, Fast singular limits of hyperbolic pdes, J. Differential Equations 114 (2) (1994) 476–512.

[123] P.F. Embid, A.J. Majda, Averaging over fast gravity waves for geophysical flows with arbitrary potential vorticity, Comm. Partial Differential Equations 21 (3–4) (1996) 619–658.

[124] P.F. Embid, A.J. Majda, Low froude number limiting dynamics for stably stratified flow with small or finite rossby numbers, Geophys. Astrophys. Fluid Dyn. 87 (1–2) (1998) 1–50.

[125] B.A. Wingate, P. Embid, M. Holmes-Cerfon, M.A. Taylor, Low rossby limiting dynamics for stably stratified flow with finite froude number, J. Fluid Mech. 676 (2011) 546–571.

[126] Y. Lu, Homogenization of Stokes equations in perforated domains: a unified approach, J. Math. Fluid Mech. 22 (3) (2020) 13, Paper No.(44).

[127] A. Mikelić, Homogenization of nonstationary Navier-Stokes equations in a domain with a grained boundary, Ann. Mat. Pura Appl. 158 (4) (1991) 167–179.

[128] A. Roberts, J. Bunder, Slowly varying, macroscale models emerge from microscale dynamics over multiscale domains, IMA J. Appl. Math. 82 (5) (2017) 971–1012.

[129] M. Engel, M.A. Gkogkas, C. Kuehn, Homogenization of coupled fast-slow systems via intermediate stochastic regularization, J. Stat. Phys. 183 (2) (2021) 34, Paper No. 25.

[130] D. Kelly, I. Melbourne, Deterministic homogenization for fast-slow systems with chaotic noise, J. Funct. Anal. 272 (10) (2017) 4063–4102.

[131] J. Wouters, G.A. Gottwald, Edgeworth expansions for slow–fast systems with finite time-scale separation, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 475 (2223) (2019) 20180358.

[132] P. Kloeden, M. Rasmussen, Nonautonomous Dynamical Systems, AMS, 2011.

[133] P. Ashwin, S. Wieczorek, R. Vitolo, P. Cox, Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system, Phil. Trans. R. Soc. A 370 (1962) (2012) 1166–1184.

[134] C. Kuehn, I. longo, Estimating rate-induced tipping via asymptotic series and a Melnikov-Like method, 2020, pp. 1–25, arXiv:2011.04031.

[135] H. Alkhayuon, P. Ashwin, L.C. Jackson, C. Quinn, R.A. Wood, Basin bifurcations, oscillatory instability and rate-induced thresholds for atlantic meridional overturning circulation in a global oceanic box model, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 475 (2225) (2019) 20190051.

[136] S. Wieczorek, C. Xie, P. Ashwin, Rate-induced tipping: Thresholds, edge states and connecting orbits, 2021, arXiv preprint arXiv:2111.15497.

[137] H. Mori, Transport, collective motion, and brownian motion, Progr. Theoret. Phys. 33 (3) (1965) 423–455.

[138] R. Zwanzig, Memory effects in irreversible thermodynamics, Phys. Rev. 124 (4) (1961) 983.

[139] R. Zwanzig, Nonlinear generalized Langevin equations, J. Stat. Phys. 9 (3) (1973) 215–220.

[140] A.J. Chorin, O.H. Hald, R. Kupferman, Optimal prediction with memory, Physica D 166 (3–4) (2002) 239–257.

[141] J. Wouters, V. Lucarini, Multi-level dynamical systems: Connecting the ruelle response theory and the Mori-Zwanzig approach, J. Stat. Phys. 151 (5) (2013) 850–860.

[142] S.K. Falkena, C. Quinn, J. Sieber, J. Frank, H.A. Dijkstra, Derivation of delay equation climate models using the mori-zwanzig formalism, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 475 (2227) (2019) 20190075.

[143] S.K. Falkena, C. Quinn, J. Sieber, H.A. Dijkstra, A delay equation model for the atlantic multidecadal oscillation, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 477 (2246) (2021) 20200659.

[144] C. Franzke, A.J. Majda, E. Vanden-Eijnden, Low-order stochastic mode reduction for a realistic barotropic model climate, J. Atmos. Sci. 62 (6) (2005) 1722–1745.

[145] L. Socha, Linearization Methods for Stochastic Dynamic Systems, Springer, 2008.

[146] C. Levermore, Moment closure hierarchies for kinetic theories, J. Stat. Phys. 83 (5) (1996) 1021–1065.

[147] T. Gross, H. Sayama (Eds.), Adaptive Networks: Theory, Models and Applications, Springer, 2009.

[148] C. Kuehn, Moment closure—a brief review, in: Control of Self-Organizing Nonlinear Systems, in: Underst. Complex Syst., Springer, Cham, 2016, pp. 253–271.

[149] B. Nadler, S. Lafon, R. Coifman, I. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Appl. Comput. Harmonic Anal. 21 (1) (2006) 113–127.

[150] J.N. Kutz, Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data, Oxford University Press, 2013.

[151] P.R. Vlachas, W. Byeon, Z.Y. Wan, T.P. Sapsis, P. Koumoutsakos, Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks, Proc. R. Soc. A: Math., Phys. Eng. Sci. 474 (2213) (2018) 20170844.

[152] S. Scher, Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning, Geophys. Res. Lett. 45 (22) (2018) 12–616.

[153] T. Sapsis, Z.Y. Wan, B. Dodov, H. Dijkstra, C. Lessig, Data-assisted reduced-order modeling of climate dynamics, in: Geophysical Research Abstracts, Volume 21, 2019.

[154] N. Berglund, B. Gentz, Noise-Induced Phenomena in Slow-Fast Dynamical Systems, Springer, 2006.

[155] M. Freidlin, A. Wentzell, RandOm Perturbations of Dynamical Systems, Springer, 1998.

[156] Y. Kabanov, S. Pergamenshchikov, Two-Scale Stochastic Systems, Springer, 2003.

[157] P. Boxler, A stochastic version of center manifold theory, Probab. Theory Related Fields 83 (4) (1989) 509–545.

[158] J. Duan, K. Lu, B. Schmalfuss, Invariant manifolds for stochastic partial differential equations, Ann. Probab. 31 (4) (2003) 2109–2135.

[159] C. Kuehn, A. Neamtu, Rough center manifolds, SIAM J. Math. Anal. 53 (4) (2021) 3912–3957.

[160] B. Schmalfuss, K. Schneider, Invariant manifolds for random dynamical systems with slow and fast variables, J. Dynam. Differential Equations 20 (1) (2008) 133–164.

[161] Z. Schuss, Theory and Applications of Stochastic Processes: An Analytical Approach, Springer, 2009.

[162] M. Chekroun, H. Liu, S. Wang, Non-Markovian reduced systems for stochastic partial differential equations: The additive noise case, 2013, pp. 1–11, arXiv:1311.3069.

[163] M. Chekroun, H. Liu, S. Wang, On stochastic parameterizing manifolds: Pullback characterization and Non-Markovian reduced equations, 2013, pp. 1–143, arXiv:1310.3896.

[164] D. Blömker, Amplitude Equations for Stochastic Partial Differential Equations, World Scientific, 2007.

[165] H.A. Dijkstra, A. Tantet, J. Viebahn, E. Mulder, M. Hebbink, D. Castellana, H. van den Pol, J. Frank, S. Baars, F. Wubs, M. Chekroun, C. Kuehn, A numerical framework to understand transitions in high-dimensional stochastic dynamical systems, Dyn. Statist. Clim. Syst. 1 (1) (2016).

[166] F. Bouchet, A. Venaille, Statistical mechanics of two-dimensional and geophysical flows, Phys. Rep. 515 (5) (2012) 227–295.

[167] V.M. Gálfi, V. Lucarini, J. Wouters, A large deviation theory-based analysis of heat waves and cold spells in a simplified model of the general circulation of the atmosphere, J. Stat. Mech. Theory Exp. 2019 (3) (2019) 033404.

[168] V.M. Galfi, V. Lucarini, Fingerprinting heatwaves and cold spells and assessing their response to climate change using large deviation theory, Phys. Rev. Lett. 127 (5) (2021) 058701.

[169] G. Margazoglou, T. Grafke, A. Laio, V. Lucarini, Dynamical landscape and multistability of a climate model, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 477 (2250) (2021) 20210019.

[170] V. Lucarini, T. Bódai, Edge states in the climate system: exploring global instabilities and critical transitions, Nonlinearity 30 (7) (2017) R32.

[171] A.J. Majda, I. Timofeyev, E.V. Eijnden, Models for stochastic climate prediction, Proc. Natl. Acad. Sci. 96 (26) (1999) 14687–14691.

[172] A.J. Chorin, A.P. Kast, R. Kupferman, Optimal prediction of underresolved dynamics, Proc. Natl. Acad. Sci. 95 (8) (1998) 4094–4098.

[173] T. Palmer, G. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, M. Leutbecher, Representing model uncertainty in weather and climate prediction, Annu. Rev. Earth Planet. Sci. 33 (2005) 163–193.

[174] F. Kwasniok, Data-based stochastic subgrid-scale parametrization: an approach using cluster-weighted modelling, Phil. Trans. R. Soc. A 370 (1962) (2012) 1061–1086.

[175] R.R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (1) (2006) 5–30.

[176] C.J. Dsilva, R. Talmon, C.W. Gear, R.R. Coifman, I.G. Kevrekidis, Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems, SIAM J. Appl. Dyn. Syst. 15 (3) (2016) 1327–1351.

[177] A. Singer, From graph to manifold Laplacian: the convergence rate, Appl. Comput. Harmon. Anal. 21 (1) (2006) 128–134.