July 24, 2017

MASTER THESIS

ON STOCHASTIC GEOPHYSICAL FLUID DYNAMICS

Erwin Luesink

Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)
Multiscale Modeling and Simulation

Exam committee:
Prof.dr.ir. B.J. Geurts
Prof.dr. D.D. Holm
Dr. C. Brune

Documentnumber
Department of Mathematics —

UNIVERSITY OF TWENTE.

Abstract

The concept of transport noise is introduced and studied. It is shown that two different types of multiplicative noise, transport noise and so-called fluctuation-dissipation noise, behave qualitatively different when studied in terms of their effects on the Lorenz system. In particular the sum of the Lyapunov exponents for this system is different for the two types of noise. Also a stochastic version of a robust, deterministic numerical algorithm for the determination of Lyapunov exponents is posed. It computes the deterministic values for the individual Lyapunov exponents with reasonable accuracy considering the numerical methods used to solve the underlying equations. Finally, a stochastic variational principle is used to derive stochastic rotating shallow water equations and it is shown that they have the same conservation laws as the deterministic version.

Preface

During the writing of this work, I was hosted at Imperial College London. Under the wings of Darryl Holm and Bernard Geurts, a year of research culminated in the following report. At Imperial College, I was allowed to use a desk in the Mathematics of Planet Earth (MPE) section of the ESPRC Centre for Doctoral Training (CDT), where the MPE PhD students do a number of courses and introductory research in order to prepare for their PhD project. It was wonderful to work with these students, since their topics, mindset and goals were similar to mine. This resulted in a good working environment and allowed me to experience the MPE CDT. In addition to the students surrounding me in the office, in the research group led by Darryl Holm, there was another group of students always open to discussion of research and problems. Being able to discuss ideas and problems immediately is one of the most pleasant things in research. Many good ideas sprung from discussion with other students.

I would like to express gratitude towards a large number of people. In particular Darryl Holm, for allowing me to do a year of research at Imperial College and work with him. He is an amazing researcher and a great source of inspiration. Our discussions over lunch or over coffee were very fruitful and amusing. I also want to thank Bernard Geurts, who was always available for a Skype meeting in which we would discuss the numerical problems and thoroughly question the notation. His eye and ear for details go unsurpassed. I am very grateful to the all the people associated to the MPE program, it was very comforting to be surrounded by such a motivated group of researchers and students. The research group led by Darryl Holm was another source of inspiration and a great help whenever I got stuck. From this group, I particularly want to thank So Takao, for his time and effort in explaining several concept of differential geometry to me as well as reading through my work. Before this year of research, I was doing coursework. In many of the courses it was required to team up with another student in order to complete projects. Whenever this was the case, I teamed up with Jeroen de Cloet and our teamwork was great. We complemented each other in terms of mathematical skills and continued discussing research throughout the final project. I want to thank Martin Rasmussen, Maximilian Engel and Valerio Lucarini for several discussions on Lyapunov exponents and the underlying framework, Valentin Resseguier and Etienne Mémin for their point of view on fractal dimensions and Dan Crisan for his helpful remarks and insight into stochastic analysis. And last but not least, my family for their support throughout the year.

Contents

1 Introduction

Weather, climate and ocean prediction relies heavily on our understanding of fluid dynamics. Since the existence and uniqueness for the Navier-Stokes equations are still an open problem, we do not know what is the best model for fluid dynamics and additionally we do not have a complete understanding of the small-, subgrid scale processes such as turbulence. Also Lorenz showed that weather and climate models suffer from sensitive dependence on initial data. The sensitive dependence is boosted by numerical limitations. Altogether, our incomplete understanding of the underlying processes, the limitations of the numerical models and the sensitive dependence on initial data make it so that deterministic modeling is inaccurate. These problems may be compensated for by modeling in a stochastic way. For these reasons forecasts nowadays are expressed in probabilistic terms. Hence, instead of relying on a single deterministic prediction, an ensemble of stochastic predictions can give a lot more insight into the uncertainty quantification of weather. We will show here that different types of multiplicative noise can have different effects on the model behavior. In particular, we shall study the effects of so-called transport noise on the Rayleigh-Bénard convection model, from which, by Fourier projection, the famous Lorenz equations may be obtained. This is a low-dimensional model that can exhibit chaos for certain parameter values and has been studied intensively for decades. The effects of two different types of noise on the Lorenz equations shall then be studied by means of Lyapunov exponents, both on a theoretical level as well as on a numerical level. For the numerical analysis of the Lyapunov exponents, particularly their sum, we propose a robust algorithm.

The third section introduces the concept of transport noise accompanied by the differential geometric framework that is necessary to insert it into fluid dynamics. The concept of a Lie derivative is introduced, which greatly generalizes and simplifies a number of calculations. In [Hol15], variational principles are used to introduce the noise in mechanical systems. A result from this theory is the Kelvin circulation theorem, that can be used to introduce noise as well, when the deterministic case is understood.

The fourth section employs the Kelvin circulation theorem to insert noise in the Rayleigh-Bénard convection model. From this model, by Fourier projection, the famous Lorenz equations can be found [Lor63]. Hence from a convection model with transport noise we will derive a set of stochastic Lorenz equations. In a paper by [CSG11] the Lorenz system is perturbed with a different type of noise. In what follows, we will compare the two systems both analytically as well as numerically.

The fifth section presents the random dynamical system theory [Arn03] that is necessary to look at Lyapunov exponents. In this framework, the stability of the stochastic Lorenz systems can be studied. In particular, the sum of the Lyapunov exponents is analyzed in detail, because it can be computed exactly for the Lorenz system. The sum describes the average rate contraction or expansion of phase-space volume. It will be proved that the sum for the two types of noise in the Lorenz system is different.

The sixth section is dedicated towards the numerical verification of the analytical statements. For the numerical calculation of Lyapunov exponents, a deterministic, robust algorithm by [UvB01] is adapted to a stochastic version and used to compute the exponents. As a test, we verify the exponents for the deterministic case and we find that the values computed by our method are similar to the ones found in existing literature.

The seventh section discusses stochastic rotating shallow water model. It relates to the previous sections in that the same type of noise is introduced, but this time via the rigorous route, the variational principle. It shall be shown that familiar conservation laws remain for this model compared to the deterministic case. The motivation for studying these equation in particular is as follows: Weather processes occur at a vast number of different timescales and due to numerical cost, it is often necessary to only model the slow time-scales. In the 1980s, there was a huge discussion among meteorologists and mathematicians about whether or not a so-called slow manifold exists for the Lorenz-1986 (L86) model [Lor86]. It consists of 5 ordinary differential equations and has two different timescales, one fast and one slow timescale. A slow manifold is a set of initial conditions from which the dynamics does not develop any fast timescale motion. Around 1995, most researchers were in favor of the nonexistence of such a set for the L86 model. Inspired by this discussion and its results, we wish to investigate the influence of the fast motion on the slow motion on the level of partial differential equations. By means of the variational principle as given by [Hol15], noise is introduced into the model. In particular, these equations possess slow- and fast timescale motion and by a change of variables, the rotating shallow water equations can be written into an alternative form in which there is a clear split between the different timescales.

2 Transport Noise

The concept of transport noise as it shall be used in this work comes from [Hol15]. The name comes from its purpose. In the Lagrangian variational principles, it is possible to constrain movement of mechanical systems along certain paths. A simple example of this is the spherical pendulum, constrained to move on the sphere. In fluid dynamics, this movement is called advection. By taking a Lagrangian and constraining it to stochastic Lagrangian paths, the fluid momentum and other advected quantities are transported along that path. It has been shown in [CFH17] that the Euler equations for an incompressible ideal fluid with transport noise have the same analytical properties as the deterministic Euler equations. We are interested in seeing whether properties are preserved by this type of noise on a lower dimensional scale as well. In [Hol15] the noise is introduced into the the dynamics by using a Clebsch constraint in a variational principle.

A note on the notation. Since a lot of tools from differential geometry shall be used, in which d denotes the exterior derivative with the special property that $\mathrm{d}^2\alpha = 0$ for any tensor α. Stochastic analysis is done in integral form because the derivative of Wiener process is not defined. Therefore the stochastic evolution operator is denoted d. This can become very confusing, so we shall denote the stochastic evolution operator as d to distinguish between the two. To derive, for instance, the stochastic Euler equations, one takes the deterministic Lagrangian and constrains the advected quantities to be advected by a stochastic velocity field. In the Lagrangian description of fluid dynamics, this boils down to constraining the motion of fluid particles to move along a stochastic curve. The Stratonovich stochastic process that arises is given by

$$\mathsf{d}\eta_t(X) = \mathbf{u}(\eta_t(X), t)\,\mathrm{d}t + \sum_{i=1}^{n} \xi_i(\eta_t(X)) \circ \mathrm{d}W_t^i \tag{1}$$

where W^i are scalar, independent Wiener processes (or Brownian motions), the construction of such a process can be found in the appendix, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the ξ_i are spatially smooth functions that represent spatial correlations, which are related to a velocity-velocity correlation matrix C_{ij} as $C_{ij} = \xi_i \xi_j^T$. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a triple, where Ω is the sample space, $\omega \in \Omega$ is a sample, $\mathcal{F}$ is the family of events or σ-algebra, and $\mathbb{P}$ is a probability measure. By definition the probability measure of the sample space $\mathbb{P}(\Omega) = 1$. The σ-algebra determines which events can occur, this includes events that have probability zero of happening. The number n of eigenvectors and Wiener processes is arbitrary. The multiplication symbol $\circ$ in the context of stochastic integrals implies that the stochastic integral is of the Stratonovich type. This is the type of stochastic integral that admits the standard chain rule and is therefore an invaluable concept throughout the derivations. The Eulerian description is in terms of vector fields, that are consructed as

$$\mathrm{d}\mathbf{x}_t(x) = \mathsf{d}\eta_t\eta_t^{-1} = \mathbf{u}(x, t)\,\mathrm{d}t + \sum_{i=1}^{n} \xi_i(x) \circ \mathrm{d}W_t^i \tag{2}$$

so that the stochastic process $\mathsf{d}\eta_t$ is related to the stochastic vector field $\mathrm{d}\mathbf{x}_t$ by pullback as $\eta_t^*\,\mathrm{d}\mathbf{x}_t = \mathsf{d}\eta_t$. Hereafter Einstein's summation convention shall be used, so summation of repeated indices should be understood. The pullback is a concept from differential geometry, that is defined for arbitrary tensors as

Definition 2.1 *If $\phi : M \to N$ is a diffeomorphism that maps manifold M to N and $t \in T_s^r(M)$ is an r, s-tensor, let $\phi_* t := (T\phi)_s^r \circ t \circ \phi^{-1}$ be the **pushforward** of t by ϕ. Here $T\phi$ is the tangent of ϕ and $\circ$ means composition. If $t \in T_s^r(N)$, the **pullback** is given by the inverse operation $\phi^* t = (\phi^{-1})_* t$.*

For an excellent introductory overview on manifold theory see [Tu10], especially for the construction of a smooth manifold and its associated bundles, in [Hol08] one can find a very accessible introduction to geometric mechanics that includes a large number of illustrative examples and for a more advanced overview including applications see [AM78]. This is an abstract definition for operations that are quite intuitive. In the following figure, let $f : N \to \mathbb{R}$ be a 0-form (a function) and let $\phi : M \to N$ be a diffeomorphism, then the pullback of that function is simply the composition $\phi^* f = f \circ \phi$

$$\begin{tikzcd}
M \arrow[rr, "\phi"] \arrow[dr, "f\circ\phi"'] & & N \arrow[dl, "f"] \\
& \mathbb{R} &
\end{tikzcd}$$

Figure 1: Pullback of a function

Let $\phi : M \to N$ be a diffeomorphism and $\boldsymbol{\omega} : TN \to \mathbb{R}^n$ be a 1-form. Here TN is related to the tangent bundle to the manifold N.

$$\begin{tikzcd} TM \arrow[rr, "T\phi"'] \arrow[dr, "\phi^*\boldsymbol{\omega}"'] & & TN \arrow[dl, "\boldsymbol{\omega}"] \\ & \mathbb{R}^n & \end{tikzcd}$$

Figure 2: Pullback of a 1-form

In [Hol08] one can find the following intuitive definition of the pullback and pushforward of a k-form.

Definition 2.2 *Let $\phi : M \to N$ be a smooth invertible map from the manifold M to the manifold N and let α be a k-form on N. The pullback $\phi * \alpha$ of α by ϕ is defined as the $k - form$ on M given by*

$$\phi^* \alpha_m = \alpha_{i_1 \dots i_k}(\phi(m))(T_m\phi \cdot \mathrm{d}x)^{i_1} \wedge \dots \wedge (T_m\phi \cdot \mathrm{d}x)^{i_k},$$

with $i_1 < i_2 < \ldots < i_k$. If the map ϕ is a diffeomorphism, the pushforward $\phi_*\alpha$ of a k-form α by ϕ is defined by the inverse of the pullback $\phi_*\alpha = (\phi^*)^{-1}\alpha$.

The following example of a 1-form is also given in [Hol08]. In the previous definition, $T_m\phi$ expresses the chain rule for change of variables in local coordinates. For example

$$(T_m\phi \cdot \mathrm{d}x)^{i_1} = \frac{\partial \phi^{i_1}(m)}{\partial x^{i_A}} \, \mathrm{d}x^{i_A}.$$

Thus, the pullback of a 1-form is given by

$$\begin{aligned} \phi^*(\mathbf{v}(\mathbf{x}) \cdot \mathrm{d}\mathbf{x}) &= \mathbf{v}(\phi(\mathbf{x})) \cdot \mathrm{d}\phi(\mathbf{x}) \\ &= v_{i_1}(\phi(\mathbf{x})) \left(\frac{\partial \phi^{i_1}(\mathbf{x})}{\partial x^{i_A}} \, \mathrm{d}x^{i_A} \right) \\ &= \mathbf{v}(\phi(\mathbf{x}) \cdot (T_{\mathbf{x}}\phi \cdot \mathrm{d}\mathbf{x}). \end{aligned}$$

The pullback is valuable operation that will allow us to switch between the Eulerian and Lagrangian description and allows us to define the Lie derivative. The Lie derivative evaluates the rate of change of a tensor field along a certain vector field. It shall become clear that advection in fluid dynamics is a Lie derivative. To define it, first the flow of a vector field is introduced.

Definition 2.3 (Flow of a vector field) *The flow of Y is the differentiable map $\phi : U \times I \to M$, where $I \subset \mathbb{R}$ is an interval containing 0 and U is an open subset of manifold M, such that, for any $z \in U$, the map $\phi_z(t) := \phi(z,t)$ is an integral curve of Y with $\phi_z(0) = z$.*

So the flow of a vector field is an integral curve of that vector field. Hence, when differentiated with respect to time and evaluated at the identity, one recovers the vector field. This sets us up for the first definition of the Lie derivative.

Definition 2.4 (Dynamical defintion of the Lie derivative) *Given a differentiable tensor field T and a differentiable vector field Y defined on a differentiable manifold M, we can calculate the change of T along Y. Let ϕ be the flow of Y, then the Lie derivative of T with respect to Y at a point $p \in M$ is defined as*

$$(\pounds_Y T)_p := \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} (\phi_t^* T)_p \tag{3}$$

where ϕ_t^* denotes the pull-back.

There is a second definition of the Lie derivative, sometimes referred to as "Cartan's magic formula", which is incredibly useful for straight computations. The dynamical definition is more useful for general proofs that require Lie derivatives.

Definition 2.5 (Cartan's formula for the Lie derivative) *Given a differentiable tensor field T and a differentiable vector field Y defined on a differentiable manifold M, we can calculate the change of T along Y. Cartan's formula states that*

$$(\pounds_Y T) := Y \lrcorner \, \mathrm{d}T + \mathrm{d}(Y \lrcorner \, T) \tag{4}$$

where the hook notation $A \lrcorner B := B(A)$ denotes the insertion of A into B.

Upon equating Lie derivatives for 1-forms in both definitions, the fundamental vector identity of fluid dynamics is derived. It is important because it allows us to write fluid dynamics in an alternative form.

Theorem 2.6 (Fundamental vector identity of fluid dynamics) *Let* $\mathbf{v}$ *be a 1-form and* $\mathbf{Y}$ *be a vector field defined on a manifold* M*, then the following identity is true*

$$(\mathbf{Y}\cdot\nabla)\mathbf{v} + v_j\nabla Y^j = \nabla(\mathbf{Y}\cdot\mathbf{v}) - \mathbf{Y}\times\operatorname{curl}\mathbf{v} \tag{5}$$

Proof. The Lie derivative of $\mathbf{v}\cdot\mathrm{d}\mathbf{x}$ with respect to some vector field Y is, according to the dynamical definition

$$\begin{aligned}
\pounds_Y(\mathbf{v}\cdot\mathrm{d}\mathbf{x}) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\phi_t^*(\mathbf{v}\cdot\mathrm{d}\mathbf{x})\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} v_i(\phi_t(X))\,\mathrm{d}\phi_t^i(X)\\
&= \left[\frac{\partial v_i}{\partial\phi_t^k(X)}\frac{\partial\phi_t^k(X)}{\partial t}\,\mathrm{d}\phi_t^i(X) + v_i(\phi_t(X))\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\phi_t^i(X)}{\partial X^j}\,\mathrm{d}X^j\right]_{t=0}\\
&= \frac{\partial v_j}{\partial x^k}Y^k\,\mathrm{d}x^j + v_i\frac{\partial Y^i}{\partial x^j}\,\mathrm{d}x^j\\
&= ((\mathbf{Y}\cdot\nabla)\mathbf{v} + v_j\nabla Y^j)\cdot\mathrm{d}\mathbf{x}
\end{aligned}$$

where the diffeomorphism ϕ_t maps between coordinates X and x and $\phi_t(X)|_{t=0} = x$. We have also used that the derivative of a flow at the identity recovers the vector field. According to Cartan's formula

$$\begin{aligned}
\pounds_Y(\mathbf{v}\cdot\mathrm{d}\mathbf{x}) &= Y \lrcorner\ \mathrm{d}(\mathbf{v}\cdot\mathrm{d}\mathbf{x}) + \mathrm{d}(Y\lrcorner\,(\mathbf{v}\cdot\mathrm{d}\mathbf{x}))\\
&= Y \lrcorner\ \mathrm{d}(\mathbf{v}\cdot\mathrm{d}\mathbf{x}) + \nabla(\mathbf{Y}\cdot\mathbf{v})\cdot\mathrm{d}\mathbf{x}\\
&= Y^m\partial_m \lrcorner \left(\epsilon_{ijk}\frac{\partial v_k}{\partial x_j}\,\mathrm{d}S^i\right) + \nabla(\mathbf{Y}\cdot\mathbf{v})\cdot\mathrm{d}\mathbf{x}\\
&= \epsilon_{ijk}\frac{\partial v_k}{\partial x_j}Y^m\partial_m\lrcorner\ \mathrm{d}S^i + \nabla(\mathbf{Y}\cdot\mathbf{v})\cdot\mathrm{d}\mathbf{x}\\
&= \epsilon_{ijk}\frac{\partial v_k}{\partial x_j}\epsilon_{imn}Y^m\,\mathrm{d}x^n + \nabla(\mathbf{Y}\cdot\mathbf{v})\cdot\mathrm{d}\mathbf{x}\\
&= \operatorname{curl}\mathbf{v}\cdot\mathbf{Y}\times\mathrm{d}\mathbf{x} + \nabla(\mathbf{Y}\cdot\mathbf{v})\cdot\mathrm{d}\mathbf{x}\\
&= (-\,\mathbf{Y}\times\operatorname{curl}\mathbf{v} + \nabla(\mathbf{Y}\cdot\mathbf{v}))\cdot\mathrm{d}\mathbf{x}.
\end{aligned}$$

We denote $\frac{\partial}{\partial x^m}$ by ∂_m. By definition $\partial_i \lrcorner\ \mathrm{d}x^j = \delta_i^j$. The ϵ_{ijk} denotes the totally antisymmetric tensor (or Levi-Civita symbol). We have also used identities for $\mathrm{d}(\mathbf{v}\cdot\mathrm{d}\mathbf{x})$ and $\mathrm{d}(Y\lrcorner\,(\mathbf{v}\cdot\mathrm{d}\mathbf{x}))$, which are shown in the appendix. Identifying the two definitions gives rise to the identity. ■

The motion equation in fluid dynamics describes the evolution of a velocity field $\mathbf{u}$. In the fundamental vector identity 1-forms $\mathbf{v}$ appear. Any momentum equation in fluid dynamics features both. Namely, these quantities both have dimensions of velocity, but in terms of Riemannian geometry, the velocity $\mathbf{u} = u^i\partial_i$ is contravariant with indices up and transports fluid properties, such as temperature or density. The momentum per unit mass $\mathbf{v} = v_i\,\mathrm{d}x^i$ is covariant and has indices down. Hence, in general, these two velocities are different, in that their physical meanings are different and their transformation under diffeomorphisms are different. In the special case where the kinetic energy is given by the L^2 metric and the coordinate system is Cartesian with an Euclidean metric, then the components of the two velocities can be set equal. The Euler equations for an incompressible ideal fluid are such a special case.

2.1 Table of Lie derivatives

In fluid dynamics a number of Lie derivatives appear frequently. For a quick overview, they are listed here. Each Lie derivative is calculated along vector field X.

Tensor	Lie derivative	$\mathbb{R}^3$ expression
function f	$\pounds_X f$	$\mathbf{X}\cdot\nabla f$
1-form $\mathbf{v}\cdot\mathrm{d}\mathbf{x}$	$\pounds_X(\mathbf{v}\cdot\mathrm{d}\mathbf{x})$	$(\mathbf{X}\cdot\nabla\mathbf{v} + v_j\nabla X^j)\cdot\mathrm{d}\mathbf{x}$ or $(-\mathbf{X}\times\operatorname{curl}\mathbf{v} + \nabla(\mathbf{X}\cdot\mathbf{v}))\cdot\mathrm{d}\mathbf{x}$
2-form $\omega\cdot\mathrm{d}\mathbf{S}$	$\pounds_X(\omega\cdot\mathrm{d}\mathbf{S})$	$(\operatorname{curl}(\omega\times\mathbf{X}) + \mathbf{X}\operatorname{div}\omega)\cdot\mathrm{d}\mathbf{S}$ or $(-\omega\cdot\nabla\mathbf{X} + \mathbf{X}\cdot\nabla\omega + \omega\operatorname{div}\mathbf{X})\cdot\mathrm{d}\mathbf{S}$
top-form $f\,\mathrm{d}^3x$	$\pounds_X(f\,\mathrm{d}^3x)$	$\operatorname{div}(f\mathbf{X})\,\mathrm{d}^3x$

Table 1: A list of Lie derivatives for differential forms in $\mathbb{R}^3$.

The proofs of these identities can be found in the appendix. The Lie derivative of a 1-form has already been shown in the derivation for the fundamental vector identity of fluid dynamics. To appropriately introduce stochasticity to fluid dynamics, one has to start from a Clebsch constrained variational principle as described in [Hol15]. A less formal way is to use the Kelvin circulation theorem, which is a result from the general variational principle, to introduce the noise. Similar to the Kelvin filtered NS-α model [FHT02], [Geu04], by means of adapting the fluid loop velocity in the Kelvin theorem, it is possible to derive new equations of motion.

2.2 Kelvin Circulation Theorem

As an example, the Euler equations for an incompressible, ideal fluid are considered. The familiar, deterministic equations are given by

$$\begin{aligned} \partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} &= -\nabla p, \\ \operatorname{div}\mathbf{u} &= 0. \end{aligned} \tag{6}$$

Since the equations are incompressible, the density is constant. For simplicity, it is set to unity. Before going to the Kelvin theorem, the Lie derivative formula is introduced.

Lemma 2.7 *Consider an arbitrary time dependent 1-form $\mathbf{v}(\boldsymbol{x},t)\cdot \mathrm{d}\mathbf{x}$ and let η be the flow of a vector field $\mathrm{d}\mathbf{x}_t$ so that $\mathsf{d}\eta = \eta_t^* \,\mathrm{d}\mathbf{x}_t$ and $\eta_t(X) = \eta(X,t) = \mathbf{x}$ maps the Lagrangian coordinates to the Eulerian coordinates, then*

$$\mathsf{d}\eta_t^*(\mathbf{v}\cdot \mathrm{d}\mathbf{x}) = \eta_t^*(\mathsf{d} + \pounds_{\mathrm{d}\mathbf{x}_t})(\mathbf{v}\cdot\mathrm{d}\mathbf{x}). \tag{7}$$

Proof. By definition of the pullback

$$\eta_t^*(\mathbf{v}\cdot\mathrm{d}\mathbf{x}) = \mathbf{v}(\eta_t(X),t)\cdot \mathrm{d}\eta_t(X).$$

Important to note once again is that the 1-form is time dependent. The pullback acts on the spatial coordinate, not on the time. Computing the stochastic evolution of the previous expression yields

$$\begin{aligned}
\mathsf{d}\eta_t^*(\mathbf{v}\cdot\mathrm{d}\mathbf{x}) &= \mathsf{d}\mathbf{v}(\eta_t(X),t)\cdot \mathrm{d}\eta(X) + \mathbf{v}(\eta_t(X),t)\cdot \mathrm{d}\mathsf{d}\eta_t(X)\\
&= \left(\mathsf{d}\mathbf{v}(\eta_t(X),t) + \frac{\partial \mathbf{v}(\eta_t(X),t)}{\partial \eta_t(X)}\cdot \mathrm{d}\mathbf{x}_t(X)\right)\cdot \mathrm{d}\eta_t(X) + \mathbf{v}(\eta_t(X),t)\cdot \mathrm{d}\,\mathrm{d}\mathbf{x}_t(\eta_t(X))\\
&= \eta_t^*\left(\left(\mathsf{d}\mathbf{v}(\mathbf{x},t) + \frac{\partial \mathbf{v}(\mathbf{x},t)}{\partial \mathbf{x}}\cdot \mathrm{d}\mathbf{x}_t(\mathbf{x})\right)\cdot \mathrm{d}\mathbf{x}\right) + v_i(\eta_t(X),t)\frac{\partial\, \mathrm{d}x_t^i(\eta_t(X))}{\partial \eta_t^j(X)}\,\mathrm{d}\eta_t^j(X)\\
&= \eta_t^*\left(\mathsf{d}v(\mathbf{x},t) + \mathrm{d}\mathbf{x}_t\cdot\nabla v + v_i\nabla\,\mathrm{d}x_t^i\right)\cdot \mathrm{d}\mathbf{x}\\
&= \eta_t^*(\mathsf{d} + \pounds_{\mathrm{d}\mathbf{x}_t})(\mathbf{v}\cdot\mathrm{d}\mathbf{x}).
\end{aligned}$$

In the last step we have used the Lie derivative of a 1-form as presented in Table 1. ■

The deterministic version of the Lie derivative formula is recovered when the stochastic evolution operator d is replaced by the partial time derivative and $\partial_t\eta = \mathbf{u}$. The same computation as in the previous proof then leads to $\partial_t\eta_t^*(v\cdot\mathrm{d}\mathbf{x}) = \eta_t^*(\partial_t + \pounds_{\mathbf{u}})(v\cdot\mathrm{d}\mathbf{x})$. It is now a simple task to prove the Kelvin theorem. The loop integral in the Kelvin theorem moves with the velocity $\mathbf{u}$, so the domain is moving. By using the pullback, the Eulerian frame is transformed into the Lagrangian frame. This makes the integration domain stationary and allows for the partial time derivative or stochastic evolution operator to be pulled inside the integral.

Theorem 2.8 (Kelvin's circulation theorem for the Euler equations) *The Euler equations for an ideal fluid preserve the circulation integral*

$$I(t) = \oint_{c(t)} \mathbf{v}\cdot\mathrm{d}\mathbf{x},$$

where $c(t)$ is closed loop moving with velocity $\mathbf{u}$.

Proof.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} I(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \oint_{c(t)} \mathbf{v} \cdot \mathrm{d}\mathbf{x} \\
&= \oint_{c(0)} \frac{\mathrm{d}}{\mathrm{d}t} \eta_t^* (\mathbf{v} \cdot \mathrm{d}\mathbf{x}) \\
&= \oint_{c(0)} \eta_t^* (\partial_t + \pounds_{\mathbf{u}})(\mathbf{v} \cdot \mathrm{d}\mathbf{x}) \\
&= \oint_{c(t)} (\partial_t + \pounds_{\mathbf{u}})(\mathbf{v} \cdot \mathrm{d}\mathbf{x}) \\
&= \oint_{c(t)} (\partial_t \mathbf{v} + \mathbf{u} \cdot \nabla \mathbf{v} + v_j \nabla u^j) \cdot \mathrm{d}\mathbf{x} \\
\text{(setting } \mathbf{u} = \mathbf{v}) \qquad &= \oint_{c(t)} (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + u_j \nabla u^j) \cdot \mathrm{d}\mathbf{x} \\
&= \oint_{c(t)} \left(-\nabla p + \frac{1}{2} \nabla |\mathbf{u}|^2 \right) \cdot \mathrm{d}\mathbf{x} \\
&= 0,
\end{aligned} \tag{8}$$

where in the last step the fundamental theorem of calculus was used. Identifying $\mathbf{u}$ and $\mathbf{v}$ is possible because the Euler equations is the special case as mentioned earlier. The identity $u_j \nabla u^j = \frac{1}{2} \nabla |\mathbf{u}|^2$ is used. This identity is only true if the components of the two velocities are equal, which in the case of deterministic fluid dynamics, is satisfied. ∎

If we let the closed loop $c(t)$ move with velocity $\mathrm{d}\mathbf{x}_t = \mathbf{u}\,\mathrm{d}t + \xi_i \circ \mathrm{d}W_t^i$ instead of $\mathbf{u}$, we can introduce stochasticity in the Euler equations as follows

$$\begin{aligned}
\mathsf{d}\mathbf{u} + \mathrm{d}\mathbf{x}_t \cdot \nabla \mathbf{u} + u_j \nabla \, \mathrm{d}x_t^j &= -\nabla p \, \mathrm{d}t, \\
\operatorname{div} \mathrm{d}\mathbf{x}_t &= 0,
\end{aligned} \tag{9}$$

The assumption is made that $\operatorname{div} \xi_i = 0$ for all $i = 1, \ldots, n$. It is for this set of equations that [CFH17] show that the analytical properties are not worse than for the deterministic Euler equations. The stochastic version of the Kelvin theorem is valid for these equations.

Theorem 2.9 (Kelvin's circulation theorem for the stochastic Euler equations) *The stochastic Euler equations for an ideal fluid preserve the circulation integral*

$$I(t) = \oint_{c(t)} \mathbf{v} \cdot \mathrm{d}\mathbf{x},$$

where $c(t)$ is closed loop moving with velocity $\mathrm{d}\mathbf{x}_t$.

Proof. Now that there is no confusion between $\mathbf{u}$ and $\mathbf{v}$, we immediately start with $\mathbf{u}$ and do the coordinate transformations involving the pullback in a single step. Letting the stochastic evolution operator act on the circulation integral gives rise to

$$\begin{aligned}
\mathsf{d} I(t) &= \mathsf{d} \oint_{c(t)} \mathbf{u} \cdot \mathrm{d}\mathbf{x} \\
&= \oint_{c(t)} (\mathsf{d} + \pounds_{\mathrm{d}\mathbf{x}_t})(\mathbf{u} \cdot \mathrm{d}\mathbf{x}) \\
&= \oint_{c(t)} (\mathsf{d}\mathbf{u} + \mathrm{d}\mathbf{x}_t \cdot \nabla \mathbf{u} + u_j \nabla \, \mathrm{d}x_t^j) \cdot \mathrm{d}\mathbf{x} \\
&= \oint_{c(t)} -\nabla p \, \mathrm{d}t \cdot \mathrm{d}\mathbf{x} \\
&= 0
\end{aligned}$$

∎

The stochastification of fluid dynamics using transport noise changes the advective velocity field. Advected quantities are moved around by the same velocity field as the motion itself. The rigorous framework for this, using variational principles, can be found in [Hol15]. The result is that in the Eulerian framework the vector field in the Lie derivative becomes the stochastic vector field $\mathrm{d}\mathbf{x}_t$.

3 Rayleigh-Bénard Convection

By considering the Rayleigh-Bénard convection process with transport noise, it is possible to study what happens to a paradigm example in chaos theory, pattern formation and fully developed turbulence [Kad01]. It describes convective motion of a fluid between two plates with different temperatures. The bottom plate is heated and the top plated is cooled. On the two plates, the velocity field satisfies no-slip boundary conditions and impermeability in the vertical direction. In the horizontal directions have periodic boundary conditions. The constant temperature on bottom plate is T_b and on the top plate is T_t with $T_b > T_t$.

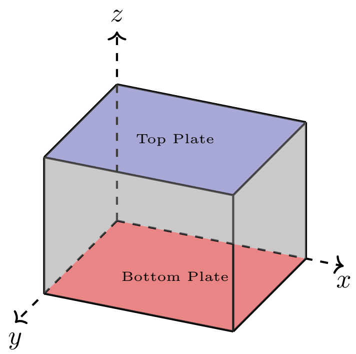


Figure 3: The color shading indicates the temperature difference. The bottom plate is being heated and the top plate is being cooled.

In the Oberbeck-Boussinesq approximation, the density is assumed to depend linearly on the temperature. The equations of motion for the Rayleigh-Bénard process are then

$$\begin{aligned} \partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} &= -\nabla p + \nu \Delta \mathbf{u} + \mathbf{F}, \\ \partial_t T + \mathbf{u} \cdot \nabla T &= \gamma \Delta T, \\ \operatorname{div} \mathbf{u} &= 0. \end{aligned} \tag{10}$$

The fluid described by these equations is incompressible and affected by a buoyancy force $\mathbf{F} = \alpha g T \hat{\mathbf{e}}_k$, which is linearly dependent on the temperature and by viscosity, the strength of which is governed by the kinematic viscosity ν. It is assumed that the specific heat per unit mass c_p is constant. The heat equation can therefore be written in terms of temperature, as the constant may be divided out in each term, but the equation should still be read as an advection-diffusion equation in heat. The diffusion of heat is governed by the heat diffusivity constant γ. The buoyancy force acts only in the vertical direction and depends on thermal expansion coefficient α, gravity g and the temperature T. This is the convection process that Lorenz studied [Lor63], given rise to the famous Lorenz system. The stochastic version of the momentum equation is obtained by adding viscosity and a body force to the stochastic Euler equations (9). To properly introduce the transport noise into the heat equation, it is necessary to go back to the general theory. In [Hol15] for an arbitrary Lagrangian a set of advected quantities is considered. The constraint in the variational principle is that fluid properties are advected along stochastic Lagrangian paths. This argument dictates that the heat should satisfy the advection equation, given by

$$(\mathsf{d} + \pounds_{\mathrm{d}\mathbf{x}_t}) q = 0,$$

where q is the collection of advected quantities (in the Rayleigh-Bénard convection problem this is just the heat). This gives the advection of the heat by the stochastic velocity field. Additionally, the heat diffuses over time, so the equation gets an additional diffusive term. The heat in terms of the specific heat per unit mass times temperature is a scalar function, satisfies the advection equation with a dissipative term

$$\begin{aligned} \gamma \Delta T \, \mathrm{d}t &= (\mathsf{d} + \pounds_{\mathrm{d}\mathbf{x}_t}) T \\ &= \mathsf{d} T + \mathrm{d}\mathbf{x}_t \cdot \nabla T. \end{aligned}$$

Here we have used the identity for a scalar function from Table 1. Thus the noisy convection process defined on the domain $[0, T] \times \mathbb{R}^3 \times \Omega$ is described by

$$\begin{aligned} \mathsf{d}\mathbf{u} + \mathrm{d}\mathbf{x}_t \cdot \nabla \mathbf{u} + u_j \nabla \, \mathrm{d}x_t^j &= (-\nabla p + \nu \Delta \mathbf{u} + \mathbf{F}) \, \mathrm{d}t, \\ \mathsf{d} T + \mathrm{d}\mathbf{x}_t \cdot \nabla T &= \gamma \Delta T \, \mathrm{d}t, \\ \operatorname{div} \mathrm{d}\mathbf{x}_t &= 0, \\ \mathrm{d}\mathbf{x}_t = \mathbf{u} \, \mathrm{d}t + \sum_{i=1}^{n} \xi_i(\mathbf{x}) \circ \mathrm{d}W_t^i, & \end{aligned} \tag{11}$$

where ξ_i are related to the velocity-velocity correlation matrix and W_t^i is a sequence of scalar, independent, Wiener processes. The fluid motion is constrained to convective rolls in the xz-plane, which makes the model 2 dimensional and allows for a number of simplifications. Firstly, instead of considering heat, a quantity that we will call temperature profile shall be used.

3.1 Temperature Profile

The temperature $T(x, z, t)$ can be expanded into a horizontal mean value and a departure from the mean [Sal62]. This gives

$$T(x, z, t) = T_{av}(z, t) + T'(x, z, t),$$

where T_{av} is the horizontal mean and T' is perturbation therefrom. Additionally the mean can be expanded into two parts, the first part represents a linear difference between the lower and upper boundary and the second part is a perturbation of this linear difference.

$$T_{av}(z, t) = T_{av}(0, t) - \frac{T_\Delta}{H} z + T''_{av}(z, t)$$

where T''_{av} is the perturbation from the linear difference, $T_\Delta = |T_b - T_t|$ is the constant temperature difference between the lower and upper plate and H is the height between them. This leads to the following equation

$$T(x, z, t) = \left(T_{av}(0, t) - \frac{T_\Delta}{H} z \right) + T'(x, z, t) + T''_{av}(x, z, t) \tag{12}$$

In this model it shall be assumed that there is some external heating to maintain the constant temperature difference. Introducing what we will call the temperature profile $\phi(x, z, t) := T'(x, z, t) + T''_{av}(x, z, t)$ allows us to write the Rayleigh-Bénard convection problem in same way as in [Lor63]. Substituting (12) into the heat equation leads to

$$\mathsf{d}\phi + \mathrm{d}\mathbf{x}_t \cdot \nabla\phi = \left(\frac{T_\Delta}{H} w + \gamma\Delta\phi \right) \mathrm{d}t. \tag{13}$$

where w is the $\hat{\mathbf{z}}$-component of the velocity field.

3.2 Vorticity

The momentum equation can be simplified as well. By going to vorticity formulation, we can remove the pressure term and by numerous observations, the vorticity equation undergoes a number of simplifications. The vorticity is defined as $\omega = \operatorname{curl}\mathbf{u}$, so taking the curl of the momentum equation

$$\operatorname{curl}(\mathsf{d}\mathbf{u} + \mathrm{d}\mathbf{x}_t \cdot \nabla\mathbf{u} + u_j \nabla \, \mathrm{d}x_t^j) = \operatorname{curl}\left(\frac{1}{\rho}\nabla p + \nu\Delta\mathbf{u} + \frac{1}{\rho}\mathbf{F} \right)$$

The Laplacian Δ commutes with the curl and so does the stochastic evolution operator, so the vorticity can be identified. Substituting in the buoyancy force for $\mathbf{F}$ and taking the curl then results in

$$\mathsf{d}\omega + \operatorname{curl}(\mathrm{d}\mathbf{x}_t \cdot \nabla\mathbf{u} + u_j \nabla \, \mathrm{d}x_t^j) = \nu\Delta\omega + \frac{1}{\rho}\operatorname{curl}(\alpha g T \hat{\mathbf{e}}_k)$$

It is here that we shall use the fundamental vector identity of fluid dynamics (5). This identity allows us to rewrite the advection terms into their curl form, which simplifies the vector calculus operations that are necessary to derive the vorticity formulation. Expanding the curl of the buoyancy and rewriting the equation as

$$\mathsf{d}\omega + \operatorname{curl}(\nabla(\mathrm{d}\mathbf{x}_t \cdot \mathbf{u}) - \mathrm{d}\mathbf{x}_t \times \operatorname{curl}\mathbf{u}) = \nu\Delta\omega + \alpha g \phi_x.$$

The curl of a gradient is zero, so the first term drops. The vorticity equation then becomes

$$\mathsf{d}\omega - \operatorname{curl}(\mathrm{d}\mathbf{x}_t \times \omega) = \nu\Delta\omega + \alpha g \phi_x.$$

The curl of the cross product of the stochastic vector field with the vorticity can be expanded as

$$-\operatorname{curl}(\mathrm{d}\mathbf{x}_t \times \omega) = -(\mathrm{d}\mathbf{x}_t(\operatorname{div}\omega) - \omega(\operatorname{div}\mathrm{d}\mathbf{x}_t) + (\omega \cdot \nabla)\,\mathrm{d}\mathbf{x}_t - (\mathrm{d}\mathbf{x}_t \cdot \nabla)\omega)$$

where the divergence of ω is zero because ω is defined as the curl of a vector field. Upon making the assumption that ξ_i are divergence free for all $i = 1, \ldots, n$, the second term also drops. This then yields

$$-\operatorname{curl}(\mathrm{d}\mathbf{x}_t \times \omega) = \mathrm{d}\mathbf{x}_t \cdot \nabla\omega - \omega \cdot \nabla\,\mathrm{d}\mathbf{x}_t = [\mathrm{d}\mathbf{x}_t, \omega].$$

Here $[\mathrm{d}\mathbf{x}_t, \omega]$ is the commutator for vector fields $\mathrm{d}\mathbf{x}_t$ and ω. for incompressible fluid problems. The motion shall be restricted to convective rolls in the xz-plane, making the problem 2 dimensional. The curl of a 2 dimensional velocity field $\mathbf{u} = (u, 0, w)$ is then in the $\hat{y}$-direction, so $\omega = (0, \omega, 0)$. From here onward, we will always speak about vorticity as a scalar function, instead of a vector field. This allows for further reduction of terms in the vorticity equation. The second term in the commutator, the vortex stretching term is equal to zero. So finally the vorticity equation becomes

$$\mathrm{d}\omega + \mathrm{d}\mathbf{x}_t \cdot \nabla\omega = (\nu\Delta\omega + \alpha g \phi_x)\,\mathrm{d}t. \tag{14}$$

3.3 Fourier Mode Projection

The advection terms in the temperature profile and vorticity equations can be written in terms of the stochastic stream function $\tilde{\psi}$ by using

$$(\hat{\mathbf{x}} \cdot \mathrm{d}\mathbf{x}_t, 0, \hat{\mathbf{z}} \cdot \mathrm{d}\mathbf{x}_t) = \left(\frac{\partial\tilde{\psi}}{\partial z}, 0, -\frac{\partial\tilde{\psi}}{\partial x}\right)$$

to write the advection term as the determinant of the Jacobian

$$\mathrm{d}\mathbf{x}_t \cdot \nabla\omega = \left(\hat{\mathbf{x}} \cdot \mathrm{d}\mathbf{x}_t \frac{\partial\omega}{\partial x}\right) + \left(\hat{\mathbf{z}} \cdot \mathrm{d}\mathbf{x}_t \frac{\partial\omega}{\partial z}\right) = \frac{\partial\tilde{\psi}}{\partial z}\frac{\partial\omega}{\partial x} - \frac{\partial\tilde{\psi}}{\partial x}\frac{\partial\omega}{\partial z} = \left|\frac{\partial(\tilde{\psi}, \omega)}{\partial(x, z)}\right|.$$

The equations for Rayleigh-Bénard convection restricted to xz-plane (13) and (14) can then be written as

$$\begin{aligned}
\mathrm{d}\omega + \left|\frac{\partial(\tilde{\psi}, \omega)}{\partial(x, z)}\right| &= (\nu\Delta\omega + \alpha g \phi_x)\,\mathrm{d}t, \\
\mathrm{d}\phi + \left|\frac{\partial(\tilde{\psi}, \phi)}{\partial(x, z)}\right| &= (\gamma\Delta\phi - \frac{T_\Delta}{H}\psi_x)\,\mathrm{d}t, \\
\omega &= -\Delta\psi.
\end{aligned} \tag{15}$$

Thus the new system of equations is comprised of a vorticity equation and an equation determining the temperature profile. Furthermore, ψ is the stream function, $\tilde{\psi}$ is the noisy stream function, T_Δ is the constant temperature difference between the two plates and H is the distance between those plates. The relation between the vorticity and the stream function is given by a Poisson equation. To derive the Lorenz system, the truncated Fourier series is adapted to include stochasticity. This is possible because the transport noise only appears in terms that have spatial derivative operators acting on them and the transport noise vector field is assumed to be smooth in space, but is not differentiable in time. The Fourier expansions for the terms without noise are identical to the ones Lorenz used in his famous 1963 article [Lor63],

$$\begin{aligned}
\frac{k}{\gamma(1+k^2)}\psi &= X\sqrt{2}\sin\left(\frac{k\pi x}{H}\right)\sin\left(\frac{\pi z}{H}\right), \\
\frac{\pi R_a T_\Delta}{R_c}\phi &= Y\sqrt{2}\cos\left(\frac{k\pi x}{H}\right)\sin\left(\frac{\pi z}{H}\right) - Z\sin\left(\frac{2\pi z}{H}\right), \\
\frac{k}{\gamma(1+k^2)}\tilde{\psi} &= (X\sqrt{2}\,\mathrm{d}t + \beta\sqrt{2}\circ\mathrm{d}W_t)\sin\left(\frac{k\pi x}{H}\right)\sin\left(\frac{\pi z}{H}\right).
\end{aligned} \tag{16}$$

Here, k is the wave number, $R_a = \alpha g H^3 T_\Delta \nu^{-1}\gamma^{-1}$ is the Rayleigh number and $R_c = \pi^4 k^{-2}(1+k^2)^3$ is the critical value of the Rayleigh number. These scaling constants have been introduced in order to be able to write the resulting equations in a compact form. The reason for using this Fourier expansion is in certain cases, when the Rayleigh number exceeds a critical value, using the full Fourier series reduces to exactly these three terms [Sal62]. Due to the orthogonality of the Fourier basis functions, from a mathematical point of view, the only sensible choice for the noise in terms of its Fourier series expansion is to have the exact same Fourier series expansion as the stream function, as the projection step will eliminate all other terms. From a physical point of view, we do not want the stochasticity to give rise to types of motion other than rolls between the two plates. The projection then formally yields

$$\begin{aligned}
X_\tau &= \sigma(Y - X), \\
Y_\tau &= -\tilde{X}Z + rX - Y, \\
Z_\tau &= \tilde{X}Y - bZ,
\end{aligned} \tag{17}$$

where $\sigma = \gamma\nu^{-1}$ is the Prandtl number, $r = R_a R_c^{-1}$ is a scaled Rayleigh number, $b = (4(1+k^2))^{-1}$ is parameter related to the wavenumber and $\tilde{X} = X\,\mathrm{d}t + \beta\circ\mathrm{d}\dot{W}_t$ is the X variable with noise. The time τ is dimensionless and related to the time t in (15) by $\tau = \pi^2(1+k^2)\gamma t H^{-2}$. From this point onward, the time τ will just be written as t. In terms of stochastic differential equations (SDEs), in the proper notation

$$\begin{aligned} \mathrm{d}X &= \sigma(Y-X)\,\mathrm{d}t, \\ \mathrm{d}Y &= (rX - XZ - Y)\,\mathrm{d}t - \beta Z \circ \mathrm{d}W_t, \\ \mathrm{d}Z &= (XY - bZ)\,\mathrm{d}t + \beta Y \circ \mathrm{d}W_t. \end{aligned} \tag{18}$$

As can be seen clearly in the "formal, inappropriate form" (17), the noise appears only in the nonlinear terms, similar to the stochastic partial differential equations (15), where the nonlinearity is in the transport terms. On this low-dimensional scale, the nonlinear terms represent rotation, the physical interpretation of the stochasticity is that it is a stochastic angular velocity. This shows that the transport noise, when carried down through the Fourier projection, is of multiplicative nature. Upon setting the noise amplitude β to zero, the deterministic Lorenz equations are recovered. In case of the parameter values $r = 28$, $\sigma = 10$ and $b = 8/3$, almost all initial conditions will tend to an invariant set, an **attractor** set. This attractor set is **strange**, meaning that set has a fractal structure, which implies that the solution to the system of equations is chaotic. Fractal refers to the fact that a set can have integer topological dimension, but the space it actually takes up may be noninteger higher dimensional. However, upon the introduction of noise, there is no longer an attractor set, as the noise will push the trajectory out of any bounded set with probability 1, due to the unbounded variation of the Wiener process. [CSG11] added linear multiplicative noise to the deterministic Lorenz equations. The SDEs that describe that system are given by

$$\begin{aligned} \mathrm{d}X &= \sigma(Y-X)\,\mathrm{d}t + \beta X\,\mathrm{d}W_t, \\ \mathrm{d}Y &= (rX - XZ - Y)\,\mathrm{d}t + \beta Y\,\mathrm{d}W_t, \\ \mathrm{d}Z &= (XY - bZ)\,\mathrm{d}t + \beta Z\,\mathrm{d}W_t. \end{aligned} \tag{19}$$

We will refer to this type of noise as **fluctuation-dissipation** noise. It serves a similar purpose as the transport noise that we have introduced, in that in both cases the goal is to improve the models used in weather, ocean and climate prediction. These two models will be compared by analyzing their properties as random dynamical systems.

4 Lyapunov Exponents

The system of stochastic differential equations (SDE) with transport noise and the system of SDEs with fluctuation-dissipation noise satisfy the local Lipschitz continuity condition because the partial derivatives of the vector fields are all continuously differentiable functions, but is not globally Lipschitz continuous. Also, since the noise is linear and multiplicative, the growth condition is satisfied. These two conditions together are sufficient for local existence and uniqueness of solutions of the systems of SDEs [vRS10], [Sep12]. In general Stratonovich SDEs are written as

$$\mathrm{d}x_t = f_0(x_t)\,\mathrm{d}t + \sum_{j=1}^{m} f_j(x_t) \circ \mathrm{d}W_t^j = \sum_{j=0}^{m} f_j(x_t) \circ \mathrm{d}W_t^j \tag{20}$$

with the convention $\mathrm{d}W_t^0 = \mathrm{d}t$ to allow for this shorthand. Additionally, it is shown that the deterministic part is globally attracting except in a bounded set, by means of a Lyapunov function. Using several theorems from [Arn03], it is possible to generate a random dynamical system (RDS) from a Stratonovich stochastic differential equation. A RDS is a tuple (ϕ, ϑ), where ϕ is a cocycle, the solution of the dynamical system ϑ. In this text, ϑ will be the set of SDEs that are being considered. If additionally, an integrability criterion is met, then Oseledet's multiplicative ergodic theorem (MET) [Ose68] implies the regularity and the existence of Lyapunov exponents. The following theorem sets up the RDS framework in which the MET can be applied,

Theorem 4.1 (RDS from Stratonovich SDE) *Let $f_0 \in \mathcal{C}_b^{k,\delta}$, $f_1, \dots, f_m \in \mathcal{C}_b^{k+1,\delta}$ and $\sum_{j=1}^m \sum_{i=1}^d f_j^i \frac{\partial}{\partial x_i} f_j \in \mathcal{C}_b^{k,\delta}$ for some $k \geq 1$ and $\delta > 0$. Here $\mathcal{C}_b^{k,\delta}$ is the Banach space of $\mathcal{C}^k$ vector fields on $\mathbb{R}^d$ with linear growth and bounded derivatives up to order k and the k-th derivative is δ-Hölder continuous. Then:*

i)

$$\mathrm{d}x_t = \sum_{j=0}^{m} f_j(x_t) \circ \mathrm{d}W_t^j, \qquad t \in \mathbb{R} \tag{21}$$

generates a unique $\mathcal{C}^k$ RDS φ over the dynamical system (DS) describing Brownian Motion (the background theory for this can be found in [Arn03],[Elw78]). For any $\epsilon \in (0,\delta)$, φ is a $\mathcal{C}^{k,\epsilon}$-semimartingale cocycle and $(t,x) \to \varphi(t,\omega)x$ belongs to $\mathcal{C}^{0,\beta;k,\epsilon}$ for all $\beta < \frac{1}{2}$ and $\epsilon < \delta$.

ii) The RDS φ has stationary independent (multiplicative) increments, i.e. for all $t_0 \leq t_1 \leq \ldots \leq t_n$, the random variables

$$\varphi(t_1) \circ \varphi(t_0)^{-1}, \quad \varphi(t_2) \circ \varphi(t_1)^{-1}, \quad \ldots, \quad \varphi(t_n) \circ \varphi(t_{n-1})^{-1}$$

are independent and the law of $\varphi(t+h) \circ \varphi(t)^{-1}$ is independent of t. Here $\circ$ means composition.

iii) If $D\varphi(t, \omega, x)$ denotes the Jacobian of $\varphi(t, \omega)$ at x, then $(\varphi, D\varphi)$ is a $\mathcal{C}^{k-1}$ RDS uniquely generated by (21) together with

$$\mathrm{d}v_t = \sum_{j=0}^{m} Df_j(x_t)v_t \circ \mathrm{d}W_t^j, \qquad t \in \mathbb{R} \tag{22}$$

Hence $D\varphi$ uniquely solves the variational Stratonovich SDE on $\mathbb{R}$

$$D\varphi(t,x) = I + \sum_{j=0}^{m} \int_0^t Df_j(\varphi(s)x)D\varphi(s,x) \circ \mathrm{d}W_s^j, \qquad t \in \mathbb{R} \tag{23}$$

and is thus a matrix cocycle over $\Theta = (\vartheta, \varphi)$.

iv) The determinant $\det D\varphi(t, \omega, x)$ satisfies Liouville's equation on $\mathbb{R}$

$$\det D\varphi(t,x) = \exp\left(\sum_{j=0}^{m} \int_0^t trace(Df_j(\varphi(s)x) \circ \mathrm{d}W_s^j\right) \tag{24}$$

and is thus a scalar cocycle over Θ.

The proof of this theorem can be found in [Arn03]. A cocycle is a solution of the underlying SDE. The conditions for the theorem are (local) Lipschitz continuity and linear growth, since these imply (local) existence and uniqueness of solutions. We will require i), iii) and iv): i) gives us the required random dynamical system over the metric dynamical system describing Brownian motion, iii) gives us the variational equation from which Lyapunov exponents are computed and iv) gives the means to compute the sum of the Lyapunov exponents. Point ii) guarantees the independence of increments of the solution to the SDE and shows that it is a process without memory. Oseledet's MET requires the integrability condition

$$\log^+ \|D\varphi(t,\omega,x)\| \in L^1$$

which makes sure that the integrals given in iii) and iv) are well defined. The operation $\log^+$ is defined as $\log^+ x := \max(0, \log x)$. For all finite systems of SDEs (thus no stochastic partial differential equations), the Jacobian of the dynamics is square matrix. Since in $\mathbb{R}^{d\times d}$ all norms are equivalent, the condition is satisfied or dissatisfied for all norms simultaneously. If the integrability condition is satisfied, the multiplicative ergodic theorem states that $\lim_{t\to\infty}(v_t(\omega)^T v_t(\omega))^{1/2t} =: \Phi(\omega) \geq 0$ exits and logarithm of the eigenvalues of Φ are the Lyapunov exponents. By definition of the Lyapunov exponents and using Liouville's equation (24) (also called Abel-Jacobi-Liouville formula), the following important fact is derived.

Lemma 4.2 *If the trace of the Jacobian Df_0 is constant and the trace of Df_j for $j \geq 1$ is zero, then the sum of the Lyapunov exponents is equal to the trace of Df_0.*

Proof. Taking the determinant of Φ

$$\lim_{t\to\infty}\left(\det(v_t^T v_t)^{1/2}\right)^{1/t} = \lim_{t\to\infty}(\det v_t)^{1/t} = \lim_{t\to\infty}\left(\prod_{i=1}^{n} e^{\gamma_i}\right)^{1/t} \tag{25}$$

by using several properties of the determinant for square matrices. Firstly, $\det(A^T) = \det(A)$, secondly $\det(AB) = \det(A)\det(B)$. These properties allow the first step. Additionally, the determinant is related to the eigenvalues of the matrix it is acting on by $\det(A) = \prod_i \lambda_i$, where λ_i are the eigenvalues of A. Thus, let e^{γ_i} be the eigenvalues of the matrix v_t, where γ_i are the unaveraged Lyapunov exponents. Using Liouville's equation and the right hand side of (25)

$$\lim_{t\to\infty}(\det v_t)^{1/t} = \lim_{t\to\infty}\exp\left(\sum_{j=0}^{m}\int_0^t \mathrm{trace}(Df_j) \circ \mathrm{d}W_s^j\right)^{1/t} = \lim_{t\to\infty}\left(\prod_{i=1}^{n} e^{\gamma_i}\right)^{1/t} = \lim_{t\to\infty}\exp\left(\sum_{i=1}^{n}\gamma_i\right)^{1/t}.$$

Finally, using the trace conditions that were set and taking the logarithm yields

$$\sum_{i=1}^{n} \lambda_i = \lim_{t\to\infty} \left(\mathrm{trace}(Df_0)^t\right)^{1/t} = \mathrm{trace}(Df_0).$$

■

The λ_i are the Lyapunov exponents, by definition. In the notation of the theorem, the deterministic part is given by f_0

$$f_0(\mathbf{X}) = \begin{pmatrix} -\sigma & \sigma & 0 \\ r-Z & -1 & -X \\ Y & X & -b \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

where the $f_0(\mathbf{X})$ is written as the product of a matrix and a vector. This makes computing the Jacobian particularly easy. The stochastic part f_1 for the transport noise can be written as

$$f_1(\mathbf{X}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\beta \\ 0 & \beta & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

which has zero trace. The stochastic part $\bar{f}_1$ for the fluctuation-dissipation noise is given by

$$\bar{f}_1(\mathbf{X}) = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

and has nonzero trace. It is this fact that will lead to different qualitative properties of the two types of noise.

4.1 Lyapunov Function

To satisfy the integrability condition, an additional observation is required. The deterministic Lorenz equations have a global attractor set. Together with the local existence and uniqueness of strong solutions to the system of SDEs, this implies that solutions cannot blow up. To prove the existence of a globally attracting set, consider the Lyapunov function [Spa12]

$$V(\mathbf{X}) = rX^2 + \sigma Y^2 + \sigma(Z-2r)^2. \tag{26}$$

Then the time derivative is

$$\dot{V}(\mathbf{X}) = 2rX\dot{X} + 2\sigma Y\dot{Y} + 2\sigma Z\dot{Z} - 4r\sigma\dot{Z}.$$

Recall that the deterministic Lorenz equations are given by

$$\begin{aligned} \dot{X} &= \sigma(Y-X), \\ \dot{Y} &= rX - XZ - Y, \\ \dot{Z} &= XY - bZ. \end{aligned}$$

Inserting the deterministic Lorenz equations into the time derivative of the Lyapunov function leads to

$$\begin{aligned} \dot{V}(\mathbf{X}) &= 2rX(\sigma(Y-X)) + 2\sigma Y(-XZ + rX - Y) + 2\sigma Z(XY - bZ) - 4r\sigma(XY - bZ) \\ &= 2r\sigma XY - 2r\sigma X^2 - 2\sigma XYZ + 2r\sigma XY - 2\sigma Y^2 + 2\sigma XYZ - 2\sigma bZ^2 - 4r\sigma XY + 4r\sigma bZ \\ &= -2r\sigma X^2 - 2\sigma Y^2 - 2\sigma bZ^2 + 4r\sigma bZ. \end{aligned}$$

Dividing by $2r^2\sigma b$ yields the equation for an ellipsoid

$$\frac{\dot{V}(\mathbf{X})}{2r^2\sigma b} = -\frac{X^2}{br} - \frac{Y^2}{br} - \frac{(Z-r)^2}{r^2} + 1.$$

This shows that $\dot{V}$ is negative outside of the ellipsoid and positive inside the ellipsoid given by

$$\frac{X^2}{br} + \frac{Y^2}{br} + \frac{(Z-r)^2}{r^2} = 1$$

So inside the ellipsoid the dynamics are unstable, as there is no converging behavior. Outside of the ellipsoid, where $\dot{V} < 0$, the dynamics converge towards the ellipsoid. Hence $V(\mathbf{X})$ is a Lyapunov function outside of an ellipsoid. This proves that no finite time blow-up can occur for the deterministic case. Since the transport noise and fluctuation-dissipation noise Lorenz systems both satisfy linear growth, also the stochastic versions do not blow up.

4.2 Integrability Condition

We can now verify the integrability condition. The Jacobian of the system of SDEs for the Lorenz equations with transport noise is

$$Df_0 + Df_1 = \begin{pmatrix} -\sigma & \sigma & 0 \\ r-Z & -1 & -X-\beta \\ Y & X+\beta & -b \end{pmatrix}$$

and the Jacobian of the system of SDEs for the Lorenz equations with fluctuation-dissipation noise is

$$Df_0 + D\bar{f}_1 = \begin{pmatrix} -\sigma+\beta & \sigma & 0 \\ r-Z & -1+\beta & -X \\ Y & X & -b+\beta \end{pmatrix}$$

So we check whether

$$\log^+ \left\| \begin{matrix} -\sigma & \sigma & 0 \\ r-Z & -1 & -X-\beta \\ Y & X+\beta & -b \end{matrix} \right\| \in L^1$$

and

$$\log^+ \left\| \begin{matrix} -\sigma+\beta & \sigma & 0 \\ r-Z & -1+\beta & -X \\ Y & X & -b+\beta \end{matrix} \right\| \in L^1$$

which is true if all of the elements of the matrices are in L^1. This condition is violated if any of the elements of the matrix is unbounded, since then the argument of the logarithm would become unbounded. The dynamics have a global attractor and local existence and uniqueness of strong solutions, so for any initial condition, the dynamics stay bounded. Hence the integrability condition is satisfied and Oseledet's MET guarantees the existence of Lyapunov exponents. Using Liouville's equation, it can be shown that for the transport noise Lorenz system the sum of the Lyapunov exponents is equal to the deterministic case

$$\sum_{i=1}^{3} \lambda_i = -\sigma - 1 - b$$

whereas for the fluctuation-dissipation noise

$$\sum_{i=1}^{3} \lambda_i = -\sigma - 1 - b + 3\beta \lim_{t\to\infty} (W_t)^{1/t}.$$

The sum of the Lyapunov exponents resembles the average rate of expansion or contraction of phase-space volume. Hence this result shows on a theoretical level that the phase-space contraction (or expansion) of the two systems is different.

5 Computation of Lyapunov Exponents

Numerically determining the Lyapunov exponents requires the simultaneous solving of the governing dynamics and the corresponding variational equation. When the dynamics is multidimensional, the variational equation becomes a matrix differential equation. Usually, one takes the identity matrix as the initial condition for the variational equation. This corresponds evolving the unit ball along the linearized dynamics. The unit ball changes shape and it is the average deformation that is associated to the Lyapunov exponents. Directly solving the variational equation will not provide a satisfactory answer, as the vectors associated to the different Lyapunov exponents tend to all align along the direction of largest increase. Regularly orthonormalizing avoids this issue, but makes the solution procedure a bit more involved. A QR-decomposition of the matrix in the variational equation is a means to incorporate the orthonormalization. Consider the Stratonovich SDE on $\mathbb{R}^n$ given by

$$\mathsf{d}Y_t = \sum_{j=0}^{m} f_j(Y_t) \circ \mathrm{d}W_t^j \tag{27}$$

where the functions f_j are as presented in the theorem above. Here the convention $\mathrm{d}W_t^0 = \mathrm{d}t$ is used. Then the corresponding variational equation is given by

$$\mathsf{d}v_t = \sum_{j=0}^{m} Df_j(Y_t) v_t \circ \mathrm{d}W_t^j, \qquad v_0 = I, \quad v_t \in \mathbb{R}^{n\times n}, \tag{28}$$

where $Df_j(Y_t) =: J_j$ is the Jacobian of the dynamical system and $I \in \mathbb{R}^{n\times n}$ is the identity matrix. The Lyapunov exponents are defined to be the logarithm of the eigenvalues of the matrix

$$\Phi = \lim_{t\to\infty} (v_t^T v_t)^{1/2t}$$

The QR-method dictates that v is decomposed into an orthogonal matrix $Q \in O(n) := \{Q \in \mathbb{R}^{n\times n} : \det Q = \pm 1\}$ and an upper triangular matrix R, such that $v_t = QR$. It must be noted that the orthogonal matrices are allowed to have a determinant of +1 or -1. If, when solving the variational equation, the sign changes, then by a continuity argument, the matrix Q at some time between the sign change and the previous timestep would need to have a determinant that is equal to zero. This singularity can cause the algorithm to break down. Another issue with this method is that the orthogonality of Q cannot be guaranteed throughout the time integration. For this reason the so-called Cayley method [UvB01] is posed, which is an adaptation of the QR-method. The following is the stochastic generalization of the deterministic Cayley method. It is completely analogous to the deterministic case, since the stochasticity does not affect the properties of the matrices. The only change is that the differential equations become stochastic and hence have to be solved with a different method.

5.1 Standard QR Method

Setting $v_t = QR$ and multiplying from the left with Q^T and from the right with R^{-1} leads to

$$Q^T \mathsf{d}Q + \mathsf{d}RR^{-1} = \sum_{j=0}^{m} Q^T J_j Q \circ \mathrm{d}W_t^j, \qquad Q(0) = I, \quad R(0) = I. \tag{29}$$

By definition of the orthogonal matrices $Q^T Q = I$. Hence

$$0 = \mathsf{d}I = \mathsf{d}Q^T Q + Q^T \mathsf{d}Q = (Q^T \mathsf{d}Q)^T + Q^T \mathsf{d}Q$$

shows that $Q^T \mathsf{d}Q$ is skew-symmetric. Let $R^{-1} = [y_1 \dots y_n]$, where y_k for $1 \le k \le n$ is an $n \times 1$ column vector. Now, since $RR^{-1} = I = [e_1 \dots e_n]$, where e_k is the column vector with one in the k-th entry and the rest zeros, it obviously has zeros below the k-th entry. Also $RR^{-1} = R[y_1 \dots y_n] = [Ry_1 \dots Ry_n] = [e_1 \dots e_n]$. Since R is upper triangular and $Ry_k = e_k$, y_k must also have zeros below the k-th entry. This leaves to conclude that R^{-1} is upper triangular. It is known that a product of two upper triangular matrices is upper triangular, so this proves that $\mathsf{d}RR^{-1}$ is upper triangular. The procedure of solving for R starts with considering the lower triangular part of (29). Since $\mathsf{d}RR^{-1}$ is upper triangular, it does not feature here. Let

$$S_{ab} = \begin{cases} \sum_{j=0}^{m} (Q^T J_j Q)_{ab} \circ \mathrm{d}W_t^j & \text{for} \quad a > b \\ 0 & \text{for} \quad a = b \\ -\sum_{j=0}^{m} (Q^T J_j Q)_{ba} \circ \mathrm{d}W_t^j & \text{for} \quad a < b \end{cases}$$

which implies that $S = Q^T \mathsf{d}Q$. This gives a differential equation in Q only, namely

$$\mathsf{d}Q = QS, \qquad Q(0) = I.$$

The upper triangular matrix R matters because it will supply the Lyapunov exponents. This can be seen from the following calculation. By definition the Lyapunov exponents are

$$\lambda = \ln \mathrm{eig}\Big(\lim_{t\to\infty} \big((v_t^T v_t)^{1/2t}\big)\Big) = \ln \mathrm{eig}\Big(\lim_{t\to\infty} \big((R^T Q^T Q R)^{1/2t}\big)\Big) = \ln \mathrm{eig}\Big(\lim_{t\to\infty} \big((R^T R)^{1/2t}\big)\Big)$$

For any triangular matrix, its eigenvalues are on the diagonal, hence the only part of R that is important is its diagonal. Taking the transpose does not change the diagonal, so the eigenvalues of $(R^T R)^{1/2}$ are the same as the eigenvalues of R. Therefore

$$\lambda = \ln \mathrm{eig}\Big(\lim_{t\to\infty} \big(R^{1/t}\big)\Big)$$

The variable $\rho_a := \ln(R_{aa})$ is introduced, since $\mathsf{d}\rho_a = \mathsf{d}R_{aa} R_{aa}^{-1}$. So the Lyapunov exponents are determined from the solution of

$$\mathsf{d}\rho_a = \sum_{j=0}^{m} (Q^T J_j Q)_{aa} \circ \mathrm{d}W_t^j, \qquad \rho_a(0) = 0$$

as $\lambda_a = \lim_{t\to\infty} \frac{\rho_a}{t}$.

5.2 Cayley Method

The Cayley method builds upon the fact that a (special) orthogonal matrix can be constructed from a skew-symmetric matrix by the Cayley transform. A property of the transformation is that only orthogonal matrices with a determinant equal to +1 can be constructed. This greatly increases the robustness of the algorithm and solves the sign of the determinant issue. The Cayley method however relies on the Cayley transform, which is not applicable when Q has an eigenvalue close or equal to -1. The orthogonality issue is solved by restarting the calculation as soon as some condition is violated, which will be introduced later. This restarting procedure is possible due to the following lemma.

In (28), for $t > t_0$, set $v_{t_0} = Q_0 R_0$ where Q_0 is orthogonal and R_0 is upper triangular with all diagonal elements positive. As in [UvB01], the real line is divided into subintervals $t_i \leq t \leq t_{i+1}$ for $i = 1, 2, \ldots$, so that each interval has length $\Delta t_i = t_{i+1} - t_i$. The solution to the variational equation (28) at time t_i can be decomposed as $v_{t_i} = Q_i R_i$ for $i = 0, 1, 2, \ldots$. This is the preparation necessary to introduce the following lemma

Lemma 5.1 *At any time $t = t_i + \tau$, $0 \leq \tau \leq \Delta t_i$, for $i = 0, 1, 2, \ldots$, the solution of the variational equation can be expressed as*

$$v_t = v(t_i + \tau) = Q_i \tilde{v}_\tau R_i = Q_i \tilde{Q}_\tau \tilde{R}_\tau R_i, \qquad 0 \leq \tau \leq \Delta t_i, \quad t_i \leq t \leq t_{i+1},$$

where $\tilde{v}_\tau$ is the solution to the differential equation

$$\mathsf{d}\tilde{v}_\tau = \sum_{j=0}^{m} \tilde{J}_j(\tau) \tilde{v}_\tau \circ \mathrm{d}W_\tau^j, \qquad 0 \leq \tau \leq \Delta t_i, \quad \tilde{v}_0 = I, \quad i = 0, 1, 2, \ldots$$

with $Q_0 = I$, $R_0 = I$ and $\tilde{J}_j(\tau) = Q_i^T J_j(t_i + \tau) Q_i$.

The proof of this lemma can be found in [UvB01], where the variational equation is deterministic. The stochastic case is straightforwardly found from the deterministic one, as the only change is the variational equation itself. The Cayley transformation is defined as

$$Q = (I - K)(I + K)^{-1}$$

where $I \in \mathbb{R}^{n \times n}$ is the identity matrix and $K \in \mathbb{R}^{n \times n}$ is a skew-symmetric matrix. An important property of the matrices $(I - K)$ and $(I + K)^{-1}$ is that they commute. This transformation is valid as long as none of the eigenvalues of Q are equal to -1. Now a differential equation for K shall be derived, where the initial condition is determined by $Q(0) = I$, leading to $K(0) = 0$. Taking the stochastic evolution differential of Q and using the definition of the Cayley transform, the following is found

$$\mathsf{d}Q = -\mathsf{d}K(I + K)^{-1} - (I - K)(I + K)^{-1}\mathsf{d}K(I + K)^{-1}$$

Hence $Q^T \mathsf{d}Q$ is given by

$$Q^T \mathsf{d}Q = -(I + K)^{-T}(I - K)^T \mathsf{d}K(I + K)^{-1} - (I + K)^{-T}(I - K)^T(I - K)(I + K)^{-1}\mathsf{d}K(I + K)^{-1}$$

Since K is skew symmetric, for any invertible matrix $(A^T)^{-1} = (A^{-1})^T$ and using the distributive property of the transpose, the previous equation can be rewritten as

$$Q^T \mathsf{d}Q = -(I - K)^{-1}(I + K)\mathsf{d}K(I + K)^{-1} - (I - K)^{-1}(I + K)(I - K)(I + K)^{-1}\mathsf{d}K(I + K)^{-1}$$

It is here that the commutative property is necessary. Changing the order of the matrices, one obtains

$$Q^T \mathsf{d}Q = -\big((I - K)^{-1}(I + K) + I\big)\mathsf{d}K(I + K)^{-1}$$

Finally, writing $(I + K) = -(I - K) + 2I$ and setting $H := (I + K)^{-1}$ yields

$$\begin{aligned} Q^T \mathsf{d}Q &= -2(I - K)^{-1}\mathsf{d}K(I + K)^{-1} \\ &= -2H^T \mathsf{d}K H. \end{aligned} \tag{30}$$

It is not difficult to see that when $G := (I - K)$ and H as before

$$\sum_{j=0}^{m} Q^T J_j Q = \sum_{j=0}^{m} H^T G^T J_j G H \tag{31}$$

Substituting (30) and (31) in equation (29) then gives

$$-2H^T\mathrm{d}KH + \mathrm{d}RR^{-1} = \sum_{j=0}^{m} H^T G^T J_j GH \circ \mathrm{d}W_t^j \tag{32}$$

Similar to the QR-method, the first matrix on the left hand side of (32) is skew-symmetric and the second matrix is upper triangular, by the same arguments as before. Hence the solution method is also very similar, but the skew-symmetry of the matrix K provides additional advantages. Define $S := H^T\mathrm{d}KH$ so that

$$S_{ab} = \begin{cases} \frac{1}{2}\left(\sum_{j=0}^{m} H^T G^T J_j GH\right)_{ab} \circ \mathrm{d}W_t^j & \text{for} \quad a > b \\ 0 & \text{for} \quad a = b \\ -\frac{1}{2}\left(\sum_{j=0}^{m} H^T G^T J_j GH\right)_{ab} \circ \mathrm{d}W_t^j & \text{for} \quad a < b \end{cases}$$

This constitutes the differential equation for K as follows

$$\mathrm{d}K = H^{-T}SH^{-1} = \begin{cases} (G^T SG)_{ab} & \text{for} \quad a > b \\ 0 & \text{for} \quad a = b \\ -(G^T SG)_{ab} & \text{for} \quad a < b \end{cases}$$

Observe that since K is skew-symmetric, it is determined by the lower triangular part of $G^T SG$. In the QR-method, it was required to first construct S and then solve a full matrix differential equation in Q, so some computational cost is saved here. Now that K is known, the Lyapunov exponents are determined as the averages of the solutions of the differential equation for $\rho_a := \ln(R_{aa})$,

$$\mathrm{d}\rho_a = \sum_{j=0}^{m} h_a^T G^T J_j G h_a, \qquad \rho_a(0) = 0 \tag{33}$$

where h_a are the columns of $H = [h_1 \quad h_2 \quad \dots \quad h_n]$. The Lyapunov exponents are then found as

$$\lambda_a = \lim_{t\to\infty} \frac{\rho_a}{t}.$$

As a remark in [UvB01], this method of computing Lyapunov exponents is valid as long as the eigenvalues of Q do not equal -1. As the initial condition of $Q(0) = I$, there is always an interval of time $0 \le t \le t_0$ in which the condition for the Cayley transform is not violated. The following condition for restarting the algorithm is introduced: let $\eta \in [0, 1)$ be chosen by the user of the algorithm so that $\|K\| \le \eta < 1$ for some suitable norm. At time t_0, when the norm of K equals η, $Q(t_0) =: Q_0$ is computed and stored. The algorithm is restarted at that time, where due to the lemma, we have

$$\mathrm{d}v_\tau = \sum_{j=0}^{m} Q_0^T J Q_0 v_\tau \circ \mathrm{d}W_\tau^j = \sum_{j=0}^{m} \tilde{J}_j v_\tau \circ \mathrm{d}W_\tau^j$$

which is the same equation as (28) apart from the adapted Jacobian. The same solution method applies to this equation and whenever the norm of K does not satisfy our condition anymore, the algorithm is restarted in the same way. Equation (33) is solved with $\rho_a(0) = \rho_a(t_0)$ as the initial condition. The Lorenz system has been studied intensively with the standard parameter values $\sigma = 10$, $r = 28$ and $b = 8/3$, [Lor63],[Kel96],[AS01], though in the latter two for an adapted version of the Lorenz system. [WSSV85] studied the Lyapunov exponents for the deterministic Lorenz system with nonstandard parameter values $\sigma = 16$, $r = 45.92$ and $b = 4$. In particular Lorenz shows that for the standard values the deterministic Lorenz system has a strange attractor. Upon introduction of random effects in the form of Wiener processes, an attractor set as in the deterministic sense is no longer apparent, as the noise pushes the dynamics out of a bounded set almost surely, due to the unbounded variation of the Wiener process. As a result, the notion of fractal dimension or box-counting dimension etc. is not applicable to stochastic dynamical systems. The initial condition for the Lorenz system is chosen to be $(X(0), Y(0), Z(0)) = (0, 1, 0)$. The Lorenz system is then evolved for 50000 time steps and that sets the initial condition for the determination of the Lyapunov exponents. The SDEs in the Cayley method are solved with the Euler-Maruyama method with a time step size of $\Delta t = 0.001$ for 10^5 iterations in total. The norm tolerance of the matrix K is set to $\eta = 0.8$. The Euler-Maruyama method in the deterministic case is the forward Euler method. It is known that these methods have a bad convergence (1/2 for Euler-Maruyama and 1 for forward Euler), so the individual exponents can be calculated more accurately by improving the numerical schemes. Here the individual exponents for the deterministic case are given to show their values agree reasonably well with existing literature.

5.3 Deterministic Case

When there is no noise, the Liouville equation guarantees that for the Lorenz system the sum of the Lyapunov exponents is equal to the trace of the Jacobian of the dynamics. For the standard parameter values $r = 28$, $\sigma = 10$ and $b = 8/3$, the sum is given by

$$\sum_{i=1}^{3} \lambda_i = -\sigma - 1 - b = -10 - 1 - \frac{8}{3} \approx -13.6667 \tag{34}$$

and for the nonstandard parameter values $r = 45.92$, $\sigma = 16$ and $b - 4$, as used by [WSSV85],

$$\sum_{i=1}^{3} \lambda_i = -\sigma - 1 - b = -16 - 1 - 4 = -21 \tag{35}$$

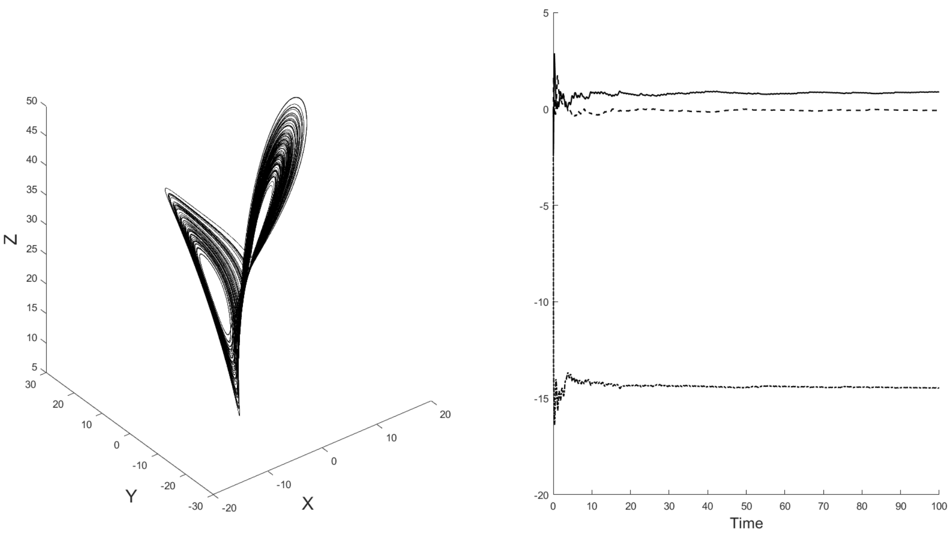


Figure 4: The deterministic Lorenz equations generate the famous butterfly shaped attractor for the standard parameter values, shown in the left figure. The figure on the right shows the convergence of the Lyapunov exponents.

The Lyapunov exponents and the sum they constitute are given in the following table. They are compared against the values computed by [Spr03].

	λ_1	λ_2	λ_3	$\sum_{i=1}^{3} \lambda_i$
Cayley method (forward Euler)	0.8739	-0.0798	-14.4606	-13.6665
Values according to [Spr03],[Spa12]	0.9056	0	-14.5721	-13.6665

Table 2: The individual Lyapunov exponents and sum for $\sigma = 10$, $r = 28$ and $b = 8/3$ as computed with the Cayley method and those found in literature.

The individual values are not exactly the same, which is due to the bad convergence of the numerical methods used here, but the sum is the same in all decimal places. The values as shown in the table are computed using a 4th order Runge-Kutta method with a fixed step size of 0.001, performed over 10^9 iterations. As an additional test, the individual values are also compared with the ones calculated by [WSSV85] for the parameter values $\sigma = 16$, $r = 45.92$ and $b = 4$. It has to be noted that in that paper the exponents are expressed in base 2, instead of in base e. After a conversion, the following values are found.

	λ_1	λ_2	λ_3	$\sum_{i=1}^{3} \lambda_i$
Cayley method (forward Euler)	1.4858	-0.0721	-22.4135	-20.9998
Values according to [WSSV85]	1.50	0	-22.46	-20.96

Table 3: The individual Lyapunov exponents and sum for $\sigma = 16$, $r = 45.92$ and $b = 4$ as computed with the Cayley method and those found in literature.

The Cayley method with forward Euler as its numerical scheme computes the sum of the Lyapunov exponents very robustly and is agreement with both the theory as well as computations in existing literature. The individual values that are found using our method are in good agreement with the values for the Lyapunov exponents in literature.

5.4 Transport Noise

Here the Lorenz equations with transport noise are studied for a noise amplitude of $\beta = 0.5$. The numerical scheme to solve the stochastic differential equations is the Euler-Maruyama method.

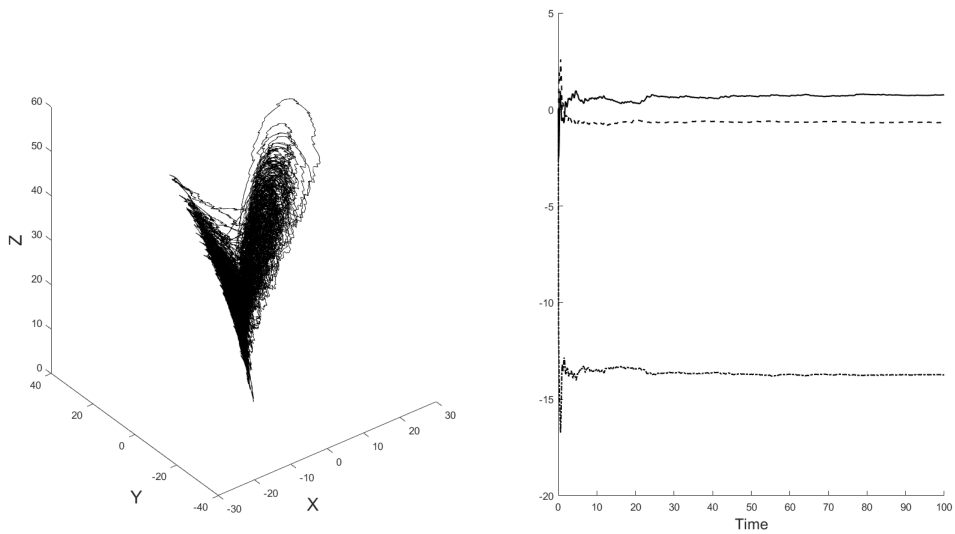


Figure 5: The Lorenz equations with transport noise ($\beta = 0.5$). The left figure shows a single realization of the stochastic dynamics. The right figure shows the convergence of the Lyapunov exponents.

The sum is -13.6665, equal in all digits to the deterministic sum. From the convergence plots it can be seen that the individual exponents have not converged completely yet. The stochastic differential equations that determine the motion and those in the Cayley method are solved with the Euler-Maruyama method, which has a convergence of order 1/2. In the deterministic case, the order of convergence of the numerical method is 1. Hence to determine the individual exponents accurately, one has to run for much longer. The sum however is obtained accurately very quickly.

5.5 Fluctuation - Dissipation Noise

The Lorenz equations are made stochastic using fluctuation - dissipation noise with amplitude $\beta = 0.5$.

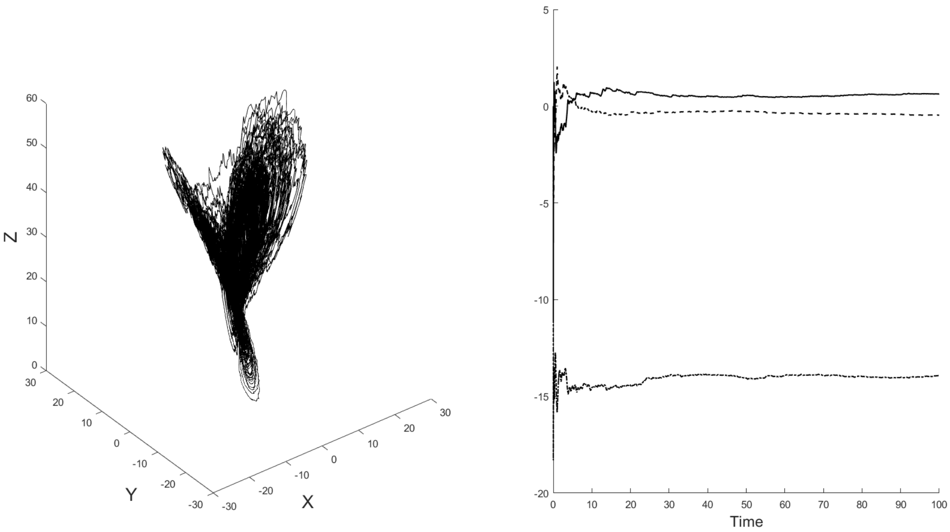


Figure 6: The Lorenz equations with fluctuation - dissipation noise ($\beta = 0.5$). The left figure shows a single realisation of the stochastic dynamics. The right figure shows the convergence of the Lyapunov exponents.

Similarly to the transport noise version, the individual exponents change for each realization of the Wiener process. The sum is -13.7636. This is a change in the first decimal place compared to the deterministic and the transport noise case. This confirms the theory for a single realization.

5.6 Sum against Noise Amplitude

The constancy of the sum for the transport noise becomes especially clear in the following plot.

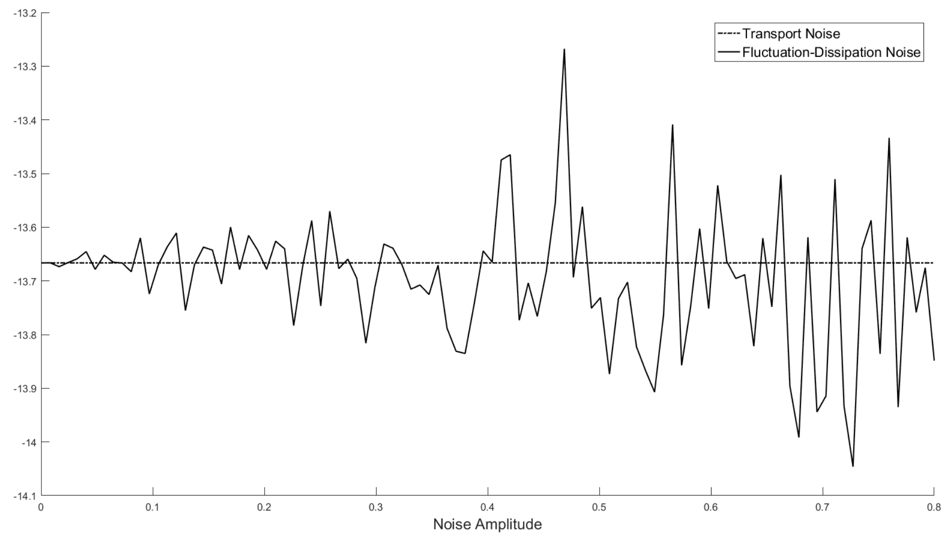


Figure 7: Sum of the Lyapunov exponents for the different types of noise for varying noise amplitude. Each step in noise amplitude is for a different realization of the Wiener process.

The plot shows 100 computations for increasing noise amplitude. At each computation there is a different path of the Wiener process. As expected from the theory, the difference between the sum for the transport noise case and the fluctuation-dissipation noise case increases with increasing noise amplitude. The theory shows a linear relationship between sum and noise amplitude for a fixed path of the Wiener process. The next plot shows that this is indeed the case.

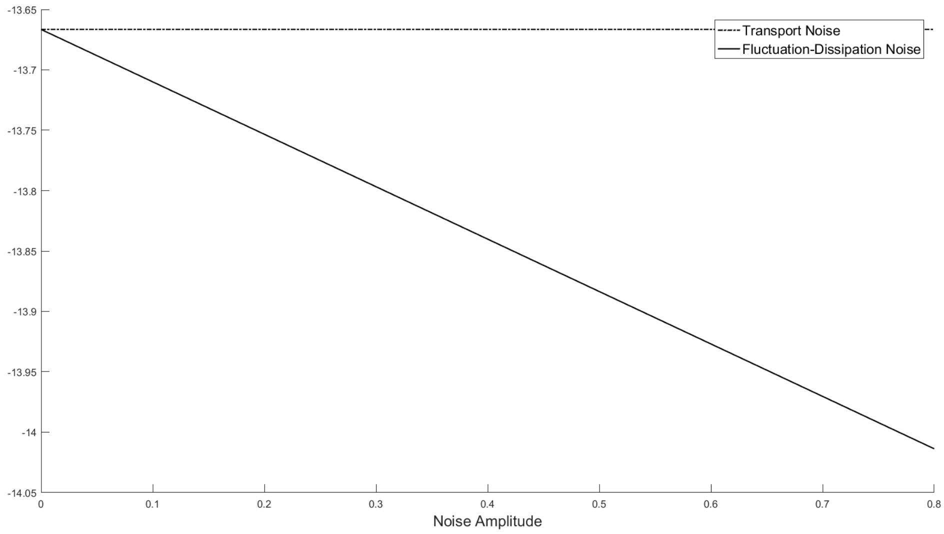


Figure 8: Sum of the Lyapunov exponents for the different types of noise for varying noise amplitude. Each step in noise amplitude has the same realization of the Wiener process.

This means that the transport noise does not affect the average contraction-expansion rate of the underlying deterministic system. In the special case of a Hamiltonian system, which is symplectic and hence preserves phase space volume, the Lyapunov exponents sum up to zero. Adding a type of noise that affects the sum of the Lyapunov exponents thus destroys the Hamiltonian structure. For the Lorenz system, the transport noise does not change its dissipative properties, whereas the fluctuation-dissipation noise does.

5.7 Individual Exponents

The individual exponents for the two noisy Lorenz systems are analyzed using the same methods as before. Although the accuracy can be improved by using better numerical methods for solving the stochastic differential equations, we have shown that the individual exponents closely resemble the values found in literature for the deterministic case. Here the individual Lyapunov exponents are computed for the two types of stochastic Lorenz equations. For both computations are done using the same realization of the Wiener process. The individual exponents versus the noise amplitude for the transport noise Lorenz system are given in the following figure.

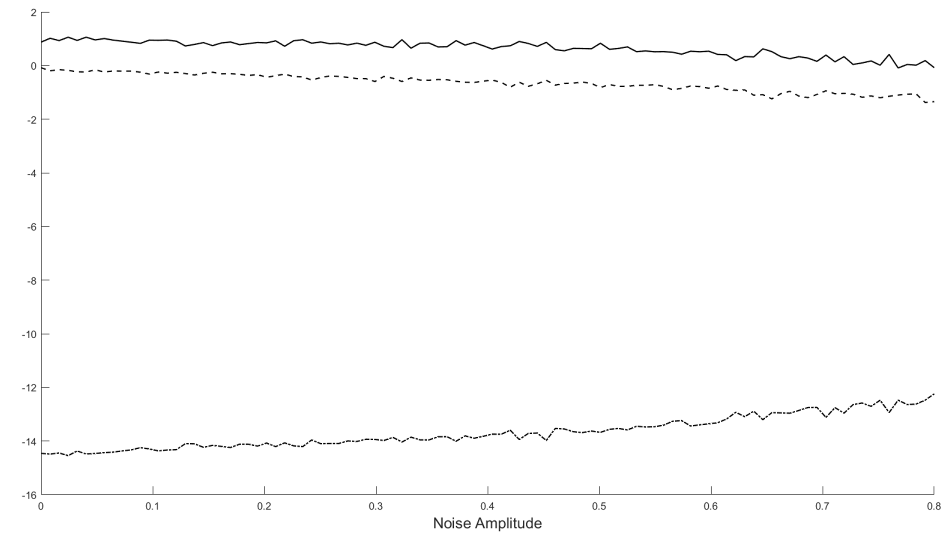


Figure 9: The individual Lyapunov exponents for the Lorenz system with transport noise for a fixed realization of the Wiener process. The bottom exponent increases to compensate for the decrease of the top two exponents. This maintains the constant value of the sum.

The noise brings the individual exponents closer together, in such a way that the sum remains the same. The average rate of phase-space volume contraction hence is constant as a function of the noise amplitude, but the individual exponents change. The fluctuation-dissipation noise behaves differently, though the Lyapunov exponents themselves have a similar behavior compared to the transport noise, as can be seen in the following figure.

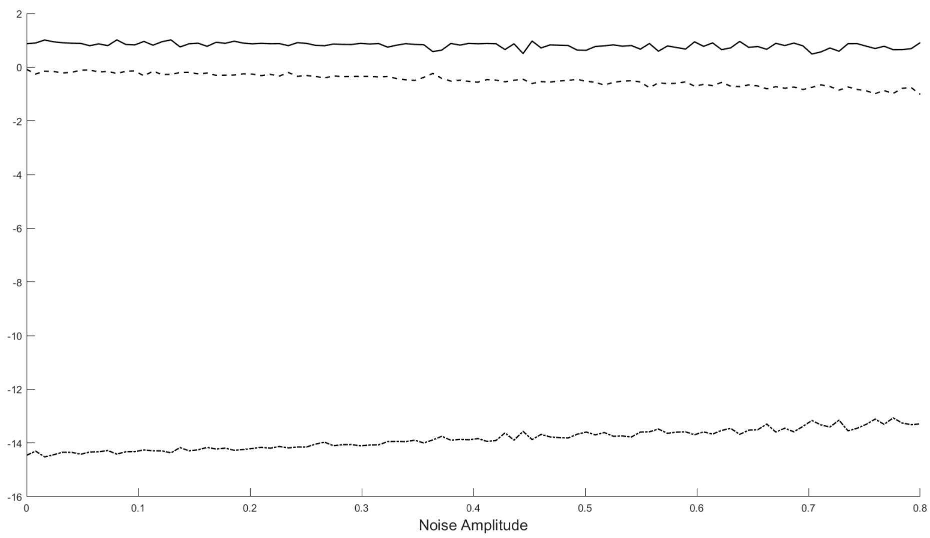
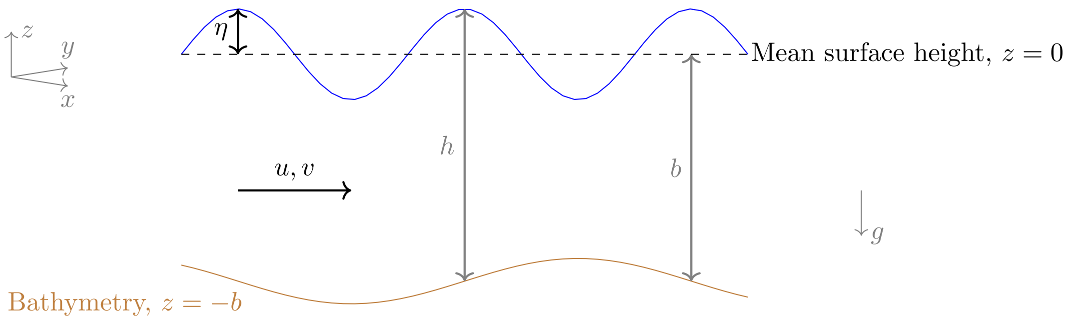


Figure 10: The individual Lyapunov exponents for the Lorenz system with fluctuation-dissipation noise for a fixed realization of the Wiener process. The bottom exponent does not increase enough to compensate for the decrease of the top two and causes the sum to decrease.

In both cases the noise changes the individual exponents. The transport noise decreases the amplitude of the two highest exponents and increases the lowest exponent to keep the sum constant. The fluctuation-dissipation noise has a similar effect, though it is a lot weaker, the two highest exponents decrease, but not as strongly as with transport noise. The lowest exponent increases, but not fast enough to keep the sum constant.

6 Stochastic Rotating Shallow Water

The rotating shallow water (RSW) equations apply when modeling fluids in domains where the horizontal scales are much larger than the depth scale. They can be derived by using this shallowness approximation and then depth integrating the Navier-Stokes equations or the Euler equations, depending on whether viscosity is important in the model.



The method of deriving the RSW equations from another fluid model is not very helpful when we want to introduce stochasticity. We will use the stochastically constrained variational principle was introduced in [Hol15] to have a means of deriving equations in continuum mechanics rigorously with stochasticity. The Lagrangian $\ell(\mathbf{u}, \eta)$ for rotating shallow water is given by

$$\ell(\mathbf{u}, \eta) = \int \frac{\epsilon}{2}\eta|\mathbf{u}|^2 + \eta\mathbf{u}\cdot\mathbf{R} - \frac{(\eta - b)^2}{\epsilon\mathcal{F}}\,\mathrm{d}^2x \tag{36}$$

where $\mathbf{u} \in \mathfrak{X}(\mathbb{R}^2)$ is a vector field on $\mathbb{R}^2$ and $\eta \in V$ is the depth; an advected quantity. The vector space V contains the advected quantities of all types. In the most general case V is the set of differential forms of all degrees, which for $\mathbb{R}^2$ would be $V := \{a, \mathbf{b}\cdot \mathrm{d}x, d\,\mathrm{d}^2x\}$, where $a, \mathbf{b}, d$ are scalar and vector valued functions on $\mathbb{R}^2$. In this Lagrangian, the only advected quantity is η, the depth, which is a density. Furthermore, ϵ denotes the Rossby number, b describes the bottom topography, $\mathcal{F}$ is the Froude number and $\mathbf{R}$ is the Coriolis vector field. The top form in the domain of the rotating shallow water equations is d^2x, so here $\eta = \eta\,\mathrm{d}^2x$, by abuse of the notation. It satisfies the advection relation

$$(\mathsf{d} + \pounds_{\mathrm{d}\mathbf{x}_t})(\eta\,\mathrm{d}^2x) = 0$$

Here $\pounds_{\mathrm{d}\mathbf{x}_t}(\eta\,\mathrm{d}^2x)$ is the Lie derivative of the depth with respect to the stochastic vector field $\mathrm{d}\mathbf{x}_t := \mathbf{u}\,\mathrm{d}t + \xi_i \circ \mathrm{d}W_t^i$. Summing over repeated indices is understood. Taking the Lie derivative of different types of tensors leads to different expressions and for that reason it is necessary to keep the abstract notation through the application of Hamilton's principle. In the advection relation, the type of tensor is known (the depth is a top form), so the Lie derivative can be computed

$$\big(\mathsf{d}\eta + \operatorname{div}(\eta\,\mathrm{d}\mathbf{x}_t)\big)\,\mathrm{d}^2x = 0 \tag{37}$$

The stochastically constrained action as in [Hol15] with the RSW Lagrangian is

$$S(\mathbf{u}, \eta, p) = \int_a^b \ell(\mathbf{u}, \eta) + \langle p, \mathsf{d}\eta + \pounds_{\mathrm{d}\mathbf{x}_t}\eta\rangle_V\,\mathrm{d}t \tag{38}$$

Hamilton's variational principle applied on the action S leads to the so-called Euler-Poincaré equations, which may be used in deriving the stochastic rotating shallow water (SRSW) equations. Hamilton's variational principle implies that $\delta S = 0$, where the δ operator means to take a variational derivative. First we write S into a more convenient form, where the action is split into a deterministic (Lebesgue) integral and a stochastic (Stratonovich) integral. This step requires the diamond operation.

Definition 6.1 (Diamond operation) *The diamond operation $\diamond : T^*V \to \mathfrak{X}^*$ is defined for a vector space V with $(a, b) \in T^*V$ and a vector field $w \in \mathfrak{X}$ is defined using the Lie derivative as*

$$\langle b \diamond a, w\rangle_V := \langle b, -\pounds_w a\rangle_{\mathfrak{X}} \tag{39}$$

The diamond operation depends on the Lie derivative, which changes form depending on what type of tensor a is. The diamond operation greatly simplifies taking variations. It allows us to change pairing and thereby grants the possibility to take variations of the vector field along which the Lie derivative is evaluated. Rewriting yields

$$S(\mathbf{u}, \eta, p) = \int_a^b \left(\ell(\mathbf{u}, \eta) + \left\langle p, \frac{\mathrm{d}\eta}{\mathrm{d}t} + \pounds_{\mathbf{u}}\eta\right\rangle_V\right)\mathrm{d}t - \int_a^b \langle p \diamond \eta, \xi_i\rangle_{\mathfrak{X}} \circ \mathrm{d}W_t^i \tag{40}$$

Applying Hamilton's variational principle leads to

$$
\begin{aligned}
0 = \delta S &= \delta \int_a^b \left(\ell(\mathbf{u},\eta) + \left\langle p, \frac{\mathrm{d}\eta}{\mathrm{d}t} + \pounds_{\mathbf{u}}\eta \right\rangle_V \right) \mathrm{d}t + \delta \int_a^b \langle p \diamond \eta, \xi_i \rangle_{\mathfrak{X}} \circ \mathrm{d}W_t^i \\
&= \int_a^b \left[\left\langle \frac{\delta \ell}{\delta \mathbf{u}}, \delta \mathbf{u} \right\rangle_{\mathfrak{X}} + \left\langle \frac{\delta \ell}{\delta \eta}, \delta \eta \right\rangle_V + \left\langle \frac{\mathrm{d}\eta}{\mathrm{d}t} + \pounds_{\mathbf{u}}\eta, \delta p \right\rangle_V + \left\langle -\frac{\mathrm{d}p}{\mathrm{d}t} + \pounds_{\mathbf{u}}^T p, \delta\eta \right\rangle_V + \langle -p \diamond \eta, \delta \mathbf{u} \rangle_{\mathfrak{X}} \right] \mathrm{d}t \\
&\quad - \int_a^b \left[\langle -\pounds_{\xi_i}\eta, \delta p \rangle_V - \langle \pounds_{\xi_i}^T p, \delta \eta \rangle_V \right] \circ \mathrm{d}W_t^i \\
&= \int_a^b \left[\left\langle \frac{\delta \ell}{\delta \mathbf{u}}\,\mathrm{d}t - p \diamond \eta \, \mathrm{d}t, \delta \mathbf{u} \right\rangle_{\mathfrak{X}} + \left\langle \frac{\delta \ell}{\delta \eta}\,\mathrm{d}t - \mathsf{d}p + \pounds_{\mathrm{d}\mathbf{x}_t}^T p, \delta\eta \right\rangle_V + \langle \mathsf{d}\eta + \pounds_{\mathrm{d}\mathbf{x}_t}\eta, \delta p \rangle_V \right] \mathrm{d}t
\end{aligned}
$$

In this derivation the diamond operation was used a number of times, as well as integration by parts and the fact that the adjoint of the Lie derivative is its transpose. The notation $\frac{\delta \ell}{\delta \mathbf{u}}$ denotes a partial derivative arising from a variation. As the functions $\delta\mathbf{u}, \delta\eta$ and δp are arbitrary, to assure that $\delta S = 0$, the terms that they multiply have to be zero. This yields the following set of equations

$$
\begin{aligned}
\delta \mathbf{u} :&\quad \frac{\delta \ell}{\delta \mathbf{u}} = \epsilon \eta \mathbf{u} + \eta \mathbf{R} = p \diamond \eta \\
\delta \eta :&\quad \frac{\delta \ell}{\delta \eta} = \frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - \frac{\eta - b}{\epsilon \mathcal{F}}\,\mathrm{d}t = \mathsf{d}p - \pounds_{\mathrm{d}\mathbf{x}_t}^T p \\
\delta p :&\quad \mathsf{d}\eta + \pounds_{\mathrm{d}\mathbf{x}_t}\eta = 0
\end{aligned}
\tag{41}
$$

The momentum m, defined as

$$
m := \frac{\delta \ell}{\delta \mathbf{u}} = \epsilon \eta \mathbf{u} + \eta \mathbf{R}
$$

is dual to $\delta\mathbf{u}$, which is a vector field. This can be seen from Hamilton's variational principle, where the momentum is paired with $\delta\mathbf{u}$. This makes the momentum part of the 1-form densities. The momentum satisfies the momentum equation, which is a result from the general theory presented in [Hol15].

$$
\mathsf{d}m + \pounds_{\mathrm{d}\mathbf{x}_t} m = \frac{\delta \ell}{\delta \eta} \diamond \eta \, \mathrm{d}t \tag{42}
$$

At this point is possible to explicitly write the action of the diamond operator by going back to its definition. The formal adjoint of the gradient is minus the divergence on any $\mathbb{R}^n$ space, so

$$
\begin{aligned}
\left\langle \frac{\delta \ell}{\delta \eta} \diamond \eta, w \right\rangle_V &= -\int \frac{\delta \ell}{\delta \eta} \cdot \pounds_w (\eta \, \mathrm{d}^2 x) \\
&= -\int \frac{\delta \ell}{\delta \eta} \cdot \operatorname{div}(\eta w)\, \mathrm{d}^2 x \\
&= -\int \left\langle \frac{\delta \ell}{\delta \eta}, \operatorname{div}(\eta w) \right\rangle_{\mathbb{R}^2} \mathrm{d}^2 x \\
&= \int \left\langle \nabla \frac{\delta \ell}{\delta \eta}, \eta w \right\rangle_{\mathbb{R}^2} \mathrm{d}^2 x
\end{aligned}
$$

Since η is a scalar, the diamond operation is

$$
\frac{\delta \ell}{\delta \eta} \diamond \eta = \eta \nabla \frac{\delta \ell}{\delta \eta}
$$

Using the momentum equation and the advection equation for the depth η, we derive the momentum equation per unit depth, which provides us with the equation of motion for stochastic rotating shallow water. The advection equation for the 1-form $\frac{m}{\eta}$ satisfies

$$
\begin{aligned}
\mathsf{d}\left(\frac{m}{\eta}\right) + \pounds_{\mathrm{d}\mathbf{x}_t}\left(\frac{m}{\eta}\right) &= \frac{1}{\eta}\left(\mathsf{d}m + \pounds_{\mathrm{d}\mathbf{x}_t} m - \frac{m}{\eta}\left(\mathsf{d}\eta + \pounds_{\mathrm{d}\mathbf{x}_t}\eta\right)\right) \\
&= \frac{1}{\eta}(\mathsf{d}m + \pounds_{\mathrm{d}\mathbf{x}_t} m) \\
&= \frac{1}{\eta}\frac{\delta \ell}{\delta \eta} \diamond \eta
\end{aligned}
$$

As m is a 1-form density, dividing it by the depth, which is a density, yields the 1-form $\frac{m}{\eta}$. This is the information that is necessary to calculate what the Lie derivative of $\frac{m}{\eta}$ is. Using the identity for diamond operation, we find the momentum (per unit depth) equation for SRSW.

$$\mathsf{d}(\epsilon\mathbf{u}+\mathbf{R}) + \mathrm{d}\mathbf{x}_t\cdot\nabla(\epsilon\mathbf{u}+\mathbf{R}) + (\epsilon u+R)_j\nabla\,\mathrm{d}x_t^j = \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - \frac{\eta-b}{\epsilon\mathcal{F}}\right)\mathrm{d}t \tag{43}$$

6.1 Alternative Formulation

The fundamental vector identity of fluid dynamics allows us to rewrite equation (43) in an alternative, equivalent form that is convenient when deriving the vorticity equation. Setting $h := \frac{\eta-b}{\epsilon\mathcal{F}}$ and applying the fundamental vector identity gives

$$\epsilon\mathsf{d}\mathbf{u} - \mathrm{d}\mathbf{x}_t\times\mathrm{curl}(\epsilon\mathbf{u}+\mathbf{R}) + \nabla(\mathrm{d}\mathbf{x}_t\cdot(\epsilon\mathbf{u}+\mathbf{R})) = \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - h\right)\mathrm{d}t$$

The terms on the right hand side allow for a remarkable cancellation with the deterministic part of the advection terms on the left hand side. By expanding the the stochastic vector field into its deterministic part and its stochastic part, the previous equation may be rewritten as

$$\epsilon\mathsf{d}\mathbf{u} - \mathrm{d}\mathbf{x}_t\times\mathrm{curl}(\epsilon\mathbf{u}+\mathbf{R}) + \nabla(\mathbf{u}\,\mathrm{d}t\cdot(\epsilon\mathbf{u}+\mathbf{R})) + \nabla(\xi_i\circ\mathrm{d}W_t^i\cdot(\epsilon\mathbf{u}+\mathbf{R})) = \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - h\right)\mathrm{d}t$$

The cancellation eliminates removes the Coriolis term on the right hand side and using $\nabla(\mathbf{u}\cdot\epsilon\mathbf{u}) = \epsilon\nabla|\mathbf{u}|^2$, the curl form is obtained

$$\epsilon\mathsf{d}\mathbf{u} - \mathrm{d}\mathbf{x}_t\times\mathrm{curl}(\epsilon\mathbf{u}+\mathbf{R}) + \nabla(\xi_i\circ\mathrm{d}W_t^i\cdot(\epsilon\mathbf{u}+\mathbf{R})) + \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + h\right)\mathrm{d}t = 0 \tag{44}$$

Introducing the stochastic constraint in the variational principle makes the Lie derivative in the momentum equation being evaluated along a different vector field. It is a random vector field, but that does not change the structure or geometry of the problem. Therefore, including stochasticity by this variational principle will lead to everything being advected along a random vector field.

The stochastic rotating shallow water equations are given by the momentum equation (43) or alternatively, the equivalent curl form (44) and the depth equation (37). In an overview, this is

$$\begin{aligned}\epsilon\mathsf{d}\mathbf{u} + \mathrm{d}\mathbf{x}_t\cdot\nabla(\epsilon\mathbf{u}+\mathbf{R}) + (\epsilon u+R)_j\nabla\,\mathrm{d}x_t^j &= \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - h\right)\mathrm{d}t,\\ \epsilon\mathsf{d}\mathbf{u} - \mathrm{d}\mathbf{x}_t\times\mathrm{curl}(\epsilon\mathbf{u}+\mathbf{R}) + \nabla(\xi_i\circ\mathrm{d}W_t^i\cdot(\epsilon\mathbf{u}+\mathbf{R})) + \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + h\right)\mathrm{d}t &= 0,\\ \mathsf{d}\eta + \mathrm{div}(\eta\,\mathrm{d}\mathbf{x}_t) &= 0.\end{aligned} \tag{45}$$

6.2 Validity

The stochastic rotating shallow water equations have been derived using a variational principle. The validity of these equations can be tested once more by removing the noise and seeing whether the familiar deterministic equations can be obtained. The depth equation and the curl form are quite obvious, but the momentum equation follows from a number of nontrivial cancellations. The deterministic equations are

$$\begin{aligned}\epsilon\partial_t\mathbf{u} + \epsilon\mathbf{u}\cdot\nabla\mathbf{u} + f\hat{\mathbf{z}}\times\mathbf{u} + \nabla h &= 0,\\ \epsilon\partial_t\mathbf{u} - \mathbf{u}\times\mathrm{curl}(\epsilon\mathbf{u}+\mathbf{R}) + \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + h\right) &= 0,\\ \partial_t\eta + \mathrm{div}(\eta\mathbf{u}) &= 0.\end{aligned}$$

The following calculation shows that the stochastic equations indeed equal the deterministic equations upon removal of the noise. Without noise, $\mathrm{d}\mathbf{x}_t = \mathbf{u}\,\mathrm{d}t$, which turns the equation of motion into

$$\epsilon\mathsf{d}\mathbf{u} + \mathbf{u}\,\mathrm{d}t\cdot\nabla(\epsilon\mathbf{u}+\mathbf{R}) + (\epsilon u+R)_j\nabla u^j\,\mathrm{d}t = \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - h\right)\mathrm{d}t.$$

It is now possible to switch from the stochastic integral notation back to the deterministic differential notation. The equation of motion is then

$$\epsilon\partial_t\mathbf{u} + \epsilon\mathbf{u}\cdot\nabla\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{R} + \epsilon u_j\nabla u^j + R_j\nabla u^j = \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - h\right)$$

where the $\epsilon\mathbf{u} + \mathbf{R}$ terms have been split up. The term $\epsilon u_j \nabla u^j = \frac{\epsilon}{2}\nabla|\mathbf{u}|^2$, hence

$$\epsilon\partial_t \mathbf{u} + \epsilon\mathbf{u}\cdot\nabla\mathbf{u} + \frac{\epsilon}{2}\nabla|\mathbf{u}|^2 + \mathbf{u}\cdot\nabla\mathbf{R} + R_j\nabla u^j = \frac{\epsilon}{2}\nabla|\mathbf{u}|^2 + \nabla(\mathbf{u}\cdot\mathbf{R}) - \nabla h$$

The fundamental vector identity of fluid dynamics allows the transformation $\mathbf{u}\cdot\nabla\mathbf{R} + R_j\nabla u^j - \nabla(\mathbf{u}\cdot\mathbf{R}) = -\mathbf{u}\times\operatorname{curl}\mathbf{R}$. We obtain

$$\epsilon\partial_t \mathbf{u} + \epsilon\mathbf{u}\cdot\nabla\mathbf{u} - \mathbf{u}\times\operatorname{curl}\mathbf{R} + \nabla h = 0$$

The curl of the Coriolis vector field is the Coriolis force in the vertical direction, $\operatorname{curl}\mathbf{R} = f\hat{\mathbf{z}}$. Substitution of this relation into the previous equation leads to the familiar deterministic form of the RSW equations

$$\epsilon\partial_t \mathbf{u} + \epsilon\mathbf{u}\cdot\nabla\mathbf{u} + f\hat{\mathbf{z}}\times\mathbf{u} + \nabla h = 0$$

6.3 Fast-Slow Split and Conservation Laws

The rotating shallow water equations possess slow and fast time-scale dynamics. To identify these time-scales, the equations have to be reformulated in terms of potential vorticity, divergence of the velocity field and the divergence of the balance condition.

The potential vorticity can be derived algebraically from the equation for the vertical component of the total vorticity. To derive this equation, take the the dot product of the vertical unit vector with the curl of the curl form of the SRSW equations. This yields, when $\varpi = \hat{\mathbf{z}}\cdot\operatorname{curl}(\epsilon\mathbf{u} + \mathbf{R})$,

$$\begin{aligned}
0 &= \hat{\mathbf{z}}\cdot\operatorname{curl}\left(\epsilon \mathsf{d}\mathbf{u} - \mathrm{d}\mathbf{x}_t\times\operatorname{curl}(\epsilon\mathbf{u}+\mathbf{R}) + \nabla(\xi_i\circ \mathrm{d}W_t^i\cdot(\epsilon\mathbf{u}+\mathbf{R})) + \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + h\right)\mathrm{d}t\right)\\
&= \mathsf{d}\varpi - \hat{\mathbf{z}}\cdot\operatorname{curl}(\mathrm{d}\mathbf{x}_t\times\operatorname{curl}(\epsilon\mathbf{u}+\mathbf{R}))\\
&= \mathsf{d}\varpi + \varpi\operatorname{div}\mathrm{d}\mathbf{x}_t + \mathrm{d}\mathbf{x}_t\cdot\nabla\varpi - \varpi\cdot\nabla\,\mathrm{d}\mathbf{x}_t
\end{aligned}$$

The SRSW equations describe the horizontal velocity field, whereas the total vorticity is in the vertical direction. This results in that the vortex stretching term is zero as it is the dot product between the gradient of the horizontal velocity and the (vertical) total vorticity, hence the total vorticity equation can be written as

$$\mathsf{d}\varpi + \mathrm{d}\mathbf{x}_t\cdot\nabla\varpi + \varpi\operatorname{div}\mathrm{d}\mathbf{x}_t = 0$$

The depth equation can be rewritten directly, to yield

$$\mathsf{d}\eta + \mathrm{d}\mathbf{x}_t\cdot\nabla\eta + \eta\operatorname{div}\mathrm{d}\mathbf{x}_t = 0$$

6.3.1 Potential Vorticity

It is now possible to prove that the SRSW equations conserve potential vorticity $q := \frac{\varpi}{\eta}$ along stochastic Lagrangian paths.

$$\begin{aligned}
\mathsf{d}q + \mathrm{d}\mathbf{x}_t\cdot\nabla q &= \mathsf{d}\left(\frac{\varpi}{\eta}\right) + \mathrm{d}\mathbf{x}_t\cdot\nabla\left(\frac{\varpi}{\eta}\right)\\
&= \frac{1}{\eta}\left(\mathsf{d}\varpi + \mathrm{d}\mathbf{x}_t\cdot\nabla\varpi - \frac{\varpi}{\eta}(\mathsf{d}\eta + \mathrm{d}\mathbf{x}_t\cdot\nabla\eta)\right)\\
&= \frac{1}{\eta}\left(-\varpi\operatorname{div}\mathrm{d}\mathbf{x}_t + \frac{\varpi}{\eta}\eta\operatorname{div}\mathrm{d}\mathbf{x}_t\right)\\
&= 0
\end{aligned}$$

The conservation of potential vorticity is crucial for atmospheric models, as disturbances of potential vorticity gives rise to Rossby waves, which define the high and low pressure zones on earth and the jet stream. They have a major influence on the day-to-day weather patterns at mid-latitudes. Rossby waves also occur in the ocean.

6.3.2 Integral Quantities

The conservation of potential vorticity implies the conservation of infinitely many integral quantities.

$$\begin{aligned}
\mathsf{d}\int \eta\Phi(q)\,\mathrm{d}^2x &= \int \mathsf{d}\eta\Phi(q) + \eta\Phi'(q)\mathsf{d}q\,\mathrm{d}^2x \\
&= -\int \mathrm{div}(\eta\,\mathrm{d}\mathbf{x}_t)\Phi(q) + \eta\Phi'(q)(\mathrm{d}\mathbf{x}_t\cdot\nabla q)\,\mathrm{d}^2x \\
&= -\int \mathrm{div}(\eta\,\mathrm{d}\mathbf{x}_t)\Phi(q) + \eta\,\mathrm{d}\mathbf{x}_t\cdot\nabla\Phi(q)\,\mathrm{d}^2x \\
&= -\int \mathrm{div}(\eta\Phi(q)\,\mathrm{d}\mathbf{x}_t)\,\mathrm{d}^2x \\
&= -\oint \eta\Phi(q)\hat{\mathbf{n}}\cdot\mathrm{d}\mathbf{x}_t\,\mathrm{d}S \\
&= 0
\end{aligned}$$

The integral over the boundary of the domain vanishes due to the velocity being tangent to the boundary. Among these conserved integral quantities, there is one that stands out. The enstrophy q^2 is conserved as it is a special case of this result and is important because it is the quantity that is directly related to the dissipation in the kinetic energy of the flow model. It is particularly useful in the study of turbulence.

6.3.3 Kelvin Circulation Theorem

To prove the Kelvin theorem for SRSW, we use lemma 2.7. The SRSW equations satisfy the Kelvin theorem with 1-form $m/\eta = (\epsilon\mathbf{u}+\mathbf{R})\cdot\mathrm{d}\mathbf{x}$ for a loop $c(t)$ moving in the stochastic vector field $\mathrm{d}\mathbf{x}_t$

$$\begin{aligned}
\mathsf{d}\oint_{c(t)} (\epsilon\mathbf{u}+\mathbf{R})\cdot\mathrm{d}\mathbf{x} &= \mathsf{d}\oint_{c(0)} \phi_t^*(\epsilon\mathbf{u}+\mathbf{R})\cdot\mathrm{d}\mathbf{x} \\
&= \oint_{c(0)} \phi_t^*(\mathsf{d}+\pounds_{\mathrm{d}\mathbf{x}_t})\big((\epsilon\mathbf{u}+\mathbf{R})\cdot\mathrm{d}\mathbf{x}\big) \\
&= \oint_{c(t)} (\mathsf{d}+\pounds_{\mathrm{d}\mathbf{x}_t})\big((\epsilon\mathbf{u}+\mathbf{R})\cdot\mathrm{d}\mathbf{x}\big) \\
&= \oint_{c(t)} \big(\epsilon\mathsf{d}\mathbf{u} + \mathrm{d}\mathbf{x}_t\cdot\nabla(\epsilon\mathbf{u}+\mathbf{R}) + (\epsilon u+R)_j\nabla\,\mathrm{d}x_t^j\big)\cdot\mathrm{d}\mathbf{x} \\
&= \oint_{c(t)} \nabla\left(\frac{\epsilon}{2}|\mathbf{u}|^2 + \mathbf{u}\cdot\mathbf{R} - h\right)\mathrm{d}t\cdot\mathrm{d}\mathbf{x} \\
&= 0
\end{aligned}$$

In this calculation, the essential step is to transform to Lagrangian coordinates, so that the domain of the integral becomes stationary. This then allows the stochastic differential to be pulled through the integral sign. The final step is due to the fundamental theorem of calculus. The Kelvin theorem is a result of Noether's theorem, which implies that to every continuous symmetry, there is an associated conservation law. The symmetry from which the Kelvin theorem arises is that the Lagrangian particles are invariant under relabeling. This can be done in uncountably many ways and hence gives rise to uncountably many conservation laws.

In contrast to the deterministic case, the energy is not conserved. This is because the noise acts as an explicitly time-dependent force, which makes the system nonautonomous and hence cannot conserve energy.

6.3.4 Fast-Slow Split

The SRSW equations can be split into fast and slow motion. The slow motion will be the potential vorticity, of which disturbances lead to Rossby waves. The fast motion will generated by the divergence of the velocity field and the divergence of the balance condition, of which disturbances lead to gravity waves. These particular variables are interesting as they will turn out to be canonically conjugate variables in a Hamiltonian formulation of these equations. The divergence of the velocity field $\mathcal{D} := \mathrm{div}\,\mathbf{u}$ is obtained by taking the divergence of the motion equation. It is convenient to split up the deterministic and random parts on the left hand side, since this allows us to introduce the balance condition.

$$\mathrm{div}\left(\epsilon\mathsf{d}\mathbf{u} + \epsilon\mathbf{u}\,\mathrm{d}t\cdot\nabla\mathbf{u} + \xi_i\circ\mathrm{d}W_t^i\cdot\nabla(\epsilon\mathbf{u}+\mathbf{R}) + (\epsilon u+R)_j(\xi_i\circ\mathrm{d}W_t^i)^j\right) = -\,\mathrm{div}\left(f\hat{\mathbf{z}}\times\mathbf{u}+\nabla h\right)\mathrm{d}t$$

The divergence of the advection term $\epsilon \mathbf{u}\,\mathrm{d}t \cdot \nabla \mathbf{u} = \epsilon\,\mathrm{div}(\mathcal{D}\mathbf{u})\,\mathrm{d}t - 2\epsilon J(u,v)\,\mathrm{d}t$, where J is the Jacobian. Note that the third and the fourth term on the left hand side constitute the Lie derivative of a 1-form with respect to $\xi_i \circ \mathrm{d}W_t^i$. For notation purposes, the vector field $\Gamma := \xi_i \circ \mathrm{d}W_t^i$ will be used for the computations. The divergence of the Lie derivative can be expanded further by noting that the divergence can be written as the exterior derivative, which commutes with the Lie derivative

$$\begin{aligned}\mathrm{d}\pounds_\Gamma\big((\epsilon\mathbf{u}+\mathbf{R})\cdot \mathrm{d}\mathbf{x}\big) &= \pounds_\Gamma\,\mathrm{d}\big((\epsilon\mathbf{u}+\mathbf{R})\cdot\mathrm{d}\mathbf{x}\big)\\ &= \pounds_\Gamma\,\mathrm{div}(\epsilon\mathbf{u}+\mathbf{R})\,\mathrm{d}^2x\\ &= \mathrm{div}(\mathrm{div}(\epsilon\mathbf{u}+\mathbf{R})\Gamma)\,\mathrm{d}^2x\end{aligned}$$

This recasts the divergence equation into

$$\epsilon\mathrm{d}\mathcal{D} + \epsilon\,\mathrm{div}(\mathcal{D}\mathbf{u})\,\mathrm{d}t - 2\epsilon J(u,v)\,\mathrm{d}t + \mathrm{div}(\mathrm{div}(\epsilon\mathbf{u}+\mathbf{R})\Gamma) = -\,\mathrm{div}\left(f\hat{\mathbf{z}}\times\mathbf{u}+\nabla h\right)\mathrm{d}t.$$

The Coriolis vector field R can be taken to be divergence free, as it is the vector potential of the divergence free rotation about the vertical direction. The fundamental theorem of vector calculus states that any vector field can be decomposed into a solenoidal and an irrotational part. This means that

$$\begin{aligned}\mathbf{R} &= \mathbf{R}_{\text{solenoidal}} + \mathbf{R}_{\text{irrotational}}\\ ,\mathrm{curl}\,\mathbf{R} &= \mathrm{curl}\,\mathbf{R}_{\text{solenoidal}} = f\hat{\mathbf{z}}\end{aligned}$$

Hence it is sensible to choose $\mathbf{R} = \mathbf{R}_{\text{solenoidal}}$, as the irrotational part does not contribute to the curl and therefore is divergence-free. This observation allows us to further simplify the divergence equation to finally obtain

$$\epsilon\mathrm{d}\mathcal{D} + \epsilon\,\mathrm{div}(\mathcal{D}\,\mathrm{d}\mathbf{x}_t) = -\,\mathrm{div}\left(f\hat{\mathbf{z}}\times\mathbf{u}+\nabla h\right)\mathrm{d}t + 2\epsilon J(u,v)\,\mathrm{d}t. \tag{46}$$

We now introduce the imbalance $\Omega := -\,\mathrm{div}(f\hat{\mathbf{z}}\times\mathbf{u}+\nabla h)$, which describes divergence of the geostrophic balance condition. It is assumed that the velocity field has an ϵ-weighted Helmholtz decomposition $\mathbf{u} = \hat{\mathbf{z}}\times\nabla\psi + \epsilon\nabla\chi$. It follows that

$$\hat{\mathbf{z}}\times\mathbf{u} = -\nabla\psi + \epsilon\hat{\mathbf{z}}\times\nabla\chi$$

which allows us to rewrite the imbalance as

$$\Omega = -\,\mathrm{div}(f\hat{\mathbf{z}}\times\mathbf{u}+\nabla h) = \mathrm{div}(f\nabla\psi) + \epsilon J(f,\chi) - \Delta h$$

We will now compute the stochastic evolution of Ω. Together with the equations for potential vorticity, depth and divergence, the imbalance equation forms a model for rotating shallow water in insightful variables. It will turn out that the pair $\mathcal{D},\Omega$ form canonically conjugate variables in a Hamiltonian. In correspondence with the derivation of quasi-geostrophy (QG), the following assumptions are made:

- $f(\mathbf{x}) = 1 + \epsilon f_1(\mathbf{x})$
- $b(\mathbf{x}) = 1 + \epsilon b_1(\mathbf{x})$

The stochastic evolution of Ω is

$$\begin{aligned}\mathrm{d}\Omega &= \mathrm{d}\big(\mathrm{div}(f\nabla\psi) + J(f,\chi) - \Delta h\big)\\ &= \underbrace{\epsilon\,\mathrm{div}\left(f_1\nabla\mathrm{d}\psi\right) + \epsilon^2 J(f_1,\mathrm{d}\chi)} + \underbrace{\Delta\mathrm{d}\psi} + \underbrace{-\Delta\mathrm{d}h}\end{aligned}$$

This equation shall be split up in the three parts, as shown above, and these terms will be evaluated individually to yield the final equation. The derivation of the following equations is shown in the appendix.

- The equation for $-\Delta\mathrm{d}h$ is given by

$$-\,\Delta\mathrm{d}h = \frac{1}{\epsilon\mathcal{F}}\Delta\,\mathrm{div}(\mathrm{d}\mathbf{x}_t) + \frac{1}{\mathcal{F}}\Delta\,\mathrm{div}\left((b_1+\mathcal{F}h)\,\mathrm{d}\mathbf{x}_t\right) \tag{47}$$

- The equation for $\epsilon\,\mathrm{div}(f_1\nabla\mathrm{d}\psi) + \epsilon^2 J(f_1,\mathrm{d}\chi)$ is given by

$$\epsilon\,\mathrm{div}(f_1\nabla\mathrm{d}\psi) + \epsilon^2 J(f_1,\mathrm{d}\chi) = \epsilon f_1\Delta\mathrm{d}\psi - \varpi\Delta f_1\cdot\mathrm{d}\mathbf{x}_t - J\left(f_1,(h+\frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t + \xi_i\circ\mathrm{d}W_t^i\cdot(\epsilon\mathbf{u}+\mathbf{R})\right) \tag{48}$$

- The equation for $\Delta\mathrm{d}\psi$ is given by

$$\Delta\mathrm{d}\psi = -\,\mathrm{div}\left((\Delta\psi+f_1)\,\mathrm{d}\mathbf{x}_t\right) - \frac{\mathrm{div}\,\mathrm{d}\mathbf{x}_t}{\epsilon} \tag{49}$$

So $\mathsf{d}\Omega$ is given by

$$\mathsf{d}\Omega + \frac{1}{\epsilon\mathcal{F}}(\Delta - \mathcal{F})\,\mathrm{div}\,\mathrm{d}\mathbf{x}_t = \frac{1}{\mathcal{F}}\Delta\,\mathrm{div}\,\big((b_1 + \mathcal{F}h)\,\mathrm{d}\mathbf{x}_t\big) - \varpi\Delta f_1\cdot\mathrm{d}\mathbf{x}_t - J\left(f_1, (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t + \xi_i \circ \mathrm{d}W_t^i\cdot(\epsilon\mathbf{u} + \mathbf{R})\right)$$
$$- f\,\mathrm{div}\,\big((\Delta\psi + f_1)\,\mathrm{d}\mathbf{x}_t\big) - f_1\,\mathrm{div}\,\mathrm{d}\mathbf{x}_t$$

The SRSW equations formulated in potential vorticity, depth, divergence and imbalance are thus

$$\begin{aligned}
\mathsf{d}q + \mathrm{d}\mathbf{x}_t\cdot\nabla q &= 0,\\
\mathsf{d}\eta + \mathrm{d}\mathbf{x}_t\cdot\nabla\eta + \eta\,\mathrm{div}\,\mathrm{d}\mathbf{x}_t &= 0,\\
\mathsf{d}\mathcal{D} - \frac{1}{\epsilon}\Omega\,\mathrm{d}t &= -\,\mathrm{div}(\mathcal{D}\,\mathrm{d}\mathbf{x}_t) + 2J(u,v)\,\mathrm{d}t,\\
\mathsf{d}\Omega + \frac{1}{\epsilon\mathcal{F}}(\Delta - \mathcal{F})\,\mathrm{div}\,\mathrm{d}\mathbf{x}_t &= \frac{1}{\mathcal{F}}\Delta\,\mathrm{div}\,\big((b_1 + \mathcal{F}h)\,\mathrm{d}\mathbf{x}_t\big) - \varpi\Delta f_1\cdot\mathrm{d}\mathbf{x}_t\\
&\quad - J\left(f_1, (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t + \xi_i\circ\mathrm{d}W_t^i\cdot(\epsilon\mathbf{u} + \mathbf{R})\right)\\
&\quad - f\,\mathrm{div}\,\big((\Delta\psi + f_1)\,\mathrm{d}\mathbf{x}_t\big) - f_1\,\mathrm{div}\,\mathrm{d}\mathbf{x}_t
\end{aligned} \tag{50}$$

Rescaling time in the equations for $\mathcal{D}$ and Ω to $t \mapsto t/\epsilon$ and expanding at leading order reveals their conjugacy in terms of their role in the following Hamiltonian

$$\mathcal{H} = \frac{1}{2}\int\left(\Omega^2\,\mathrm{d}t + \mathcal{D}^2\,\mathrm{d}t + \frac{1}{\mathcal{F}}|\nabla\mathcal{D}|^2\,\mathrm{d}t + \frac{\mathcal{D}}{\mathcal{F}}(\Delta - \mathcal{F})\,\mathrm{div}\,\xi_i\circ\mathrm{d}W_t^i\right)\mathrm{d}x\,\mathrm{d}y$$

7 Conclusion

We have introduced the concept of transport noise, along with the differential geometry that is necessary to work with the equations that follow, which is a novel way to do uncertainty quantification, for instance in geophysical fluid dynamics, and applied it to the Rayleigh-Bénard convection problem. On the level of partial differential equations, transport noise preserves many conserved quantities and is therefore appealing for uncertainty quantification in geophysical fluid dynamics, which relies heavily on these conserved quantities, examples being potential vorticity, enstrophy and the Kelvin circulation theorem. The example of the Kelvin theorem for the Euler equations shows this. It is then shown that the Euler equations with transport noise still satisfy a Kelvin theorem.

We can then add viscosity and a heat equation, which results in a stochastic version of the Rayleigh-Bénard convection problem, modeled by a number of stochastic partial differential equations. Using a specific truncated Fourier series expansion, the Lorenz system with transport noise was obtained. This low-dimensional system of stochastic differential equations can be compared with alternative stochastic Lorenz systems. In particular we studied the fluctuation-dissipation noise (linear multiplicative noise in each variable) Lorenz system, as this type of noise was introduced also with the purpose of doing uncertainty quantification and stochastic parametrization. With methods from random dynamical systems theory, we were able to show that the two types of systems have different qualitative properties, in that the system with linear multiplicative noise in each variable changes the average rate of contraction or expansion of phase space volume, whereas the transport noise conserves this rate with respect to the deterministic system. This implies that the type of noise introduced to low-dimensional dynamical systems can affect properties of the underlying deterministic system and one should consider the effects of stochasticity on a qualitative level. In particular, when a system of equations is Hamiltonian, introducing arbitrary noise may destroy the Hamiltonian structure completely by altering the average rate of phase-space volume contraction, whereas transport noise conserves this.

For the numerical verification of analytical results for the Lorenz system, we have introduced the stochastic generalization of the so-called Cayley method for the numerical computation of Lyapunov exponents. This method is a QR-based algorithm in which the orthogonal matrix is determined via the Cayley transform. This method turns out the be as robust and stable as it is in the deterministic case. Improvements to the numerical calculations can be made by using more advanced numerical schemes to solve the various stochastic differential equations in the method. The numerical results are in agreement with the analytical statements and it is shown that the method calculates the deterministic values of the Lyapunov exponents with reasonable accuracy.

Using the stochastic variational principle, the stochastic rotating shallow water equations are derived. The variational principle is used to insert stochastic transport noise into the rotating shallow water dynamics. It has been shown that these equations conserve potential vorticity, an infinite amount of integral quantities and that the Kelvin theorem holds. Energy is not conserved. It was then shown that these equations possess different timescales and by choosing alternative variables (potential vorticity, depth, the divergence of the velocity field and the imbalance) it is possible to split up the dynamics in fast and slow motion without approximation.

Appendix I: Wiener Process/ Brownian Motion

Definition 7.1 *A stochastic process $\{W_t, t \geq 0\}$ is called a* ***standard Wiener process*** *or* ***standard Brownian Motion*** *if it has the following properties:*

a) $W_0 = 0$.

b) $(W_{t_2} - W_{t_1}) \sim \mathcal{N}(0, t_2 - t_1)$.

c) $(W_{t_2} - W_{t_1})$ *and* $(W_{t_3} - W_{t_2})$ *are independent for all* $t_1 < t_2 < t_3$.

d) For almost all ω*, the sample function* $t \mapsto W_t(\omega)$ *is continuous.*

If a process W_t does not have property a), but does have the others, then we call it a ***Brownian Motion*** *(or* ***Wiener Process****).*

Theorem 7.2 *Brownian motion exists in the L_2 sense.*

Although existence has been proven for much more general (filtered) function spaces, this construction is intuitive and clear. A generalization of this theorem is by Wiener (1923). This theorem will be proved by constructing a stochastic process that has the listed properties. The construction will use concepts from functional analysis. Brownian motion is named after its discoverer Robert Brown, who described the erratic motion of a particle trapped in cavities inside pollen grains in water. The mathematical description of this physical process is called a Wiener process. Hence in mathematics, there exists the preference towards calling it Wiener process instead of Brownian motion.

Proof. Let $\{\phi_i\}$ be an arbitrary complete orthonormal basis of $L_2([0,t])$ and let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables that are defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with each $X_i \sim \mathcal{N}(0,1)$. Here Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on Ω. For $n = 1, 2, \ldots$ define

$$W_t^n = \sum_{i=1}^{n} X_i \int_0^t \phi_i(s)\,\mathrm{d}s.$$

We now want to prove that the limit $(n \to \infty)$ of this process satisfies all of the properties. To do so, we must first prove that W_t^n is a Cauchy sequence. Hence, the following lemma:

Lemma 7.3 *Suppose that for each t, W_t^n is a Cauchy sequence in $L_2(\Omega, \mathcal{F}, \mathbb{P})$ whose limit W_t is a normal random variable with mean zero and variance t. For any two times t, s the expectation $\mathbb{E}[W_t W_s] = \min(t,s)$.*

Proof. Define the function $\mathbb{I}_t \in L_2([0,t])$ such that

$$\mathbb{I}_t(s) = \begin{cases} 1 & \text{if } s < t, \\ 0 & \text{if } s \geq t. \end{cases}$$

Clearly

$$\langle \mathbb{I}_t, \phi_i \rangle = \int_0^t \mathbb{I}_t(s)\phi_i(s)\,\mathrm{d}s = \int_0^t \phi_i(s)\,\mathrm{d}s.$$

Since ϕ_i is a basis function, we may express $\mathbb{I}_t$ in terms of basis functions in the following way

$$\mathbb{I}_t = \sum_{i=1}^{\infty} \langle \mathbb{I}_t, \phi_i \rangle \phi_i.$$

Using this, we find

$$\|\mathbb{I}_t\|^2_{L_2([0,t])} = \langle \mathbb{I}_t, \mathbb{I}_t \rangle_{L_2([0,t])} = \int_0^t \mathbb{I}_t^2(s)\,\mathrm{d}s = t,$$

$$\|\mathbb{I}_t\|^2_{L_2([0,t])} = \left\langle \sum_{i=1}^{\infty} \langle \mathbb{I}_t, \phi_i \rangle \phi_i, \sum_{i=1}^{\infty} \langle \mathbb{I}_t, \phi_i \rangle \phi_i \right\rangle_{L_2([0,t])} = \sum_{i=1}^{\infty} \langle \mathbb{I}_t, \phi_i \rangle^2.$$

Thus the identity $\sum_{i=1}^{\infty}\langle \mathbb{I}_t, \phi_i\rangle^2 = t$ is valid. Let $n > m$,

$$\begin{aligned}\mathbb{E}(W_t^n - W_t^m)^2 &= \mathbb{E}\left(\sum_{i=1}^{n} X_i \int_0^t \phi_i(s)\,\mathrm{d}s - \sum_{i=1}^{m} X_i \int_0^t \phi_i(s)\,\mathrm{d}s\right)^2 \\ &= \mathbb{E}\left(\sum_{i=m+1}^{n} X_i \int_0^t \phi_i(s)\,\mathrm{d}s\right)^2\end{aligned}$$

The integral inside the expectation leaves just a number, so we are allowed to pull it out of the expectation. This yields

$$\begin{aligned}\mathbb{E}(W_t^n - W_t^m)^2 = \mathbb{E}(X_i)^2\left(\sum_{i=m+1}^{n} \int_0^t \phi_i(s)\,\mathrm{d}s\right)^2 &= \left(\sum_{i=m+1}^{n} \langle \mathbb{I}_t, \phi_i\rangle\right)^2 \leq (n-m)^2 \sum_{i=m+1}^{n} \langle \mathbb{I}_t, \phi_i\rangle^2 \\ &= (n-m)^2\left(\sum_{i=1}^{n}\langle \mathbb{I}_t, \phi_i\rangle^2 - \sum_{i=1}^{m}\langle \mathbb{I}_t, \phi_i\rangle^2\right)\end{aligned}$$

where we used the Cauchy-Schwarz inequality. Taking the limit $n, m \to \infty$ and using the identity derived earlier, $\mathbb{E}(W_t^n - W_t^m)^2 \to 0$. So $\{W_t^n\}$ is a Cauchy sequence in $L_2(\Omega, \mathcal{F}, \mathbb{P})$. The mean of W_t is

$$\begin{aligned}\mathbb{E}(W_t) = \mathbb{E}(\lim_{n\to\infty} W_t^n) = \mathbb{E}\left(\lim_{n\to\infty}\sum_{i=1}^{n} X_i \int_0^t \phi_i(s)\,\mathrm{d}s\right) &= \lim_{n\to\infty}\mathbb{E}\left(\sum_{i=1}^{n} X_i \int_0^t \phi_i(s)\,\mathrm{d}s\right) \\ = \lim_{n\to\infty}\left(\sum_{i=1}^{n}\mathbb{E}(X_i)\int_0^t \phi_i(s)\,\mathrm{d}s\right) = 0.\end{aligned}$$

The variance of W_t is

$$\begin{aligned}\mathrm{var}(W_t) = \mathbb{E}(W_t^2) - (\mathbb{E}(W_t))^2 &= \mathbb{E}\left(\lim_{n\to\infty}\left(\sum_{i=1}^{n} X_i \int_0^t \phi_i(s)\,\mathrm{d}s\right)^2\right) \\ &= \mathbb{E}\left(\lim_{n\to\infty}\left(\sum_{i=1}^{n} X_i^2\left(\int_0^t \phi_i(s)\,\mathrm{d}s\right)^2 + 2\sum_{j=1}^{n}\sum_{k=1}^{j-1} X_j X_k \int_0^t \phi_j(s)\phi_k(s)\,\mathrm{d}s\right)\right) \\ &= \lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}(X_i^2)\left(\int_0^t \phi_i(s)\,\mathrm{d}s\right)^2 + 2\lim_{n\to\infty}\sum_{j=1}^{n}\sum_{k=1}^{j-1}\mathbb{E}X_j\mathbb{E}X_k\int_0^t \phi_j(s)\phi_k(s)\,\mathrm{d}s \\ &= \sum_{i=1}^{\infty}\left(\int_0^t \phi_i(s)\,\mathrm{d}s\right)^2 = t.\end{aligned}$$

Here we used the fact that the X_i's are independent, as that allows us to write the expectation of the product as the product of the expectations. The expectation $\mathbb{E}[W_t W_s]$ is

$$\begin{aligned}\mathbb{E}[W_t W_s] &= \mathbb{E}\left(\left(\lim_{n\to\infty}\sum_{i=1}^{n} X_i \int_0^t \phi_i(\tau)\,\mathrm{d}\tau\right)\left(\lim_{n\to\infty}\sum_{i=1}^{n} X_i \int_0^s \phi_i(\tau)\,\mathrm{d}\tau\right)\right) \\ &= \mathbb{E}\left(\left(\lim_{n\to\infty}\sum_{i=1}^{n} X_i \int_0^t \mathbb{I}_t(\tau)\phi_i(\tau)\,\mathrm{d}\tau\right)\left(\lim_{n\to\infty}\sum_{i=1}^{n} X_i \int_0^s \mathbb{I}_s(\tau)\phi_i(\tau)\,\mathrm{d}\tau\right)\right) \\ &= \mathbb{E}\left(\left(\lim_{n\to\infty}\sum_{i=1}^{n} X_i\langle \mathbb{I}_t, \phi_i\rangle\right)\left(\lim_{n\to\infty}\sum_{i=1}^{n} X_i\langle \mathbb{I}_s, \phi_i\rangle\right)\right)\end{aligned}$$

Taking the expectation inside the sums, we make an important remark. Since the X_is are independent and identically distributed with mean zero and variance one, the product of the sums reduces to a single sum only over i, as products $\mathbb{E}[X_i X_j] = \delta_{ij}$. Hence this reduces to

$$\mathbb{E}[W_t W_s] = \sum_{i=1}^{\infty}\langle \mathbb{I}_t, \phi_i\rangle\langle \mathbb{I}_s, \phi_i\rangle = \langle \mathbb{I}_t, \mathbb{I}_s\rangle = \min(t, s).$$

■

These are all of the identities that we will need to show that we have constructed a process that satisfies properties $a), b)$ and $c)$. We now continue with the proof of the main theorem. Property $a)$ is immediate, since

$$W_0 = \lim_{n\to\infty} W_0^n = \lim_{n\to\infty} \sum_{i=1}^n X_i \int_0^0 \phi_i(s)\,\mathrm{d}s = 0.$$

To prove property $b)$, note that W_t^n is normally distributed (Gaussian) because it is a finite sum of normally distributed random variables. The variance of W_t^n is given by

$$\mathrm{var}(W_t^n) = \mathrm{var} \sum_{i=1}^n X_i \int_0^t \phi_i(s)\,\mathrm{d}s = \sum_{i=1}^n \mathrm{var}(X_i) \int_0^t \phi_i(s)\,\mathrm{d}s = \sum_{i=1}^n \int_0^t \phi_i(s)\,\mathrm{d}s = \sum_{i=1}^n \langle \mathbb{I}_t, \phi_i\rangle =: \sigma_n^2$$

The characteristic function for W_t^n is given by

$$\chi_n(x) = \mathbb{E}[e^{ixW_t^n}] = e^{-1/2\sigma_n^2 x^2}$$

which, in the limit $n \to \infty$, converges to $\chi(x) = e^{-1/2tx^2}$. Because $\{W_t^n\}$ is a Cauchy sequence in $L_2(\Omega, \mathcal{F}, \mathbb{P})$ with limit W_t, this guarantees that there exists a subsequence $W_t^{n_k} \to W_t$ as $k \to \infty$ in the same space. The dominated convergence theorem then implies that $\mathbb{E}[e^{ixW_t^{n_k}}] \to \mathbb{E}[e^{ixW_t}]$. So the characteristic function of W_t is $\chi(x)$, the one uniquely defined for $\mathcal{N}(0,t)$. This proves $b)$. For any $s, t > 0$, $\mathbb{E}[W_t W_s] = \min(t,s)$ and any $W_t^n \sim \mathcal{N}(0, \sigma_n^2)$, so its increments are independent. Finally, to show that W_t is almost surely continuous, we turn to so-called Haar basis functions, which form a wavelet basis, defined as

$$\psi(t) := \begin{cases} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} < t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and $\psi_{i,j}(t) := \psi(2^i t - j)$. Without proof we state that these functions form a complete orthonormal basis in L_2. This specific basis in necessary to have uniform convergence of stochastic processes. To prove property $d)$, we use Lévy's construction, as described in [McK69], which is based on a well known lemma from real analysis.

Lemma 7.4 *Suppose that, for $n = 1, 2, \ldots, f_n : [0,1] \to \mathbb{R}$ is a continuous function, and that for all $\epsilon > 0$ there exists a number N such that $n \ge N$ implies $|f_n(t) - f(t)| < \epsilon$ for any $t \in [0,1]$. Then the limit f is a continuous function.*

To show that W_t has almost surely continuous paths, it is necessary to show that W_t^n converges almost surely uniformly in $(\Omega, \mathcal{F}, \mathbb{P})$. Since this is not true for an arbitrary basis in L_2, we have to redefine W_t^n. Let X_0 and $\{X_{i,j}\}$ where $i = 1, 2, \ldots$ and $j = 1, 2, \ldots, 2^{i-1}$ be independent identically distributed random variables with distribution $\mathcal{N}(0,1)$ defined on the probability space. Let $W_t^I = X_0 \int_0^t \psi(s)\,\mathrm{d}s + \sum_{i=1}^I Y_i$, where we define Y_i

$$Y_i(t,\omega) = \sum_{j=1}^{2^{i-1}} X_{i,j}(\omega) \int_0^t \psi_{i,j}(s)\,\mathrm{d}s.$$

To prove that W_t^I converges almost surely uniformly, we will show that the maximum process $M_i(t,\omega) := \max_{t\in[0,T]} Y_i(t,\omega)$ converges almost surely uniformly. The indefinite integral of the Haar basis gives rise to nonzero triangular shaped functions, with maximum $2^{-(i+1)/2}$. Hence

$$M_i = 2^{-(i+1)/2} \max_{0\le j<2^{i-1}} |X_{i,j}|,$$

is the upper bound process and 0 is the lower bound. By using the Borel-Cantelli lemma, which states if the sum of probabilities of $X_{i,j}$ is finite, then the probability that infinitely many of them occur is equal to zero. Therefore, for some constant c_i

$$\begin{aligned} \mathbb{P}[M_i > 2^{-(i+1)/2} c_i] &= \mathbb{P}\left[\max_{0\le j<2^{i-1}} |X_{i,j}| > c_i\right] \\ &= \mathbb{P} \bigcup_j [|X_{i,j}| > c_i] \\ &\le \sum_j \mathbb{P}[|X_{i,j}| > c_i] \end{aligned}$$

where the last inequality came from the definition of the probability measure. Let $x > c$, by definition of the $X_{i,j}$'s

$$\mathbb{P}[X_{i,j} > c] = \frac{1}{\sqrt{2\pi}} \int_c^\infty e^{-x^2/2}\,\mathrm{d}x < \frac{1}{\sqrt{2\pi}} \int_c^\infty \frac{x}{c} e^{-x^2/2}\,\mathrm{d}x = \frac{1}{c\sqrt{2\pi}} e^{-c^2/2}.$$

Now, using the symmetry of the normal distribution, we see that $\mathbb{P}[|X_{i,j}| > c] = 2\mathbb{P}[X_{i,j} > c]$. Thus

$$\mathbb{P}[M_i > 2^{-(i+1)/2} c_i] < \sum_j \frac{2}{c_i\sqrt{2\pi}} e^{-c_i^2/2} = \frac{2^i}{c_i\sqrt{2\pi}} e^{-c_i^2/2}.$$

Now choose $c_i = \vartheta\sqrt{2i\log 2}$ for some $\vartheta > 1$. Inserting this into the last inequality gives

$$\mathbb{P}[M_i > a_i] < \frac{1}{\vartheta\sqrt{4\pi\log 2}} \frac{2^{(1-\vartheta^2)i}}{\sqrt{i}} e^{-\vartheta^2},$$

where $a_i := 2^{-(i+1)/2}\vartheta\sqrt{2i\log 2} = \vartheta\sqrt{i2^{-i}\log 2}$. The sum $\sum_{i=1}^\infty a_i$ is convergent, as well as the sum of the right hand side of the inequality. Now using the Borel-Cantelli lemma, we have

$$\mathbb{P}[M_i > a_i \text{ infinitely often}] = 0.$$

Hence, for almost all ω, there exists a number $N(\omega)$ such that $M_i < a_i$ when $n \geq N$. This shows that M_i is an almost surely convergent sequence, which implies that W_t^I is an almost surely convergent series. By construction W_t^I is continuous, so now using the lemma from real analysis stated earlier, we have that W_t has almost surely continuous paths. ■

Appendix II: Numerical Methods

Stochastic differential equations are written in integral form, because the time derivative of a Wiener process is not defined, but the integral over the Wiener measure is. To analyze the convergence of numerical methods that arise from integral approximation, a category that includes the most used stochastic numerical methods, it is necessary to derive the Stratonovich (or Itô) Taylor expansion. For detailed exposition of stochastic numerical methods see [KPS12].

Stochastic Taylor Expansion

The most common flavors of stochasticity are Itô and Stratonovich noise. They are not equivalent, but differ by a correction term. This correction term is there because an Itô process does not satisfy the ordinary chain rule, but is a semimartingale, whereas Stratonovich processes satisfy the ordinary chain rule, but is not a semimartingale. Hence, to do statistics with Stratonovich processes, they have to transformed into Itô form. A Stratonovich stochastic differential equation is given by

$$\mathsf{d}X = \mu(X)\,\mathrm{d}t + \sigma(X)\circ\mathrm{d}W_t, \qquad X(t_0) = x_0. \tag{51}$$

Now let $g \in \mathcal{C}^\infty(\mathbb{R})$ be a function of X, then its stochastic evolution is given by the ordinary chain rule

$$\mathsf{d}g = \partial_X g(X)\mathsf{d}X. \tag{52}$$

If instead one takes an Itô stochastic differential equation,

$$\mathsf{d}X = \mu(X)\,\mathrm{d}t + \sigma(X)\,\mathrm{d}W_t, \qquad X(t_0) = x_0. \tag{53}$$

Let $g \in \mathcal{C}^\infty(\mathbb{R})$ be a function of X, then its stochastic evolution is given by Itô's lemma

$$\mathsf{d}g(X) = \partial_X g(X)\mathsf{d}X + \frac{1}{2}\partial_{XX} g(X)\,\mathrm{d}[X] \tag{54}$$

where $\mathrm{d}[X]$ is the quadratic variation of X. Since X is generated by an SDE with a Wiener process, $\mathrm{d}[X] = \sigma^2\,\mathrm{d}t$. Expanding $\mathsf{d}X$ for the Stratonovich case results in

$$\begin{aligned}\mathsf{d}g(X) &= \partial_X g(X)\big[\mu(X)\,\mathrm{d}t + \sigma(X)\circ\mathrm{d}W_t\big] \\ &= \mu(X)\partial_X g(X)\,\mathrm{d}t + \sigma(X)\partial_X g(X)\circ\mathrm{d}W_t\end{aligned}$$

It is convenient to define the differential operators $A := \mu(X)\partial_X$ and $B := \sigma(X)\partial_X$ so that

$$\mathrm{d}g(X) = Ag(X)\,\mathrm{d}t + Bg(X)\circ \mathrm{d}W_t \tag{55}$$

The Itô case is similar, but due to the extra term, the operator A for the Itô case is given by $A := \mu(X)\partial_X + \frac{1}{2}\sigma^2(X)\partial_{XX}$. The stochastic differential equation in $g(X)$ is then given by

$$\mathrm{d}g(X) = Ag(X)\,\mathrm{d}t + Bg(X)\,\mathrm{d}W_t. \tag{56}$$

We shall use equations (55) and (56) interchangeably, hence the choice for keeping the notation so very similar. The difference can always be noticed in the $\circ$ symbol that shall consistently be used to denote a Stratonovich integral. From this point onward, we shall use the Stratonovich notation, but it is straightforward to go to the Itô form, by changing the differential operator A and the integral type. The integral form of the SDE given by (55) is

$$g(X) = g(x_0) + \int_{t_0}^{t} Ag(X(s_1))\,\mathrm{d}s_1 + \int_{t_0}^{t} Bg(X(s_1))\circ \mathrm{d}W_{s_1}.$$

Here s_1 is a dummy variable and it is numbered because we will approximate the integrands continuously and this keeps track of which integral is which. The approximation of the integrands will be an iterative process. Let $g(X) = X$, then we have

$$X = x_0 + \int_{t_0}^{t} \mu(X(s_1))\,\mathrm{d}s_1 + \int_{t_0}^{t} \sigma(X(s_1))\,\mathrm{d}W_{s_1}.$$

Next we approximate the integrands, by setting $g(X) = \mu(X)$ for the first integral and $g(X) = \sigma(X)$ for the second. The integral equations for the integrands are given by

$$\begin{aligned}\mu(X(s_1)) &= \mu(x_0) + \int_{t_0}^{s_1} A\mu(X(s_2))\,\mathrm{d}s_2 + \int_{t_0}^{s_1} B\mu(X(s_2))\circ \mathrm{d}W_{s_2},\\ \sigma(X(s_1)) &= \sigma(x_0) + \int_{t_0}^{s_1} A\sigma(X(s_2))\,\mathrm{d}s_2 + \int_{t_0}^{s_1} B\sigma(X(s_2))\circ \mathrm{d}W_{s_2}\end{aligned}$$

Substitution then leads to

$$\begin{aligned}X = x_0 &+ \int_{t_0}^{t}\left[\mu(x_0) + \int_{t_0}^{s_1} A\mu(X(s_2))\,\mathrm{d}s_2 + \int_{t_0}^{s_1} B\mu(X(s_2))\circ \mathrm{d}W_{s_2}\right]\mathrm{d}s_1\\ &+ \int_{t_0}^{t}\left[\sigma(x_0) + \int_{t_0}^{s_1} A\sigma(X(s_2))\,\mathrm{d}s_2 + \int_{t_0}^{s_1} B\sigma(X(s_2))\circ \mathrm{d}W_{s_2}\right]\circ \mathrm{d}W_{s_1}.\end{aligned}$$

Upon taking the constant terms $\mu(x_0)$ and $\sigma(x_0)$ through the integral sign and defining the remainder $\mathcal{R}_1$ as

$$\begin{aligned}\mathcal{R}_1 = &\int_{t_0}^{t}\int_{t_0}^{s_1} A\mu(X(s_2))\,\mathrm{d}s_2\,\mathrm{d}s_1 + \int_{t_0}^{t}\int_{t_0}^{s_1} B\mu(X(s_2))\circ \mathrm{d}W_{s_2}\,\mathrm{d}s_1\\ &+ \int_{t_0}^{t}\int_{t_0}^{s_1} A\sigma(X(s_2))\,\mathrm{d}s_2\circ \mathrm{d}W_{s_1} + \int_{t_0}^{t}\int_{t_0}^{s_1} B\sigma(X(s_2))\circ \mathrm{d}W_{s_2}\circ \mathrm{d}W_{s_1}\end{aligned}$$

we can write the current iteration as

$$X = x_0 + \mu(x_0)(t - t_0) + \sigma(x_0)(W_t - W_{t_0}) + \mathcal{R}_1.$$

Continuing by approximating the integrands of the remainder $\mathcal{R}_1$ indefinitely, we obtain the Stratonovich-Taylor expansion. By changing the operator A and the integral type, which is allowed if one lets the process $g(X)$ be generated by an Itô SDE, the Itô-Taylor expansion can be obtained in the same way. The order of the approximation is determined by the lowest order term. As a rule of thumb, an integral with respect to time is order one and an integral with respect to a Wiener measure is order half. Multiple integrals are additive in terms of order. Hence the worst term in terms of order is in the remainder $\mathcal{R}_1$ is the last integral, which is integrated over the Wiener measure twice and has order one. If we stop approximating the solution X at this point, we obtain the first step of the Euler-Maruyama method. If we iterate on the integrand of the worst term of the remainder, we obtain the first step of the Milstein method and iterating even further brings one to the so-called strong Taylor methods. There is a trade-off when going to higher orders of convergence in that the number of integrals that have to be approximated increases very quickly. We will derive the Milstein method now, for an Itô process. By approximating the worst term in the remainder we find

$$B\sigma(X(s_2)) = B\sigma(x_0) + \int_{t_0}^{s_2} AB\sigma(X(s_3))\,\mathrm{d}s_3 + \int_{t_0}^{s_2} B^2\sigma(X(s_3))dW_{s_3}.$$

The constant term can be pulled out of the integral as before and we put the remain integrals in the previous expression in the remainder $\mathcal{R}_1$, giving the new remainder $\mathcal{R}_2$. So we have

$$\int_{t_0}^{t}\int_{t_0}^{s_1} B\sigma(X(s_2))\,\mathrm{d}W_{s_2}\,\mathrm{d}W_{s_1} = B\sigma(x_0)\int_{t_0}^{t}\int_{t_0}^{s_1}\mathrm{d}W_{s_2}\,\mathrm{d}W_{s_1} + \dots$$

where the dots represent the integrals in the approximation of $B\sigma(X(s_2))$. The double integral in the above expression is special, in the sense that it is an integral over a Wiener process with respect to the Wiener measure, for which we do not have an intuitive way of integration. We have to derive the answer. We start off by evaluating the inner integral

$$\begin{aligned}\int_{t_0}^{t}\int_{t_0}^{s_1}\mathrm{d}W_{s_2}\,\mathrm{d}W_{s_1} &= \int_{t_0}^{t} W_{s_1} - W_{t_0}\,\mathrm{d}W_{s_1}\\ &= \int_{t_0}^{t} W_{s_1}\,\mathrm{d}W_{s_1} - \int_{t_0}^{t} W_{t_0}\,\mathrm{d}W_{s_1}\end{aligned}$$

To continue, we do an educated guess. We use the Itô lemma to derive an expression for $\mathrm{d}W^2$.

$$\mathrm{d}W^2 = 2W\,\mathrm{d}W + \mathrm{d}t.$$

Rewriting the unknown integral in terms of $\mathrm{d}W^2$ yields

$$\begin{aligned}\int_{t_0}^{t}\int_{t_0}^{s_1}\mathrm{d}W_{s_2}\,\mathrm{d}W_{s_1} &= \frac{1}{2}\int_{t_0}^{t}[\mathrm{d}W^2 - \mathrm{d}t] - W_{t_0}(W_t - W_{t_0})\\ &= \frac{1}{2}(W_t^2 - W_{t_0}^2 - t + t_0) - W_{t_0}W_t + W_{t_0}^2\\ &= \frac{1}{2}(W_t - W_{t_0})^2 - \frac{1}{2}(t - t_0)\end{aligned}$$

Using this fact, we are now able to write a higher order formulation

$$X = x_0 + \mu(x_0)(t - t_0) + \sigma(x_0)(W_t - W_{t_0}) + \frac{1}{2}\sigma(x_0)\sigma'(x_0)\Big((W_t - W_{t_0})^2 - (t - t_0)\Big) + \mathcal{R}.$$

This is the Milstein method. Continuing by approximating terms in the remainder yields the strong Taylor methods, but one has to ask himself whether it is worth it to use numerically, since the number of operations at each timestep increases rapidly. If one is willing to pay the computational prize, using higher order methods will allow one to compute the individual Lyapunov exponents for the Lorenz system with greater accuracy by using higher order methods, since that is where the current crux lies.

Order of Convergence

We shall now give rigorous proofs for the order of convergence of the Euler-Maruyama method. The convergence of the higher order methods can be computed in the same way. Given the Itô SDE

$$\mathrm{d}X = \mu(X)\,\mathrm{d}t + \sigma(X)\,\mathrm{d}W_t, \qquad X(t_0) = x_0$$

The corresponding integral form is given by

$$X_t = x_0 + \int_{t_0}^{t}\mu(X_s)\,\mathrm{d}s + \int_{t_0}^{t}\sigma(X_s)\,\mathrm{d}W_s.$$

The simplest approximation to the integral form is the Euler-Maruyama method, which is acquired by partitioning the time interval $[t_0, t]$ into equally sized subintervals of length Δt and applying the Itô-Taylor expansion. This yields

$$X_{n+1} = X_n + \mu(X_n)\Delta t + \sigma(X_n)\Delta W_n,$$

where $\Delta t = (t_{n+1} - t_n)$ and $\Delta W_n = (W_{n+1} - W_n)$. By definition of the Wiener process, $\Delta W_n \sim \mathcal{N}(0, \Delta t)$. To analyze the order of convergence, we consider geometric Brownian motion, since we can analytically solve the SDE that generates this process, namely the SDE

$$dY = \mu Y \,\mathrm{d}t + \sigma Y \,\mathrm{d}W_t, \qquad Y(t_0) = y_0$$

has analytical solution

$$Y_t = y_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t}.$$

The Euler-Maruyama method for this SDE is

$$Y_{n+1} = Y_n + \mu Y_n \Delta t + \sigma Y_n \Delta W_n = (1 + \mu\Delta t + \sigma \Delta W_n) Y_n.$$

From which it is quite straightforward to see that the numerical solution at time $T = n\Delta t$ is given by

$$Y_n = y_0 \prod_{i=0}^{n-1} (1 + \mu\Delta t + \sigma\Delta W_i)$$

In stochastic numerical methods, there are two ways of evaluating the error. One may take the mean of the difference of the numerical and the analytical solution, which is called strong convergence, or one could take the difference of the mean of the numerical and the mean of the analytical solution, which is called weak convergence. It is the strong convergence that will allow us to compare the order of convergence of stochastic numerical methods with that of deterministic numerical methods. Thus consider

$$\mathbb{E}|Y_n - Y_t(t_n)| = y_0 \mathbb{E}\left|\prod_{i=0}^{n-1}(1 + \mu\Delta t + \sigma\Delta W_i) - e^{(\mu - \frac{1}{2}\sigma^2)t_n + \sigma W_{t_n}}\right|$$

The Taylor series expansion of the analytical solution is just the Taylor series expansion of the exponential function. Around Δt, we have

$$\begin{aligned}
e^{\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_n} &= 1 + \left[\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_n\right] + \frac{1}{2}\left[\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_n\right]^2 \\
&\quad + \frac{1}{6}\left[\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_n\right]^3 + \dots \\
&= 1 + \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_n + \frac{1}{2}\left(\mu - \frac{1}{2}\sigma^2\right)^2 \Delta t^2 + \sigma\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t\Delta W_n \\
&\quad + \frac{1}{2}\sigma^2\Delta W_n^2 + \frac{1}{6}\sigma^3\Delta W_n^3 + \mathcal{O}(\Delta t^2) \\
&= 1 + \mu\Delta t + \sigma\Delta W_n + \sigma\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t\Delta W_n + \frac{1}{6}\sigma^3\Delta W_n^3 + \mathcal{O}(\Delta t^2)
\end{aligned}$$

In this computation we have used that $\Delta W_n^2 = \Delta t$. Rearranging yields

$$1 + \mu\Delta t + \sigma\Delta W_n = e^{\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_n} - \sigma\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t\Delta W_n - \frac{1}{6}\sigma^3\Delta W_n^3 - \mathcal{O}(\Delta t^2).$$

This gives us an expression for the numerical solution in terms of the analytical solution. If we consider the previous expression at an intermediate time step t_i instead of final time t_n,

$$\begin{aligned}
\prod_{i=0}^{n-1}(1 + \mu\Delta t + \sigma\Delta W_i) &= \prod_{i=0}^{n-1}\left(e^{\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_i} - \sigma\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t\Delta W_i - \frac{1}{6}\sigma^3\Delta W_i^3 - \mathcal{O}(\Delta t^2)\right) \\
&= e^{\left(\mu - \frac{1}{2}\sigma^2\right)n\Delta t + \sigma\Delta W_n} + n\mathcal{O}(\Delta t\Delta W) + n\mathcal{O}(\Delta W^3) + n\mathcal{O}(\Delta t^2)
\end{aligned}$$

Now we can evaluate the mean of the difference at final time $T = n\Delta t$.

$$\begin{aligned}
\mathbb{E}\,|Y_n - Y_t(T)| &= \mathbb{E}|n\mathcal{O}(\Delta t\Delta W) + n(\mathcal{O}(\Delta W^3) + n\mathcal{O}(\Delta t^2)| \\
&= \mathbb{E}\left|\frac{T}{\Delta t}\mathcal{O}(\Delta t\Delta W) + \frac{T}{\Delta t}(\mathcal{O}(\Delta W^3) + \frac{T}{\Delta t}\mathcal{O}(\Delta t^2)\right| \\
&= T\mathbb{E}\left|\underbrace{\frac{1}{\Delta t}\mathcal{O}(\Delta t\Delta W)}_{\mathcal{O}(\Delta t^{1/2})} + \underbrace{\frac{1}{\Delta t}(\mathcal{O}(\Delta W^3)}_{\mathcal{O}(\Delta t^{1/2})} + \underbrace{\frac{1}{\Delta t}\mathcal{O}(\Delta t^2)}_{\mathcal{O}(\Delta t)}\right| \\
&= \mathcal{O}(\Delta t^{1/2}).
\end{aligned}$$

So this shows that the strong convergence of the Euler-Maruyama method is of order 1/2. Similarly it can be shown that the weak convergence of the method is of order 1. The Milstein method has strong and weak convergence order 1 [Hig01].

Appendix III: Lie Derivatives

To prove the identities for the Lie derivatives, a substantial amount of exterior calculus is necessary. The differential basis elements dx^i and $dS_i = \frac{1}{2}\epsilon_{ijk}dx^j \wedge dx^k$, for $i,j,k = 1,2,3$, in vector notation are denoted as

$$\begin{aligned} d\mathbf{x} &:= (dx^1, dx^2, dx^3)\,, \\ d\mathbf{S} &= (dS_1, dS_2, dS_3) \\ &:= (dx^2 \wedge dx^3,\, dx^3 \wedge dx^1,\, dx^1 \wedge dx^2)\,, \\ dS_i &:= \frac{1}{2}\epsilon_{ijk}dx^j \wedge dx^k\,, \\ d^3x &= d\mathrm{Vol} := dx^1 \wedge dx^2 \wedge dx^3\,. \end{aligned}$$

Contraction, here denoted by the hook notation ($\partial_j \mathbin{\lrcorner} \mathrm{d}x^i = \delta^i_j$) with the vector field $X = X^j\partial_j =: \mathbf{X}\cdot\nabla$ recovers the following familiar operations among vectors

$$\begin{aligned} X \mathbin{\lrcorner} d\mathbf{x} &= \mathbf{X}\,, \\ X \mathbin{\lrcorner} d\mathbf{S} &= \mathbf{X} \times d\mathbf{x}\,, \\ (\text{or,}\quad X \mathbin{\lrcorner} dS_i &= \epsilon_{ijk}X^j dx^k) \\ Y \mathbin{\lrcorner} X \mathbin{\lrcorner} d\mathbf{S} &= \mathbf{X} \times \mathbf{Y}\,, \\ X \mathbin{\lrcorner} d^3x &= \mathbf{X}\cdot d\mathbf{S} = X^k dS_k\,, \\ Y \mathbin{\lrcorner} X \mathbin{\lrcorner} d^3x &= \mathbf{X}\times\mathbf{Y}\cdot d\mathbf{x} = \epsilon_{ijk}X^iY^jdx^k\,, \\ Z \mathbin{\lrcorner} Y \mathbin{\lrcorner} X \mathbin{\lrcorner} d^3x &= \mathbf{X}\times\mathbf{Y}\cdot\mathbf{Z}\,. \end{aligned}$$

The exterior derivative and wedge product satisfy the following relations in components and in three-dimensional vector notation

$$\begin{aligned} df &= f_{,j}\,dx^j =: \nabla f \cdot d\mathbf{x} \\ 0 = d^2 f &= f_{,jk}\,dx^k \wedge dx^j \\ df \wedge dg &= f_{,j}\,dx^j \wedge g_{,k}\,dx^k =: (\nabla f \times \nabla g)\cdot d\mathbf{S} \\ df \wedge dg \wedge dh &= f_{,j}\,dx^j \wedge g_{,k}\,dx^k \wedge h_{,l}\,dx^l =: (\nabla f \cdot \nabla g \times \nabla h)\,d^3x \end{aligned}$$

Now, for a 1-form and a 2-form, it can be shown that

$$\begin{aligned} d(\mathbf{v}\cdot d\mathbf{x}) &= (\mathrm{curl}\,\mathbf{v})\cdot d\mathbf{S} \\ d(\mathbf{A}\cdot d\mathbf{S}) &= (\mathrm{div}\,\mathbf{A})\,d^3x\,. \end{aligned}$$

A number of familiar vector calculus identities follow from the compatibility condition $d^2 = 0$ as

$$\begin{aligned} 0 = d^2 f = d(\nabla f \cdot d\mathbf{x}) &= (\mathrm{curl}\,\mathrm{grad}\,f)\cdot d\mathbf{S}\,, \\ 0 = d^2(\mathbf{v}\cdot d\mathbf{x}) = d\big((\mathrm{curl}\,\mathbf{v})\cdot d\mathbf{S}\big) &= (\mathrm{div}\,\mathrm{curl}\,\mathbf{v})\,d^3x\,. \end{aligned}$$

It is now possible, using the previous calculations, to show that

$$\begin{aligned} &d(X \mathbin{\lrcorner} \mathbf{v}\cdot d\mathbf{x}) = d(\mathbf{X}\cdot\mathbf{v}) = \nabla(\mathbf{X}\cdot\mathbf{v})\cdot d\mathbf{x}, \\ &d(X \mathbin{\lrcorner} \boldsymbol{\omega}\cdot d\mathbf{S}) = d(\boldsymbol{\omega}\times\mathbf{X}\cdot d\mathbf{x}) = \mathrm{curl}\,(\boldsymbol{\omega}\times\mathbf{X})\cdot d\mathbf{S}, \\ &d(X \mathbin{\lrcorner} f\,d^3x) = d(f\mathbf{X}\cdot d\mathbf{S}) = \mathrm{div}\,(f\mathbf{X})\,d^3x. \end{aligned}$$

Using Cartan's formula, it is now not difficult to show

$$\begin{aligned} \pounds_X f &= X \mathbin{\lrcorner} df = \mathbf{X}\cdot\nabla f, \\ \pounds_X\,(\mathbf{v}\cdot d\mathbf{x}) &= \big(-\,\mathbf{X}\times\mathrm{curl}\,\mathbf{v} + \nabla(\mathbf{X}\cdot\mathbf{v})\big)\cdot d\mathbf{x}, \\ \pounds_X(\boldsymbol{\omega}\cdot d\mathbf{S}) &= \big(\mathrm{curl}\,(\boldsymbol{\omega}\times\mathbf{X}) + \mathbf{X}\,\mathrm{div}\,\boldsymbol{\omega}\big)\cdot d\mathbf{S}, \\ &= \big(-\,\boldsymbol{\omega}\cdot\nabla\mathbf{X} + \mathbf{X}\cdot\nabla\boldsymbol{\omega} + \boldsymbol{\omega}\,\mathrm{div}\,\mathbf{X}\big)\cdot d\mathbf{S}, \\ \pounds_X(f\,d^3x) &= (\mathrm{div}\,f\mathbf{X})\,d^3x. \end{aligned}$$

The second equality for the Lie derivative of a 2-form follows from the vector calculus identity for the curl of the cross product of two vectors.

Appendix IV: Imbalance Equation

We start by listing a couple of identities that are necessary en route. The weighted Helmholtz decomposition of $\mathbf{u}$ is given by

$$\mathbf{u} = \hat{\mathbf{k}} \times \nabla\psi + \epsilon\nabla\chi.$$

From here we can derive a couple of relations that will be very useful. We start by taking the curl of $\mathbf{u}$

$$\nabla \times \mathbf{u} = \det\begin{pmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \partial_x & \partial_y & \partial_z \\ -\psi_y & \psi_x & 0 \end{pmatrix} = \Delta\psi = \omega.$$

Next we take the cross product of $\hat{\mathbf{z}}$ and $\mathbf{u}$

$$\hat{\mathbf{z}} \times \mathbf{u} = -\nabla\psi + \epsilon\hat{\mathbf{z}} \times \nabla\chi.$$

We work towards the derivation of $\mathrm{d}\Omega$. By definition

$$\Omega := -\operatorname{div}\left[f\hat{\mathbf{z}} \times \mathbf{u} + \nabla h\right] = \operatorname{div}(f\nabla\psi) + \epsilon J(f,\chi) - \Delta h.$$

So, taking the partial derivative with respect to time

$$\begin{aligned}
\mathrm{d}\Omega =& \mathrm{d}\left[\operatorname{div}(f\nabla\psi) + \epsilon J(f,\chi) - \Delta h\right], \\
=& \operatorname{div}(f\nabla \mathrm{d}\psi) + J(f, \epsilon\mathrm{d}\chi) - \Delta\mathrm{d}h, \qquad \text{We use the rule } \operatorname{div}(\phi\mathbf{A}) = \phi\operatorname{div}\mathbf{A} + \mathbf{A}\cdot\nabla\phi \\
=& f\operatorname{div}(\nabla\mathrm{d}\psi) + (\nabla\mathrm{d}\psi)\cdot\nabla f + \hat{\mathbf{z}}\cdot\nabla f \times \epsilon\nabla\mathrm{d}\chi - \Delta\mathrm{d}h, \\
=& f\Delta\mathrm{d}\psi + \nabla f\cdot(\nabla\mathrm{d}\psi + \epsilon\nabla\mathrm{d}\chi\times\hat{\mathbf{z}}) - \Delta\mathrm{d}h, \\
=& (1+\epsilon f_1)\Delta\mathrm{d}\psi + \epsilon\nabla f_1\cdot(\nabla\mathrm{d}\psi + \epsilon\nabla\mathrm{d}\chi\times\hat{\mathbf{z}}) - \Delta\mathrm{d}h, \\
=& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla f_1\cdot(\nabla\mathrm{d}\psi + \epsilon\nabla\mathrm{d}\chi\times\hat{\mathbf{z}}) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h, \\
=& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla f_1\cdot(-\hat{\mathbf{z}}\times\mathrm{d}\mathbf{u}) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h, \qquad \text{Substitute } \hat{\mathbf{z}}\times\mathbf{u} \\
=& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla f_1\cdot(\nabla\mathrm{d}\psi - \epsilon\hat{\mathbf{z}}\times\nabla\mathrm{d}\chi) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h, \\
=& \epsilon\operatorname{div}(f_1\nabla\mathrm{d}\psi) - \epsilon^2(\nabla f_1\cdot\hat{\mathbf{z}}\times\nabla\mathrm{d}\chi) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h, \\
=& \epsilon\operatorname{div}(f_1\nabla\mathrm{d}\psi) - \epsilon^2(\hat{\mathbf{z}}\cdot\nabla\mathrm{d}\chi\times\nabla f_1) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h, \qquad \text{Here use } \mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A} \\
=& \epsilon\operatorname{div}(f_1\nabla\mathrm{d}\psi) + \epsilon^2 J(f_1,\mathrm{d}\chi) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h.
\end{aligned}$$

Now that we know which terms take place in the equation for $\mathrm{d}\Omega$, we will find expressions for each term in segments. We start with $-\Delta\mathrm{d}h$.

$$\begin{aligned}
h :=& \frac{\eta - b}{\epsilon\mathcal{F}}, \\
-\Delta\mathrm{d}h =& -\Delta\mathrm{d}\left(\frac{\eta-b}{\epsilon\mathcal{F}}\right), \\
=& -\frac{1}{\epsilon\mathcal{F}}\Delta\mathrm{d}\eta, \qquad \text{Now we can use equation (45)} \\
=& \frac{1}{\epsilon\mathcal{F}}\Delta(\operatorname{div}(\eta\,\mathrm{d}\mathbf{x}_t)), \\
=& \frac{1}{\epsilon\mathcal{F}}\Delta(\operatorname{div}((b+\epsilon\mathcal{F}h)\,\mathrm{d}\mathbf{x}_t)), \qquad \text{Here we apply } \operatorname{div}(\phi\mathbf{A}) = \phi\operatorname{div}\mathbf{A} + \mathbf{A}\cdot\nabla\phi \\
=& \frac{1}{\epsilon\mathcal{F}}\Delta\left((1+\epsilon b_1+\epsilon\mathcal{F}h)\operatorname{div}\mathrm{d}\mathbf{x}_t + \nabla(1+\epsilon b_1+\epsilon\mathcal{F}h)\cdot\mathrm{d}\mathbf{x}_t\right), \\
=& \frac{1}{\epsilon\mathcal{F}}\Delta\operatorname{div}\mathrm{d}\mathbf{x}_t + \frac{1}{\mathcal{F}}\Delta(b_1+\mathcal{F}h)\operatorname{div}\mathrm{d}\mathbf{x}_t + \frac{1}{\mathcal{F}}\Delta\big(\nabla(b_1+\mathcal{F}h)\cdot\mathrm{d}\mathbf{x}_t\big), \\
=& \frac{1}{\epsilon\mathcal{F}}\Delta\operatorname{div}\mathrm{d}\mathbf{x}_t + \frac{1}{\mathcal{F}}\Delta\Big(\operatorname{div}\left((b_1+\mathcal{F}h)\,\mathrm{d}\mathbf{x}_t\right)\Big)
\end{aligned}$$

Next up are $\epsilon\operatorname{div}(f_1\nabla\mathrm{d}\psi) + \epsilon^2 J(f_1,\mathrm{d}\chi)$. We start by expanding the first term again.

$$\begin{aligned}
\epsilon\nabla\cdot(f_1\nabla\mathrm{d}\psi) + \epsilon^2 J(f_1,\mathrm{d}\chi) =& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla\mathrm{d}\psi\cdot\nabla f_1 + \epsilon J(f_1,\epsilon\mathrm{d}\chi), \\
=& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla\mathrm{d}\psi\cdot\nabla f_1 + \epsilon\left(\hat{\mathbf{z}}\cdot\nabla f_1\times\epsilon\nabla\mathrm{d}\chi\right), \\
=& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla f_1\cdot\left(\nabla\mathrm{d}\psi - \hat{\mathbf{z}}\times\epsilon\nabla\mathrm{d}\chi\right), \\
=& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla f_1\cdot\left(-\hat{\mathbf{z}}\times(\hat{\mathbf{z}}\times\nabla\mathrm{d}\psi + \epsilon\nabla\mathrm{d}\chi)\right), \\
=& \epsilon f_1\Delta\mathrm{d}\psi + \epsilon\nabla f_1\cdot\left(-\hat{\mathbf{z}}\times\mathrm{d}\mathbf{u}\right), \\
=& \epsilon f_1\Delta\mathrm{d}\psi - \nabla f_1\cdot\left(\hat{\mathbf{z}}\times\epsilon\mathrm{d}\mathbf{u}\right).
\end{aligned}$$

To continue, we insert the curl form, the second equation in (45), into the last equation derived. Further, we will use the relation $\varpi = \epsilon\omega + f$, where $\omega = \nabla \times \mathbf{u}$.

$$
\begin{aligned}
\epsilon \mathrm{d}\mathbf{u} &= \mathrm{d}\mathbf{x}_t \times \operatorname{curl}(\epsilon\mathbf{u} + \mathbf{R}) + \nabla(\xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R})) - \nabla(\frac{\epsilon}{2}|\mathbf{u}|^2 + h)\,\mathrm{d}t, \\
&= \mathrm{d}\mathbf{x}_t \times (\varpi\hat{\mathbf{z}}) - \nabla\left(\xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right), \\
\hat{\mathbf{z}} \times \epsilon\mathrm{d}\mathbf{u} &= \hat{\mathbf{z}} \times \left[\mathrm{d}\mathbf{x}_t \times (\varpi\hat{\mathbf{z}}) - \nabla\left(\xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right)\right], \\
&= \hat{\mathbf{z}} \times \mathrm{d}\mathbf{x}_t \times (\varpi\hat{\mathbf{z}}) - \hat{\mathbf{z}} \times \nabla\left(\xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right).
\end{aligned}
$$

We will now substitute this into the appropriate place

$$
\begin{aligned}
\epsilon\nabla \cdot (f_1\nabla\mathrm{d}\psi) + \epsilon^2 J(f_1, \mathrm{d}\chi) &= \epsilon f_1 \Delta\mathrm{d}\psi - \nabla f_1 \cdot \left(\hat{\mathbf{z}} \times \mathrm{d}\mathbf{x}_t \times (\varpi\hat{\mathbf{z}}) - \hat{\mathbf{z}} \times \nabla\left(\xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right)\right), \\
&= \epsilon f_1 \Delta\mathrm{d}\psi - \nabla f_1 \cdot \Big(\mathrm{d}\mathbf{x}_t(\hat{\mathbf{z}} \cdot \varpi\hat{\mathbf{z}}) - \varpi\hat{\mathbf{z}}(\hat{\mathbf{z}} \cdot \mathrm{d}\mathbf{x}_t) \\
&\qquad - \hat{\mathbf{z}} \times \nabla\left(\xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right)\Big), \qquad \text{use } \mathrm{d}\mathbf{x}_t \perp \hat{\mathbf{z}} \\
&= \epsilon f_1 \Delta\mathrm{d}\psi - \nabla f_1 \cdot \mathrm{d}\mathbf{x}_t\varpi - \hat{\mathbf{z}} \cdot \nabla f_1 \times \nabla\left(\xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right), \\
&= \epsilon f_1 \Delta\mathrm{d}\psi - \varpi\nabla f_1 \cdot \mathrm{d}\mathbf{x}_t - J\left(f_1, \xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right).
\end{aligned}
$$

What is left is to derive an expression for $\Delta\mathrm{d}\psi$. We know that $\varpi = \epsilon\Delta\psi + f$ so for the time derivative, we can use the SRSW potential vorticity equation. This gives us

$$
\begin{aligned}
\mathrm{d}\varpi &= -\operatorname{div}(\varpi\,\mathrm{d}\mathbf{x}_t), \\
\epsilon\Delta\mathrm{d}\psi &= -\operatorname{div}\left((\epsilon\Delta\psi + f)\,\mathrm{d}\mathbf{x}_t\right), \\
\Delta\mathrm{d}\psi &= -\operatorname{div}(\Delta\psi\,\mathrm{d}\mathbf{x}_t) - \frac{1}{\epsilon}\operatorname{div}\left((1 + \epsilon f_1)\,\mathrm{d}\mathbf{x}_t\right), \\
&= -\operatorname{div}(\Delta\psi\,\mathrm{d}\mathbf{x}_t) - \frac{\operatorname{div}\mathrm{d}\mathbf{x}_t}{\epsilon} - \operatorname{div}(f_1\,\mathrm{d}\mathbf{x}_t)
\end{aligned}
$$

Now we substitute all our derivations into the main equation and we find

$$
\begin{aligned}
\mathrm{d}\Omega =& \epsilon\nabla \cdot (f_1\nabla\mathrm{d}\psi) + \epsilon^2 J(f_1, \mathrm{d}\chi) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h, \\
=& \epsilon f_1 \Delta\mathrm{d}\psi - \varpi\nabla f_1 \cdot \mathrm{d}\mathbf{x}_t - J\left(f_1, \xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right) + \Delta\mathrm{d}\psi - \Delta\mathrm{d}h, \\
=& \epsilon f_1\Big(-\operatorname{div}(\Delta\psi\,\mathrm{d}\mathbf{x}_t) - \frac{\operatorname{div}\mathrm{d}\mathbf{x}_t}{\epsilon} - \operatorname{div}(f_1\,\mathrm{d}\mathbf{x}_t)\Big) - \varpi\nabla f_1 \cdot \mathrm{d}\mathbf{x}_t - J\left(f_1, \xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right) \\
& - \operatorname{div}(\Delta\psi\,\mathrm{d}\mathbf{x}_t) - \frac{\operatorname{div}\mathrm{d}\mathbf{x}_t}{\epsilon} - \operatorname{div}(f_1\,\mathrm{d}\mathbf{x}_t) + \frac{1}{\epsilon\mathcal{F}}\Delta\mathcal{D} + \frac{1}{\mathcal{F}}\Delta\Big(\operatorname{div}\left((b_1 + \mathcal{F}h)\,\mathrm{d}\mathbf{x}_t\right)\Big), \\
=& -f\operatorname{div}\left((\Delta\psi + f_1)\,\mathrm{d}\mathbf{x}_t\right) - f_1\operatorname{div}\mathrm{d}\mathbf{x}_t - \varpi\nabla f_1 \cdot \mathrm{d}\mathbf{x}_t - J\left(f_1, \xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right) \\
& - \frac{\operatorname{div}\mathrm{d}\mathbf{x}_t}{\epsilon} + \frac{1}{\epsilon\mathcal{F}}\Delta\operatorname{div}\mathrm{d}\mathbf{x}_t + \frac{1}{\mathcal{F}}\Delta\Big(\operatorname{div}\left((b_1 + \mathcal{F}h)\,\mathrm{d}\mathbf{x}_t\right)\Big).
\end{aligned}
$$

Rearranging gives the final result

$$
\begin{aligned}
\mathrm{d}\Omega - \frac{1}{\epsilon\mathcal{F}}(\Delta - \mathcal{F})\operatorname{div}\mathrm{d}\mathbf{x}_t =& \frac{1}{\mathcal{F}}\Delta(\operatorname{div}(b_1 + \mathcal{F}h)\,\mathrm{d}\mathbf{x}_t) - \varpi\nabla f_1 \cdot \mathrm{d}\mathbf{x}_t \\
& - J\left(f_1, \xi_i \circ \mathrm{d}W_t^i \cdot (\epsilon\mathbf{u} + \mathbf{R}) + (h + \frac{\epsilon}{2}|\mathbf{u}|^2)\,\mathrm{d}t\right) \\
& - f\operatorname{div}\left((\Delta\psi + f_1)\,\mathrm{d}\mathbf{x}_t\right) - f_1\operatorname{div}\mathrm{d}\mathbf{x}_t.
\end{aligned}
$$

References

[AM78] R Abraham and JE Marsden. *Foundations of Mechanics*, volume 36. Benjamin/Cummings Publishing Company Reading, Massachusetts, 1978.

[Arn03] L Arnold. *Random Dynamical Systems.* Springer-Verlag Berlin Heidelberg 1998, 2003.

[AS01] L Arnold and B Schmalfuss. Lyapunov's second method for random dynamical systems. *Journal of Differential Equations*, 177(1):235–265, 2001.

[CFH17] D Crisan, F Flandoli, and DD Holm. Solution properties of a 3d stochastic Euler fluid equation. *arXiv preprint arXiv:1704.06989*, 2017.

[CSG11] MD Chekroun, E Simonnet, and M Ghil. Stochastic climate dynamics: Random attractors and time-dependent invariant measures. *Physica D: Nonlinear Phenomena*, 21:1685–1700, 2011.

[Elw78] KD Elworthy. Stochastic dynamical systems and their flows. *Stochastic analysis*, pages 79–95, 1978.

[FHT02] C Foias, DD Holm, and ES Titi. The three dimensional viscous Camassa-Holm equations, and their relation to the Navier-Stokes equations and turbulence theory. *Journal of Dynamics and Differential Equations*, 14(1):1–35, 2002.

[Geu04] B Geurts. *Elements of direct and large-eddy simulation.* RT Edwards, 2004.

[Hig01] DJ Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43:525–546, 2001.

[Hol08] DD Holm. *Geometric mechanics.* Imperial College Press, 2008.

[Hol15] DD Holm. Variational principles for stochastic fluid dynamics. In *Proc. R. Soc. A*, volume 471, page 20140963. The Royal Society, 2015.

[Kad01] LP Kadanoff. Turbulent heat flow: Structures and scaling. *Physics Today*, 54(8):34–39, 2001.

[Kel96] H Keller. *Attractors and bifurcations of the stochastic Lorenz system.* Citeseer, 1996.

[KPS12] PE Kloeden, E Platen, and H Schurz. *Numerical solution of SDE through computer experiments.* Springer Science & Business Media, 2012.

[Lor63] EN Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20:130–141, 1963.

[Lor86] EN Lorenz. On the existence of a slow manifold. *Journal of the Atmospheric Sciences*, 43(15):1547–1558, 1986.

[McK69] HP McKean. *Stochastic Integrals.* Academic Press, 1969.

[Ose68] VI Oseledec. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc*, 19(2):197–231, 1968.

[Sal62] B Saltzman. Finite amplitude free convection as an initial value problem - i. *Journal of the Atmospheric Sciences*, 4:329–341, 1962.

[Sep12] T Seppäläinen. Basics of stochastic analysis. *University of Wisconsin–Madison. Available at the University of Wisconsin-Madison: http://www. math. wisc. edu/˜ seppalai/courses/735/notes. pdf*, 2012.

[Spa12] C Sparrow. *The Lorenz equations: bifurcations, chaos, and strange attractors*, volume 41. Springer Science & Business Media, 2012.

[Spr03] JC Sprott. *Chaos and time-series analysis*, volume 69. Oxford University Press, 2003.

[Tu10] LW Tu. *An introduction to manifolds.* Springer Science & Business Media, 2010.

[UvB01] FE Udwadia and HF von Bremen. An efficient and stable approach of Lyapunov characteristic exponents of continuous dynamical systems. *Applied Mathematics and Computation*, 121(1):219–259, 2001.

[vRS10] MK von Renesse and M Scheutzow. Existence and uniqueness of solutions of stochastic functional differential equations. *Random Operators and Stochastic Equations*, 18(3):267–284, 2010.

[WSSV85] A Wolf, JB Swift, HL Swinney, and JA Vastano. Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, 16(3):285–317, 1985.