

**Numerical study of an energy consistent
stochastic parameterization for climate
models**

Dissertation

with the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Earth Sciences
at Universität Hamburg

submitted by

Federica Gugole

Hamburg, 2019

Accepted as Dissertation at the Department of Earth Sciences

Day of doctoral defense: 01.11.2019

Reviewers: PD Dr. Christian Franzke
Jun.-Prof. Dr. Gualtiero Badin

Chair of the Subject Doctoral Committee: Prof. Dr. Dirk Gajewski

Dean of Faculty of MIN: Prof. Dr. Heinrich Graener

Abstract

Numerical models are essential tools to study the underlying drivers of the climate, although many errors are currently present in realistic models, due to the lack of knowledge of some physical mechanisms and the limitations imposed by the numerics. Since it is not always possible to determine the relations between cause and effect, it is useful to use simplified models, for which the restrictions due to computational time are much less influential and the dynamics is better understood. In this thesis I dealt with the issue of the numerical truncation of the resolved scales. Whenever the size of the numerical grid box and of the numerical time step are set, phenomena evolving in a smaller physical space or in a shorter time window cannot be reproduced by the numerics. These are called sub-grid processes. Since the various atmospheric phenomena interact nonlinearly with each other, the position of the numerical cut strongly influences the model outcomes. In particular the energy of the small unresolved scales can no longer be backscattered into the large resolved scales and there is a depletion of energy on the resolved modes. In order to reduce the error due to the numerical truncation, parameterizations representing the sub-grid scales are being introduced into the models.

I developed an energy conserving stochastic parameterization in the framework of the 2-layer Quasi-Geostrophic (QG) model. To some extent this can be considered as a realistic model of the atmosphere in the mid-latitudes, but it does not demand prohibitive amounts of computational time. Hence it is a perfect test bed to analyze how to define and structure the parameterization in order to include the effect of the sub-grid processes and reduce the numerical error. In particular I ascertained that the projection operator approach outlined by Frank and Gottwald (2013) can be applied also in case of high dimensional systems. Furthermore the importance of the noise covariance structure within a Eulerian framework has been emphasized. In this setup it is essential to define a dynamically consistent spatial correlation structure for the sub-grid processes in order to preserve important statistical properties. For the definition of such a structure I used tools from statistical analysis and dynamical systems theory. The former detects patterns with long-lasting validity, while the modes identified by the latter need to be regularly updated. The comparison between the two revealed that the use of climatic patterns has larger uncertainties, and the individual ensemble members depart from each other. On the other hand a dynamically adapted noise covariance is able to keep

track of the large scale movements while introducing the effect of the sub-grid scales. The main outcome of this thesis is a new method for the construction of the noise covariance, with improved performances with respect to the currently most adopted method in the literature.

Zusammenfassung

Numerische Modelle sind unerlässliche Instrumente für die Untersuchung des Klimasystems, obwohl sogar realistische Klimamodellen immer noch viele Fehler beinhalten, einerseits wegen mangelnder Kenntnisse bestimmter physikalischer Mechanismen und andererseits wegen numerischer Begrenzungen. Da die Verbindung zwischen Ursache und Wirkung nicht immer bestimmt werden kann, ist es nützlich, vereinfachte Modelle zu verwenden. Solche Modelle werden durch die Begrenzungen aufgrund der Rechenzeit viel weniger beeinflusst und die Dynamik des simulierten Systems kann besser verstanden werden. In meiner Doktorarbeit beschäftigte ich mich mit dem Thema der numerischen Auflösung der simulierten Skalen. Wenn die Größe des numerischen Gitters und des Zeitschritts festgelegt sind, können Phänomene, die sich in einem kleineren physikalischen Raum oder in einem kürzeren Zeitfenster entwickeln, nicht durch die Numerik reproduziert werden. Diese Prozesse werden als subskalige Prozesse bezeichnet. Die numerische Auflösung beeinflusst die Modellergebnisse stark, weil die verschiedenen atmosphärischen Phänomene nichtlinear miteinander interagieren. Insbesondere kann die Energie der kleinen unaufgelösten Skalen nicht mehr in die großen aufgelösten Skalen zurückgestreut werden, und ein Teil der Energie der aufgelösten Skalen verschwindet. Um die Ungenauigkeit aufgrund der numerischen Auflösung zu verringern, werden die subskaligen Prozesse in den Modellen parametrisiert.

Ich habe eine energieerhaltende stochastische Parametrisierung für ein quasi-geostrophisches (QG) Zwei-Schichten-Modell entwickelt. Dieses Modell kann zu einem gewissen Grad als realistisches Atmosphärenmodell für die mittleren Breiten betrachtet werden, erfordert jedoch keine unerschwinglichen Mengen an Rechenzeit. Daher ist es ein perfektes Instrument, um die Struktur der Parametrisierung zu überprüfen, um den Effekt der subskaligen Prozesse wiederzugeben, um die numerische Ungenauigkeit zu verringern. Insbesondere habe ich festgestellt, dass der von Frank and Gottwald (2013) beschriebene, auf einen Projektionsoperator basierende, Ansatz auch bei hochdimensionalen Systemen angewendet werden kann. Darüber hinaus wurde die Bedeutung der Struktur der Rauschkovarianz aufgrund der Eulerschen Betrachtung hervorgehoben. In dieser Anordnung ist es wichtig, eine dynamisch konsistente räumliche Korrelationsstruktur für die subskaligen Prozesse zu definieren, um wichtige statistische Eigenschaften zu erhalten. Für die Definition einer solchen Struktur verwendete ich Methoden der statistischen Analyse und der Theorie der dynamischen Systeme. Während die Erste für die Identi-

fizierung langlebiger Muster verwendet werden kann, die Moden der Zweiten müssen regelmäßig aktualisiert werden. Der Vergleich der beiden Methoden ergab einerseits, dass die Verwendung von klimatischen Mustern größere Fehler zeigt und die einzelnen Ensemblemitglieder voneinander abweichen. Andererseits kann eine dynamisch angepasste Rauschkovarianz sowohl die großskaligen Bewegungen als auch den Effekt der subskaligen Prozesse wiedergeben.

Das Hauptergebnis dieser Dissertation ist eine neue Methode für die Entwicklung der Rauschkovarianz, die eine verbesserte Leistung im Vergleich zu der derzeit am häufigsten verwendeten Methode in der Literatur zeigt.

Author's contribution

The topic of this PhD project was proposed by my supervisor PD Dr. Christian L. E. Franzke. In agreement with him, I decided to focus and investigate the numerical aspects of energy conserving stochastic parameterizations for climate models. In particular, I wrote the numerical code for the 2-layer QG models here employed, performed the analyses and contributed to elaborate the concepts of the papers, and to the write-up. Dr. C. L. E. Franzke cooperated in writing and developing the concepts of both papers. Paper PII is a collaboration with Prof. G. A. Gottwald, who took part in the concept development and write-up.

List of papers

- PI: **Gugole, F.** and Franzke, C. L. E. (2019), Numerical Development and Evaluation of an Energy Conserving Conceptual Stochastic Climate Model, *Mathematics for Climate and Weather Forecasting*, 5:45-64.
- PII: **Gugole, F.**, Gottwald, G. A. and Franzke, C. L. E. (2019), Spatial covariance modeling for stochastic sub-grid scales parameterizations, *manuscript*.

The following paper is not part of this thesis:

- Gottwald, G. A. and **Gugole, F.** (2019), Detecting regime transitions in time series using dynamic mode decomposition, *submitted to Journal of Statistical Physics*.

Contents

Abstract	i
Zusammenfassung	iii
Author's contribution	v
List of papers	v
List of acronyms	ix
1 Introduction	1
1.1 Numerical modeling	3
1.2 The governing equations and the CFL condition	4
1.3 Parameterizations	6
1.4 Stochastic differential equations	8
2 Summary of papers	11
3 Paper I	13
4 Paper II	35
5 Conclusions	59
6 Outlook	63
Acknowledgments	i
References	iii

List of acronyms

CFL Courant-Friedrichs-Lewy

DMD Dynamic Mode Decomposition

EOF Empirical Orthogonal Function

GCM Global Circulation Model

iid independent and identically distributed

PDE Partial Differential Equation

PI Paper I

PII Paper II

SDE Stochastic Differential Equation

QG Quasi-Geostrophic

1 Introduction

The Earth climate system is driven by the solar radiation. To prevent the temperature of the planet from continuously increasing, the net solar energy absorbed by the atmosphere and the earth must equal in the mean the infrared energy radiated back to space by the planet. Although the annual averaged solar heating is strongly dependent on the latitude, with a maximum at the equator and minima at the poles, the latitude dependency of the outgoing infrared radiation is very weak. As a consequence there is a differential heating due to a radiation surplus at the equator and a radiation deficit in the polar region. This differential heating warms the equatorial atmosphere relative to higher latitudes and creates a pole-to-equator temperature gradient. The latter produces available potential energy, and enables processes, e.g. winds and convection, with the aim of transporting heat poleward, in order to balance the radiation deficit until the pole-to-equator gradient ceases to exist. At the same time, these phenomena convert potential energy into kinetic energy, thereby maintaining the kinetic energy of the atmosphere against the effects of frictional dissipation Holton and Hakim (2012).

This is a basic mechanism that every climate model, also known as general circulation model (GCM), should resolve in order to correctly reproduce also the other dynamical mechanisms. Computationally this is a great challenge. As a matter of fact the discrete equivalent of a continuous system does not preserve, in general, the same properties. This means that while the continuous system is energy conserving, this might not be the case for its discrete counterpart. Arakawa (1966) provides an illustrative example for the case of the Jacobian operator and its discrete formulations. Therefore it is crucial for the models to correctly get the energy budget, i.e. the net flow of energy into and out of the Earth system, particularly in the context of global warming and climate change. Moreover whenever a continuous multi-scale system, as the atmosphere and the oceans (see Figure 1.1 for a scheme of the different spatial and temporal ranges covered by phenomena contributing to the atmospheric dynamics), is translated into a discrete numerical model, a truncation due to the size of the spatial grid and of the time step is introduced. Hence any dynamical mechanism developing in a physical space smaller than the chosen grid box, or in a time frame smaller than the selected time step, is not captured by the numerical model. This induces errors into the model dynamics, which grow with the size of the spatial grid and of the time step. Because of these and other issues, GCMs are currently not able to reproduce correctly all the different dynamics

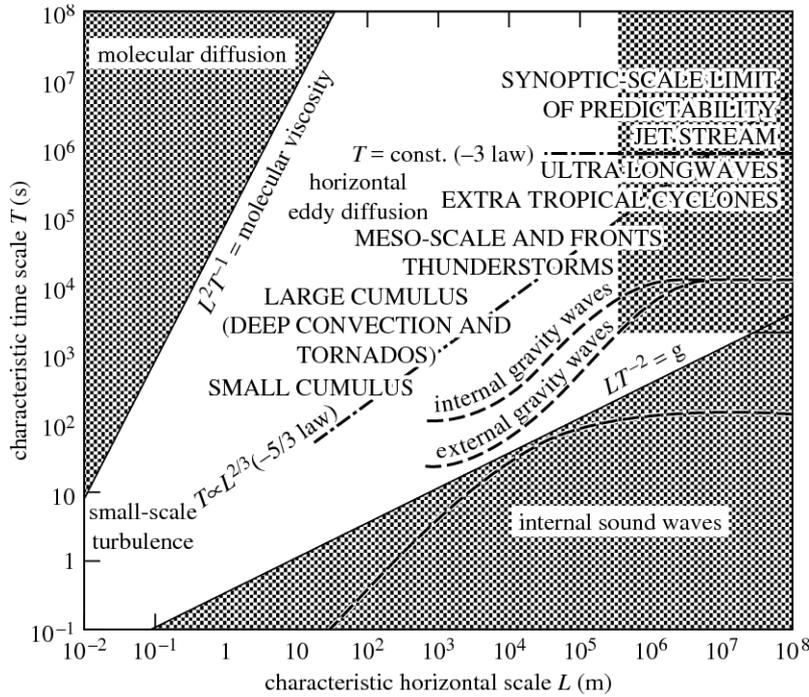


Figure 1.1: Spatial and temporal scales of typical atmospheric processes. Figure taken from Cullen and Brown (2009).

involved in these complex systems.

One way to improve the numerical models is to include the effect of the sub-grid scales, i.e. those processes not resolved by the model due to the numerical truncation, on the resolved scales. It is generally assumed that the former develop much faster than the latter, see also Figure 1.1. Hence their dynamics at $t = t_n + \Delta t$, where t_n denotes an instant in time and Δt is the time step of the numerical model, is uncorrelated with respect to their state at $t = t_n$. This is a fundamental assumption that allows them to be represented by means of stochastic processes, as first suggested by Hasselmann (1976).

In my thesis I further extend and analyze the method proposed by Frank and Gottwald (2013) to parametrize the sub-grid processes through a stochastic formulation which conserves the total energy of the stochastic system. They validated their formulation on a 4-dimensional toy model of the atmosphere, but it is unclear whether this formulation could be implemented also in case of more complex models. In particular I address the following research questions:

1. can the projection operator approach, outlined by Frank and Gottwald (2013), be employed also in case of high dimensional systems?
2. how should the noise covariance be defined such that it is dynamically consistent and representative of the sub-grid processes?

3. Vannitsem (2017) showed that, in case of spectral models, the error dynamics is dependent on the scale where it is introduced; is it possible to choose the noise scale also in case of grid point models?
4. how much do a priori assumptions on the noise covariance affect the outcomes of the numerical model?

The first two research questions are addressed in PI, while the last two are answered in PII. Furthermore PII proposes an alternative method for the noise covariance definition and compares it to the most often used procedure in the context of stochastic parameterizations for climate models. Before giving a summary about the setups and the main findings of the two papers, I briefly motivate why these questions need to be addressed. In particular I will discuss from the point of view of numerical analysis and modeling.

1.1 Numerical modeling

Numerical models can be divided into two main categories: grid point and spectral models. In what follows I give a short overview over the basic aspects and differences between these two approaches. More details about the underlying functional analysis as well as numerical analysis can be found, for instance, in Quarteroni (2017).

Let us consider a general partial differential equation (PDE) in the $d + 1$ independent variables $\mathbf{x} = (x_1, \dots, x_d)^\top$ and t

$$\mathcal{P}(u, g) = \left(\mathbf{x}, t, u, \frac{\partial u}{\partial t}, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_d}, \dots, \frac{\partial^{p_1 + \dots + p_d + p_t} u}{\partial x_1^{p_1} \dots \partial x_d^{p_d} \partial t^{p_t}}, g \right) = 0, \quad (1.1)$$

where u denotes the unknown function in the space variable \mathbf{x} and in the time variable t , g is the set of data on which the PDE depends, and $p_1, \dots, p_d, p_t \in \mathbb{N}$. In order to convert equation (1.1) into a numerical model, initial and boundary conditions must be defined. Once this is done, the spatial domain can be approximated by means of a grid, and the modeler can decide to employ either a grid point based or a spectral discretization of the continuous equations.

In case of a grid point model the solution u is represented by its value at the grid points. The system is solved for each point of the grid and derivatives are computed through an approximation by means of Taylor expansions of

$$\begin{aligned} \frac{\partial u}{\partial t} &= \lim_{h \rightarrow 0} \frac{u(t+h) - u(t)}{h}, \\ \frac{\partial u}{\partial x_i} &= \lim_{h \rightarrow 0} \frac{u(x_i+h) - u(x_i)}{h} \quad i \in \{1, \dots, d\}, \end{aligned}$$

given that the regularity conditions are satisfied. In the numerical model h is strictly bigger than zero and it would represent either the distance between two points in the i^{th} coordinate, or the time step in case of the time derivative. This determines a possible source of errors in the model, which can be reduced either by considering more grid points and higher order terms in the Taylor expansions, or by decreasing h and hence increasing the total number of grid points and the amount of computations.

Spectral models instead use a set of basis functions in \mathbb{R}^d , e.g. spectral harmonics, spanning the phase space, i.e. a space where all possible solutions are represented Strogatz (1994), and the field \mathbf{u} is written as a linear combination of these basis functions. Spatial derivatives are easy to compute since the derivatives of the basis functions are known, while temporal derivatives are computed as above. The model resolution is increased by considering a larger set of basis functions. In this framework the computations are done both in phase space and on the grid domain, since local processes must be computed in the real space. Hence a lot of computational time is required to transform the variables of interest from one space to the other. This is a significant limitation in case of multi-scale systems, like the ocean and the atmosphere, with many local processes.

Given the huge variety of spatial and temporal resolutions of the numerous phenomena interacting one with each other in the ocean or in the atmosphere (see Figure 1.1), most GCMs are grid point based. Despite the difficulties in approximating the spatial derivatives, this framework allows an easy inclusion of local processes, and hence of most physical parameterizations.

1.2 The governing equations and the CFL condition

The water in the oceans and the air in the atmosphere are viscous fluids, whose motion in a rotating framework is described by the Navier-Stokes equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (1.2a)$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\frac{\nabla p}{\rho} + \nu \nabla^2 \mathbf{u} \quad (1.2b)$$

where t stands for time, ρ for the fluid density, \mathbf{u} is the three-dimensional fluid velocity, $\boldsymbol{\Omega}$ the three-dimensional angular velocity, p represents pressure and ν the kinematic viscosity. ∇ and ∇^2 denote the gradient and the Laplacian operators respectively. Equation (1.2a) is derived by imposing the conservation of mass, which is by definition conserved in time; while equation (1.2b) describes the conservation of momentum, in analogy with the second law of Newton Vallis (2006); Badin and Crisciani (2018). Together with an equation for the energy of the system, equations 1.2 are often referred as primitive

1 Introduction

equations. If we focus on equation (1.2b), we can see that it is a set of three non-linear hyperbolic balance laws, whose non-linearity is due to the advection term $\mathbf{u} \cdot \nabla \mathbf{u}$.

Many simplifying assumptions can be formulated and applied to the Navier-Stokes equations. For instance in case of water it can be assumed that its density does not change, i.e. it is incompressible, hence equation (1.2a) simplifies into

$$\nabla \cdot \mathbf{u} = 0 ,$$

or, in case of a static fluid or with small vertical accelerations compared to gravity, the momentum equation for the vertical coordinate can be approximated by the hydrostatic balance relation:

$$\frac{\partial p}{\partial z} = -\rho g ,$$

where z denotes the vertical coordinate and g the gravitational force. Scaling analysis on the magnitude of the physical quantities involved in system (1.2) can also be performed leading, for example, to the planetary geostrophic equations. In this framework the Rossby number

$$\text{Ro} = \frac{U}{fL}$$

(with U typical velocity scale, L typical horizontal length and f the Coriolis force) and the deformation radius

$$L_d = \frac{\sqrt{gH}}{f}$$

(with H typical vertical extent) are defined. When considering scales much larger than the deformation radius, the time derivative and the advection terms are order Rossby number smaller ($\text{Ro} \approx 0.1$ for the atmosphere and $\text{Ro} \approx 0.01$ for the oceans) than the Coriolis and pressure terms, hence they can be neglected Vallis (2006). Nonetheless, aside from this extremely simplified case, the time derivative and the advection terms cannot be excluded and hence a set of hyperbolic equations has to be solved.

When solving hyperbolic partial differential equations, the numerical analysis of the time integration schemes reveals a constraint, that the time step needs to satisfy for the numerical scheme to converge to the exact solution. This constraint is called Courant-Friedrichs-Lewy (CFL) condition, after the authors who first described it Courant et al. (1928). In case of one-dimensional dynamics, the CFL condition reads

$$C = \frac{u\Delta t}{\Delta x} \leq C_{\max} \quad (1.3)$$

where C is a dimensionless quantity called the CFL (or simply Courant) number, u is the magnitude of the one-dimensional velocity, Δt indicates the time step and Δx the length interval. C_{\max} is an upper bound which varies according to the solver considered, in particular if it is explicit or implicit. In case of complex models explicit schemes

are generally preferred, since implicit methods require to calculate the inverses of large matrices (operations which may not be well posed and might require lots of computations). For explicit schemes, we usually have $C_{\max} = 1$. It is easy to see that in case of global circulation models, in order to consider a wider and wider range of scales, and hence resolve faster and faster processes, prohibitive small values for Δt and Δx are required. Hence the numerical model would need excessive amounts of time for the computations, and it would not be useful for practical purposes.

1.3 Parameterizations

Due to its much smaller vertical extent compared to its horizontal area, the atmosphere can be considered in the limit as a two-dimensional system. In case of two dimensional turbulence enstrophy transfers from larger to smaller scales, until it may be dissipated at the eddy level, while energy propagates from smaller to larger scales Vallis (2006). In Figure 1.2 a scheme of this mechanism is reported. As briefly mentioned above, in any

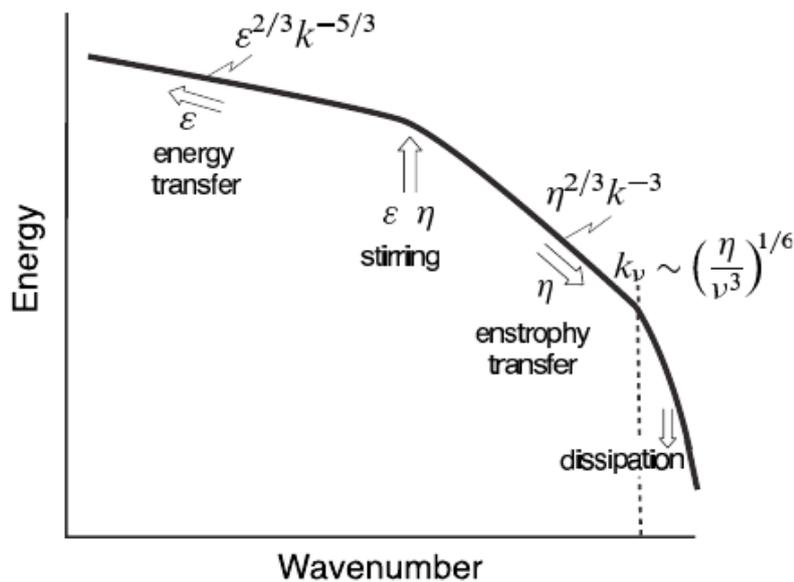


Figure 1.2: Plot of the energy and enstrophy transfer for two-dimensional turbulence. The energy supplied at some rate ϵ is transferred to large scales, whereas the enstrophy supplied at some rate η is transferred to small scales, where it may be dissipated by viscosity. Figure taken from Vallis (2006).

GCM there is a numerical truncation to the scale of resolved phenomena. This truncation interrupts the energy and enstrophy cascades, leading to a pile up of enstrophy at the truncation level and preventing the energy to transfer back into the large scales. In particular the growth of enstrophy at the truncation scale makes the model unstable and more likely to blow up in finite time. As a consequence enstrophy has to be dissi-

1 Introduction

pated also at larger scales when making use of coarse resolutions. This leads also to an increased dissipation of energy Danilov et al. (2019). Therefore, due to the limited computational capabilities, it is of fundamental importance to model the sub-grid processes in an energetically consistent fashion.

As briefly mentioned above, stochastic processes are a useful tool to parameterize, i.e. represent in terms of the resolved scales, the sub-grid processes. Thanks to the difference in the temporal evolution between the slow resolved processes and the fast unresolved scales, the latter can be regarded as noise with respect to the former. In fact the time step in numerical models is tailored to resolve the slow modes, while in the same time interval the fast sub-grid processes fully evolve and decorrelate; see Figure 1.3. There

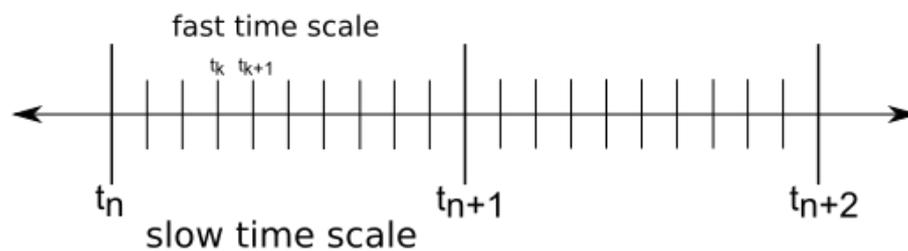


Figure 1.3: Scheme of the different evolution times for slow and fast phenomena. The evolution time scale of the small scales is much faster than the one of the large scales, hence in one time step of the large scales, there are many smaller time steps for the fast scales.

are several advantages in using stochastic parameterizations; the most important are: gain in computational time compared to simulations resolved on a finer grid, reduction of model errors, and systematic representation of uncertainties and model errors Palmer et al. (2009); Berner et al. (2017). On the other hand, the introduction of extra terms into the equations of motions might break the symmetries and the conservation laws of the system, and hence also of the numerical model.

Since the 1970s much research on the topic has been carried out. Examples of some of the most important outcomes are given by the works of Mémin (2014) and Holm (2015), who derived stochastic formulations of the Navier-Stokes equations conserving either the kinetic energy or the Kelvin circulation theorem, respectively. Although these works have great potential, they also suffer from the numerical restrictions. Moreover in order to include them in current GCMs, the models dynamical core might have to be rewritten. Hence detailed analyses on their pros and cons are required. Other parameterizations with a more straightforward numerical implementation, also in case of more complex models, and with different purposes have been developed too. The range of phenomena that need to be parametrized in the global circulation models is very wide and ranges from scales proportional to the size of the model grid box to processes occurring at the molecular level. To give a few examples, some of the parametrized phenomena

are the representation of eddies, e.g. Porta Mana and Zanna (2014); Berloff (2015), the re-injection of energy into the large scale, e.g. Jansen and Held (2014), and the description of clouds formation and aggregation, e.g. Dorrestijn et al. (2016). Most of these parameterizations are either data driven or they look for an analytical expression describing the dynamics of interest Gottwald et al. (2017). The parameterization I used has been analytically derived such that the total energy of the system is conserved, but the definition of the noise covariance is data driven, i.e. data analysis techniques have been used to detect and approximate the main modes of motion, which have in turn been used to construct the noise covariance.

1.4 Stochastic differential equations

Here I briefly introduce a few basic notions about stochastic differential equations (SDEs) and their integral interpretation. More details can be found, for instance, in Pavliotis and Stuart (2008); Gardiner (2009).

Before introducing the integration rules for SDE, I state here the definition of a Brownian motion, also known as Wiener process:

A 1-dimensional standard *Brownian motion* $W(t) : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a real-valued stochastic process with the following properties:

- $W(0) = 0$;
- $W(t)$ is continuous;
- $W(t)$ has increments $W(t) - W(s)$ that are independent on non-overlapping intervals. Furthermore, for every $t > s \geq 0$ $W(t) - W(s)$ has a Gaussian distribution with mean 0 and variance $t - s$;
- a d -dimensional standard Brownian motion $W(t) : \mathbb{R}^+ \rightarrow \mathbb{R}^d$ is a collection of d independent 1-dimensional Brownian motions.

I wish to remark that, intuitively, since the variance of a Wiener process scales with time t , dW scales approximately as \sqrt{dt} ; see Gardiner (2009) for a rigorous proof.

SDEs can be interpreted by use of two integral forms: Itô and Stratonovich. Let us consider first the Itô SDE for $z(t) : \mathbb{R}^+ \rightarrow \mathcal{Z} \subseteq \mathbb{R}^d$

$$\frac{dz}{dt} = \gamma(z) + \sigma(z) \frac{dW}{dt}, \quad z(0) = z_0, \quad (1.4)$$

with $\gamma : \mathcal{Z} \rightarrow \mathbb{R}^d$ a smooth vector valued function, $W(t)$ a standard m -dimensional Brownian motion, and $\sigma : \mathcal{Z} \rightarrow \mathbb{R}^{d \times m}$ a smooth matrix-valued function. The function γ can be referred as the *drift* and σ as the *diffusion coefficient*. The term dW/dt is thought to

1 Introduction

represent Gaussian white noise. The latter can informally be thought of as a Gaussian process with zero mean and correlation $\delta(t-s)\mathbb{1}$ (with $\delta(0) = 1$ and $\delta(t) = 0$ otherwise, and $\mathbb{1}$ identity function). Such a process exists only as a distribution, thus the interpretation of (1.4) is as an integral equation for $z(t)$

$$z(t) = z_0 + \int_0^t \gamma(z(s))ds + \int_0^t \sigma(z(s))dW(s) . \quad (1.5)$$

For (1.5) to have a meaning, we need to define the stochastic integral against $dW(s)$. The Itô interpretation of

$$I(t) = \int_0^t f(s)dW(s)$$

is defined as

$$I(t) := \lim_{K \rightarrow \infty} \sum_{k=1}^{K-1} f(t_{k-1}) (W(t_k) - W(t_{k-1})) , \quad (1.6)$$

where $t_k = k\Delta t$ and $K\Delta t = t$. The Stratonovich integral instead is determined by

$$I_{\text{strat}}(t) := \lim_{K \rightarrow \infty} \sum_{k=1}^{K-1} \frac{1}{2} (f(t_{k-1}) + f(t_k)) (W(t_k) - W(t_{k-1})) , \quad (1.7)$$

with t_k and $K\Delta t$ as above. Note that in case of Itô's integral the function f is evaluated only at the left end of each interval $[t_{k-1}, t_k]$, which satisfies the martingale property, i.e. given all prior events the expectation value of future stochastic events equals the present value. In contrast, the Stratonovich approach evaluates f at both endpoints, which abandons the martingale property and maintains the usual rules of calculus Moon and Wettlaufer (2014). When using the Stratonovich interpretation, the resulting integral form is written as

$$I_{\text{strat}}(t) = \int_0^t f(s) \circ dW(s) .$$

Since a Brownian motion does not have bounded variations, the limits (1.6) and (1.7) differ. However when f and W are correlated through an SDE, as in (1.4), a formula exists to convert between them. It can be shown that the solution of the Stratonovich SDE

$$\frac{dz}{dt} = \gamma(z) + \sigma(z) \circ \frac{dW}{dt} , \quad z(0) = z_0 ,$$

satisfies also the following Itô SDE:

$$\frac{dz}{dt} = \gamma(z) + \frac{1}{2} \nabla \cdot (\sigma(z)\sigma(z)^T) - \frac{1}{2} \sigma(z) \nabla \cdot (\sigma(z)^T) + \sigma(z) \frac{dW}{dt} , \quad z(0) = z_0 .$$

In general, it is easier to characterize the class of integrands for which the Itô integral can be defined. Hence it is also the most often used in applications. Lastly, under some regularity conditions on the the drift and the diffusion coefficients, and in case of an

initial condition z_0 independent of the Brownian motion, it can be proved that equation (1.4) has a unique solution $z(t)$ with

$$\mathbb{E} \left[\int_0^T |z(t)|^2 dt \right] < \infty \quad \forall T < \infty ,$$

where \mathbb{E} denotes the expectation operator.

2 Summary of papers

Paper I

An important aspect of a parameterization is its applicability. In fact, parameterizations giving excellent results but only in the cases of few simplified models might not be as useful as other parameterizations with less good outcomes but much wider applicability. Following this line of thought, the objective in PI is to extend the projection operator approach outlined in Frank and Gottwald (2013) to high dimensional system and analyze the issues related to its numerical implementation. The equations of a stochastic energy conserving 2-layer Quasi-Geostrophic model with no external forcing and damping have been derived and turned into a grid-point based numerical model. The results show that the choice of using a Eulerian, instead of Lagrangian, description affects also the assumptions that could be done on the noise covariance matrix. As a matter of fact in this setup the use of independent and identically distributed noise leads to unphysical results, while the adoption of a spatial correlation structure, determined by means of Empirical Orthogonal Functions (EOFs), retains important statistical properties and improves the eddy length. Finally the total energy of the system is conserved with fluctuations of less than 1% of the total energy, and this level of accuracy is not affected by the introduction of the noise.

Paper II

As a consequence of the results of PI, the focus of PII is on the influence of the definition of the noise covariance matrix on the dynamics of the system. This aspect is here further investigated by analyzing the outcomes of the forced and damped 2-layer QG model when either EOF or Dynamic Mode Decomposition (DMD) are used to determine the noise covariance. While EOF focuses on the variance of the field, DMD is a data driven procedure for the approximation of the Koopman operator, which encodes the dynamics at a specific time. The main practical difference is that EOF looks at the statistics of the dynamics, and hence requires very long time series and finds statistical climatic patterns. On the other hand DMD considers much shorter time intervals, hence its modes do not have long lasting validity and they should be periodically recomputed. The patterns defined by the two techniques are similar, but also significantly different.

The comparison revealed that DMD is able to follow the meridional shift of the jet, while the use of the EOF eigenvectors for the definition of the noise covariance does not have as accurate performances. Moreover the uncertainties grow faster in case of a constant covariance stencil hence, particularly in case of long run simulations, dynamically adapted noise covariances should be adopted.

3 Paper I

This chapter consists of the paper *Numerical Development and Evaluation of an Energy Conserving Conceptual Stochastic Climate Model*, by myself and my supervisor, published on the journal *Mathematics for Climate and Weather Forecasting*. The journal is open access and the paper can be found at the following link:

<https://www.degruyter.com/view/j/mcwf.2019.5.issue-1/mcwf-2019-0004/mcwf-2019-0004.xml>



Research Article

Open Access

Federica Gugole* and Christian L. E. Franzke

Numerical Development and Evaluation of an Energy Conserving Conceptual Stochastic Climate Model

<https://doi.org/10.1515/mcwf-2019-0004>

Received November 8, 2018; accepted April 18, 2019

Abstract: In this study we aim to present the successful development of an energy conserving conceptual stochastic climate model based on the inviscid 2-layer Quasi-Geostrophic (QG) equations. The stochastic terms have been systematically derived and introduced in such a way that the total energy is conserved. In this proof of concept study we give particular emphasis to the numerical aspects of energy conservation in a high-dimensional complex stochastic system and we analyze what kind of assumptions regarding the noise should be considered in order to obtain physical meaningful results. Our results show that the stochastic model conserves energy to an accuracy of about 0.5% of the total energy; this level of accuracy is not affected by the introduction of the noise, but is mainly due to the level of accuracy of the deterministic discretization of the QG model. Furthermore, our results demonstrate that spatially correlated noise is necessary for the conservation of energy and the preservation of important statistical properties, while using spatially uncorrelated noise violates energy conservation and gives unphysical results. A dynamically consistent spatial covariance structure is determined through Empirical Orthogonal Functions (EOFs). We find that only a small number of EOFs is needed to get good results with respect to energy conservation, autocorrelation functions, PDFs and eddy length scale when comparing a deterministic control simulation on a 512×512 grid to a stochastic simulation on a 128×128 grid. Our stochastic approach has the potential to seamlessly be implemented in comprehensive weather and climate prediction models.

Keywords: stochastic parameterization, energy conservation, projection operator, spatial noise structure, Empirical Orthogonal Functions

MSC: 65C20, 68U20

1 Introduction

The dynamics of the atmosphere and the oceans are by nature complex. Processes with different time and length scales interact with each other affecting the system as a whole. While climate and ocean models have considerably improved over the last few decades, we still cannot resolve all important physical scales and processes, see for instance [20, 6, 22]. The discretization of the continuous governing equations of motion is limited by the model resolution, which determines the size of the smallest resolvable scale. Despite the continued increase of computer power and, thus, of resolution, there are still many important processes in the atmosphere and in the oceans that cannot be explicitly resolved. These include turbulent motions with scales ranging from a few centimeters to the size of the model grid box, as well as processes that occur at a molecular

*Corresponding Author: **Federica Gugole:** Meteorological Institute and Center for Earth System Research and Sustainability, University of Hamburg, Hamburg, Germany; E-mail: federica.gugole@uni-hamburg.de

Christian L. E. Franzke: Meteorological Institute and Center for Earth System Research and Sustainability, University of Hamburg, Hamburg, Germany

scale, like condensation and evaporation. Any numerical forecaster or modeller has to make a decision, based on the targeted objectives, regarding the spatial and temporal scales to resolve. As a consequence of this decision, each numerical scheme inevitably fails to resolve subgrid-scale processes.

These unresolved processes and scales cause many of the observed differences between models and observations. In order to represent these unresolved processes, so-called parameterizations are necessary which take into account the influence that the unresolved have on the resolved processes, if they would be resolved in high-resolution simulations [20]. Most parameterizations, however, are damping and do not take account of the energy and momentum fluxes from the unresolved to the resolved scales [34, 57]. This is a likely source of many of the observed biases in climate and ocean models. Without the added dissipation, however, energy and enstrophy would accumulate at the truncation scale and lead to a blow up in finite time. Hence, it is of fundamental importance to find systematic ways to parameterize the unresolved scales and processes of models, and to improve the model performance and reduce the model biases at coarser resolutions.

As suggested already in 1976 by Hasselmann [31], fast waves can be considered as noise with respect to the slowly evolving large-scale modes and, therefore, can be parameterized by stochastic processes [23]. Hence, to ameliorate this problem of too large damping, stochastic parameterization schemes have been developed (see recent reviews [20, 6, 22, 27]). There are several advantages in using stochastic parameterizations; the most important are: gain in computational time compared to higher resolved simulations; reduction of model errors and systematic representation of uncertainties and model errors. Most operational stochastic parameterization schemes are rather *ad hoc* developments [7] and do not conserve energy or momentum. As a consequence, current schemes have the disadvantage of the forfeiture of conservation laws and a likely loss of important symmetries in the model equations. For climate simulations conservation properties are of importance because they are leading to stable and realistic climate simulations, and should be considered also for stochastic parameterizations, not only for stability reasons, but also to respect the underlying dynamics of geophysical fluid flows.

From a mathematical perspective, there has been a growing interest over the last few decades in finding appropriate techniques to develop systematic methods to accurately and efficiently represent fast variables in multi-scale systems. Much fundamental work has already been done, e.g. Mémin [37] derived energy conserving geophysical fluid equations assuming that the velocity can be written as a mean state plus some perturbations, while Holm [32] used stochastic variational principles to obtain new stochastic fluid equations conserving helicity and the Kelvin circulation theorem. Numerical models following these theoretical approaches have been developed and show good performance and improved results with respect to the deterministic counterpart at the same resolution, see for example [51, 52] for numerical implementations and results of [37] and [12, 11] for applications of [32].

Furthermore, Majda, Timofeyev and Vanden-Eijnden [43, 42, 44, 39, 38] developed a systematic strategy for stochastic mode reduction starting from the assumption that the explicit nonlinear self-interaction of the fastest scales involved can be represented by a linear stochastic operator. This procedure is mathematically justified only for large time scale separation but showed good performances also in case of a less pronounced time scale separation. For its application to more complex atmospheric models see also [26, 19, 18, 50, 41]. A less theoretical, but still efficient, approach is given by the stochastic kinetic energy backscatter scheme (SKEBS) which is often used to represent model uncertainty arising from unresolved subgrid-scale processes and their interactions with larger scales [56, 5, 22, 47, 21]. Connected to the idea of backscatter, different deterministic and stochastic parameterizations aiming at representing the upscale turbulent cascades in eddy-permitting simulations have been developed. Among others, noticeable examples are given by [33, 49, 62, 28, 15], which also showed that the stochastic backscatter is, in general, a more efficient eddy parameterization than its deterministic counterpart.

In this paper we systematically further develop the projector approach outlined by Frank and Gottwald [16]. Here the stochastic noise is projected onto the energy manifold. This approach has the main advantage that it can be straight forwardly implemented in existing models while the above approaches [37, 32, 43] derive new equations of motions which are harder to implement in already existing and operational models. Frank and Gottwald [16] tested their scheme with a 4-dimensional toy-model for the large-scale dynamics of the atmosphere by means of a Lagrangian description of the dynamics. Here instead we apply it to a high-

dimensional conceptual climate model, i.e. the inviscid 2-layer QG model, in a Eulerian framework. The purpose of our study is a proof of concept whether this scheme can also be applied to high-dimensional models. Thus, the main research questions we aim to address in this study is: (i) Can we accurately conserve energy in a high-dimensional QG model? and (ii) What conditions do we need to impose on the spatial noise covariance matrix for this? Hence, in this work we focus on the technical aspects of this approach, analyzing potential issues due to the discretization of the continuous equations or to the numerical implementation in general. With this intention, we choose to apply this projection operator scheme to an energy conserving QG model as a hard numerical test case. While our particular set up might not seem interesting from a geophysical fluid dynamics point of view, we still consider it numerically challenging and hence a valuable benchmark in testing the numerical aspects and accuracy of our stochastic system in a high-dimensional geophysical model.

The remainder of this paper is organized as follows: in Section 2 we present the inviscid 2-layer energy conserving QG model both in its deterministic and stochastic formulations. Details on the numerics, like the choice of the numerical solvers and the definition of the spatio-temporal noise structure are provided in Section 3. Section 4 displays and discusses the outcomes of our stochastic model experiments. Finally in Section 5, we give a brief summary of our findings and outlook of future research directions.

2 The 2-Layer QG Model

2.1 The deterministic equations

As mentioned above, we start from the non-dimensional inviscid 2-layer QG equations presented in [59] on a β -plane with double-periodic boundary conditions

$$\frac{\partial q_B}{\partial t} = -J(\psi_B, q_B) - J(\psi_T, q_T), \quad (1a)$$

$$\frac{\partial q_T}{\partial t} = -J(\psi_T, q_B) - J(\psi_B, q_T), \quad (1b)$$

$$q_B = \nabla^2 \psi_B + \beta y, \quad (2a)$$

$$q_T = \nabla^2 \psi_T - k_d^2 \psi_T, \quad (2b)$$

where $q_B = q_B(\mathbf{x}, t)$, $\psi_B = \psi_B(\mathbf{x}, t)$ and $q_T = q_T(\mathbf{x}, t)$, $\psi_T = \psi_T(\mathbf{x}, t)$ represent, respectively, potential vorticity (PV) and streamfunction of the barotropic and baroclinic mode on the horizontal plane $\mathbf{x} \in \mathbb{R}^2$ at time $t \in \mathbb{R}$, ∇ is the horizontal gradient, ∇^2 the Laplacian and J the Jacobian operator

$$J(A, B) = \frac{\partial A}{\partial x} \frac{\partial B}{\partial y} - \frac{\partial A}{\partial y} \frac{\partial B}{\partial x}.$$

Since we employ a non-dimensional description, the domain has been rescaled to a $2\pi \times 2\pi$ square. Double periodic QG models on a β -plane have been widely used in theoretical studies [8, 9, 29, 45]. Here we consider layers of equal thickness and the parameter k_d (given by the relation $k_d^2/2 = (2f_0/Nh)^2$ where $N = 1.2 \cdot 10^{-2}$ is the Brunt-Väisälä frequency, h the mean depth of the layers and $f \approx f_0 + \beta y$ the approximate Coriolis term where $f_0 = 1$ and $\beta = 0.509$) determines the strength of the shear between the two layers and hence also the intensity of the baroclinic instability. For most simulations in this study we are using a deformation radius of about 0.042 non-dimensional units and, thus, are in an ocean like eddy-permitting regime. In this setting, one non-dimensional time unit corresponds to roughly 2.5 days.

Since we want to focus on the energy conservation properties of our numerical scheme we do not introduce terms accounting for external forcing or eddy dissipation, and instead consider an initial value problem. We want to stress, that we want to focus in this study on the numerical and accuracy aspects of energy conservation in a stochastic setting and not on geophysical flow realism (see section 2.2 below).

The system given by equations (1)-(2) conserves its total energy E and enstrophy Z :

$$E(q_B, q_T) = \frac{1}{2} \int_A [(\nabla\psi_B)^2 + (\nabla\psi_T)^2 + k_d^2 \psi_T^2] dA, \quad Z(q_B, q_T) = \frac{1}{2} \int_A (q_B^2 + q_T^2) dA.$$

Details about conservation properties and how to derive them can be found in [59].

The Hamiltonian H of the system is given by its total energy, thus it reads

$$H(q_B, q_T) = \frac{1}{2} \int_A [(\nabla\psi_B)^2 + (\nabla\psi_T)^2 + k_d^2 \psi_T^2] dA. \quad (3)$$

It can be shown that

$$\begin{aligned} \delta H &= + \int_A (\nabla\psi_B \cdot \delta\nabla\psi_B + \nabla\psi_T \cdot \delta\nabla\psi_T + k_d^2 \psi_T \cdot \delta\psi_T) dA \\ &= - \int_A (\psi_B \cdot \delta\nabla^2\psi_B + \psi_T \cdot \delta\nabla^2\psi_T - \psi_T \cdot k_d^2 \delta\psi_T) dA \\ &= - \int_A (\psi_B \cdot \delta q_B + \psi_T \cdot \delta q_T) dA, \end{aligned}$$

which implies

$$\frac{\partial H}{\partial q_B} = -\psi_B, \quad \frac{\partial H}{\partial q_T} = -\psi_T.$$

For a general review of Hamiltonian mechanics and its application to geophysical fluid dynamics see [2] and [55]. The following notation will be employed

$$A : B = a_{ij} b_{ij} = \text{Tr}(AB^T).$$

2.2 The stochastic formulation

In this section we derive a stochastic energy conserving version of the 2-layer QG equations. In our formulation we include unresolved fast sub-grid processes by means of a stochastic forcing, modeled as an Ornstein-Uhlenbeck process, which we assume to act first on the baroclinic mode and then, because of the coupling between the two modes, to affect also the slower barotropic mode. For this choice we relate to the idea of backscatter, where energy goes from the smaller scales back into the larger processes. Therefore, we add a simple 2-dimensional stochastic field only to the fast baroclinic mode. The source terms so introduced would lead the dynamics to leave the manifold of constant energy on which the deterministic model (1)-(2) evolves. In order to balance the stochastic fluctuations that would affect the aforementioned manifold, we introduce an auxiliary 2-dimensional stochastic process dY_t . This procedure follows the method introduced by Frank and Gottwald in [16], with the difference that we consider a high-dimensional system in a Eulerian framework, and not a 4-dimensional system with a Lagrangian description. The following set of equations is therefore proposed:

$$dq_B = (-J(\psi_B, q_B) - J(\psi_T, q_T)) dt, \quad (4a)$$

$$dq_T = (-J(\psi_T, q_B) - J(\psi_B, q_T)) dt - \Gamma q_T dt + \Sigma dW_t + dY_t, \quad (4b)$$

$$dY_t = S_t dW_t + B_t dt, \quad (4c)$$

where dW_t denotes a 2-dimensional Wiener process, $\Gamma, \Sigma, S_t \in \mathbb{R}^{2 \times 2}$ and $B_t \in \mathbb{R}^2$. The choice of adding the stochastic terms on the equation of the baroclinic PV not only reconnects to the concept of backscatter, but allows potentially also the application of stochastic mode reduction, as Frank and Gottwald did in their work

[16]. For instance, one could derive a reduced order stochastic model for just the barotropic modes [43, 42, 44, 19, 18].

Instead of dealing with two different stochastic processes, we want to write S_t and B_t as functions of Σ and Γ . For that purpose, we write the increment of H as a sum of two parts, a deterministic part μ_H including all the terms multiplied by dt , and a stochastic part σ_H containing those with the Wiener process. By Ito's theorem we have

$$\begin{aligned} dH &= \frac{\partial H}{\partial q_B} \cdot dq_B + \frac{\partial H}{\partial q_T} \cdot dq_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : dq_T dq_T^T \\ &= -\psi_B \cdot (-J(\psi_B, q_B) - J(\psi_T, q_T)) dt \\ &\quad - \psi_T \cdot (-J(\psi_B, q_T) - J(\psi_T, q_B) + B_t) dt \\ &\quad + \left(\frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : (\Sigma + S_t)(\Sigma + S_t)^T \right) dt \\ &\quad - \psi_T \cdot (-\Gamma q_T) dt - \psi_T \cdot (\Sigma + S_t) dW \\ &= \mu_H dt + \sigma_H dW_t, \end{aligned}$$

where the transposed is denoted by the superscript T , and

$$\begin{aligned} \mu_H &= \psi_B \cdot (J(\psi_B, q_B) + J(\psi_T, q_T)) + \psi_T \cdot (J(\psi_T, q_B) + J(\psi_B, q_T) - B_t) \\ &\quad + \psi_T \cdot \Gamma q_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : (\Sigma + S_t)(\Sigma + S_t)^T \\ &= -\nabla_{q_B} H \cdot (J(\psi_B, q_B) + J(\psi_T, q_T)) - \nabla_{q_T} H \cdot (J(\psi_T, q_B) + J(\psi_B, q_T)) \\ &\quad + \nabla_{q_T} H \cdot B_t - \nabla_{q_T} H \cdot \Gamma q_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : (\Sigma + S_t)(\Sigma + S_t)^T \\ \sigma_H &= -\psi_T \cdot (\Sigma + S_t) \\ &= \nabla_{q_T} H \cdot (\Sigma + S_t). \end{aligned} \tag{5}$$

Since we want to conserve the total energy, dH has to be zero. Therefore we impose both μ_H and σ_H to be zero. Following the reasoning outlined in [16], the auxiliary stochastic process dY_t should not perturb the dynamics on the tangent space and should be constructed only to counterbalance those components of the OU process which are orthogonal to the manifold of constant energy, thus we define a projection operator $\mathbb{P} \in \mathbb{R}^{2 \times 2}$. Since the Wiener process affects only the evolution equation of the baroclinic PV, it will be sufficient to project with respect to the manifold of constant baroclinic energy:

$$\begin{aligned} \mathbb{P} &= \mathbf{I} - \frac{1}{|\nabla_{q_T} H|^2} \nabla_{q_T} H (\nabla_{q_T} H)^T \\ &= \mathbf{I} - \frac{1}{|\psi_T|^2} \psi_T \psi_T^T, \end{aligned}$$

where $\mathbf{I} \in \mathbb{R}^{2 \times 2}$ stands for the identity matrix. Since $\mathbb{P}(\nabla_{q_T} H) = 0$, \mathbb{P} projects onto the tangent space of the baroclinic kinetic energy surface. Consequently, we want S_t and B_t to satisfy

$$\mathbb{P}S_t = 0, \quad \mathbb{P}B_t = 0. \tag{6}$$

From the assignment $\sigma_H = 0$ we can easily determine S_t . In fact, since $\nabla_{q_T} H$ is in the kernel of \mathbb{P} , imposing $\sigma_H = 0$ is equivalent to requiring $\Sigma + S_t = \mathbb{P}(\Sigma + S_t)$. Thus, using Eq. (6), we obtain

$$S_t = -(\mathbf{I} - \mathbb{P})\Sigma. \tag{7}$$

Substituting Eq. (7) into Eq. (5) and considering only the terms arising from the inclusion of the stochastic processes into the deterministic set of equations (since the deterministic model conserves energy, the other terms do not contribute to the variation of total energy), we can determine B_t from the requirement $\mu_H = 0$:

$$B_t = (I - \mathbb{P}) \Gamma q_T + \frac{1}{2 |\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P} \Sigma \Sigma^T \mathbb{P} \right) \psi_T. \quad (8)$$

Placing Eq. (7) and Eq. (8) into Eq. (4), after some manipulations, we get the following set of equations

$$dq_B = (-J(\psi_B, q_B) - J(\psi_T, q_T)) dt, \quad (9a)$$

$$dq_T = (-J(\psi_T, q_B) - J(\psi_B, q_T)) dt + \mathbb{P} \Sigma dW_t - \left(\mathbb{P} \Gamma q_T - \frac{1}{2 |\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P} \Sigma \Sigma^T \mathbb{P} \right) \psi_T \right) dt. \quad (9b)$$

Equations (2)-(9) constitute our stochastic energy conserving 2-layer QG system. A detailed derivation of equations (8) and (9) is reported in Appendix A. As can be seen, the resulting set of equations contains multiplicative noise and nonlinear damping, due to the specific definition of the projection operator. The multiplicative noise is in fact a correlated additive multiplicative (CAM) noise [38, 53, 17].

3 Numerical implementation

Since we aim to analyze possible applications of this approach to climate and ocean models, which are typically formulated in terms of finite volumes or finite differences, we discretize equations (2)-(9) in terms of finite differences in the framework of a grid-point based code. Our discretization of the QG model is based on the energy and enstrophy conserving discretization scheme by Arakawa [1]. This scheme ensures that energy and enstrophy are conserved for all truncations. Especially this scheme does not require any numerical diffusion or dissipation for numerical stability. This property affects the energy and enstrophy cascades by automatically redistributing the energy and enstrophy at the truncation scales, making the model simulations unrealistic. However, using this discretization scheme will allow us to focus on the accuracy of the energy conservation of our stochastic approach.

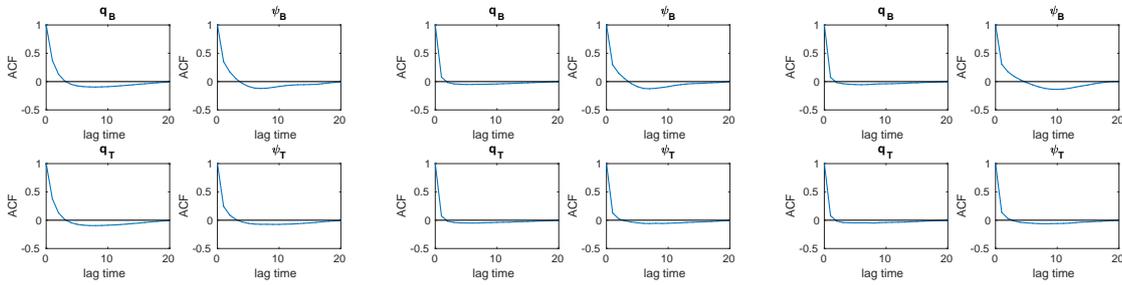
For the time stepping we employ explicit Runge-Kutta (RK) methods (whose order will be a matter of discussion in the following section), and we use a Fast Fourier Transform (FFT) to invert the Laplacian operator and obtain the streamfunctions from the corresponding PV. Since FFT is an exact numerical method and the Arakawa scheme is designed to conserve energy and enstrophy for any truncation, the only spurious effect on the energy due to the numerics is given by the RK method, which is known to be to a small extent dissipative in time. When dealing with the stochastic terms, we integrate them using either the Euler-Maruyama or Milstein schemes [48]. Finally, the initial distributions of the barotropic and baroclinic streamfunctions are generated using a pseudo-random number generator, i.e. no a-priori structure is given as input, and we define the corresponding PV by equations (2). We do not change the seed when defining the initial condition, thus all simulations at resolution 128×128 start from the same initial condition. Once the initial condition is defined, we set a new seed for the noise generator. Even though we do not have any forcing the model does not settle into a barotropic regime; the baroclinic modes are still active with active barotropic and baroclinic mode interactions (not shown). Furthermore, the probability density functions of the barotropic-baroclinic energy transfer terms are symmetric (not shown), suggesting an active interaction between barotropic and baroclinic modes.

3.1 Deterministic model

Before moving to the stochastic set of equations and related results, we test different orders of accuracy of our numerical scheme in the implementation of the deterministic 2-layer QG model (equations (1)-(2)) in order to find the optimal balance between accuracy and computational time.

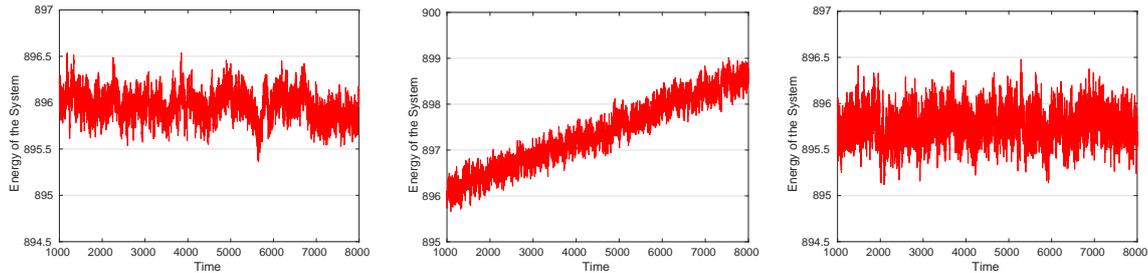
In our code, we solve the above evolution equations (1) for the PVs and then we compute the corresponding streamfunctions through equations (2). We use explicit Runge-Kutta 2^{nd} and 4^{th} order methods for the time integration, Arakawa 2^{nd} and 4^{th} order discretizations of the Jacobian [1] and a Fast Fourier Transform to invert the Laplacian operator. While performing our tests, we also consider different values of the mean depth of the fluid h ; in particular we consider the cases $h = 1, 10, 100$. These tests are performed over a 128×128 spatial grid with a time step of $\Delta t = 10^{-3}$.

We do not report here all the statistics and energy graphs obtained with the different combinations of solvers, but show only those motivating our choice to employ RK4 and Arakawa 4^{th} order in the following.



(a) $h = 100$; RK2 & Arakawa 2^{nd} order. (b) $h = 100$; RK2 & Arakawa 4^{th} order. (c) $h = 100$; RK4 & Arakawa 4^{th} order.

Figure 1: ACF for the case with $h = 100$ and different combinations of deterministic numerical solvers. By using the second order method both for RK and Arakawa schemes, the baroclinic mode and the barotropic PV decorrelate more slowly.



(a) $h = 1$; RK2 & Arakawa 2^{nd} order. (b) $h = 1$; RK2 & Arakawa 4^{th} order. (c) $h = 1$; RK4 & Arakawa 4^{th} order.

Figure 2: Total energy graphs with $h = 1$ and different combinations of deterministic numerical solvers. It can be seen that also after the initial spin up period, which has been here neglected, energy increases in time when solving with RK2 and 4^{th} order Arakawa.

Figure 1 shows that the lower order combination of RK2 with Arakawa 2^{nd} order does not reproduce accurately the autocorrelation function (ACF) in the case $h = 100$ and that just increasing the order of Arakawa's discretization is enough to capture correctly the ACF. However, when combined with RK2 it does not conserve energy in the case $h = 1$ also after the initial spin up period (Fig. 2, where the spin up period has been neglected). More generally, we found that RK4 with Arakawa 4^{th} order is more reliable and that the scenario with $h = 10$ has a less discernible spin up period and it is the best reproduced case study with all the con-

sidered solvers, therefore we decided to employ this higher order numerical implementation and we fixed $h = 10$. As a consequence of this choice, the Rossby deformation radius $1/k_d$ is approximately 0.042 and we are in a ocean-like regime with small eddies. Furthermore, while the energy fluctuates around a mean value the fluctuations are relatively small; the energy fluctuation amplitude is less than 1% of the total energy. With the chosen numerical solvers also enstrophy is conserved by the system to a similar accuracy, as in the continuous scenario (not shown).

3.2 Stochastic equations

As in the deterministic case, we first solve the stochastic evolution equations (9) for the PVs and then we get the corresponding streamfunctions through equations (2). As a consequence of the analyses in the previous paragraph, we use an explicit RK 4th order method for the time integration, Arakawa 4th order discretization of the Jacobian [1] and FFT to invert the Laplacian operator. The stochastic terms are integrated using either the Euler-Maruyama or Milstein schemes [48]. Later we will analyze differences in the outcomes due to the stochastic solver. In the stochastic simulations we employ a 128×128 spatial grid with a time step of size $\Delta t = 10^{-3}$. As a consequence $\Delta x \approx 0.049$ and the model is in the eddy permitting regime.

As we will demonstrate below, for the dynamical consistency between deterministic and stochastic models it is crucial to consider spatially correlated noise. To demonstrate this, we consider two scenarios: in the first we assume that the noise on each grid point behaves as independent and identically distributed (iid) random variables, while in the second we allow for correlation between different grid points. In the following subsection a more detailed description of how the correlation matrix of the noise is constructed can be found. Finally, in order to generate the noise, we first produce uniformly distributed random numbers using the Mersenne-Twister algorithm [46], and then we reshape them through the Box-Muller procedure in such a fashion that they are sampled from a Gaussian distribution with the desired mean and variance, which in our case is $\mathcal{N}(\mu = 0, \sigma^2 = \Delta t)$. We compare the outcomes with a reference solution given by a deterministic simulation with 512×512 grid points and $\Delta t = 10^{-4}$. For a fair comparison, we project the fine grid data onto a grid with the same resolution as for the stochastic simulations.

3.3 Spatial noise structure

For allowing spatial correlations among different grid points, we need to determine the elements of the matrix Σ . We do this using eigenvectors obtained from a dimension reduction technique. Here we employ Empirical Orthogonal Functions (EOFs) [58]. We derive the EOFs from the high-dimensional deterministic control simulation.

Once the eigenvectors and the corresponding eigenvalues are computed, we select a number of EOFs and define Σ as a convex combination of the chosen eigenvectors E_i . Such technique has been applied already in [25, 24]. A more general linear combination could be used and it would be easy to modify this constraint in order to attribute a stronger (or weaker) amplitude to the noise. The weights ω_i are selected as uniformly distributed random numbers, i.e.,

$$\Sigma = \sum_i \omega_i E_i, \quad \sum_i \omega_i = 1, \quad \omega_i \sim \mathcal{U}\{0, 1\},$$

where the eigenvectors are matrices with dimensions equal to the grid size. Coefficients related to the redundant eigenvectors are set to be zero. In view of the fact that the noise is only in the equation of the baroclinic mode, we use ψ_T -EOFs computed with respect to the L^2 norm using the data of the high resolution run projected onto the coarser grid. Considering that the weights ω_i are chosen randomly in each simulation, no particular direction is preferred with respect to the others reducing possible biases in the results due to how the EOFs have been computed.

4 Results of the stochastic simulations

4.1 Space-time independent noise

For reasons that will become evident later, in this scenario we perform the stochastic integration only with the Euler-Maruyama scheme. As assessment criteria, we first look at the conservation of energy and then at other statistical properties like the autocorrelation function (ACF) and probability density function (PDF). Here we consider iid white noise with zero mean and variance equal to the time step. Therefore, the matrices Σ and Γ can be written as

$$\Sigma = \sigma I, \quad \Gamma = \gamma I$$

where $I \in \mathbb{R}^{2 \times 2}$ stands for the identity matrix and $\sigma, \gamma \in \mathbb{R}$. We show the results for the case $\sigma = 1$ and $\gamma = 1$.

4.1.1 Numerical results

In this case, after a positive jump at the beginning of the simulation (because of the forcing, the system moves to the closest stable state, which has a different amount of total energy), energy is conserved (see Figure 3a) with very small fluctuations in time (see Figure 3b). In fact, when compared to a deterministic run at the same resolution (see Figure 3c), quite unexpectedly, variations in the total energy graph are smaller in the stochastic simulation. Still, looking at Figure 3b, it is possible to notice a small decreasing trend suggesting that the parameterization is damping in the long run which is likely due to the Runge-Kutta scheme, which is known to be damping for.

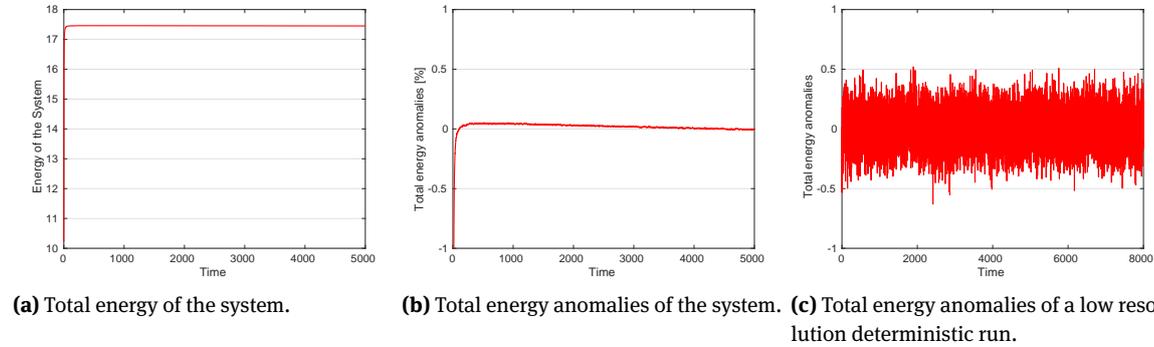
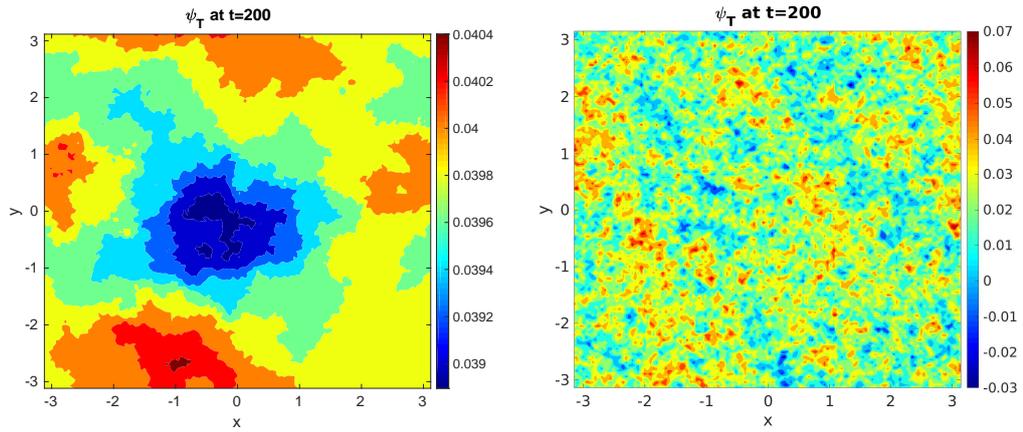


Figure 3: Total energy graph for the 2-layer stochastic QG model, with iid white noise and $\sigma = 1$, $\gamma = 1$. After an initial positive jump, energy is conserved. Graph (c) shows the total energy anomalies of a deterministic run at the same resolution. It can be noticed that, surprisingly enough, oscillations are smaller in the case of the stochastic system.

A less reassuring result is given by the contour plot of the baroclinic mode. In Figure 4 we show the baroclinic streamfunction (but a very similar result can be observed also for the baroclinic PV) at time $t = 200$ (left) and the plot of the same field at the same time given by our reference solution (right). What immediately stands out is the different pattern presented by the two figures. Furthermore, at a closer look it can also be noticed that the different colors in Figure 4a represent differences in the order of 10^{-4} and they become even smaller when looking at the contour plot for later times (not shown), meaning that the field is moving towards a constant state in space. This explains the smaller amplitude of the fluctuations in the energy graph with respect to the deterministic scenario, and it is reflected also in the ACF and in the PDF (see Figures 5-6). The former displays longer decorrelation times in general and, more specifically, the baroclinic streamfunction

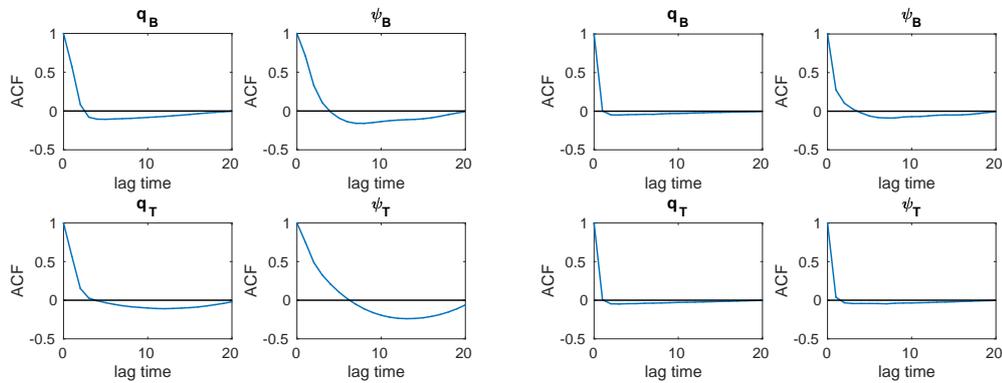


(a) Baroclinic streamfunction.

(b) Baroclinic streamfunction: reference solution.

Figure 4: Contour plot for the baroclinic streamfunction given by the stochastic system with iid white noise, $\sigma = 1, \gamma = 1$ (left) and by the high resolution deterministic simulation (right). The reader will immediately notice the different patterns displayed by the two pictures and, at a closer look, that the colors in the left graph represent differences in the order of 10^{-4} , which get even smaller with the developing of the simulation, implying that the field is moving towards a constant state in space.

seems to require a longer decorrelation time with respect to the barotropic streamfunction which is in contrast with the physics. The latter instead shows no Gaussian distribution for the baroclinic mode and smaller variance for the barotropic mode.

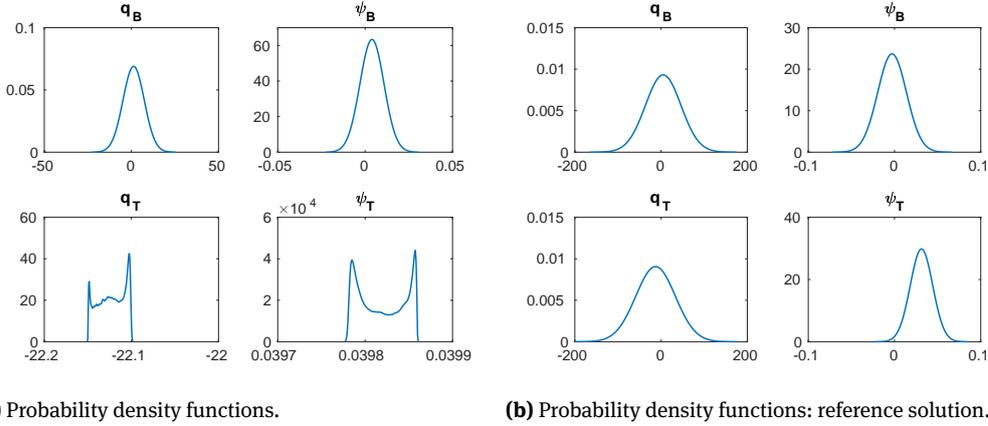


(a) Auto-correlation functions.

(b) Auto-correlation functions: reference solution.

Figure 5: ACF resulting from the stochastic system with iid white noise, $\sigma = 1, \gamma = 1$ (left) and from the high resolution deterministic simulation (right). The stochastic simulation does not well reproduce the ACF of the reference solution but displays longer decorrelation times.

In their paper [16] Frank and Gottwald used iid noise obtaining conservation of energy and physically reasonable outcomes. As already stated earlier, they employ a Lagrangian discretization of the system, while we consider the dynamics from a Eulerian point of view. Since Lagrangian descriptions of motion follow the trajectories of the single particles, and not the fluid as a whole in a fixed domain, in this frame the main purpose of the noise is simply to perturb the trajectory while remaining on the manifold of constant energy. Hence any spatial information added to the noise is not strictly necessary. On the other hand, Eulerian descriptions focus on what happens inside a well-defined domain and do not care about the behavior of the single particles. Thus in this framework spatial iid noise means that each grid point does not feel the influence of its



(a) Probability density functions.

(b) Probability density functions: reference solution.

Figure 6: PDF rising from the stochastic system with iid white noise, $\sigma = 1$, $\gamma = 1$ (left) and from the high resolution deterministic simulation (right). The baroclinic mode of the stochastic set of equations does not have a Gaussian distribution. Moreover the barotropic mode displays less variance.

neighbors, hence the noise would represent phenomena which fully evolve and decorrelate inside the cell; the gap between the large resolved scales and such small phenomena is too big to be correctly resolved by the numerics. Therefore it appears crucial to define a spatial structure of the noise in order to characterize how the noise should behave inside the domain and interact with the deterministic dynamics. The next section discusses in more detail our results.

4.2 Space-time correlated noise

To ease computations, we neglected the Γ term in equation 9b, i.e. $\Gamma = \mathbf{0}$. For both the Euler-Maruyama and Milstein schemes, we run an ensemble of 40 simulations using a convex combination of the first two EOFs to build the covariance matrix Σ . We tried also combinations with a different number of EOFs. When considering up to the first 10 EOFs, similar results to those we report here are obtained. With 20 or more EOFs we noted slightly worse performances of the scheme. Because of the constraint $\sum_i \omega_i = 1$, when considering relatively many EOFs, each of them has a small amplitude and then the patterns contrast with each other resulting in a not well-defined structure. On the other hand, a combination of a smaller number of EOFs can still maintain the individual patterns while allowing interaction with each other. In what follows, we opted for using only the first two main patterns.

As evaluation criteria, we employ the same analyses as before. Regarding the PDF, we also compute the first and second moments of the centers in order to investigate the ensemble variance. In addition we compare to the reference solution: the total variance and eddy length (computed through space correlations as presented in [3]).

4.2.1 Energy conservation

In each simulation the total energy fluctuates in time around a constant value. Differently from the previous case, there is no jump to a different stable state at the beginning of the time integration, meaning that our stochastic system keeps its evolution on the manifold defined by the initial condition. In both ensembles, if we compare the amplitude of the anomalies A_{Anom} with respect to the mean value of the energy μ_{En} , we see that A_{Anom} is, for most of the running time, around 0.5 % of μ_{En} with spikes no greater than 0.7 %. We would like to point out that, even though for each individual simulation the evolution of the field variables is different (and this is shown by the fact that the PDFs of each individual run are centered in different locations),

the total energy of the system is almost the same at each time step, with differences in the order of 10^{-6} , for each ensemble member. This implies that, starting from the same initial condition the model is exploring different possible configurations available with the defined amount of total energy. In Figure 7 we show the time-evolution of the energy anomalies for an individual Euler-Maruyama (Milstein) ensemble member together with the total energy anomalies graph of a deterministic simulation at the same resolution. It can be seen that fluctuations are roughly of the same amplitude for both the deterministic and the stochastic system. Thus, the energy fluctuations are mainly a result of the deterministic numerical scheme and not of the used stochastic scheme. This shows that the projection operator works very well in high-dimensional models and suggests that one should improve the discretization of the deterministic part in order to ameliorate energy conservation of the stochastic system.

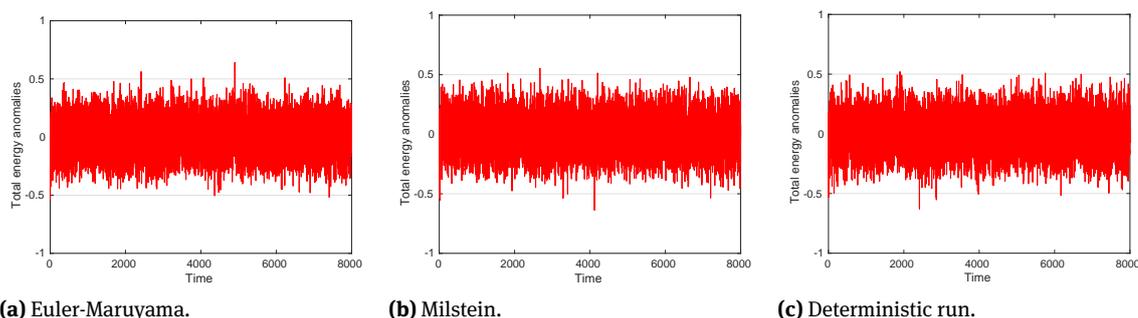


Figure 7: Total energy anomalies, displayed as a percentage of the mean total energy value, of an individual member of the Euler-Maruyama (Milstein) ensemble. On the right we also show the total energy anomalies graph for a deterministic run at the same resolution. The reader can notice that the fluctuations have roughly the same amplitude.

4.2.2 ACF

Every ensemble member shows roughly the same ACF pattern, independent of the stochastic solver. Differently from the previous case with iid white noise, in both ensembles we obtain decorrelation times very close to the reference. The barotropic streamfunction displays a longer decorrelation time with respect to the baroclinic streamfunction, suggesting that in future work a stochastic mode reduction might be performed for eliminating the baroclinic modes and having a stochastic barotropic model as in [43, 42, 44, 19, 18]. In Figure 8 we show the ACF for one stochastic simulation of the Euler-Maruyama (Milstein) ensemble together with the reference solution.

4.2.3 PDF

In contrast to the case with iid noise, in each run of both ensembles we recover the Gaussian behavior of the baroclinic mode displayed by the reference solution and more variance for the barotropic, see Figure 9 for the PDF graph of an individual Euler-Maruyama (Milstein) ensemble member and the reference solution. On the other hand, the PDF of an individual stochastic run shows less variance with respect to the reference solution except for the barotropic streamfunction which, in the comparison, shows more (this can be noticed in Figure 9). Hence we decided to investigate the variance of the ensemble by looking at the first and second moment of the center of the PDFs of the ensemble members. While we are not too much interested in the exact value taken by the first moment, due to the chaotic nature of the system, we would like to point out that, the Milstein ensemble displays more variance with respect to the Euler-Maruyama scheme. In Table 1 we report the 95%

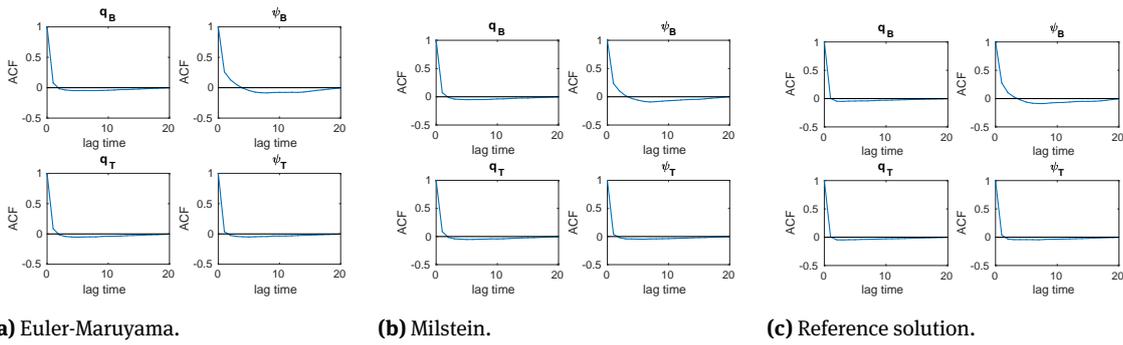


Figure 8: ACF of a particular Euler-Maruyama (Milstein) ensemble member and reference solution. Both ensembles well catch and reproduce the ACF shown by the high resolution deterministic simulation.

confidence interval (CI) of the centers of the PDFs; those related to the Milstein scheme span a wider range of values. We would like to remind the reader that the Euler-Maruyama and Milstein schemes have the same order of weak convergence (i.e. it is 1 for both schemes), but different order of strong convergence (0.5 for Euler-Maruyama and 1 for Milstein). Hence, since we are considering long time simulations, statistical properties of the field variables are more sensitive to weak convergence, while the evolution of trajectory paths is more sensitive to strong convergence. This explains why both ensembles catch the right shape of the PDF but at the same time the Milstein ensemble displays more variance. It could be argued that 0.5 might not be a meaningful difference; on the other side, high order stochastic integration methods include complicated correcting terms which might be hard to implement, see for instance [54] for an example of necessary conditions that have to be satisfied by a class of stochastic integration methods with (strong) order 1.5. Hence here we tried to analyze two of the most likely employed methods in complex climate models. We also checked the total variance of the stochastic ensemble and compared to a deterministic run at the same resolution and to the high resolution deterministic simulation. Unfortunately our stochastic parameterization is still not able to mend for the variance lost with the coarsening of the grid (see Table 1).

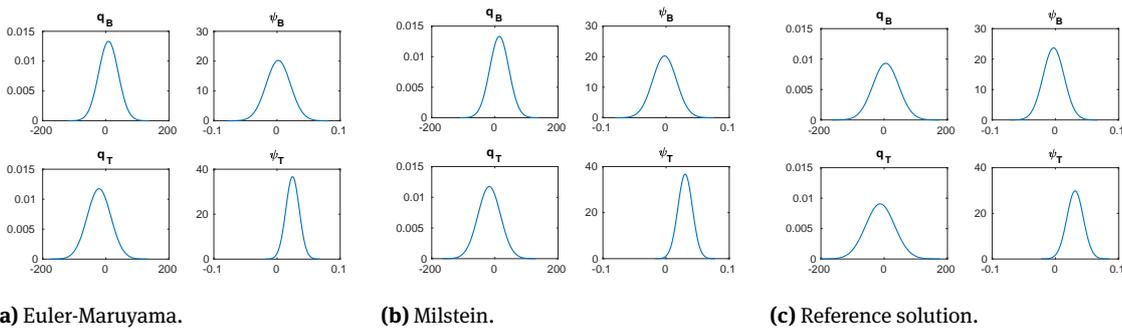


Figure 9: PDF of an individual member of Euler-Maruyama (Milstein) ensemble and reference solution. The careful reader might notice that an individual ensemble run displays less variance with respect to reference for almost all field variables, inducing us to investigate the ensemble variance (see Table 1).

4.2.4 Eddy length

When looking at the contour plots for the baroclinic (but also the barotropic) mode, we can see that in this scenario the stochastic simulations reproduce a similar pattern as for the case of the reference solution (see Figure 10 for a representation of the barotropic and baroclinic streamfunctions in the different setups). In order to have a more objective comparison criterion, we computed the eddy length for the two streamfunctions as a measure of the correlation among different grid points at a fixed time and looked at the e-folding scale; see for instance [3] for a more detailed description. In Table 1 we report the outcomes for the reference solution, a coarse deterministic simulation, Euler-Maruyama and Milstein ensemble. When considering the baroclinic mode, our stochastic parameterization does not improve the eddy length. In both ensembles it is about half of the baroclinic eddy length displayed by the high resolution simulation, remaining close to the outcome of a coarse deterministic run (see Table 1). A different conclusion is valid for the barotropic mode. In fact, in the case of the low resolution deterministic run we get an eddy length of circa 0.524 both in zonal and meridional direction, while in the stochastic Euler-Maruyama (Milstein) ensemble it can vary between ≈ 0.511 (≈ 0.510) and ≈ 0.570 (≈ 0.560). Comparing these results with the reference solution, we can notice that we still did not manage to reproduce the high resolution eddy length ($\approx 0.714 - 0.716$) but we obtained an improvement of circa 9% with respect to the low resolution deterministic simulation (see Table 1). This result suggests that the perturbation induced by the noise is still not strong enough, but we are heading in the right direction. Furthermore, as has already been shown in [60] and references therein, the error dynamics considerably depends on the specific scale at which it is introduced, with a faster growth when located at small spatial scales. Here the noise structure is built with the first two EOFs, hence it can be regarded as a perturbation on the large spatial scales, which is in good agreement with the improved eddy length for the slower mode. Nevertheless, as stated in [58], if the first EOF can be associated with a definite physical process, this is more difficult already with the second (and even harder for higher-order) EOF because of the orthogonality constraint. On the other hand, real-world processes might not have orthogonal patterns. In fact, the patterns that most efficiently represent variance do not necessarily have anything to do with the underlying dynamical structure.

5 Conclusions and perspectives

We described the numerical implementation and evaluation of an energy conserving high-dimensional stochastic conceptual climate model. Our main focus here is the proof of concept whether the projection operator approach [16] can be applied to high-dimensional complex geophysical flow systems in a Eulerian setting ([16] developed this approach in a Lagrangian setting for a low-order model). Furthermore we investigate which assumptions regarding the noise should be considered in order to obtain not only energy conservation but also dynamically consistent results. For this purpose we used our QG model without forcing, dissipation and hyperdiffusion. Even though, the resulting circulations are less realistic when compared with the atmosphere and the oceans, this setup provides an ideal test bed for the numerical evaluation of energy conservation of our numerical scheme.

In their paper, Frank and Gottwald [16] derived an energy conserving stochastic formulation for a 4-dimensional multi-scale toy model of the atmosphere. In order to preserve the conservation of energy they projected the noise with respect to the manifold of constant energy in such a fashion that those components of the noise, which would lead the trajectory to leave this manifold, are eliminated. In this paper we brought forward this approach and applied it to the high-dimensional 2-layer QG model through its Hamiltonian formulation. With the idea of analyzing the applicability of this procedure, not just to simple models but to a wider range of models with different degrees of complexity, we discretized the evolving equations in a Eulerian framework by means of finite differences, i.e. in a similar setup as most climate and ocean models. We could also introduce a time-scale separation parameter ε , depending on the different time scales of barotropic and baroclinic modes, to account for the time scale separation between the two modes. Even though here we

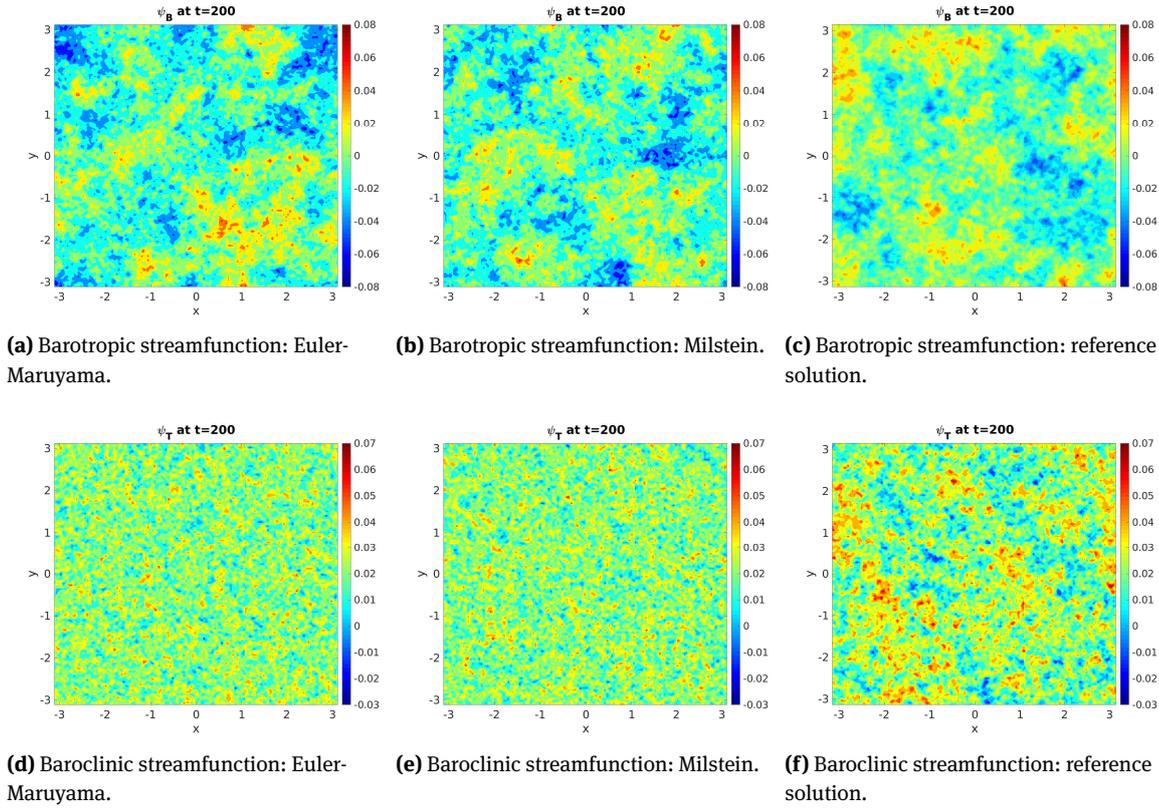


Figure 10: Contour plot for the barotropic (top) and baroclinic (bottom) streamfunctions for an individual member of Euler-Maruyama (Milstein) ensemble and for the reference simulation. The patterns are similar, even though the stochastic simulations still do not have the same eddy length as the high resolution deterministic run (see Table 1).

focused on other issues and did not consider any time scale separation in our numerical simulations, stochastic mode reduction is a possible research direction to be followed.

In particular, we investigated the delicate step from a continuous to a discrete formulation and found that the numerics can be sensitive to the mean depth of the fluid, and hence chose solvers that reproduce correctly the properties of the system, e.g. conservation of energy, in the most general scenario. Once this aspect had been settled, we analyzed the effects on the system dynamics and statistics due to the introduction of the stochastic process. Mainly we compared the results for two different scenarios: in the first, the noise of each grid point behaves like an iid random variable while, in the second, we considered spatio-temporal correlations. We found that employing iid noise leads to either that energy is not conserved or to unphysical results and hence defining a spatio-temporal structure is important to respect the underlying dynamics of geophysical flows and for the conservation of energy and the preservation of important statistical properties, e.g. PDF. This is due to the Eulerian nature of our implementation. Frank and Gottwald employed a Lagrangian description which follows the trajectories of the single elements. Hence in their model the noise had the unique purpose of perturbing the trajectories while remaining on the manifold of constant energy. On the other hand, a Eulerian point of view looks at a well-defined domain and considers the fluid as a whole. Therefore in this frame the noise should perturb the dynamics while conserving energy and preserving the main properties of the fluid.

In the present work a convex combination of the first two EOFs, computed on the data of a high resolution deterministic run, have been used to define the spatio-temporal correlations consistent with the behavior of the deterministic system; other dimension reduction techniques, such as [4, 10, 14], could be used too. We did not recover the same amount of variance as in the high resolution simulation, but the eddy length in the barotropic mode is improved. This suggests that the stochastic perturbations are not strong or spatially

Table 1: Summary of some of the previously discussed analyses in the cases of, from left to right: the reference solution, a deterministic run at the coarse resolution, Euler-Maruyama and Milstein ensembles. In particular we report: location of the center of the PDF and, in case of the ensembles, its 95% CI, total variance and eddy length (the latter computed only for barotropic and baroclinic streamfunctions).

Field variable	512 × 512 det.	128 × 128 det.	Euler-Maruyama	Milstein
Mean				
ψ_B	$-3.5687 \cdot 10^{-3}$	$-2.4223 \cdot 10^{-3}$	$-6.3485 \cdot 10^{-4}$	$-4.4880 \cdot 10^{-4}$
ψ_T	$2.6234 \cdot 10^{-2}$	$1.987 \cdot 10^{-2}$	$2.5108 \cdot 10^{-2}$	$2.4466 \cdot 10^{-2}$
q_B	1.0412	-5.1523	1.1737	-0.2762
q_T	-14.6414	-32.4428	-14.9759	-12.0254
95% CI				
ψ_B			[-0.0078, 0.0066]	[-0.0070, 0.0061]
ψ_T			[0.0212, 0.0290]	[0.0199, 0.0291]
q_B			[-9.5180, 11.8654]	[-13.8563, 13.3039]
q_T			[-28.8209, -1.1308]	[-28.0849, 4.0342]
Total variance				
ψ_B	$2.6302 \cdot 10^{-4}$	$3.5852 \cdot 10^{-4}$	$3.6169 \cdot 10^{-4}$	$3.6100 \cdot 10^{-4}$
ψ_T	$1.6564 \cdot 10^{-4}$	$1.0931 \cdot 10^{-4}$	$1.0931 \cdot 10^{-4}$	$1.0931 \cdot 10^{-4}$
q_B	1698.6409	830.2853	829.9909	829.9071
q_T	1796.2310	1066.7904	1067.1422	1067.1993
Eddy length				
ψ_B zonal	$7.1428 \cdot 10^{-1}$	$5.2422 \cdot 10^{-1}$	[0.51124, 0.56958]	[0.50976, 0.56011]
ψ_B merid.	$7.1632 \cdot 10^{-1}$	$5.2406 \cdot 10^{-1}$	[0.51151, 0.56854]	[0.50994, 0.56006]
ψ_T zonal	$1.4918 \cdot 10^{-1}$	$7.4597 \cdot 10^{-2}$	[0.07452, 0.07459]	[0.07451, 0.07459]
ψ_T merid.	$1.4918 \cdot 10^{-1}$	$7.4577 \cdot 10^{-2}$	[0.07451, 0.07460]	[0.07451, 0.07460]

coherent enough. Another possible explanation is the lack of temporal memory in our scheme [27]. Memory terms have been included in many techniques, such as multi-level regression models [36, 35, 40, 30], showing encouraging results. However they might be rather complicated to implement and might lead to unstable and diverging simulations as reported in [13] in the case of the Wouters and Lucarini parameterization [61]. Investigating the impact of spatial coherence and memory in the noise will be part of our future research.

Two basic arguments are that the constraint $\sum_i \omega_i = 1$ for the noise amplitude was arbitrary and it could be changed in order to attribute a stronger (or weaker) amplitude to the noise; moreover we computed the EOFs with respect to the Euclidian norm and not to the total energy norm. A more philosophical discussion regards instead the usage of the EOF technique itself. In fact it is sensible that using the first two EOFs improves the dynamics of the large scales since the first EOFs can be easily associated to large scale dynamics. Going down the ladder, because of the orthogonality constraint, it becomes harder and harder to associate EOFs to well-defined physical phenomena and hence also to the smaller scales [58]. As has already been shown in [60] and references therein, the error dynamics is considerably dependent on the specific scale at which it is introduced, with a faster growth when located at small spatial scales. In spectral models this obstacle is easily resolved, since the wavenumber where the noise should be introduced (choosing therefore its spectral properties) can be selected directly. In a grid-point framework this is not the case. Further studies in this direction will be done in order to gain this ability also when using a grid-point discretization since most climate and ocean models are based on this type of numerics and will be reported elsewhere.

Author contributions: F.G. and C.F. developed the stochastic model formulation and wrote the manuscript. F.G. developed the model code and performed the simulations and data analysis.

Supplementary material: The model code and data are available upon request.

Funding: This paper is a contribution to the project M2 of the Collaborative Research Center TRR181 *Energy Transfers in Atmosphere and Ocean* funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 274762653.

Conflict of interest: The authors declare no conflicts of interest.

Acknowledgement: We thank Profs. M. Oliver and J. Rademacher for useful and constructive discussions throughout the preparation of this work. We thank also two anonymous reviewers who helped us to improve the quality of the paper.

A Derivation of the stochastic 2-layer QG model

We report here a more detailed description of the procedure to derive equations (8) and (9) starting from the system (4), and making use of (7) and of the energy conservation property of the deterministic system (1)-(2).

As already mentioned, once (7) is obtained we substitute it into Eq. (5). Now we notice that the increment of total energy of the system (1)-(2) is

$$\begin{aligned} dH_{det} &= \frac{\partial H}{\partial q_B} \cdot dq_B + \frac{\partial H}{\partial q_T} \cdot dq_T \\ &= -\psi_B \cdot (-J(\psi_B, q_B) - J(\psi_T, q_T)) - \psi_T \cdot (-J(\psi_T, q_B) - J(\psi_B, q_T)) \\ &= 0, \end{aligned}$$

due to energy conservation of the deterministic equations. Hence we need to consider only the remaining terms arising from the inclusion of the stochastic processes in the deterministic equations; we impose them to equal zero. Computations proceed as follows

$$\begin{aligned} -\psi_T \cdot B_t + \psi_T \cdot \Gamma q_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : (\Sigma + S_t)(\Sigma + S_t)^T &= 0 \\ -\psi_T \cdot B_t + \psi_T \cdot \Gamma q_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : (\Sigma - \Sigma + \mathbb{P}\Sigma)(\Sigma - \Sigma + \mathbb{P}\Sigma)^T &= 0 \\ -\psi_T \cdot B_t + \psi_T \cdot \Gamma q_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T\mathbb{P} &= 0 \\ -\psi_T^T B_t + \psi_T^T \Gamma q_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T\mathbb{P} &= 0. \end{aligned}$$

At this stage we multiply from the left by $\frac{1}{|\psi_T|^2} \psi_T$ obtaining

$$-\frac{1}{|\psi_T|^2} \psi_T \psi_T^T B_t + \frac{1}{|\psi_T|^2} \psi_T \psi_T^T \Gamma q_T + \frac{1}{|\psi_T|^2} \psi_T \left(\frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T\mathbb{P} \right) = 0.$$

Here we note that $\frac{1}{|\psi_T|^2} \psi_T \psi_T^T = \mathbb{I} - \mathbb{P}$ and that the term $\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T\mathbb{P}$ is a scalar. Hence we can rewrite the previous equation as

$$-(\mathbb{I} - \mathbb{P})B_t + (\mathbb{I} - \mathbb{P})\Gamma q_T + \frac{1}{|\psi_T|^2} \left(\frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T\mathbb{P} \right) \psi_T = 0.$$

Now, making use of properties (6) and solving for B_t we recover

$$B_t = (\mathbb{I} - \mathbb{P})\Gamma q_T + \frac{1}{2|\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T\mathbb{P} \right) \psi_T,$$

which is Eq. (8). Once relations (7)-(8) have been derived, we can include Eq. (4c) in the evolution equation of the baroclinic PV (4b) and drop it, since drift and diffusion of the process dY can now be written as a function of Σ and Γ . Then, replacing S and B with the corresponding expressions as functions of Σ and Γ , manipulations of Eq. (4b) proceed as follows:

$$\begin{aligned}
dq_T &= (-J(\psi_T, q_B) - J(\psi_B, q_T)) dt - \Gamma q_T dt + \Sigma dW + S_t dW + B_t dt \\
&= (-J(\psi_T, q_B) - J(\psi_B, q_T)) dt - \Gamma q_T dt + \Sigma dW - (I - \mathbb{P}) \Sigma dW \\
&\quad + \left((I - \mathbb{P}) \Gamma q_T + \frac{1}{2|\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P} \Sigma \Sigma^T \mathbb{P} \right) \psi_T \right) dt \\
&= (-J(\psi_T, q_B) - J(\psi_B, q_T)) dt - \Gamma q_T dt + \Sigma dW - \Sigma dW + \mathbb{P} \Sigma dW \\
&\quad + \Gamma q_T dt - \mathbb{P} \Gamma q_T dt + \frac{1}{2|\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P} \Sigma \Sigma^T \mathbb{P} \right) \psi_T dt \\
&= - (J(\psi_T, q_B) + J(\psi_B, q_T)) dt + \mathbb{P} \Sigma dW_t \\
&\quad - \left(\mathbb{P} \Gamma q_T - \frac{1}{2|\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P} \Sigma \Sigma^T \mathbb{P} \right) \psi_T \right) dt,
\end{aligned}$$

while the equation of the barotropic PV (4a) remains unchanged. The above derived stochastic evolution equation of the baroclinic PV, together with Eq. (4a) and the corresponding streamfunctions described by Eq. (2), defines our stochastic 2-layer QG system.

References

- [1] A. Arakawa. “Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I.” In: *Journal of Computational Physics* 1.1 (1966), pp. 119–143.
- [2] G. Badin and F. Crisciani. *Variational formulation of fluid and geophysical fluid dynamics - Mechanics, symmetries and conservation laws*. Springer Berlin, 2018.
- [3] E. A. Barnes and D. L. Hartmann. “The Global Distribution of Atmospheric Eddy Length Scales.” In: *Journal of Climate* 25.9 (2012), pp. 3409–3416.
- [4] M. Belkin and P. Niyogi. “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation.” In: *Neural Computation* 15.6 (2003), pp. 1373–1396.
- [5] J. Berner, G. Shutts, M. Leutbecher, and T. Palmer. “A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system.” In: *J. Atmos. Sci.* 66 (2009), pp. 603–626.
- [6] J. Berner, U. Achatz, L. Batte, L. Bengtsson, A. Camara, H. Christensen, M. Colangeli, D. Coleman, D. Crommelin, S. Dolaptchiev, C. Franzke, P. Friederichs, P. Imkeller, H. Jarvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. Palmer, C. Penland, M. Sakradzija, J. von Storch, A. Weisheimer, M. Weniger, P. Williams, and J. Yano. “Stochastic parameterization: Toward a new view of weather and climate models.” In: *Bull. Amer. Meteorol. Soc.* 98.3 (2017), pp. 565–588.
- [7] R. Buizza, M. Miller, and T. Palmer. “Stochastic representation of model uncertainties in the ECMWF ensemble prediction system.” In: *Quart. J. R. Meteorol. Soc.* 125 (1999), pp. 2887–2908.
- [8] G. F. Carnevale. “Statistical features of the evolution of two-dimensional turbulence.” In: *Journal of Fluid Mechanics* 122 (1982), 143–153.
- [9] G. F. Carnevale and J. S. Frederiksen. “Nonlinear stability and statistical mechanics of flow over topography.” In: *Journal of Fluid Mechanics* 175 (1987), 157–181.
- [10] R. R. Coifman and S. Lafon. “Diffusion maps.” In: *Applied and Computational Harmonic Analysis* 21 (2006), pp. 5–30.
- [11] C. J. Cotter, D. Crisan, D. D. Holm, W. Pan, and I. Shevchenko. “Modelling uncertainty using circulation-preserving stochastic transport noise in a 2-layer quasi-geostrophic model.” In: *arXiv:1802.05711* (2018).
- [12] C. J. Cotter, D. Crisan, D. D. Holm, W. Pan, and I. Shevchenko. “Numerically modelling stochastic Lie transport in fluid dynamics.” In: *arXiv:1801.09729* (2018).
- [13] J. Demaeyer and S. Vannitsem. “Comparison of stochastic parameterizations in the framework of a coupled ocean-atmosphere model.” In: *Nonlinear Processes in Geophysics* 25.3 (2018), pp. 605–631.
- [14] 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, and C. Kuehn. “A numerical framework to understand transitions in high-dimensional stochastic dynamical systems.” In: *Dynamics and Statistics of the Climate System* 1.1 (2016).

- [15] S. Dwivedi, C. L. E. Franzke, and F. Lunkeit. “Energetically Consistent Stochastic and Deterministic Kinetic Energy Backscatter Schemes for Atmospheric Models.” In: *Q. J. Roy. Meteorol. Soc.* (2019), submitted.
- [16] J. E. Frank and G. A. Gottwald. “Stochastic homogenization for an energy conserving multi-scale toy model of the atmosphere.” In: *Physica D* 254 (2013), pp. 46–56.
- [17] C. L. E. Franzke. “Extremes in dynamic-stochastic systems.” In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.1 (2017), p. 012101.
- [18] C. L. E. Franzke and A. J. Majda. “Low-Order Stochastic Mode Reduction for a Prototype Atmospheric GCM.” In: *J. Atmos. Sci.* 63 (2006), pp. 457–479.
- [19] C. L. E. Franzke, A. J. Majda, and E. Vanden-Eijnden. “Low-order stochastic mode reduction for a realistic barotropic model climate.” In: *J. Atmos. Sci.* 62 (2005), pp. 1722–1745.
- [20] C. L. E. Franzke, T. J. O’Kane, J. Berner, P. D. Williams, and V. Lucarini. “Stochastic Climate Theory and Modelling.” In: *WIREs Climate Change* 6 (2015), pp. 63–78.
- [21] J. S. Frederiksen, T. J. O’Kane, and M. J. Zidikheri. “Stochastic subgrid parameterizations for atmospheric and oceanic flows.” In: *Physica Scripta* 85.6 (2012), p. 068202.
- [22] J. S. Frederiksen, V. Kitsios, T. J. O’Kane, and M. J. Zidikheri. “Stochastic Subgrid Modelling for Geophysical and Three-Dimensional Turbulence.” In: *Nonlinear and Stochastic Climate Dynamics*. Ed. by C. L. E. Franzke and T. J. Editors O’Kane. Cambridge University Press, 2017, 241–275.
- [23] C. W. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Vol. 4. Springer Berlin, 2009.
- [24] D. Giannakis and A. J. Majda. “Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis.” In: *Geophysical Research Letters* 39.10 ().
- [25] D. Giannakis and A. J. Majda. “Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability.” In: *Proceedings of the National Academy of Sciences* 109.7 (2012), pp. 2222–2227.
- [26] D. Givon, R. Kupferman, and A. Stuart. “Extracting macroscopic dynamics: model problems and algorithms.” In: *Nonlinearity* 17.6 (2004), R55.
- [27] G. A. Gottwald, D. Crommelin, and C. L. E. Franzke. “Stochastic Climate Theory.” In: *Nonlinear and Stochastic Climate Dynamics*. Ed. by C. L. E. Franzke and T. O’Kane. Cambridge: Cambridge University Press, 2017.
- [28] I. Grooms and L. Zanna. “A note on ‘Toward a stochastic parameterization of ocean mesoscale eddies’.” In: *Ocean Modelling* 113 (2017), pp. 30–33.
- [29] M. J. Grote, A. J. Majda, and C. Grotta Ragazzo. “Dynamic Mean Flow and Small-Scale Interaction through Topographic Stress.” In: *Journal of Nonlinear Science* 9.1 (Feb. 1999), pp. 89–130.
- [30] J. Harlim, A. Mahdi, and A. J. Majda. “An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models.” In: *Journal of Computational Physics* 257 (2014), pp. 782–812.
- [31] K. Hasselmann. “Stochastic climate models part I. Theory.” In: *Tellus* 28.6 (1976), pp. 473–485.
- [32] D. D. Holm. “Variational principles for stochastic fluid dynamics.” In: *Proc. R. Soc. A* 471 (2015), p. 20140963.
- [33] M. F. Jansen and I. M. Held. “Parameterizing subgrid-scale eddy effects using energetically consistent backscatter.” In: *Ocean Modelling* 80 (2014), pp. 36–48.
- [34] E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.
- [35] D. Kondrashov, S. Kravtsov, and M. Ghil. “Empirical mode reduction in a model of extratropical low-frequency variability.” In: *J. Atmos. Sci.* 63.7 (2006), pp. 1859–1877.
- [36] S. Kravtsov, D. Kondrashov, and M. Ghil. “Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability.” In: *J. Climate* 18 (2005), pp. 4404–4424.
- [37] E. Mémin. “Fluid flow dynamics under location uncertainty.” In: *Geophysical and Astrophysical Fluid Dynamics* 108 (2014), pp. 119–46.
- [38] A. J. Majda, C. L. E. Franzke, and D. Crommelin. “Normal forms for reduced stochastic climate models.” In: *Proc. Natl. Acad. Sci. USA* 106 (2009), pp. 3649–3653.
- [39] A. J. Majda, C. L. E. Franzke, and B. Khouider. “An Applied Mathematics Perspective on Stochastic Modelling for Climate.” In: *Phil. Trans. R. Soc. A* 366 (2008), pp. 2429–2455.
- [40] A. J. Majda and J. Harlim. “Physics constrained nonlinear regression models for time series.” In: *Nonlinearity* 26.1 (2013), p. 201.
- [41] A. J. Majda and D. Qi. “Strategies for Reduced-Order Models for Predicting the Statistical Responses and Uncertainty Quantification in Complex Turbulent Dynamical Systems.” In: *SIAM Review* 60.3 (2018), pp. 491–549.
- [42] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden. “A mathematical framework for stochastic climate models.” In: *Communications on Pure and Applied Mathematics* 54.8 (2001), pp. 891–974.
- [43] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden. “Models for stochastic climate prediction.” In: *Proceedings of the National Academy of Sciences* 96.26 (1999), pp. 14687–14691.
- [44] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden. “Systematic strategies for stochastic mode reduction in climate.” In: *J. Atmos. Sci.* 60.14 (2003), pp. 1705–1722.
- [45] A. J. Majda and X. Wang. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press, 2006.

- [46] M. Matsumoto and T. Nishimura. “Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator.” In: *ACM Transactions on Modelling and Computer Simulation* 8.1 (1998), pp. 3–30.
- [47] T. J. O’Kane and J. S. Frederiksen. “Statistical dynamical subgrid-scale parameterizations for geophysical flows.” In: *Physica Scripta* 2008.T132 (2008), p. 014033.
- [48] G. A. Pavliotis and A. Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.
- [49] P. Porta Mana and L. Zanna. “Toward a stochastic parameterization of ocean mesoscale eddies.” In: *Ocean Modelling* 79 (2014), pp. 1–20.
- [50] D. Qi and A. J. Majda. “Low-Dimensional Reduced-Order Models for Statistical Response and Uncertainty Quantification: Two-Layer Baroclinic Turbulence.” In: *Journal of the Atmospheric Sciences* 73.12 (2016), pp. 4609–4639.
- [51] V. Resseguier, E. Mémin, and B. Chapron. “Geophysical flows under location uncertainty, Part I Random transport and general models.” In: *Geophysical & Astrophysical Fluid Dynamics* 111.3 (2017), pp. 149–176.
- [52] V. Resseguier, E. Mémin, and B. Chapron. “Geophysical flows under location uncertainty, Part II Quasi-geostrophy and efficient ensemble spreading.” In: *Geophysical & Astrophysical Fluid Dynamics* 111.3 (2017), pp. 177–208.
- [53] P. D. Sardeshmukh and P. Sura. “Reconciling non-Gaussian climate statistics with linear dynamics.” In: *J. Climate* 22.5 (2009), pp. 1193–1207.
- [54] M. J. Senosiain and A. Tocino. “Two-step strong order 1.5 schemes for stochastic differential equations.” In: *Numerical Algorithms* 75.4 (Aug. 2017), pp. 973–1003.
- [55] T. G. Shepherd. “Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics.” In: *Advances in Geophysics* 32 (1990), pp. 287–338.
- [56] G. Shutts. “A kinetic energy backscatter algorithm for use in ensemble prediction systems.” In: *Quart. J. Roy. Meteorol. Soc.* 131 (2005), pp. 3079–3102.
- [57] D. J. Stensrud. *Parameterization schemes*. Cambridge University Press, 2007.
- [58] H. von Storch and F. W. Zwiers. *Statistical analysis in climate research*. Cambridge University Press, 2003.
- [59] G. K. Vallis. *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press, 2006.
- [60] S. Vannitsem. “Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics.” In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.3 (2017), p. 032101.
- [61] J. Wouters and V. Lucarini. “Disentangling multi-level systems: averaging, correlations and memory.” In: *Journal of Statistical Mechanics: Theory and Experiment* 2012.03 (2012), P03003.
- [62] L. Zanna, P. Porta Mana, J. Anstey, T. David, and T. Bolton. “Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction.” In: *Ocean Modelling* 111 (2017), pp. 66–80.

4 Paper II

This chapter is composed by the manuscript of the paper *Spatial covariance modeling for stochastic sub-grid scales parameterizations*, product of a collaboration between Prof. G. Gottwald, my supervisor Dr. C. Franzke and myself. At the moment it has not been submitted for publication, but we plan to submit it to the Journal of Fluid Mechanics.

Spatial covariance modeling for stochastic sub-grid scales parameterizations

Federica Gugole^{*}, Georg A. Gottwald[†] and Christian L. E. Franzke^{*}

^{*}Meteorological Institute and Center for Earth System Research and Sustainability, University of Hamburg, Hamburg, Germany.

[†]School of Mathematics and Statistics, University of Sydney, Sydney, Australia.

Abstract

We apply an energy conserving stochastic parameterization to the forced and damped 2-layer Quasi-Geostrophic (QG) equations. In order to obtain physically feasible results with this kind of parameterization, the introduction of noise with dynamically consistent spatial covariance is of fundamental importance. In this study we investigate this aspect by comparing two different approaches for the definition of the noise covariance matrix. In particular we compare a statistical technique, i.e. Empirical Orthogonal Functions (EOFs), and a dynamical approach, i.e. Dynamic Mode Decomposition (DMD). While EOF focuses on explaining large portions of variance of the field, DMD induces an approximation of the Koopman operator, and hence encodes the dynamics at a specific time. The comparison reveals that individual DMD-forced realizations are more energy conserving. Moreover the use of EOFs leads to a significant growth of the ensembles uncertainties, and to a misplacement of the bi-modal eddy kinetic energy structure. DMD instead well catches the jet movement, and the ensemble uncertainties grow more slowly. Lastly the DMD-forced realizations show less noisy dynamics. This reveals that it is important to design stochastic parameterizations with dynamically adapted spatial correlations, rather than relying on statistical climatic spatial patterns.

Keywords: Koopman operator, Dynamic Mode Decomposition, Empirical Orthogonal Functions, stochastic parameterization, dynamically adapted noise covariance

1 Introduction

Complex dynamical systems, like the oceans and the atmosphere, involve phenomena with vastly different spatial and temporal ranges, which interact with each other. To obtain accurate long time simulations, a model should cover the whole range of scales. This poses a great computational challenge: whenever these continuous multi-scale systems are translated into a discrete numerical model, a truncation due to the size of the spatial grid and of the time step is introduced. Hence any phenomenon, developing in a physical space smaller than the chosen grid box or in a time frame smaller than the selected time step, is not resolved by the numerical model. These events are called sub-grid processes and are commonly parametrized, i.e. formulated in terms of other resolved processes, inside the models in order to represent their effect on the larger scales.

The numerical truncation has also other side effects. For instance, it is well known that in the atmosphere the enstrophy transfers from larger to smaller scales, until it reaches the dissipation scales at the eddy level, while the smaller scales re-inject kinetic energy into the larger scales [30, 8]. For the majority of models, as for instance in the case of general circulation models, the truncation due to the numerics occurs at scales much larger than the dissipation range. Subsequently, the enstrophy piles up at the truncation level, making the numerical model unstable and likely to blow up. In order to guarantee numerical stability, most parameterizations include some kind of hyperviscosity, which is expressed in terms of a power of the Laplacian of the field variable. This translates into an increased viscosity of the fluid, which dissipates also the kinetic energy. The energy input into the Earth system due to the solar radiation is the driver of all motions in the atmosphere and in the oceans, hence it is of utmost importance to resolve it as accurately as possible in climate and ocean models.

In the latest years there has been an extensive interest in the development of parameterizations for the sub-grid scale processes, and a growing concern for energy consistency. Many deterministic and stochastic parameterizations have, and are, being developed following two different approaches. The first technique is to derive an expression for extra terms that could be included in the equations in order to, for example, represent eddies or to re-inject energy, see for instance [3, 4, 13, 14, 23]. The second strategy instead is to derive new stochastic expressions of the Navier-Stokes equations such that they still conserve, for instance, energy [20] or the Kelvin circulation theorem [12]. Both procedures have pros and cons: the former is of easier and faster implementation in already existing models but might be case specific, while the latter has more general properties but it requires fundamental changes in the models dynamical core.

Here we employ the projection operator approach outlined in [9], which belongs to the first group of stochastic parameterizations. We perform our analyses in the framework of the forced and damped non-dimensional 2-layer Quasi-Geostrophic (QG) model, where we employ a Eulerian description of the dynamics with a grid point based discretization, as in most ocean and climate models. With very similar settings, [11] showed that giving a spatial structure to the noise covariance is necessary for this parameterization to have physically meaningful results. Hence we inquire into this aspect and analyze the importance of a-priori assumptions regarding the noise covariance. For this purpose we will compare two dimension reduction techniques, Empirical Orthogonal Function (EOF) [32] and Dynamic Mode Decomposition (DMD) [16]. These methods identify the main directions of the system evolution adopting two different approaches, statistical and dynamical respectively. While EOFs derive the dominant patterns of variability from a statistical field, DMD is closely related to the Koopman operator. The latter is an abstract concept in dynamical systems theory encoding the dynamics of a system, and it propagates observables from one instance of time to another instance of time [17]. These different starting points for the noise covariance definition affects deeply its construction. As we will show in the remainder of this work, the use of pieces of information about the Koopman operator requires the DMD modes, and hence also the noise covariance matrix, to be recomputed periodically. This has the advantage of using information of the system during its development, with no additional information.

The remainder of this paper is structured as follows. In Section 2 we introduce the deterministic evolving equations, while Section 3 describes the stochastic parameterization here considered. A brief overview of the theory of EOF and DMD, with details about their application to our problem, are given in 3.1 and 3.2 respectively. A description of the resulting noise covariance structure is also to be

found in the same Section. The numerical outcomes are reported in Section 4, and we conclude with a discussion in Section 5.

2 The QG model

We start from the 2-layer QG non-dimensional equations on a β plane with double periodic boundary conditions as presented in [30]. We consider also a biharmonic viscosity term representing the sub-grid eddies and the bottom friction in the lower level (i.e. $i = 2$), while in the upper layer (i.e. $i = 1$) a prescribed background-flow zonal velocity $U = 0.6$ is considered, as, for instance, in [6, 13]. Since we consider a non-dimensional description, the horizontal extensions have been rescaled to a $2\pi \times 2\pi$ square. Furthermore we assume that the two layers have equal thickness. Finally the evolution equations for the potential vorticities (PVs)

$$q_i(\mathbf{x}, t) = \nabla^2 \psi_i + (-1)^i \frac{k_d^2}{2} (\psi_1 - \psi_2) \quad i \in \{1, 2\}$$

on the horizontal plane $\mathbf{x} \in \mathbb{R}^2$ read

$$dq_1 = -J(\psi_1 - Uy, q_1) dt - \nabla^2 (\nu \nabla^4 \psi_1) dt, \quad (1a)$$

$$dq_2 = -J(\psi_2, q_2) dt - \nabla^2 (\nu \nabla^4 \psi_2) dt - \tau_f^{-1} \nabla^2 \psi_2 dt, \quad (1b)$$

where $\psi_i(\mathbf{x}, t)$ $i \in \{1, 2\}$ are the corresponding streamfunctions, $\tau_f = 10$ time units the frictional time-scale and ν the biharmonic viscosity coefficient. ∇ and ∇^2 denote, respectively, the horizontal gradient and the Laplacian operator, while J stands for the Jacobian operator

$$J(A, B) = \frac{\partial A}{\partial x} \frac{\partial B}{\partial y} - \frac{\partial A}{\partial y} \frac{\partial B}{\partial x}.$$

The term $k_d^2/2 = (2f_0/Nh)^2$ determines the strength of the shear between the two layers and hence also the intensity of the baroclinic instability ($N = 1.2 \cdot 10^{-2}$ being the Brunt-Väisälä frequency, $h = 200$ the mean depth of the layers and $f \approx f_0 + \beta y$ the approximate Coriolis term with $f_0 = 1$ and $\beta = 0.509$). Given these values of the parameters, the Rossby deformation radius is about $k_d^{-1} \approx 0.85$, corresponding to an atmosphere-like setting.

We set ν as in [13] which, following the argument of [18], defines it as

$$\nu = C_{Leith} \Delta^6 |\nabla^4 \psi_i|$$

where $C_{Leith} = 0.005$ is an empirical constant and Δ is the grid-spacing. We would like to point out that ν is non-constant. This will reveal to be an important feature when rewriting the system in terms of barotropic and baroclinic mode.

In order to have a better defined distinction between slow and fast modes, we rewrite equations (1) as barotropic and baroclinic modes by assuming that barotropic modes evolve more slowly than baroclinic modes. Barotropic and baroclinic streamfunctions, ψ_B and ψ_T , can be defined as:

$$\psi_B = \frac{1}{2}(\psi_1 + \psi_2), \quad \psi_T = \frac{1}{2}(\psi_1 - \psi_2);$$

which lead to the corresponding barotropic and baroclinic potential vorticities, q_B and q_T ,

$$q_B = \nabla^2 \psi_B + \beta y, \quad q_T = \nabla^2 \psi_T - k_d^2 \psi_T. \quad (2)$$

It can easily be shown that barotropic and baroclinic PVs can also be written as

$$q_B = \frac{1}{2}(q_1 + q_2), \quad q_T = \frac{1}{2}(q_1 - q_2),$$

and we can use these relations to determine the evolution equations for q_B and q_T from (1). After some manipulations we obtain

$$\begin{aligned} dq_B = & - \left(J(\psi_B - \frac{1}{2}Uy, q_B) + J(\psi_T - \frac{1}{2}Uy, q_T) \right) dt - \frac{1}{2}\tau_f^{-1} (\nabla^2 \psi_B - \nabla^2 \psi_T) dt \\ & - \frac{1}{2} \nabla^2 (C_{Leith} \Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) + C_{Leith} \Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt, \end{aligned} \quad (3a)$$

$$\begin{aligned} dq_T = & - \left(J(\psi_T - \frac{1}{2}Uy, q_B) + J(\psi_B - \frac{1}{2}Uy, q_T) \right) dt + \frac{1}{2}\tau_f^{-1} (\nabla^2 \psi_B - \nabla^2 \psi_T) dt \\ & - \frac{1}{2} \nabla^2 (C_{Leith} \Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) - C_{Leith} \Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt, \end{aligned} \quad (3b)$$

where the biharmonic coefficient has been decomposed in its constant and non-constant parts. The system (3) is Hamiltonian with Hamiltonian H given by

$$H(q_B, q_T) = \frac{1}{2} \int_A [(\nabla \psi_B)^2 + (\nabla \psi_T)^2 + k_d^2 \psi_T^2] dA.$$

It can be shown that

$$\delta H = - \int_A (\psi_B \cdot \delta q_B + \psi_T \cdot \delta q_T) dA,$$

which implies

$$\frac{\partial H}{\partial q_B} = -\psi_B, \quad \frac{\partial H}{\partial q_T} = -\psi_T.$$

For a general review of Hamiltonian mechanics and its application to geophysical fluid dynamics see for example [25, 28, 2].

Due to the high dimensionality of the QG model, the numerical truncation affects deeply the dynamics by introducing a larger error at coarser resolutions. Since high resolution simulations are computationally expensive, and it is not always possible to use a fine enough grid, we aim at reproducing the dynamics of equations (3)-(2) as resolved on a fine grid by correcting the numerical error at lower resolutions through the introduction of a stochastic parameterization for the sub-grid scales. The introduction of a stochastic source term breaks the energy balance of the system, hence extra care is to be taken in order to guarantee that the total energy of the stochastic QG model is conserved.

3 Energy conserving stochastic parameterization

As developed in [9] and further extended to high dimensional systems in [11], the stochastic terms are added by means of a projection operator, which ensures the conservation of the total energy of the system in absence of external forcing, viscosity and bottom friction. For sake of simplicity, we simplify here the stochastic perturbation considering white noise instead of red noise. As in [11], we represent the unresolved fast sub-grid processes by means of a stochastic forcing, which we assume to act directly on the baroclinic mode and indirectly on the barotropic mode. The underlying stochastic modeling assumption is that there are many fast baroclinic modes which are mixing, and can be efficiently represented by a stochastic Ansatz. We refer the interested reader to [11] for more details of the derivation, here we report only the main steps and results. After including a 2-dimensional Wiener process W_t enacting the sub-grid processes, we further add an auxiliary 2-dimensional stochastic process dY_t in order to balance the stochastic forcing just introduced and ensure conservation of energy. We wish to stress that energy will still be affected by the deterministic forcing and damping terms. At this stage, the evolution equations for the potential vorticities read

$$\begin{aligned}
dq_B &= - \left(J(\psi_B - \frac{1}{2}Uy, q_B) + J(\psi_T - \frac{1}{2}Uy, q_T) \right) dt - \frac{1}{2}\tau_f^{-1} (\nabla^2\psi_B - \nabla^2\psi_T) dt \\
&\quad - \frac{1}{2}\nabla^2 (C_{Leith}\Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) + C_{Leith}\Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt , \\
dq_T &= - \left(J(\psi_T - \frac{1}{2}Uy, q_B) + J(\psi_B - \frac{1}{2}Uy, q_T) \right) dt + \frac{1}{2}\tau_f^{-1} (\nabla^2\psi_B - \nabla^2\psi_T) dt \\
&\quad - \frac{1}{2}\nabla^2 (C_{Leith}\Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) - C_{Leith}\Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt \\
&\quad + \Sigma dW_t + dY_t , \\
dY_t &= B_t dt + S_t dW_t ,
\end{aligned}$$

where we included the evolution equation for dY_t ; $B_t \in \mathbb{R}^2$ and Σ , $S_t \in \mathbb{R}^{2 \times 2}$. Instead of dealing with two different stochastic processes, we want to write S_t and B_t as functions of Σ . For that purpose, we write the increment of H as a sum of two parts, a deterministic part μ_H including all the terms multiplied by dt , and a stochastic part σ_H containing those with the Wiener process. By Ito's theorem we have

$$\begin{aligned}
dH &= \frac{\partial H}{\partial q_B} \cdot dq_B + \frac{\partial H}{\partial q_T} \cdot dq_T + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : dq_T dq_T^T \\
&= \mu_H dt + \sigma_H dW_t ,
\end{aligned}$$

where $A : B = a_{ij}b_{ij} = Tr(AB^T)$, and

$$\begin{aligned}
\mu_H = & + \psi_B \cdot \left(J(\psi_B - \frac{1}{2}Uy, q_B) + J(\psi_T - \frac{1}{2}Uy, q_T) + \frac{1}{2}\tau_f^{-1} (\nabla^2\psi_B - \nabla^2\psi_T) \right) \\
& + \frac{1}{2}\psi_B \cdot \nabla^2 (C_{Leith}\Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) + C_{Leith}\Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt \\
& + \psi_T \cdot \left(J(\psi_T - \frac{1}{2}Uy, q_B) + J(\psi_B - \frac{1}{2}Uy, q_T) - \frac{1}{2}\tau_f^{-1} (\nabla^2\psi_B - \nabla^2\psi_T) - B_t \right) \\
& + \frac{1}{2}\psi_T \cdot \nabla^2 (C_{Leith}\Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) + C_{Leith}\Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt \\
& + \frac{1}{2} \frac{\partial^2 H}{\partial q_T \partial q_T} : (\Sigma + S_t)(\Sigma + S_t)^T , \\
\sigma_H = & - \psi_T \cdot (\Sigma + S_t) \\
& = \nabla_{q_T} H \cdot (\Sigma + S_t) .
\end{aligned}$$

Our aim is to control the stochastic forcing. This might be of particular importance for more complex models, such as primitive equation based climate models, where one wants to inject the stochastic forcing only into the balanced or unbalanced flow components. Hence, in order to guarantee the total energy not to be affected by the stochastic forcing, we set σ_H and the sum of those terms in μ_H due to the stochastic processes to be zero.

Following the reasoning outlined in [9], the auxiliary stochastic process dY_t should not perturb the dynamics on the tangent space, and should be constructed only to counterbalance those components of the stochastic process which are orthogonal to the manifold of constant energy. Thus we define a projection operator $\mathbb{P} \in \mathbb{R}^{2 \times 2}$. Since the Wiener process affects only the evolution equation of the baroclinic potential vorticity, it will be sufficient to project with respect to the manifold of constant baroclinic kinetic energy:

$$\begin{aligned}
\mathbb{P} &= \mathbf{I} - \frac{1}{|\nabla_{q_T} H|^2} \nabla_{q_T} H (\nabla_{q_T} H)^T \\
&= \mathbf{I} - \frac{1}{|\psi_T|^2} \psi_T \psi_T^T ,
\end{aligned}$$

where $\mathbf{I} \in \mathbb{R}^{2 \times 2}$ stands for the identity matrix. Note that $\mathbb{P}(\nabla_{q_T} H) = 0$. Consequently, we want S_t and B_t to satisfy

$$\mathbb{P}S_t = 0 , \quad \mathbb{P}B_t = 0 .$$

As in [11], the constraint $\sigma_H = 0$ provides an expression for S_t , while it is possible to determine B_t by considering only the terms of μ_H due to the introduction of the stochastic processes:

$$\begin{aligned}
S_t &= -(\mathbf{I} - \mathbb{P})\Sigma , \\
B_t &= + \frac{1}{2|\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T\mathbb{P} \right) \psi_T .
\end{aligned}$$

Finally we get the following set of equations

$$dq_B = - \left(J(\psi_B - \frac{1}{2}Uy, q_B) + J(\psi_T - \frac{1}{2}Uy, q_T) \right) dt - \frac{1}{2}\tau_f^{-1} (\nabla^2\psi_B - \nabla^2\psi_T) dt \\ - \frac{1}{2}\nabla^2 (C_{Leith}\Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) + C_{Leith}\Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt , \quad (4a)$$

$$dq_T = - \left(J(\psi_T - \frac{1}{2}Uy, q_B) + J(\psi_B - \frac{1}{2}Uy, q_T) \right) dt + \frac{1}{2}\tau_f^{-1} (\nabla^2\psi_B - \nabla^2\psi_T) dt \\ - \frac{1}{2}\nabla^2 (C_{Leith}\Delta^6 |\nabla^4(\psi_B + \psi_T)| \nabla^4(\psi_B + \psi_T) - C_{Leith}\Delta^6 |\nabla^4(\psi_B - \psi_T)| \nabla^4(\psi_B - \psi_T)) dt \\ + \mathbb{P}\Sigma dW_t + \frac{1}{2|\psi_T|^2} \left(\frac{\partial^2 H}{\partial q_T \partial q_T} : \mathbb{P}\Sigma\Sigma^T \mathbb{P} \right) \psi_T dt . \quad (4b)$$

Equations (2)-(4) constitute our stochastic energy conserving 2-layer QG system. The interested reader may find more details about the necessary steps for the derivation of (4) in [11]. As it can be seen, the resulting set of equations contains multiplicative noise and nonlinear damping, due to the specific definition of the projection operator. The multiplicative noise is in fact a correlated additive multiplicative (CAM) noise [19, 26].

In equations (4) the variable Σ is still unknown, and as shown in [11], it is crucial when using this parameterization in a Eulerian framework to give a dynamically consistent spatial structure to the noise covariance. Hence we look for

$$\Sigma = \sum_{i=1}^l \gamma_i \psi_i(q) \quad (5)$$

where $l \in \mathbb{N}$ is the total number of patterns considered, $\gamma_i \in \mathbb{R} \forall i$, and $\psi_i(q) \in \mathbb{R}^{2 \times 2} \forall i$ is a dynamically consistent pattern, which depends on the state variable q .

It is common in the context of stochastic parameterizations to use Empirical Orthogonal Functions (EOFs), see for instance [7, 24]. EOF is a statistical technique looking at stationary coherent structures explaining large portions of variance, and has been widely used in climate models, thanks to their easy and robust computation and the large availability of data. Nonetheless they have limitations. In particular, their physical interpretation is restricted. While it is possible to associate the first EOF with a known physical process, this becomes more and more complicated with higher-order eigenvectors because of the orthogonality constraint [32]. More details about EOF and their usage in the present work are provided in Section 3.1.

In this work we aim to analyze how the definition of the noise covariance, i.e. equation (5), influences the outcomes. For this purpose we employ here Dynamic Mode Decomposition (DMD) [27, 29, 16]. DMD is a data-driven algorithm for computing the Koopman modes, and it is, hence, related to the generator of the dynamics. The Koopman operator has been developed in the framework of dynamical systems theory [15]. In Section 3.2 we provide a brief exposition about the relation between the Koopman operator and dynamic mode decomposition; for a more exhaustive commentary on the topic, we invite the reader to see, among others, the review papers [5, 21]. Details about the inclusion of DMD in the stochastic QG model are also provided.

3.1 Empirical Orthogonal Functions

3.1.1 Theory

In this section we provide a brief description of the theoretical framework of EOF. More information can be found in [31, 32]. EOF, also known as Principal Component Analysis (PCA) or Proper Orthogonal Decomposition (POD), is a multivariate analysis technique that derives the dominant patterns of variability from a statistical field, usually indexed by location in space. Let \mathbf{X} be an n -dimensional random vector, whose mean is assumed to be zero; otherwise the anomalies of the field with respect to the mean should be considered. At its first stage the EOF analysis computes the vector ϕ_1 with $\|\phi_1\| = 1$ such that

$$\epsilon_1 = \mathbb{E} \left(\|\mathbf{X} - \langle \mathbf{X}, \phi_1 \rangle \phi_1\|^2 \right) \quad (6)$$

is minimized, where we denoted with \mathbb{E} the expectation operator, the vector norm by $\|\cdot\|$ and the inner product with $\langle \cdot, \cdot \rangle$. Equation (6) describes the projection of the field \mathbf{X} onto a 1-dimensional subspace spanned by the vector ϕ_1 . Minimizing ϵ_1 is equivalent to maximizing the variance of \mathbf{X} contained in this subspace, in fact it can be shown that

$$\epsilon_1 = \text{Var}(\mathbf{X}) - \text{Var}(\langle \mathbf{X}, \phi_1 \rangle) ,$$

where the variance of \mathbf{X} is defined to be the sum of the variances of its elements. Let Γ denote the covariance matrix of \mathbf{X} . Then we can write

$$\text{Var}(\langle \mathbf{X}, \phi_1 \rangle) = \phi_1^\dagger \Gamma \phi_1 ,$$

where \dagger denotes the complex transpose. The minimization of ϵ_1 under the constraint $\|\phi_1\| = 1$ leads to

$$\frac{d}{d\phi_1} \left[-\phi_1^\dagger \Gamma \phi_1 + \lambda_1 (\phi_1^\dagger \phi_1 - 1) \right] = -2\Gamma \phi_1 + 2\lambda_1 \phi_1 = 0$$

where λ_1 is the Lagrange multiplier associated with the constraint $\|\phi_1\| = 1$. This is equivalent to say that ϕ_1 is an eigenvector of the covariance matrix Γ with corresponding eigenvalue λ_1 . Therefore, the minimum of equation (6) is achieved by the vector associated to the largest eigenvalue of Γ , i.e. vector ϕ_1 .

The same procedure is repeated to find the second EOF, which is the vector ϕ_2 with $\|\phi_2\| = 1$ minimizing

$$\epsilon_2 = \mathbb{E} \left(\left\| (\mathbf{X} - \langle \mathbf{X}, \phi_1 \rangle \phi_1) - \langle \mathbf{X}, \phi_2 \rangle \phi_2 \right\|^2 \right) ,$$

and corresponding to the second largest eigenvalue λ_2 of Γ . Finally we remark that Γ is an Hermitian matrix, hence its eigenvectors are orthogonal to one another. Moreover in case of translationally invariant systems they are Fourier modes.

3.1.2 Constructing Σ using EOF

In case of EOFs we build Σ as in (5) where the weights γ_i are selected to be the square roots of the EOFs eigenvalues λ^{EOF} , i.e.

$$\Sigma = \sum_{i=1}^l \sqrt{\lambda_i^{EOF}} \phi_i^{EOF} .$$

In this way predominant patterns explaining larger portions of variance have a higher weight in the linear combination, while vectors responsible for small portions of variance, and potentially subject to noise and to a stronger numerical error, will play a less important role. EOFs are computed on the data of a low resolution deterministic run, after the dynamics settled on the attractor. Since the noise is inserted in the equation of the baroclinic mode, we use the baroclinic stream function ψ_T for their computation.

As in the majority of cases, the spectrum of the singular values decreases very fast and only few modes are physically relevant, while the others correspond to very small portions of variance ($\lambda_{10}^{EOF} \approx 0.001$) and hence might be considered in the limit as numerical noise (see Figure 1a). EOFs 1-2 (Figures 1b-1c) represent the predominant traveling wave included in the model, while EOFs 4-5 (Figures 1e-1f) denote the subsequent mode. EOF 3 (Figure 1d) clearly does not represent any wave but meridional shifts of the jet. In our numerical simulations, we use either only the first two eigenvectors, or the first five EOFs.

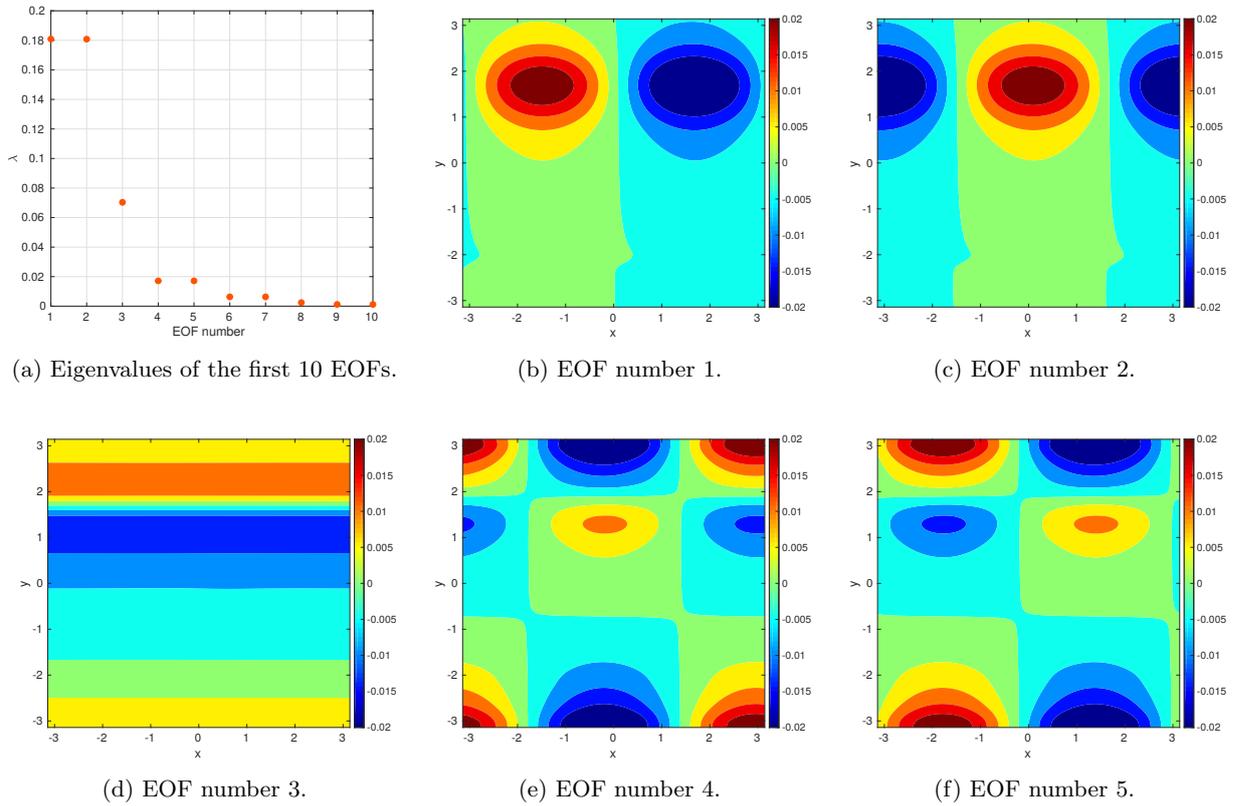


Figure 1: EOF singular values spectrum of the first 10 eigenvectors and first 5 EOF patterns.

3.2 Dynamic Mode Decomposition

3.2.1 DMD and the Koopman operator

Here we briefly present the Koopman operator and its connection with dynamic mode decomposition. Detailed reviews about the Koopman operator can be found, for instance, in [5, 21], while theory and applications of DMD are provided, among others, in [27, 29, 16].

Let $\dot{x} = f(x)$ denote a general continuous-time dynamical system with initial condition $x(0) = x_0 \in \mathbb{R}^n$. On the assumption that there exists a unique solution of this initial value problem, it is possible to introduce the flow map φ_t such that $x(t) = \varphi_t(x_0)$. Define an arbitrary observable $\psi(x)$. The value of this observable ψ , which the system starting in x_0 sees at time t , is

$$\psi(t, x_0) = \psi(\varphi_t(x_0)) .$$

The Koopman operator is a semigroup of operators \mathcal{K}_t , acting on the space of observables parametrized by time t

$$\mathcal{K}_t \psi(x_0) = \psi(\varphi_t(x_0)) .$$

If the observables coincide with the state variables, we have $\psi(x) = \text{id}(x)$. The generator of the Koopman semigroup is defined by

$$[\mathcal{K}\psi] := \lim_{t \rightarrow 0} \frac{\mathcal{K}_t \psi - \psi}{t} ,$$

provided the limit exists. It is important to underline that the operator \mathcal{K}_t is linear also in case of non-linear dynamics f , thus it makes sense to consider its spectral properties, but the eigenfunctions of the Koopman operator are not necessarily linear. Most often, there is no explicit representation of the Koopman operator, and its behavior can be determined only by its action on an observable at a finite number of initial conditions.

Dynamic mode decomposition is a data-driven technique for computing the Koopman modes. The operational definition of DMD is provided in [29]. Consider a dynamical system as above, where the observables coincide with the state variables (i.e. $\psi(x) = \text{id}(x)$), and two sets of data,

$$\mathbf{X} = \begin{pmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ | & | & \cdots & | \end{pmatrix} \quad \mathbf{X}' = \begin{pmatrix} | & | & \cdots & | \\ \mathbf{x}'_1 & \mathbf{x}'_2 & \cdots & \mathbf{x}'_m \\ | & | & \cdots & | \end{pmatrix}$$

such that

$$\begin{aligned} \mathbf{x}_k &= \mathbf{x}(t_k) \in \mathbb{R}^n , & \mathbf{x}'_k &= \mathbf{x}(t_k + \delta t) = \mathcal{K}_{\delta t} \mathbf{x}_k , \\ \mathbf{x}_k &= \mathbf{x}(t_{k-1} + \Delta t) = \mathcal{K}_{\Delta t} \mathbf{x}_{k-1} , & \mathbf{x}'_k &= \mathbf{x}(t_{k-1} + \delta t + \Delta t) = \mathcal{K}_{\Delta t} \mathbf{x}'_{k-1} , \end{aligned}$$

where $m\Delta t$ defines the time window, and $\delta t \leq \Delta t$ determines the accuracy of the reconstructed dynamics. It is important to mention that matrices \mathbf{X} and \mathbf{X}' are assumed to be *tall and skinny*, i.e. it is assumed that the size n of a snapshot is larger than the number $m - 1$ of snapshots. In the DMD algorithm the Koopman operator is approximated by means of a least square fit operator \mathbf{K} relating data $\mathbf{X}' \approx \mathbf{K}\mathbf{X}$. The DMD modes are the eigenvectors of \mathbf{K} , and each DMD mode corresponds to a particular eigenvalue of \mathbf{K} . Since the operator is linear, the decomposition gives the growth rates and

frequencies associated with each mode. The numerically stable SVD-based algorithm outlined for the first time in [27], and improved in [29], allows for a low-rank $r \leq m$ representation of the operator \mathbf{K} onto the first r EOF modes of matrix \mathbf{X} . Details about the algorithm, as well as a MATLAB[®] function, are provided in [16].

3.2.2 Defining the noise covariance by means of DMD

From the theory, we know that the eigenvalues of the Koopman operator should lay on the complex unit circle and that, aside from the eigenvalue corresponding to the mean mode, they are complex conjugate. Therefore, when using DMD to build Σ , since $|\lambda_i^{DMD}| = 1 \forall i$, it is not that straightforward to build Σ such that the noise has the same amplitude as with EOFs. Here we report the procedure we apply in order to guarantee equal amplitude to the stochastic forcing with both techniques:

1. re-order the DMD eigenmodes such that the first corresponds to the mean (i.e. its eigenvalue is the closes to $\lambda_m = 1$), and the others come according to how close they are to λ_m while remaining on the unit circle, up to a tolerance, in order to avoid numerically spurious modes;
2. after neglecting the mean mode (i.e. $i = 0$), we select the l eigenvectors, whose eigenvalues have real part closest to 1. We consider just one mode for each complex conjugate pair, and define Σ as

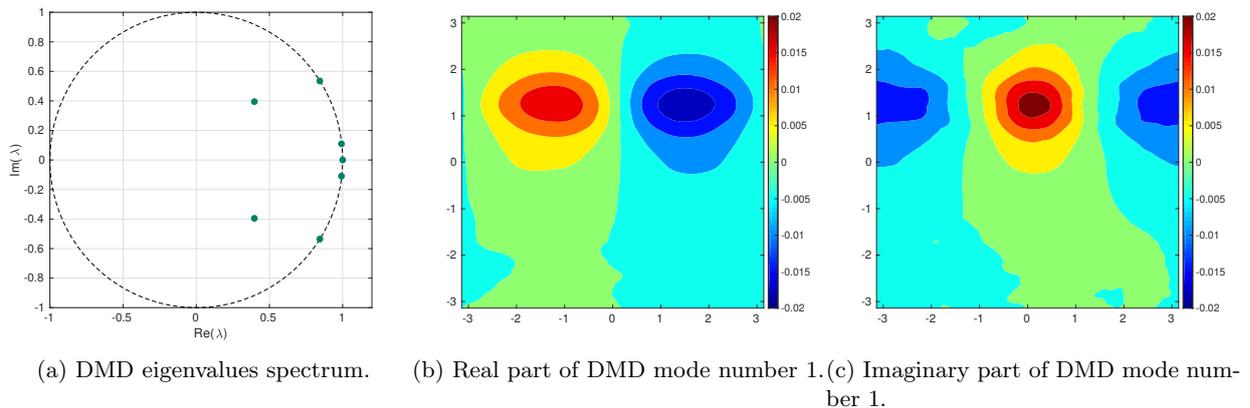
$$\Sigma = \sum_{i=1}^l (\text{Re}(\lambda_i^{DMD})\text{Re}(\phi_i^{DMD}) - \text{Im}(\lambda_i^{DMD})\text{Im}(\phi_i^{DMD})) ;$$

3. finally we normalize the trace of $\Sigma\Sigma^T$ to be the same as for EOF.

The coefficients related to the redundant eigenvectors are set to be zero, and the DMD modes are recomputed for each $m\Delta t$ time window. We would like to remark that, in principle, with the Koopman operator it is possible to propagate the DMD modes for each δt in the following $m\Delta t$ time window before new computation; in the present work we keep them constant for sake of simplicity. As for EOFs, also DMD uses the values of ψ_T , but they are computed on the fly, so the data have still low resolution but they are also perturbed by the noise introduced at previous times.

In our numerical simulations we use the following set of parameters: $l = 2$, $m = 16$, $r = 7$, $\delta t = 0.1$, $\Delta t = 3\delta t$. Hence with DMD we are considering time windows of length 4.8 time units, which corresponds to roughly half an eddy turnover time. We normalized the trace of the matrix $\Sigma\Sigma^T$ to be equal either to $\sigma^2 = \lambda_1^{EOF} + \lambda_2^{EOF} \approx 0.36$ or $\sigma^2 = \sum_{i=1}^5 \lambda_i^{EOF} \approx 0.47$. Other sets of parameters corresponding to different time windows spanning between 2 time units and 10 time units have been tested, but this particular choice was the only one among those tested which does not present two eigenvalues with null imaginary part and real part very close to 1. This second mean-mode cannot be excluded by our procedure since the module of its corresponding eigenvalue is still very close to 1, but by plotting and comparing it to the other modes it can be seen that it is numerically spurious and not dynamically meaningful. DMD is quite sensitive to the input data, hence it is very important to filter out the numerical noise, and thus the spurious scales close to the truncation. This is done by choosing an appropriate time window, so that the noise decorrelates inside the datasets, and by sub-sampling, i.e. we consider δt small such that the Koopman operator is well approximated, but

we sample consecutive snapshots in the same dataset each $\Delta t > \delta t$. We set r as the optimal value determined by the criterion developed in [10]. We also tried with $m = 48$, $r = 7$, $\Delta t \equiv \delta t = 0.1$, i.e. we considered a time window of the same length and instead of sub-sampling we chose a small value of r , but the results show that sub-sampling is more efficient in filtering out the numerical noise. When starting the DMD forced stochastic simulations the matrix Σ was initialized by means of the first two EOFs, and then regularly recomputed with DMD.



(d) Real part of DMD mode number 2. (e) Imaginary part of DMD mode number 2.

Figure 2: An example of DMD eigenvalues spectrum (a), real and imaginary parts of the first (b-c) and second (d-e) DMD mode for $m = 16$, $r = 7$, $\Delta t = 3\delta t$, $\delta t = 0.1$. The mean mode is neglected in the computations and therefore not shown here; likewise only one mode for each complex conjugate pair of eigenvalues is displayed.

In Figure 2 we show the eigenvalues and the first two DMD modes as computed in a particular time window with the aforementioned set of parameters. The mode representing the mean has been neglected and only one of the two modes corresponding to a complex conjugate pair of eigenvalues is displayed. Since DMDs are recomputed along the simulation, its modes are not exactly the same for the entire run, but they shift in the zonal direction. Besides the eddy meridional position of the first mode

changes depending on the meridional shift of the jet. The eddy length instead remains unchanged, since the system and the DMD parameters do not change and hence we are always targeting on the same waves. Real and imaginary parts of DMD mode number 1 (Figures 2b-2c) resemble closely EOFs 1-2, although in the DMD mode, and in particular in its imaginary part, the eddy patterns look smaller and less regular. Furthermore the eddies are centered in slightly, but distinctly, different meridional coordinates, this being shifted to higher values for EOFs 1-2. The DMD modes shown in Figure 2 have been computed at early stages of the simulation ($t \approx 1150$ time units). When looking at the DMD modes at later times, the eddy meridional position follows the jet and is hence shifted towards higher latitudes, more similar to EOFs 1-2 (not shown). EOFs 4-5 are the most comparable eigenvectors to the second DMD mode (Figure 2d-2e), but significant differences can be spotted for $y \in [0.8, 1.8]$, where some eddy structure is present in the EOF vectors but is absent in the DMD mode. This could be an artificial artifact due to the orthogonality constraint of the EOF algorithm. A pattern such as EOF 3 is not to be found with DMD, and for short time windows of length $m\Delta t$ it is likely included into the mean mode due to the low frequency variability of the jet.

4 Model setup and results

We discretized equations (2)-(4) by means of finite differences in a grid-point based framework. Our numerical code for the QG model is based on the energy and enstrophy conserving discretization scheme by Arakawa [1]. This scheme ensures that energy and enstrophy are conserved for all truncations. In particular this scheme does not require any numerical diffusion or dissipation for numerical stability. For the time stepping we employed a 4th order explicit Runge-Kutta method, while we used the Fast Fourier Transform to invert the Laplacian operator and the Euler-Maruyama scheme for the stochastic terms [22]. All stochastic simulations ran with a spatial resolution of 128×128 grid points and a time step of $dt = 10^{-3}$. All started from the same initial condition, which we have assured to be on the attractor by having employed a preceding integration of the deterministic equations at resolution 128×128 for the long integration time of 8000 time units. We compare the outcomes of the stochastic simulations with a deterministic low resolution and a deterministic high resolution run. For the latter, which will be referred also as reference solution, we used the same initial condition interpolated on a finer grid of 512×512 grid points, and it ran with $dt = 10^{-4}$. Its results have been projected on the coarser 128×128 spatial grid for a fairer comparison.

4.1 Total energy

Looking at the total energy graphs of the different realizations in the various setups with EOF (orange) and DMD (green) reported in Figure 3, it can be noticed that on average the energy is conserved with both techniques. Although the EOF ensemble members show more variance, when only the first two EOFs are used, the system seems to be slightly dissipative in time. This is particularly evident when looking at the ensemble mean (blue line in Figure 3a). The inclusion of EOFs 3-4-5 reduces the dissipative effect, but realizations with a clear increasing trend can be present (Figure 3c). On the other hand, the spread of the DMD ensemble members has less variance but well encloses the energy graph of the reference solution. Individual runs are more energy conserving and the system seems to be less dissipative (Figures 3b-3d). This might suggest that the usage of a dynamically adapted noise

structure may help the numerical model to remain on the manifold of constant energy. In any case deviations from the mean are less than 2%. Hence, according to the modeler target, they might be considered as negligible.

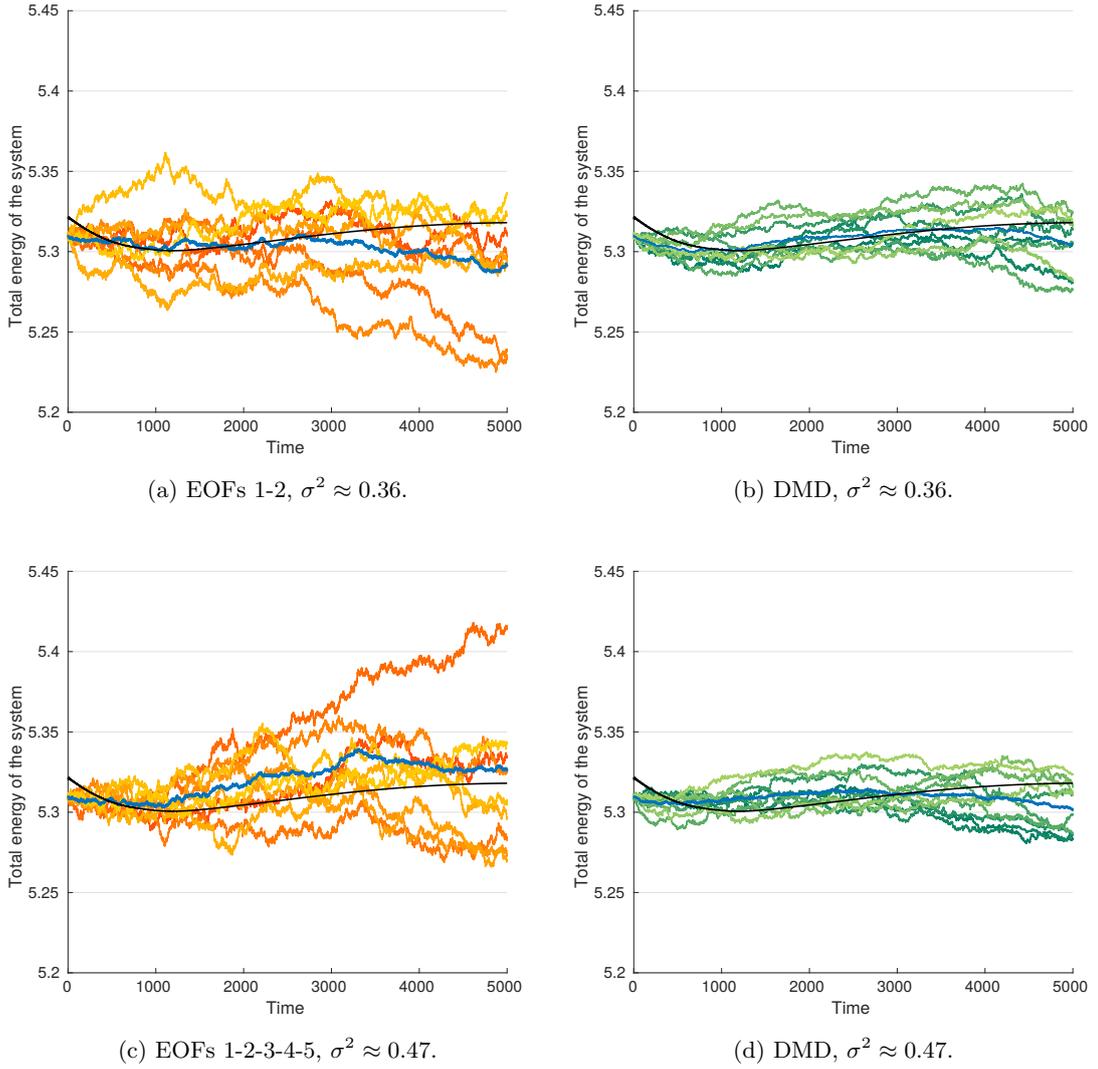


Figure 3: Total energy graphs for stochastic simulations using EOFs 1-2, $\sigma^2 \approx 0.36$ (a), EOFs 1-2-3-4-5, $\sigma^2 \approx 0.47$ (c); DMD $l = 2$, $m = 16$, $r = 7$, $\delta t = 0.1$, $\Delta t = 3\delta t$, $\sigma^2 \approx 0.36$ (b), and DMD $\sigma^2 \approx 0.47$ (d). The colored lines represent different realizations of the stochastic system in the various settings. The outcome of the high resolution deterministic run (black line) and the ensemble mean (blue line) are also shown as reference.

4.2 Eddy kinetic energy

In order to compute the eddy kinetic energy (EKE), we first computed the horizontal velocities for the barotropic and baroclinic modes from the respective streamfunctions using

$$u = -\frac{\partial\psi}{\partial y}, \quad v = \frac{\partial\psi}{\partial x},$$

where u is the zonal and v the meridional velocity. Then we considered a time window of k time units to compute the temporal mean velocities, i.e. \bar{u}_B, \bar{v}_B and \bar{u}_T, \bar{v}_T for barotropic and baroclinic modes respectively. Afterwards for each time unit we computed the deviations from the mean, e.g. $u'_B(t) = u_B(t) - \bar{u}_B$, and used these quantities to compute the EKE for each grid point for all t . As last step we either averaged in time and then also in the zonal direction, therefore the EKE is displayed simply as a function of the meridional direction y (Figures 4-5); or we averaged only in the zonal direction and looked at the time evolution of the EKE projected on y (Figure 6).

Due to the meridional movement of the jet, in our analyses we split the time series in time windows of $k = 1000$ time units and consider each time window individually. Although the time-averaged EKE shows a bi-modal behavior in all windows, the meridional location of the peaks varies according to the jet movement. Hence we want to check how well the stochastic parameterization keeps track of the jet shift. The time-averaged EKE of the baroclinic mode for $t \in [1000, 2000]$ and for $t \in [3000, 4000]$ in the different stochastic setups with EOF (orange) and DMD (green) are reported in Figures 4 and 5 respectively. The EKE of the barotropic mode shows similar results as for the baroclinic mode, hence, it is not reported here. Figure 6 shows the time evolution of the barotropic and baroclinic EKE for $t \in [3200, 3700]$ in case of one stochastic simulation with EOFs 1-2-3-4-5 and $\sigma^2 \approx 0.47$, one realization with DMDs and $\sigma^2 \approx 0.47$, and in case of the reference solution.

Both for $t \in [1000, 2000]$ and $t \in [3000, 4000]$, it can be seen that the ensemble forced by EOFs 1-2 has overshoots, which are compensated in the mean (blue line in Figures 4 and 5) by simulations with lower EKE. This is particularly evident at later times (Figure 5), where the uncertainties grow in time and the single members do not display a coherent behavior, i.e. different realizations have different meridional coordinate for the bi-modal structure and rather different EKE amplitudes. The introduction of EOFs 3-4-5 reduces the overshoots, but has also lower undershoots and does not help the ensemble members to maintain a coherent behavior for longer times. It can be further noticed in Figure 5 that, both with EOFs 1-2 and with EOFs 1-2-3-4-5, the EKE of the stochastic realizations is shifted to a too high meridional coordinate. On the other hand the DMD forced ensembles have less variance and do not always enclose the reference solution, but they follow quite well the meridional movement of the jet. This can be noticed also in the time evolution of the baroclinic EKE (Figure 6). Both with EOFs and DMDs the meridional shift of the positive EKE at $y \approx 1$ is detected, but the amplitude and other minor features, like the positive EKE at $y \approx 2$, are better resolved by the DMD forced simulation. Furthermore in the DMD ensembles, the uncertainties grow much more slowly in time, allowing the single members to display a coherent behavior also at later times. Lastly we would like to remark that to an increased noise amplitude, corresponds an increased ensemble EKE. Therefore also in the cases where the reference solution is not enclosed in the ensemble spread, it might be achieved by simply increasing the noise amplitude.

Our results suggest that the use of dynamically adapted noise covariance in stochastic parameterizations might be better suited to model phenomena, which do not reach statistical equilibrium, while

keeping track of the large scale dynamics. Moreover a dynamically adapted spatial correlation might more easily foster the system towards tipping points.

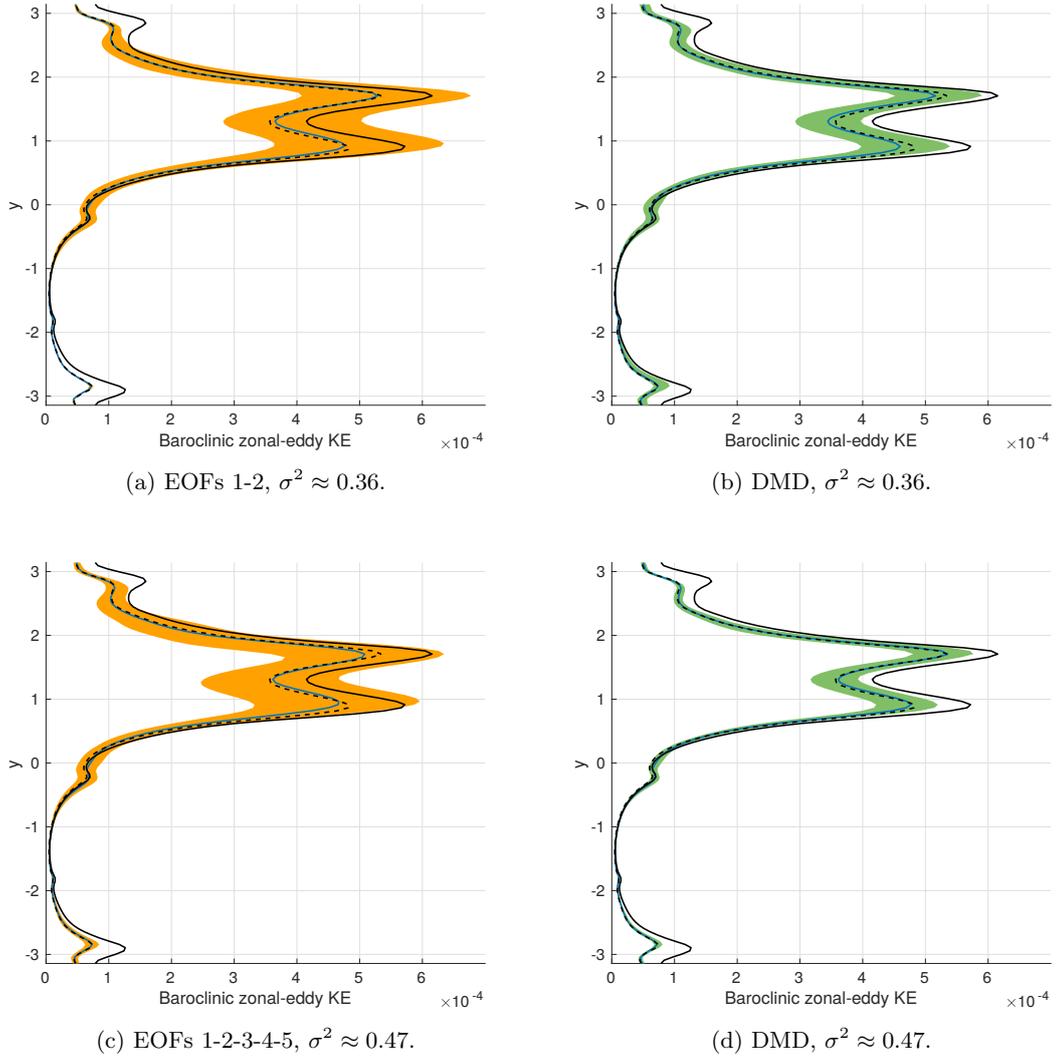


Figure 4: Baroclinic EKE for $t \in [1000, 2000]$ for stochastic simulations using EOFs 1-2, $\sigma^2 \approx 0.36$ (a), EOFs 1-2-3-4-5, $\sigma^2 \approx 0.47$ (c), or DMD $l = 2$, $m = 16$, $r = 7$, $\delta t = 0.1$ $\Delta t = 3\delta t$, $\sigma^2 \approx 0.36$ (b) and DMD, $\sigma^2 \approx 0.47$ (d). The outcomes for the high (black continuous), low (black dashed) resolution deterministic system and ensemble mean (blue) are also shown, for reference. The shaded area (orange for EOF and green for DMD) represents the area covered by different realizations of the stochastic system in the respective settings.

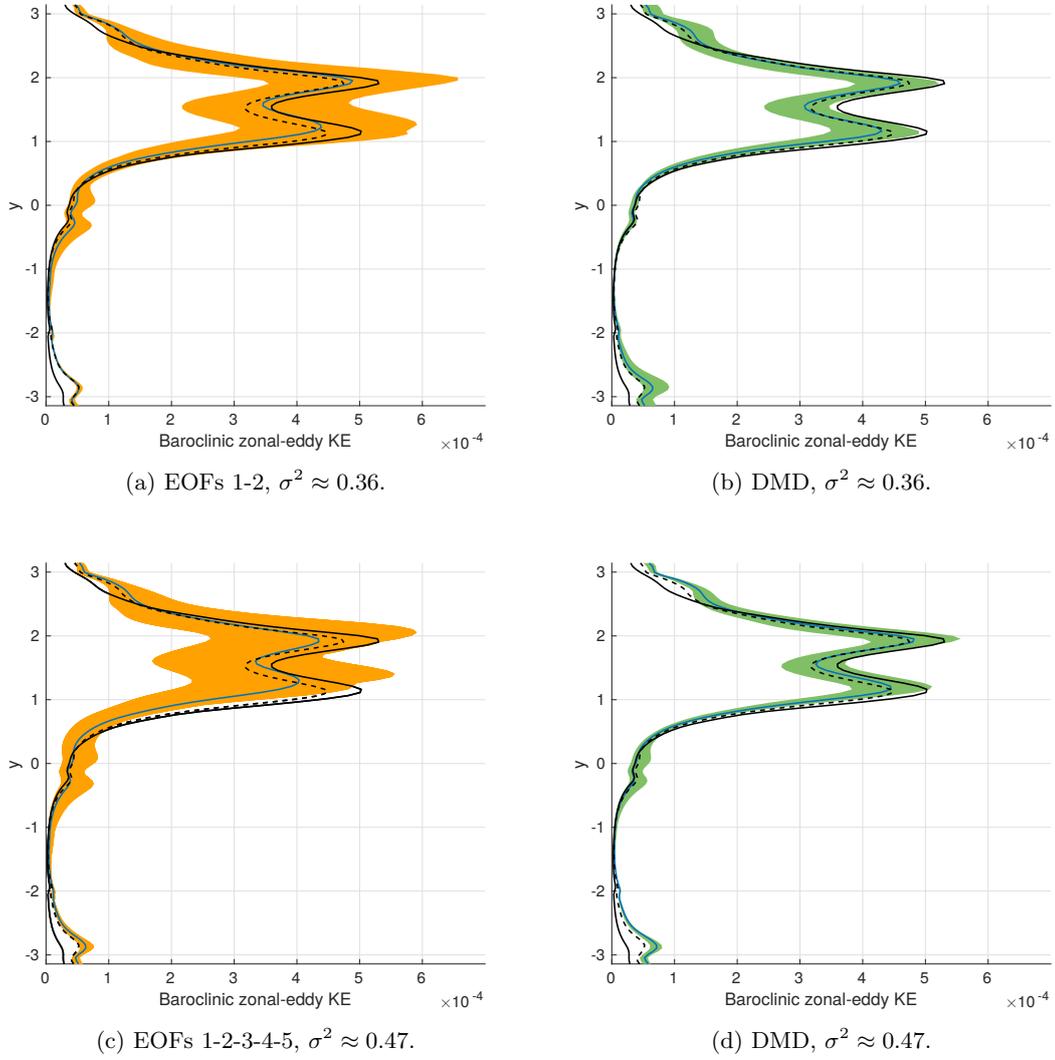
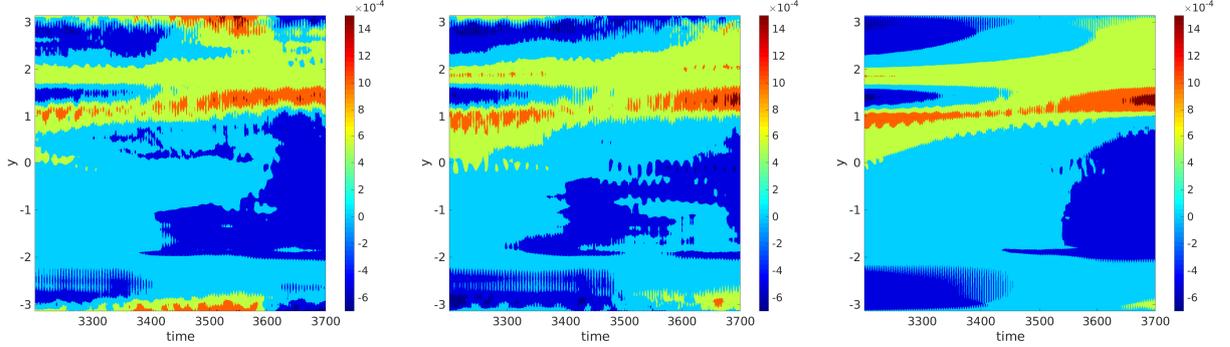
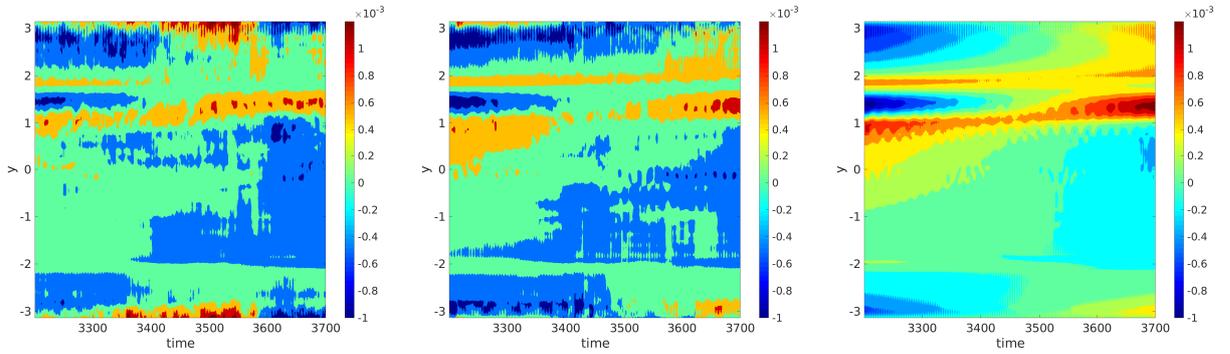


Figure 5: Baroclinic EKE for $t \in [3000, 4000]$ for stochastic simulations using EOFs 1-2 $\sigma^2 \approx 0.36$ (a), EOFs 1-2-3-4-5, $\sigma^2 \approx 0.47$ (c), or DMD $l = 2$, $m = 16$, $r = 7$, $\delta t = 0.1$ $\Delta t = 3\delta t$, $\sigma^2 \approx 0.36$ (b) and DMD, $\sigma^2 \approx 0.47$ (d). The outcomes for the high (black continuous), low (black dashed) resolution deterministic system and ensemble mean (blue) are also shown, for reference. The shaded area (orange for EOF and green for DMD) represents the area covered by different realizations of the stochastic system in the respective settings.



(a) Barotropic EKE; EOFs 1-2-3-4-5, (b) Barotropic EKE; DMD, $\sigma^2 \approx 0.47$. (c) Barotropic EKE; reference solution. $\sigma^2 \approx 0.47$.

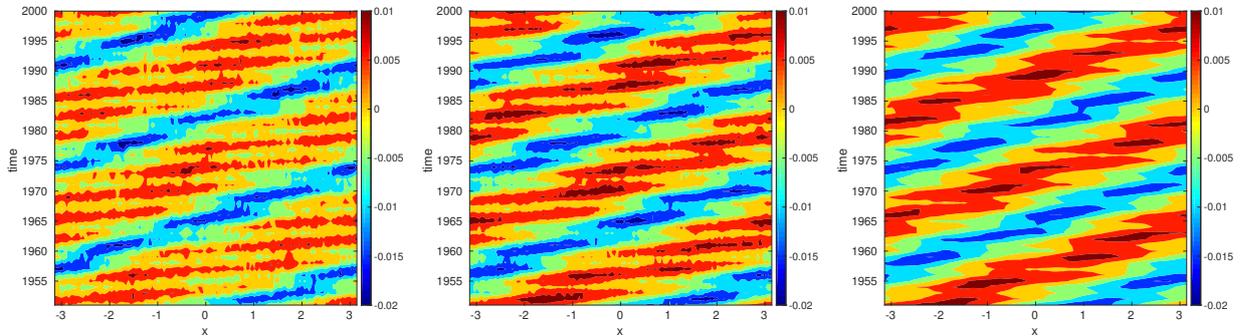


(d) Baroclinic EKE; EOFs 1-2-3-4-5, (e) Baroclinic EKE; DMD, $\sigma^2 \approx 0.47$. (f) Baroclinic EKE; reference solution. $\sigma^2 \approx 0.47$.

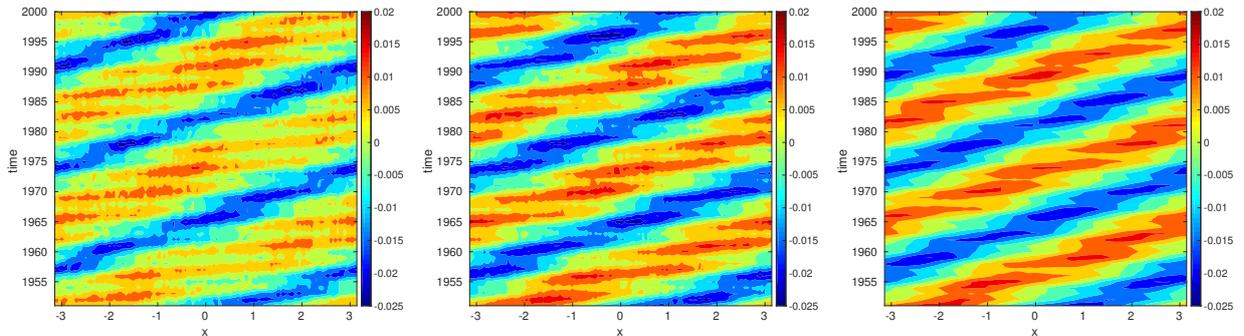
Figure 6: Time evolution of the barotropic (top) and baroclinic (bottom) EKE projected on the y -axis. From left to right: stochastic simulation with EOFs 1-2-3-4-5, $\sigma^2 \approx 0.47$; DMD forced stochastic simulation $l = 2$, $m = 16$, $r = 7$, $\delta t = 0.1$, $\Delta t = 3\delta t$, $\sigma^2 \approx 0.47$; reference solution.

4.3 Flow dynamics

In Figure 7 we show the time evolution of the projection over the zonal coordinate x of the barotropic and baroclinic potential vorticities for $t \in [1950, 2000]$. For a better comparison we removed the zonal mean flow and plot the resulting eddies. The same graph is shown for one realization with EOFs 1-2-3-4-5, $\sigma^2 \approx 0.47$ (left), one with DMD $l = 2$, $m = 16$, $r = 7$, $\delta t = 0.1$, $\Delta t = 3\delta t$, $\sigma^2 \approx 0.47$ (center), and the reference solution (right). Both techniques are able to capture the correct eddies phase speed, but DMD retains a stronger and less noisy signal. This result confirms the ability of DMD to include the sub-grid scales phenomena without weakening the signal of the larger scales.



(a) Barotropic PV; EOFs 1-2-3-4-5, (b) Barotropic PV; DMD, $\sigma^2 \approx 0.47$. (c) Barotropic PV; reference solution. $\sigma^2 \approx 0.47$.



(d) Baroclinic PV; EOFs 1-2-3-4-5, (e) Baroclinic PV; DMD, $\sigma^2 \approx 0.47$. (f) Baroclinic PV; reference solution. $\sigma^2 \approx 0.47$.

Figure 7: Time evolution of the barotropic (top) and the baroclinic (bottom) PVs anomalies with respect to the zonal mean projected on the x -axis. From left to right: stochastic simulation with EOFs 1-2-3-4-5, $\sigma^2 \approx 0.47$; DMD forced stochastic simulation $l = 2$, $m = 16$, $r = 7$, $\delta t = 0.1$ $\Delta t = 3\delta t$, $\sigma^2 \approx 0.47$; reference solution.

5 Summary and discussion

In the framework of the forced and damped 2-layer QG model we considered an energy conserving stochastic parameterization based on the projection operator approach [9], and employed two different procedures to define the noise spatial structure. As shown in [11], the definition of the latter is of fundamental importance for this parameterization to return physically meaningful results. In the present work we analyzed a statistical and a dynamical approach to define the noise covariance by using two different dimension reduction techniques, EOF and DMD. The former looks at the variance field of the fluid, while the latter is strictly linked to the Koopman operator and hence to the generator of the dynamics. EOF have been widely used in the literature, nevertheless their physical interpretation is limited because of the orthogonality constraint [32]. Moreover, being a statistical technique, it requires

long time series in order to obtain reliable patterns. In flat opposition DMD is thought to work with *tall and skinny* matrices [16], hence also with very short time series, and its modes are oscillating waves. Therefore the choice of the length of the time series, $m\Delta t$, and the temporal shift between the two input matrices, δt , are crucial and serve as scale selection. Since DMD gets pieces of information about the dynamics for possibly short time windows, its modes, and thus the noise covariance, have to be recomputed periodically. This is a new approach in stochastic parameterizations, since typically a fixed noise stencil is used during the whole realization.

Total energy graphs reveal that the EOF ensembles are either more dissipative or might include realizations with a clear increasing trend. On the other hand DMD runs are individually more energy conserving, inducing us to think that a dynamically adapted noise structure might help the numerics to stay on the manifold of constant energy. When looking at the eddy kinetic energy, it has been discovered that in case of EOF the uncertainties grow faster, which induce the single ensemble members to display the bi-modal behavior in different meridional positions. Furthermore the location of the bi-modal structure of the EKE ensemble mean is moved to too high meridional coordinates. The DMD forced ensembles instead are able to follow the jet meridional shift and well catch the meridional location of the double-peak also at later times. Moreover the uncertainties grow more slowly, allowing the individual members to display a coherent behavior also at later times. Finally, the field dynamics time evolution in the DMD ensembles retain a stronger and less noisy signal.

As regards computational time, DMD is very cheap, uses possibly short time series and does not need extra computations beforehand, but can be run alongside the main code. These aspects allow the DMD algorithm to periodically reanalyze the dynamics and redefine the noise covariance accordingly. Hence it is a very good candidate to parametrize scales undergoing phase transitions, or which do not reach statistically stable profiles. Moreover, due to its tight link to the generator operator, it might foster the system to reach tipping points.

Our results suggest that a dynamically adapted spatial structure should be considered in future developments of stochastic parameterizations. This finds further motivation in the physics. Not only the large scales are affected by the small scales, but also the fast processes are influenced by the slower motions. Hence physically correct parameterizations of the unresolved scales should allow the sub-grid processes to be influenced by the resolved modes. Future work might allow the DMD modes to be propagated by the Koopman operator for the $m\Delta t$ time window between one computation and the next, and might include a sub-grid energy model resolving the amount of energy to be backscattered at each time step. In conclusion, the propagation of the DMD modes by means of the Koopman operator might be seen as a sort of memory term, but more in detail analysis is required to establish what kind of relation, if any, exists between the propagation of the DMD modes and memory terms.

Data availability Model scripts are available upon request from the corresponding author. Data have been generated by use of the aforementioned scripts.

Funding This work is a contribution to the project M2 of the Collaborative Research Center TRR181 *Energy Transfer in Atmosphere and Ocean* funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 274762653.

Acknowledgments The authors would like to thank S. Juricke for discussions related to this work.

References

- [1] A. Arakawa. Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. part i. *Journal of Computational Physics*, 1(1):119–143, 1966.
- [2] G. Badin and F. Crisciani. *Variational formulation of fluid and geophysical fluid dynamics - Mechanics, symmetries and conservation laws*. Springer Berlin, 2018.
- [3] P. S. Berloff. Dynamically consistent parameterization of mesoscale eddies. part i: Simple model. *Ocean Modelling*, 87:1 – 19, 2015.
- [4] J. Berner, U. Achatz, L. Batte, L. Bengtsson, A. d. l. Cámara, H. M. Christensen, M. Colangeli, D. R. Coleman, D. Crommelin, S. I. Dolaptchiev, et al. Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3):565–588, 2017.
- [5] M. Budišić, R. Mohr, and I. Mezić. Applied Koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.
- [6] J. C. Cotter, D. Crisan, D. D. Holm, W. Pan, and I. Shevchenko. Modelling uncertainty using circulation-preserving stochastic transport noise in a 2-layer quasi-geostrophic model. *arXiv:1802.05711*, 2018.
- [7] J. C. Cotter, D. Crisan, D. D. Holm, W. Pan, and I. Shevchenko. Numerically modelling stochastic Lie transport in fluid dynamics. *SIAM Journal on Scientific Computing*, 2018.
- [8] B. Dubrulle. Beyond Kolmogorov cascades. *Journal of Fluid Mechanics*, 867:P1, 2019.
- [9] J. E. Frank and G. A. Gottwald. Stochastic homogenization for an energy conserving multi-scale toy model of the atmosphere. *Physica D*, 254:46–56, 2013.
- [10] M. Gavish and D. L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, Aug 2014.
- [11] F. Gugole and C. L. E. Franzke. Numerical development and evaluation of an energy conserving conceptual stochastic climate model. *Mathematics for Climate and Weather Forecasting*, 5:45–64, 2019.
- [12] D. D. Holm. Variational principles for stochastic fluid dynamics. *Proc. R. Soc., A* 471:20140963, 2015.
- [13] M. F. Jansen and I. M. Held. Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modelling*, 80:36–48, 2014.

- [14] S. Juricke, S. Danilov, A. Kutsenko, and M. Oliver. Ocean kinetic energy backscatter parametrizations on unstructured grids: Impact on mesoscale turbulence in a channel. *Ocean Modelling*, 138:51 – 67, 2019.
- [15] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.
- [16] J. Kutz, S. Brunton, B. Brunton, and J. Proctor. *Dynamic Mode Decomposition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.
- [17] A. Lasota and M. C. Mackey. *Chaos, fractals and noise - Stochastic aspects of dynamics*. Springer-Verlag, 1994.
- [18] C. Leith. Stochastic models of chaotic systems. *Physica D: Nonlinear Phenomena*, (98):481–491, 1996.
- [19] A. Majda, C. L. E. Franzke, and D. Crommelin. Normal forms for reduced stochastic climate models. *Proc. Natl. Acad. Sci. USA*, 106:3649–3653, 2009.
- [20] E. Mémin. Fluid flow dynamics under location uncertainty. *Geophysical and Astrophysical Fluid Dynamics*, 108:119–46, 2014.
- [21] I. Mezić. Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, 45(1):357–378, 2013.
- [22] G. A. Pavliotis and A. Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.
- [23] P. Porta Mana and L. Zanna. Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modelling*, 79:1–20, 2014.
- [24] V. Resseguier, W. Pan, and B. Fox-Kemper. Data-driven versus self-similar parameterizations for stochastic Lie transport and location uncertainty. *private communication*, 2019.
- [25] R. Salmon. Hamiltonian fluid mechanics. *Annual Review of Fluid Mechanics*, 20(1):225–256, 1988.
- [26] P. D. Sardeshmukh and P. Sura. Reconciling non-gaussian climate statistics with linear dynamics. *J. Climate*, 22(5):1193–1207, 2009.
- [27] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:528, 2010.
- [28] T. G. Shepherd. Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Advances in Geophysics*, 32:287–338, 1990.
- [29] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1:391, 2014.

- [30] G. K. Vallis. *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press, 2006.
- [31] H. von Storch. *Spatial Patterns: EOFs and CCA*, pages 227–257. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.
- [32] H. von Storch and F. W. Zwiers. *Statistical analysis in climate research*. Cambridge University Press, 2003.

5 Conclusions

In this thesis I extended the projection operator approach described by Frank and Gottwald (2013) to high dimensional systems in the framework of a grid-point based numerical model with a Eulerian description of the dynamics. The extension has been successfully carried out and it provided further insights into the numerical aspects of the stochastic parameterization. More in detail, the choice of the numerical setup revealed to strongly influence the structure of the noise covariance. As a matter of fact the use of independent identically distributed (iid) noise led the dynamics to lose variability and other fundamental statistical properties. This is due to the adopted Eulerian description of the dynamics. In this framework a phenomenon behaving like an iid noise is not influenced by its neighbors, which implies that it fully evolves and decorrelates inside the grid cell. When running GCMs at coarse resolutions the surface area of a grid cell can be in the order of $10^2 - 10^3$ square kilometers, so it is possible to think about phenomena bounded inside the grid box, e.g. clouds. On the other hand resolving a wave requires at least 3 – 4 grid points in order to capture its oscillating behavior. Therefore the numerical gap between the resolved scales and such sub-grid processes is too large, and the latter are dissipated by the numerics. In order to limit this gap and obtain physically consistent results a spatial covariance structure has to be defined. The usage of the first two EOF eigenvectors proved that a spatial correlation noise structure allows the dynamics to retain its underlying statistical properties and improves the eddy length for the barotropic mode. At this point one might wonder, whether EOF provides the best initialization for the noise covariance, or if more suitable techniques are available.

To answer this question DMD has been used and its results compared to EOF. As mentioned above, EOF is a statistical technique, which detects patterns explaining large portions of the fluid variance. On the other hand DMD is a data driven technique for the approximation of the Koopman operator, which encodes the fluid dynamics at a specific time. While the first technique requires long time series, the second can work also with few data. In particular the length of the time series employed by the DMD algorithm determines the scales to be detected: everything with period longer than the selected time window cannot be identified as a wave, and it is hence included in the mode representing the mean state, while processes with very high frequencies are treated as noise. Hence the length of the time series for the DMD algorithm is crucial and serves as scale selection. As a consequence, when computing the noise covariance

structure, the same structure should be kept for long running times in case of EOF, since EOFs have an asymptotic validity, while it should be regularly updated when making use of DMD. The use of a dynamically adapted spatial correlation was not a computational burden, and disclosed advantages, like the ability of following the jet shift or a more slowly increase of the uncertainties, that should be considered when developing stochastic parameterizations for the sub-grid phenomena. In fact, for a correct representation of these phenomena, not only their effect on the large scales should be modeled, but also the influence that the large scales have on the small scales.

I discuss now the main findings in terms of the research questions posed in the introduction.

1. Can the projection operator approach, outlined by Frank and Gottwald (2013), be employed also in case of high dimensional systems?

In this thesis I showed that the projection operator approach presented by Frank and Gottwald (2013) can be applied also to high dimensional system. Furthermore it is quite versatile. In fact here it has been used for the conservation of energy, but in principle it could be employed also to conserve other quantities, for example enstrophy. The requirements are: an equation describing the quantity of interest, and project with respect to the manifold where this quantity is constant. Moreover this parameterization offers quite some degrees of freedom, like the amplitude and color of the noise or the noise covariance structure, that can be tuned for a better representation of the sub-grid processes. Here the research focus was on the noise covariance, but other aspects, for instance the inclusion of a memory term, should be analyzed too.

2. How should the noise covariance be defined such that it is dynamically consistent and representative of the sub-grid processes?

The comparison between different noise covariances showed that some patterns, although physically reasonable, when included into the numerical model, lead the dynamics of the system towards unphysical results. This can be explained by the limitations of the discrete model with respect to the continuous system. Hence any parameterization should be thought not only in the context of the continuous equations, but also as part of a discrete numerical model. In particular, the results of PI showed that iid noise should not be used in the context of Eulerian dynamics at coarse resolutions but a spatial correlation for the noise should be defined. Its introduction by means of the first two EOFs improved the eddy length for the barotropic mode, but not for the baroclinic mode. Hence, the use of EOFs, despite its wide usage in the literature, might not be optimal.

3. Vannitsem (2017) showed that, in case of spectral models, the error dynamics is dependent on the scale where it is introduced; is it possible to choose the noise scale also in case of grid point

5 Conclusions

models?

The answer to this question is not trivial. It strongly depends on which data analysis techniques are employed, and hence on what the modeler thinks might work better. If the choice is to rely on statistical climatic patterns, as those found with EOF, then it is not possible to choose the scale where to introduce the noise. As shown in PII, EOFs eigenvectors do not always represent waves, but they could also be representative of other phenomena, like the jet meridional movement. Furthermore the first EOF can be associated to large scale dynamics, but because of the orthogonality constraint it is hard to associate higher order EOFs to the smaller scales. On the other hand there are other techniques, e.g. DMD whose focus is to detect the most active waves in a given time window, that give the possibility of choosing the noise scale. The drawbacks are that, in order to detect fast evolving processes, shorter time series have to be considered and hence the revealed modes do not have a long lasting validity, but they should be updated regularly. Hence, particularly in case of computationally expensive models, the choice of what kind of patterns to use will be a compromise between target, computational expense and numerical stability.

4. How much do a priori assumptions on the noise covariance affect the outcomes of the numerical model?

The a priori assumptions analyzed here regards either the choice of using climatic patterns with long lasting validity, or of inserting the noise at a specific scale. As argued in the previous paragraph, due to the characteristics of EOF and DMD, these assumptions can be stated also as either keeping the noise covariance constant during the whole simulation, or updating it regularly. Although the patterns defined by EOF and DMD look similar to a certain extent, significant differences can be spotted, which in turn led to different ensembles behaviors. In particular, individual runs with the DMD induced stochastic forcing are more energy conserving, ensemble uncertainties grow more slowly, allowing the single members to display a coherent behavior for longer time, and the parameterization is able to notice the jet meridional movement and adapts to it. Finally the DMD algorithm is numerically stable and quite fast, hence it can be run alongside the main code without significant computational burdens. These results show that care should be taken when building the noise covariance. This was still quite a simplified setup, but in more complex model it is reasonable to think that uncertainties might grow faster, due to the nonlinear interaction between different phenomena. Hence any additional error introduced by the parameterization might propagate and amplify inside the model.

Lastly I wish to conclude with a short disclaimer.

PII dealt with two representative techniques, which required two different procedures

for the noise covariance definition. The focus was on the proof of concept, that the use of climatic patterns might not be always optimal, and that other methods are available. As regards which particular technique should be used, I have to mention that the basic versions of both algorithms have been used here. More advanced versions, like complex EOFs von Storch and Zwiers (2003) or recursive DMD Noack et al. (2016), have and are being developed. Therefore it is perfectly plausible that *the* best technique has not been considered here.

6 Outlook

I conclude with some brief thoughts on research questions left open by this work. The first regards memory terms. In fact the inclusion of such a term would allow the noise to hold a dynamically consistent behavior not just with respect to the larger scales, but also with respect to its own dynamics at previous times. It would be particularly interesting in case of a dynamically adapted noise covariance, since it would allow to keep track of its different stages and hence of any trend, if present. It might be argued that the propagation of the DMD modes by means of the Koopman operator can be seen as some kind of memory term, but the relation between the two remains unclear at the moment. Hence more study on the topic is required.

In the discussion section of PII, it is stated that DMD might foster the system towards tipping points. These points represent critical transitions of the system. When the system reaches such a point, some of its basic properties may change, for instance stable solutions may become unstable, or new solutions may appear. If DMD does indeed support the system towards such points, it would unravel a wide range of applications ranging from weather forecasting to paleoclimate modeling. Therefore it is a research direction that deserve further investigations and analyses. As a side comment, this capability might be further ameliorated by the inclusion of memory terms.

The 2-layer QG model here used is still quite an idealized setting, and not many options were available as for which variable should be used to compute either the DMD or the EOF modes. In more realistic models the chosen variable and its relation with the desired sub-grid process to be modeled will play a crucial role. Teleconnections might provide some insights, but it might be also case specific.

The projection operator approach is concerned with the conservation of a desired quantity, in this case energy. As mentioned in the introduction, an important problem to be tackled is the re-injection of energy from the small scales into the large scales. Hence future research might see it coupled to a sub-grid energy model, which learns the amount of energy to be backscattered at each time step.

Finally, a dynamically adapted noise covariance has been tried as applied to just one particular parameterization. In order to better assess its validity and its pros and cons, it should be tested also when other stochastic parameterizations are employed.

Acknowledgments

The first person I would like to thank is Christian Franzke, who was always available to answer my more or less dumb questions, and allowed me to follow my own research interests also when they were not tightly connected to the project.

In this regard I wish to thank also Georg Gottwald, who has been like a second supervisor to me, and remembered me that one of the most important aspects of research is to have fun while playing around.

I am grateful to the TRR181 for the financial support. A special thanks goes to Meike Ruhнау and Jennifer Fandrich for all the help with bureaucratic formalities and problems in general, and to all the other PhD students and Post-Docs for making the various retreats and winter schools pleasant get-together occasions.

My gratitude goes also to my colleagues and friends Vera Melinda Galfi and Valerio Lembo for being always supportive, helping clearing doubts and having the patience of listening to me for five days a week. I would like to thank also my other colleagues from the Theoretical Meteorology group for their support, company and laughs.

I wish to thank also old and new friends for making me feel part of a huge family. In this regard I would like to thank also the UniSport kick-boxing and beach volley community, and in particular my kick-boxing trainer Grzegorz Zjawin, who helped me to get rid of all the frustration I might have had.

I am much obliged to my parents for their love and constant trust in me. Thousands times thanks to Raoni for his motivational support, patience and love in the darkest hours of the writing process.

References

- Arakawa, A. (1966). Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. part i. *Journal of Computational Physics*, 1(1):119–143.
- Badin, G. and Crisciani, F. (2018). *Variational formulation of fluid and geophysical fluid dynamics - Mechanics, symmetries and conservation laws*. Springer Berlin.
- Berloff, P. S. (2015). Dynamically consistent parameterization of mesoscale eddies. part i: Simple model. *Ocean Modelling*, 87:1 – 19.
- Berner, J., Achatz, U., Batte, L., Bengtsson, L., Cámara, A. d. l., Christensen, H. M., Colangeli, M., Coleman, D. R., Crommelin, D., Dolaptchiev, S. I., et al. (2017). Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3):565–588.
- Courant, R., Friedrichs, K., and Lewy, H. (1928). Über die partiellen Differenzgleichungen der mathematischen Physik. *Mathematische Annalen*, 100:32–74.
- Cullen, M. and Brown, A. (2009). Large-eddy simulation of the atmosphere on various scales. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 367:2947–56.
- Danilov, S., Juricke, S., Kutsenko, A., and Oliver, M. (2019). Towards consistent subgrid momentum closures in ocean models. In Eden, C. and Iske, A., editors, *Energy transfers in atmosphere and ocean*. Springer.
- Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., Jonker, H. J. J., and Selten, F. (2016). Stochastic convection parameterization with markov chains in an intermediate-complexity gcm. *Journal of the Atmospheric Sciences*, 73(3):1367–1382.
- Frank, J. E. and Gottwald, G. A. (2013). Stochastic homogenization for an energy conserving multi-scale toy model of the atmosphere. *Physica D*, 254:46–56.
- Gardiner, C. W. (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences*, volume 4. Springer Berlin.
- Gottwald, G., Crommelin, D., and Franzke, C. L. E. (2017). Stochastic climate theory.

- In Franzke, C. L. E. and O’Kane, T., editors, *Nonlinear and Stochastic Climate Dynamics*. Cambridge University Press, Cambridge.
- Hasselmann, K. (1976). Stochastic climate models part I. Theory. *Tellus*, 28(6):473–485.
- Holm, D. D. (2015). Variational principles for stochastic fluid dynamics. *Proc. R. Soc., A* 471:20140963.
- Holton, J. R. and Hakim, G. J. (2012). *An Introduction to Dynamic Meteorology*, volume 88. Academic Press.
- Jansen, M. F. and Held, I. M. (2014). Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modelling*, 80:36–48.
- Mémin, E. (2014). Fluid flow dynamics under location uncertainty. *Geophysical and Astrophysical Fluid Dynamics*, 108:119–46.
- Moon, W. and Wettlaufer, J. (2014). On the interpretation of stratonovich calculus. *New Journal of Physics*, 16(5):055017.
- Noack, B. R., Stankiewicz, W., Morzyński, M., and Schmid, P. J. (2016). Recursive dynamic mode decomposition of transient and post-transient wake flows. *Journal of Fluid Mechanics*, 809:843–872.
- Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., and Weisheimer, A. (2009). Stochastic parametrization and model uncertainty. Technical report, ECMWF Technical Memorandum.
- Pavliotis, G. A. and Stuart, A. (2008). *Multiscale methods: averaging and homogenization*. Springer Science & Business Media.
- Porta Mana, P. and Zanna, L. (2014). Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modelling*, 79:1–20.
- Quarteroni, A. (2017). *Numerical Models for Differential Problems*. Springer Berlin.
- Strogatz, S. H. (1994). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering*. Westview Press.
- Vallis, G. K. (2006). *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press.
- Vannitsem, S. (2017). Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(3):032101.

References

von Storch, H. and Zwiers, F. W. (2003). *Statistical analysis in climate research*. Cambridge University Press.

Eidesstattliche Versicherung

Declaration on oath

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel: "Numerical study of an energy consistent stochastic parameterization for climate models" selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Dissertation oder Teile davon vorher weder im In- noch im Ausland in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

I hereby declare, on oath, that I have written the present dissertation by myself and have not used other than the acknowledged resources and aids. The submitted written version corresponds to the electronic one. I declare that the present dissertation has not been submitted in any other previous doctorate examination procedure.

Hamburg,