Computable characterisations of stochastic uncertainty in dynamical systems

Liam Blake

June 22, 2023

Thesis submitted for the degree of
Master of Philosophy
in
Applied Mathematics
at The University of Adelaide
Faculty of Sciences, Engineering and Technology
Discipline of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

List of Tables

List of Figures

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: ..... Date: .....

Acknowledgements

Dedication

Abstract

Chapter **??** introduces

Chapter 1

Introduction

1.1 Overview of this thesis

Chapter 2

Theoretical background

In this chapter, we establish background results from dynamical systems, probability theory, and stochastic calculus that are used in the theoretical work of this thesis.

2.1 Notation

First, we establish some notational conventions that are used throughout this thesis.

The set of $n \times m$ matrices with real-valued entries is denoted as $\mathbb{R}^{n \times m}$.

The norm symbol $\|\cdot\|$ without any additional qualifiers denotes the standard Euclidean norm for a vector, and the spectral (operator) norm induced by the Euclidean norm, i.e. for an $n \times n$ matrix A

$$\|A\| = \sup \left\{ \frac{\|Av\|}{\|v\|} \,\middle|\, v \in \mathbb{R}^n, \|v\| \neq 0 \right\}.$$

2.2 Results from dynamical systems

We are interested in continuous time, continuous state-space dynamical systems which can be represented as a first-order ordinary differential equation

$$\frac{\mathrm{d}w_t}{\mathrm{d}t} = u\left(w_t, t\right), \qquad w_0 = x \in \Omega, \tag{2.1}$$

where $u : \Omega \times [0, T] \to \mathbb{R}^n$ describes the velocity at each state and time. Such systems are well-studied and used across a huge breadth of applications, including (Wiggins 2005) [citation needed] EXAMPLES. Trajectories solving (2.1) are summarised by the flow map, which is an operator mapping initial conditions to the corresponding later states, under the evolution of (2.1).

Figure 2.1: A pictorial representation of the flow map F_s^t.

Definition 2.2.1 (Flow map) *Suppose $s, t \in [0, T]$. The **flow map** $F_s^t : \mathbb{R}^n \to \mathbb{R}^n$ from time s to t associated with (2.1) is the unique solution to*

$$\frac{\partial F_s^\tau(x)}{\partial \tau} = u\left(F_s^\tau(x), \tau\right), \qquad F_s^s(x) = x, \tag{2.2}$$

solved up to time $\tau = t$. Equivalently,

$$F_s^t(x) = x + \int_s^t u\left(F_s^\tau(x), \tau\right) \mathrm{d}\tau.$$

The flow map satisfies the following properties, under ASSUMPTIONS? For any $r, s, t \in [0, T]$ and points $x, w \in \mathbb{R}^n$,

1. F_s^t is invertible with inverse

$$\left[F_s^t\right]^{-1}(w) = F_t^s(w).$$

2. $F_s^t\left(F_r^s(x)\right) = F_r^t(x)$

Moreover, the gradient of the flow map (with respect to the initial condition) satisfies a useful property; the equation of variations.

Theorem 2.2.1 *Let $F_{t_1}^t$ be the flow map corresponding to (2.1). Then, the spatial gradient $\nabla F_{t_0}^t(x)$ satisfies the equation of variations*

$$\frac{\partial \nabla F_{t_1}^t(x)}{\partial t} = \nabla u\left(F_{t_1}^t(x), t\right) \nabla F_{t_1}^t(x). \tag{2.3}$$

Proof. Taking the gradient on both sides of (2.2) and using the chain rule gives

$$\nabla\left(\frac{\partial F_{t_1}^{t_2}(x)}{\partial t}\right) = \nabla u\left(F_{t_1}^t(x), t\right) \nabla F_{t_1}^t(x).$$

SMOOTHNESS □

An important inequality for establishing bounds

Theorem 2.2.2 (Grönwall's inequality) *Let $\alpha, \beta, u : [a, b] \to \mathbb{R}$ be functions such that β and u are continuous and that the negative part of α is integrable on every closed and bounded subset of $[a, b]$. Then, if β is non-negative and for all $t \in [a, b]$,*

$$u(t) \leq \alpha(t) + \int_a^t \beta(\tau) u(\tau) \mathrm{d}\tau$$

then

$$u(t) \leq \alpha(t) + \int_a^t \alpha(\tau)\beta(\tau) \exp\left(\int_\tau^t \beta(s)\,\mathrm{d}s\right) \mathrm{d}\tau.$$

Additionally, if α is non-decreasing, then

$$u(t) \leq \alpha(t) \exp\left(\int_a^t \beta(\tau)\,\mathrm{d}\tau\right)$$

Proof.

□

2.3 Results from probability theory

2.3.1 Notions of convergence

There are several different notions of convergence for a sequence of random variables, which we briefly recall here. Consider a sequence of m-dimensional random vectors $X_1, X_2, \ldots$ and an m-dimensional random vector X. We say that:

- The sequence $X_1, X_2, \ldots$ converges *in distribution* to X if

  $$\lim_{n\to\infty} F_n(x) = F(x),$$

  where F_n is the cumulative distribution function for X_n and F is the cumulative distribution function for X, for every point $x \in \mathbb{R}^m$ where F is continuous. If this is the case, we write

  $$X_n \xrightarrow[\text{distribution}]{} X, \quad \text{as } n \to \infty.$$

- The sequence $X_1, X_2, \ldots$ converges *in probability* to X if for every $\delta > 0$

  $$\lim_{n\to\infty} P\left(\|X_n - X\| < \delta\right) = 0,$$

  in which case we write

  $$X_n \xrightarrow[\text{probability}]{} X, \quad \text{as } n \to \infty.$$

- The sequence $X_1, X_2, \ldots$ converges *almost surely* to X if

  $$P\left(\lim_{n\to\infty} X_n = X\right) = 1,$$

  in which case we write

  $$X_n \xrightarrow[\text{almost surely}]{} X, \quad \text{as } n \to \infty.$$

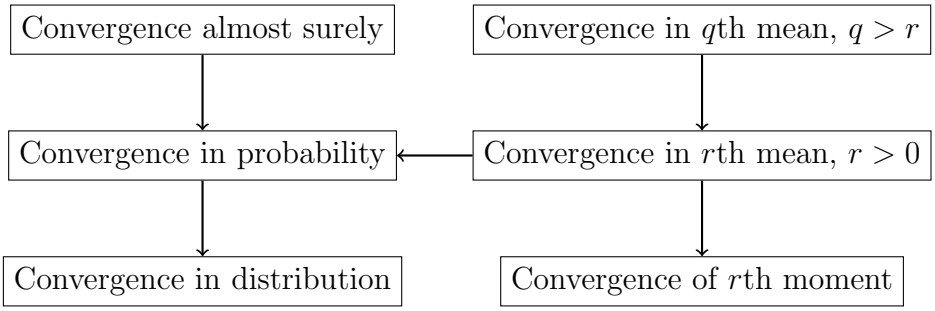


Figure 2.2: The strength of each notion of convergence in probability.

- For $r > 0$, the sequence $X_1, X_2, \ldots$ converges *in rth mean* to X if

$$\lim_{n\to\infty} \mathbb{E}\left[\|X_n - X\|^r\right] = 0,$$

in which case we write

$$X_n \xrightarrow[r\text{th mean}]{} X, \quad \text{as } n \to \infty.$$

There are implications between each notion of convergence, with convergence almost surely being the strongest and convergence in distribution the weakest. These implications are summarised in Figure 2.2, and all these results are stated in proven in Brémaud (2020), for instance.

2.3.2 The Wiener process

Definition 2.3.1 (Wiener process) *The (one-dimensional)* Wiener process *is a stochastic process B_t taking values in $\mathbb{R}$ and satisfying the following properties:*

(i) $B_0 = 0$,

(ii) for every $s > 0$, the increments $W_{s+t} - W_s$ for $t \geq 0$ are independent of W_r for all $r < s$,

(iii) $B_{s+t} - B_t \sim \mathcal{N}(0, s)$ *for all $s, t > 0$, and*

(iv) B_t is continuous in t almost surely.

*The n-*dimensional Wiener process *is a stochastic process W_t taking values in R^n such that each component of W_t is a one-dimensional Wiener process and the components of W_t are mutually independent.*

Figure 2.3: Several realisations of (left) a one-dimensional Wiener process W_t evolving through time, and (right) a two-dimensional Wiener process $\left(W_t^{(1)}, W_t^{(2)}\right)$.

Remarkably, these properties *uniquely* define the Wiener process, with the additional result that for any $t > 0$, $B_t \sim \mathcal{N}(0, t)$, a Gaussian distribution with mean zero and variance t. It follows that for the n-dimensional Wiener process W_t, at any time $t > 0$, $W_t \sim \mathcal{N}(0, tI)$, an n-dimensional Gaussian distribution with mean zero and covariance matrix tI.

A Wiener process is a type of Lévy process, which is a more general class of stochastic process satisfying only conditions (i), (ii), and (iii) in Definition 2.3.1 (Applebaum 2004).

2.3.3 The Dirac-delta distribution

The Dirac-delta probability measure has the property that for any measurable function $f : \mathbb{R}^n \to \mathbb{R}^n$,

$$\int_{\mathbb{R}^n} f \, \mathrm{d}\delta[x_0] = f(x_0).$$

2.4 Stochastic differential equations

In practice, there is uncertainty associated with a differential equation, which may arise from a variety of sources including observational error, parameter uncertainty, interpolation error and due to unresolved effects in the model. Stochastic differential equations are an extension of ordinary differential equations that include stochastic terms, which can account for this uncertainty.

For an introduction to stochastic differential equations, see Øksendal (2003) or Kallianpur and Sundar (2014).

2.4.1 The Itô integral

For our purposes, we can think of an Itô integral as being defined as the limit in probability of a sequence of sums, i.e. for a scalar but possibly random-valued function $f\colon [a, b] \to \mathbb{R}$,

$$\sum_{[t_i, t_{i+1}] \in \mathcal{P}_N} f(t_i) \left(W_{t_{i+1}} - W_{t_i}\right) \xrightarrow[\text{probability}]{} \int_a^b f(t) \, \mathrm{d}W_t, \quad \text{as } N \to \infty \tag{2.4}$$

where $\mathcal{P}_N$ is a partition of $[a, b]$ with $\lim_{N\to\infty} \mathcal{P}_N = [a, b]$, *à la* the definition of the Riemann integral. It can be shown that this limit exists for a large class of both deterministic- and

random-valued functions, by constructing appropriate approximations of the function f. This construction of the Itô integral is available in many textbooks on stochastic processes, such as Kallianpur and Sundar (2014) and Øksendal (2003), so it will not be repeated here.

The extension of the Itô integral to vector- and matrix-valued functions is straightforward. Let $g\colon [a,b] \to \mathbb{R}^{n\times m}$ be a function giving possibly random $n \times m$ matrices (take $m = 1$ to describe a vector-valued function). Then, we define the Itô integral of g with respect to the m-dimensional Wiener process W_t over the time interval $[a,b]$ as

$$\int_a^b g(t)\,\mathrm{d}W_t \coloneqq (\mathcal{I}_1,\ldots,\mathcal{I}_n)^\intercal\,, \tag{2.5a}$$

where

$$\mathcal{I}_i = \sum_{j=1}^{m} \int_a^b g_{ij}\,(t)\,\mathrm{d}W_t^{(j)}, \tag{2.5b}$$

for $i = 1,\ldots,n$ and where g_{ij} denotes the (i,j)th element of g.

2.4.2 Itô stochastic differential equations

The differential form of an n-dimensional Itô stochastic differential equation is

$$\mathrm{d}y_t = u\,(y_t,t)\,\mathrm{d}t + \sigma\,(y_t,t)\,\mathrm{d}W_t, \tag{2.6}$$

where the solution y_t is a stochastic process taking values in $\mathbb{R}^n$, $u\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is the drift and $\sigma\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n\times m}$ is the diffusivity. The driving process W_t is the canonical, m-dimensional Wiener process as defined in Definition 2.3.1. For a (possibly random) initial condition y_0, the solution y_t to (2.6) satisfies

$$y_t = y_0 + \int_0^t u\,(y_\tau,\tau)\,\mathrm{d}\tau + \int_0^t \sigma\,(y_\tau,\tau)\,\mathrm{d}W_\tau. \tag{2.7}$$

The terms of the differential form (2.6) are not all rigorously defined, and so the differential form is rather notation that is equivalent to (2.7). In the most general case, the drift u and diffusivity σ are permitted to themselves be random functions[1], but in this thesis we assume that both are deterministic.

[1]For more information, see for instance Kallianpur and Sundar (2014). The formal treatment of such stochastic differential equations remains an area of open research, such as establishing the conditions for existence and uniqueness of solutions [citation needed] , and EXAMPLE.

2.4.3 Analytical tools for Itô calculus

There are several tools available for the analytic treatment of Itô integrals and solutions to stochastic differential equations, which we make use of throughout. The first is Itô's isometry, which relates the expectation of an Itô integral to that of a deterministic one and is useful for computing moments.

Theorem 2.4.1 (Itô's Isometry) *Let $f : \Omega \times [0,T] \to \mathbb{R}$ be an Itô integrable stochastic process. Then, for any $t \in [0,T]$*

$$\mathbb{E}\left[\left(\int_0^t f(\omega,\tau)\,\mathrm{d}W_\tau\right)^2\right] = \mathbb{E}\left[\int_0^t f(\omega,\tau)^2\,\mathrm{d}\tau\right]$$

Proof. Itô's isometry typically arises in the formal construction of the Itô integral. For example, see Section 5.1 of Kallianpur and Sundar (2014). □

Next, we have Itô's Lemma (or the Itô Formula), which is a change-of-variables formula in stochastic calculus and can be thought of as a generalisation of the chain rule from deterministic calculus. We state and use the multidimensional form of the Lemma for solutions to Itô stochastic differential equations, although more general forms exist (e.g. see Theorem 5.4.1 of Kallianpur and Sundar (2014)).

Theorem 2.4.2 (Itô's Lemma) *Let X_t be the strong solution to the stochastic differential equation*

$$\mathrm{d}X_t = a(X_t,t)\,\mathrm{d}t + b(X_t,t)\,\mathrm{d}W_t,$$

where $a : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^n$, $b : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^{n\times p}$ and W_t is the canonical p-dimensional Wiener process. If $f : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^m$ is twice continuously-differentiable, then the stochastic process $Y_t \coloneqq f(X_t,t)$ is a strong solution to the stochastic differential equation

$$\begin{aligned}\mathrm{d}Y_t = \left(\frac{\partial f}{\partial t}(X_t,t) + \nabla f(X_t,t)\,a(X_t,t) + \frac{1}{2}\mathrm{tr}\left[b(X_t,t)^\intercal\,\nabla\nabla f(X_t,t)\,b(X_t,t)\right]\right)\mathrm{d}t \\ + \nabla f(X_t,t)\,b(X_t,t)\,\mathrm{d}W_t.\end{aligned}$$

Proof.

□

Our third and final result is the Burkholder-Davis-Gundy inequality, which when applied to stochastic integrals provides bounds on the expected norm.

Theorem 2.4.3 (Burkholder-Davis-Gundy Inequality) *Let M_t be an Itô-integrable stochastic process taking values in $\mathbb{R}^n$. Then, for any $p > 0$ there exists constants $c_p, C_p > 0$ independent of the stochastic process M_t such that*

$$c_p\mathbb{E}\left[\left(\int_0^t \|M_\tau\|^2\,\mathrm{d}\tau\right)^p\right] \le \mathbb{E}\left[\sup_{\tau\in[0,t]}\left\|\int_0^\tau M_s\,\mathrm{d}W_s\right\|^{2p}\right] \le C_p\mathbb{E}\left[\left(\int_0^t \|M_\tau\|^2\,\mathrm{d}\tau\right)^p\right].$$

Proof. This result is stated and proven as Theorem 5.6.3 of Kallianpur and Sundar (2014). □

2.4.4 The Stratonovich integral and Stratonovich SDEs

The Stratonovich integral is an alternative definition of a stochastic integral [citation needed] , which is arises naturally from physically considerations (). The Stratonovich integral can be written as the limit in probability

$$\sum_{t_i, t_{i+1} \in \mathcal{P}_N} \frac{f(t_{i+1}) - f(t_i)}{2} \left(W_{t_{i+1}} - W_{t_i}\right) \xrightarrow[\text{2nd mean}]{} \int_a^b f(t) \circ \mathrm{d}W_t, \quad \text{as } N \to \infty$$

where $\mathcal{P}_N$ is again a partition of $[a, b]$ and the $\circ \mathrm{d}W_t$ notation is used to distinguish the Stratonovich intepretation of the integral. The Stratonovich integral is extended to vector-valued functions and a multivariable Wiener process in the same fashion as (2.5). We can then write a Stratonovich stochastic differential equation

$$\mathrm{d}x_t = u(x_t, t)\,\mathrm{d}t + \sigma(x_t, t) \circ \mathrm{d}W_t \tag{2.8}$$

in completely the same way as an Itô SDE. In this thesis, we only consider Itô stochastic differential equations in our theoretical developments, but there is a conversion between the two interpretations that requires modifying the drift term u of the SDE. The Stratonovich SDE (2.8) is equivalent to the Itô SDE [citation needed]

$$\mathrm{d}x_t = \left[u(x_t, t) + c(x_t, t)\right]\mathrm{d}t + \sigma(x_t, t)\,\mathrm{d}W_t, \tag{2.9}$$

where $c(x_t, t) = \left(c_1(x_t, t), \ldots, c_n(x_t, t)\right)^T$ with

$$c_i(x_t, t) \coloneqq \operatorname{tr}\left(\left[\nabla \sigma_{i\cdot}(x_t, t)\right]^\intercal \sigma(x_t, t)\right),$$

where $\nabla \sigma_{i\cdot}$ denotes the Jacobian derivative of the ith row of σ.

2.4.5 The Fokker-Planck equation

The probability density function $\rho : \mathbb{R}^n \times [0, T] \to [0, \infty)$ for the solution to (2.6) at time $t \in [0, T]$ is the solution to the corresponding Fokker-Planck equation (Risken 2012)

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \nabla \cdot \nabla \cdot \left(\rho \sigma \sigma^\intercal\right) - \nabla \cdot (\rho u) \tag{2.10}$$

subject to some initial density $\rho(x, 0)$ given by the initial condition to (2.6). For a fixed and deterministic initial condition $y_0 = x$, the corresponding initial condition to (2.10) is the Dirac-delta distribution centred at x.

For certain choices of the drift u and diffusivity σ, (2.10) reduces to several other well-known partial differential equations, including:

- When $\sigma \equiv D$, a scalar constant, then (2.10) is the convection-diffusion equation with velocity field u and diffusivity D.

- When $\sigma \equiv 0$, i.e. there is no diffusion, then (2.10) is the continuity equation with velocity field u.

The connection between the Fokker-Planck equation and the SDE (2.6) means that the solutions to each of these PDEs can be equivalently thought of as the time-evolution of the probability density of an SDE.

2.4.6 Some explicitly solvable SDEs

In general, the solution to a stochastic differential equation cannot be expressed analytically, either as an explicit expression involving the Wiener process W_t or as a probability measure or density function. At best, most solutions can be written in terms of an Itô integral which can otherwise not be simplified. However, there are several simple examples for which a solution can be written, and even time-marginal probability density functions can be derived. Here, we list several examples which are used to validate theory and test algorithms throughout this thesis.

Example 2.4.1 (Homogenous and linear SDE) *Consider an n-dimensional stochastic differential equation*

$$\mathrm{d}x_t = A(t)x_t\,\mathrm{d}t + B(t)\,\mathrm{d}W_t, \tag{2.11}$$

where $A\colon [0,T] \to \mathbb{R}^{n\times n}$ is an matrix-valued function where each element is differentiable, and $B\colon [0,T] \to \mathbb{R}^{n\times m}$ is a matrix-valued function, and W_t is an m-dimensional Wiener process.

The solution to (2.11) can be written exactly as

$$x_t \sim \mathcal{N}\left(\Phi(t)x_0,\, \Phi\left(t\right)\left[\int_0^t \Phi\left(\tau\right)^{-1} B\left(t\right) B\left(t\right)^{\intercal}\left(\Phi\left(\tau\right)^{-1}\right)^{\intercal}\right]\Phi(t)^{\intercal}\right),$$

where Φ is the fundamental matrix solution to the corresponding homogeneous equation

$$\frac{\mathrm{d}\Phi(t)}{\mathrm{d}t} = A(t)\Phi(t).$$

The details of this result are provided in Appendix A.1.

Example 2.4.2 (Benê's SDE) *The 1-dimensional stochastic differential equation*

$$\mathrm{d}x_t = \tanh\left(x_t\right)\mathrm{d}t + \mathrm{d}W_t, \tag{2.12}$$

is known as Benê's stochastic differential equation *(Särkkä and Solin 2019). The probability density function of a weak solution of* (2.12) *can be derived using an appropriate change of measure with Girsanov's theorem. A proof of this, and the derivation of a weak solution to* (2.12) *are provided in Section 7.3 of Särkkä and Solin (2019). The probability density function* $p : \mathbb{R} \times (0, \infty) \to [0, \infty)$ *for the solution* x_t *at time* $t > 0$ *is given by*

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \frac{\cosh(x)}{\cosh(x_0)} \exp\left[-\frac{t}{2} - \frac{1}{2t}(x - x_0)^2\right], \tag{2.13}$$

where $x_0 \in \mathbb{R}$ *is a fixed initial condition. This probability density function is plotted, for the initial condition* $x_0 = 1/2$ *and various times, in Figure 2.4. The resulting density is not symmetric and bimodal, with the two modes moving apart in the positive and negative* x*-directions respectively as* t *increases. For fixed* t*, the probability density function can be expressed as the mixture of two Gaussians with respective means* $x_0 + t$ *and* $x_0 - t$*, with details provided in Appendix A.2. This expression allows easy calculation of the mean and expectation of* x_t*, as*

$$\mathbb{E}[x_t] = \frac{x_0 \cosh(x_0) + t \sinh(x_0)}{\cosh(x_0)},$$

and

$$\mathbb{V}[x_t] =$$

Example 2.4.3

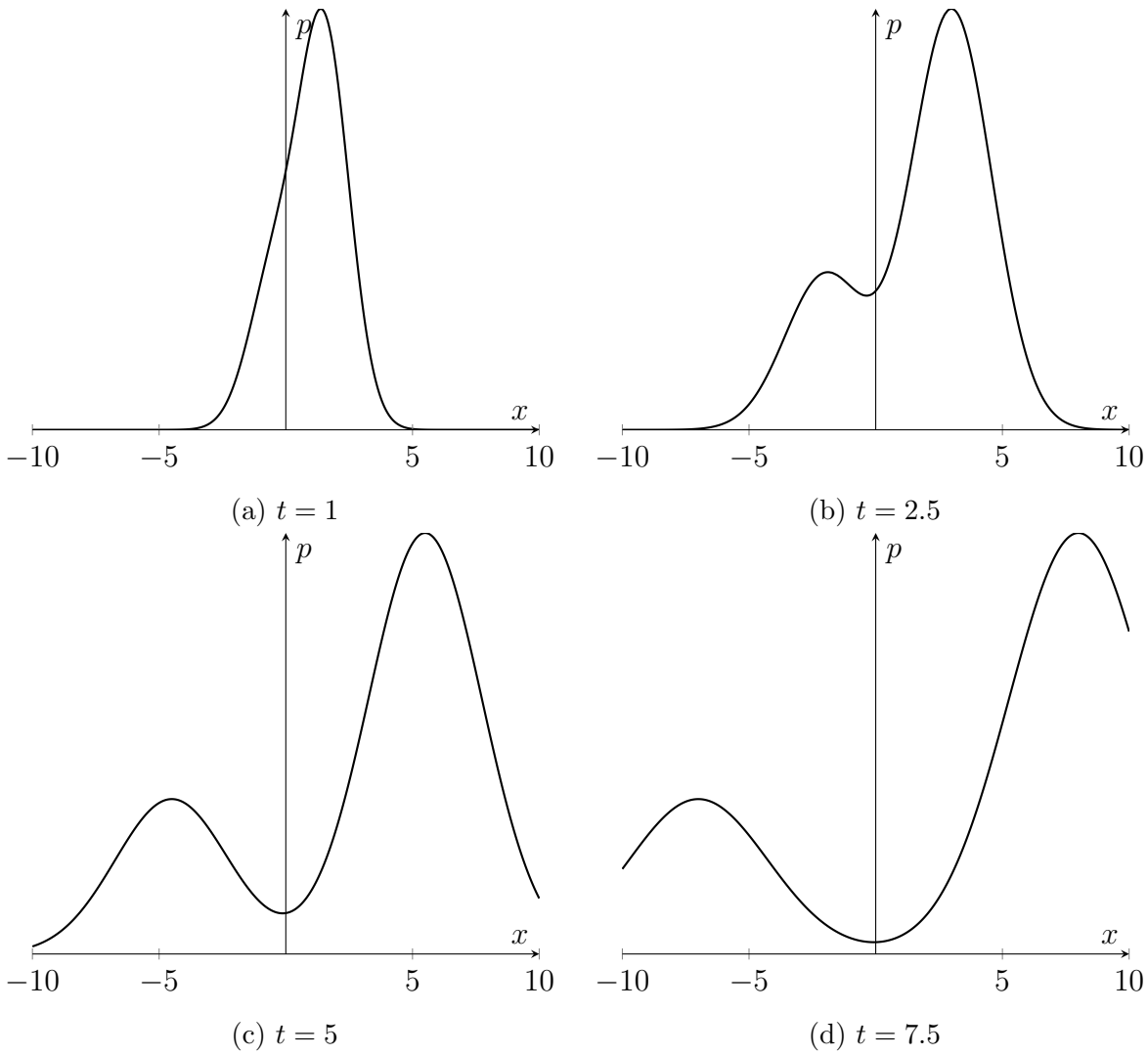
2.4.7 Numerical schemes for approximating SDEs

In general, solving a stochastic differential equation analytically is not possible, and so as with ordinary differential equations we instead look to use numerical schemes to approximate solutions. However, the solution to a stochastic differential equation is itself a random variable, so a single sample path is not sufficient. Instead, a numerical SDE scheme produces approximate *realisations* of the solution. The Euler-Maruyama (EM) method is analogous to the Euler method for ODEs, and considered by many to be the simplest method for numerically solving SDEs (Kloeden and Platen 1992). The update step of the EM scheme, with step size Δt, is

$$x_{t+\Delta t} = x_t + \Delta t u(x_t, t) + \Delta t \sigma(x_t, t) Z_t, \tag{2.14}$$

where Z_t is sampled from the standard Gaussian $\mathcal{N}(0, I)$.

There are many other schemes for generating approximate samples of a stochastic differential equation, of varying weak and strong orders. For instance, extensions of Runge-Kutta-type schemes (Roberts 2012, Rößler 2010).



(a) $t = 1$ (b) $t = 2.5$

(c) $t = 5$ (d) $t = 7.5$

Figure 2.4: The probability density function (2.13) for the time-marginal solution of Benê's SDE (2.12), for the initial condition $x_0 = 1/2$ at various times. The density function consists of two distinct modes that move further apart as t increases.

2.5 Stochastic sensitivity

Given possibly time-dependent velocity data $u : \mathbb{R}^2 \times [0,T] \to \mathbb{R}^2$, Balasuriya (2020a) considers the evolution of solutions to the differential equation

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = u\left(x_t, t\right).$$

Solutions can be summarised by the flow map $F_{t_1}^{t_2}$, as in **??**. In most practical situations, the Eulerian velocity data driving ocean and atmospheric models relies upon measurements of estimates obtained on a low resolution spatial discretisation. Balasuriya (2020a) introduces stochastic sensitivity as a new tool for directly quantifying the impact of Eulerian uncertainty on Lagrangian trajectories. The evolution of Lagrangian trajectories is modelled as solution to a Itô stochastic ordinary differential equation.

To directly account for these unresolved sources of uncertainty, the "true" Lagrangian trajectories evolve as solution to the stochastic differential equation

$$\mathrm{d}y_t = u\left(y_t, t\right)\mathrm{d}t + \varepsilon\sigma\left(y_t, t\right)\mathrm{d}W_t, , \tag{2.15}$$

where $0 < \varepsilon \ll 1$ is a parameter quantifying the scale of the noise, $\sigma : \mathbb{R}^2 \times [0,T] \to \mathbb{R}^{2\times 2}$ is the 2×2 diffusion matrix, and W_t is the canonical two-dimensional Wiener process. In the original formulation (Balasuriya 2020a), ε is a dimensionless parameter and σ is dimensional, but an alternative scaling technique relates ε to spatial and velocity uncertainty scales in the data (see the follow-up work by Badza et al. (2023), Balasuriya (2020b), Fang et al. (2020)) Since σ can vary by both space and time, the noise is multiplicative. The diffusion matrix σ is specified *a priori*, based on any knowledge of how uncertainty varies with space and time, e.g. from experimental considerations, observation error estimates. If no such prior information is known, then $\sigma \equiv I$, the 2×2 identity matrix is the default choice.

To quantify uncertainty in a way that is independent of the noise scale ε, Balasuriya (2020a) defined the random variable $z_\varepsilon\left(x,t\right)$ on $\mathbb{R}^2 \times [0,T]$ as

$$z_\varepsilon\left(x,t\right) \coloneqq \frac{y_t - F_0^t(x)}{\varepsilon}.$$

The main aim is to compute statistics of z_ε at the final time T, so that of $z_\varepsilon\left(x,T\right)$. Balasuriya (2020a) then considers the signed projection of $z_\varepsilon\left(x,T\right)$ onto a ray emanating from the deterministic position $F_0^T(x)$ in a given direction, defining

$$P_\varepsilon\left(x,\theta\right) \coloneqq \hat{n}^\intercal z_\varepsilon(x,T),$$

where $\theta \in \left[-\pi/2, \pi/2\right)$ and

$$\hat{n}(\theta) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.$$

The statistics of $z_\varepsilon(x,T)$ and $P_\varepsilon(x,\theta)$ are considered in the limit as $\varepsilon \downarrow 0$, which provides a characterisation of the uncertainty of the model *independently* of the scale of the noise. Balasuriya (2020a) provided computable expressions for the mean and variance of $P_\varepsilon(x,\theta)$ in this limit of small noise, which we summarise here. For proofs of these results, see the appendices of Balasuriya (2020a).

The first result established by Balasuriya (2020a) is that the expected location is deterministic, in the following sense.

Theorem 2.5.1 (Balasuriya 2020a) *For all $x \in \mathbb{R}^2$,*

$$\lim_{\varepsilon \downarrow 0} \mathbb{E}\left[z_\varepsilon(x,T)\right] = 0.$$

The variance of $P_\varepsilon(x,\theta)$ is used to assign a computable scalar measure of uncertainty to the trajectory.

Definition 2.5.1 (Balasuriya 2020a) *a) The **anisotropic uncertainty** is a scalar field $A : \mathbb{R}^2 \times \left[-\pi/2, \pi/2\right) \to [0,\infty)$ defined by*

$$A(x,\theta) := \sqrt{\lim_{\varepsilon \downarrow 0} \mathbb{V}\left[P_\varepsilon(x,\theta)\right]}.$$

*b) The **stochastic sensitivity** is a scalar field $S : \mathbb{R}^2 \to [0,\infty)$ defined by*

$$S^2(x) := \lim_{\varepsilon \downarrow 0} \sup_{\theta} \mathbb{V}\left[P_\varepsilon(x,\theta)\right].$$

By employing techniques from both deterministic and stochastic calculus (i.e. Grönwall's inequality, the Burkholder-Davis-Gundy inequality, Itô's Lemma), Balasuriya further established expressions for both the anisotropic uncertainty and the stochastic sensitivity that are computable given only the flow map and velocity data.

Theorem 2.5.2 (Balasuriya 2020a) *For $x \in \mathbb{R}^2$, set $w := F_0^t(x)$. Then, for any $\theta \in \left[-\pi/2, \pi/2\right)$,*

$$A(x,\theta) = \left(\int_0^T \left\| \Lambda(x,t,T)\, J\hat{n}(\theta) \right\| \mathrm{d}t \right)^{1/2},$$

where

$$\Lambda(x,t,T) := e^{\int_t^T [\nabla \cdot u]\left(F_0^\xi(x),\xi\right) \mathrm{d}\xi} \sigma\left(F_0^t(x),t\right)^{\intercal} J \nabla_w F_T^t(w),$$

with the gradients ∇_w of the flow map taken with respect to the mapped position w, and

$$J := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Additionally, stochastic sensitivity is computed as

$$S^2(x) = P(x) + N(x),$$

with

$$\begin{aligned} L(x) &\coloneqq \frac{1}{2}\sum_{i=1}^{2}\int_0^T \left[\Lambda_{i2}\left(x,t,T\right)^2 - \Lambda_{i1}\left(x,t,T\right)^2\right] \mathrm{d}t \\ M(x) &\coloneqq \sum_{i=1}^{2}\int_0^T \Lambda_{i1}\left(x,t,T\right)\Lambda_{i2}\left(x,t,T\right) \mathrm{d}t \\ N(x) &\coloneqq \sqrt{L^2(x) + M^2(x)} \\ P(x) &\coloneqq \left|\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\int_0^T \Lambda_{ij}\left(x,t,T\right)^2 \mathrm{d}t\right|, \end{aligned}$$

where Λ_{ij} is the (i,j)-element of Λ.

Definition 2.5.2 (Balasuriya 2020a) *Given a spatial resolution L_r, the* resolution-scaled stochastic sensitivity *is defined on Ω*

$$S_r(x) \coloneqq \ln\left(\frac{\sqrt{S^2(x)}}{L_r}\right).$$

2.5.1 Current applications & shortcomings

Since stochastic sensitivity is only a recent development, it has only been applied in a limited number of places so far. Here, we briefly review the literature in which the original formulation stochastic sensitivity by Balasuriya (2020a) has been applied.

- Balasuriya (2020b) uses stochastic sensitivity to compute an error bound for the finite-time Lyapunov computation.
- Fang et al. (2020)
- Badza et al. (2023) investigate the impact of velocity uncertainty on Lagrangian coherent structures (e.g. see the reviews by Balasuriya et al. (2018) and Hadjighasem et al. (2017)) extracted as robust sets with stochastic sensitivity. The stochastic model (2.15) is used to generate realisations of Lagrangian trajectories subject to noise on the velocity field. By directly capturing such uncertainty as a means of coherent set Badza et al. (2023) showed that robust sets extracted with stochastic sensitivity do SOMETHING.

There are several limitations to the work as originally presented by Balasuriya (2020a),

1. The tools are restricted to two-dimensional models, and the constructions using projections have no obvious extension to n-dimensions. Extending stochastic sensitivity to n-dimensions will enable application to a much broader class of models beyond the fluid flow context, including high-dimensional climate models.

2. Balasuriya (2020a) only computes the expectation and variance of the projections $P_\varepsilon(x, \theta)$, which does not give us the distribution under the limit as ε approaches 0.

3. The computational formula for the anisotropic uncertainty and stochastic sensitivity, as described in Theorem 2.5.2, require knowledge of the divergence $\nabla \cdot u$ of the velocity field, and computation of four integrals.

An alternative approach to uncertainty quantification in Lagrangian dynamics was recently introduced by Branicki and Uda (2023), extending results from their earlier work (Branicki and Uda 2021). However, the divergence approach does not provide any insight into the underlying probability distribution. Moreover, stochastic sensitivity

Chapter 3

The need for developments in uncertainty parameterisation

3.1 Stochastic parameterisation

3.1.1 In the atmosphere

The review by Berner et al. (2017) summarises

3.1.2 In the ocean

The mixing effect that these eddies have on the surrounding flow can be modelled with spatiotemporally-varying diffusion [citation needed] , which via the Fokker-Planck equation can be equivalently formulated as a stochastic differential equation with multiplicative noise. The Lagrangian trajectories, incorporating these unresolved eddy effects, are then modelled as solutions to the stochastic differential equation. Equivalently, through the Fokker-Planck equation we can consider the evolution of a passive tracer undergoing advection due to the deterministic drift and diffusion from both any natural diffusivity and the unresolved processes. The probability density function that solves the Fokker-Planck equation can be instead thought of as a time-varying density (with the appropriate normalisation) of the tracer. For instance, the Fokker-Planck equation has been used to model the transport of [citation needed] . Hence, understanding the evolution of solutions to a stochastic differential equation is valuable in oceanography, as a means of quantifying both observational error and unresolved subgrid processes.

There are several different methods for quantifying eddy diffusivity given either observed tracer data or a global ocean circulation model.

The simplest notion of eddy diffusivity is a defined by

A recent approach by Ying et al. (2019) uses Bayesian inference to estimate the eddy diffusivity tensor from observed Lagrangian tracer data, by numerically solving the

Fokker-Planck equation to compute a likelihood function.

3.2 Limitations of stochastic simulation

However, the most significant drawback of bulk stochastic simulation is the computational load.

In general, a large number of samples is required for convergent statistics and accurate inference, as discussed by Leutbecher (2019).

The recent review by Leutbecher et al. (2017) highlights the need to develop computationally efficient schemes for quantifying stochasticity in weather and climate forecast models.

Chapter 4

Publication: Explicit Gaussian characterisation of model uncertainty in the limit of small noise

The following is a copy of the published article by Blake et al. (2023). Sections 4.2 to 4.7 are as presented in Blake et al. (2023). The appendices include more details than in the published version.

4.1 Statement of Authorship

4.2 Abstract

Prediction via deterministic continuous-time models will always be subject to model error, for ex ample due to unexplainable phenomena, uncertainties in any data driving the model, or discretisation/resolution issues. A standard method for uncertainty quantification in such instances is to introduce noise into the system, and use stochastic simulations to empirically obtain error distributions. To supplement this computationally expensive approach, we develop an explicit and computable time-evolving uncertainty distribution for stochastic differential equations with small multiplicative noise. For any initial condition, we rigorously establish convergence bounds for all moments of the deviation of the stochastic solution from its linearised counterpart. This result extends previous work, that showed the convergence of the Kullback-Leibler divergence. We provide a characterisation of the Gaussian distribution that is the solution to the linearised equation, expressed explicitly in terms of solutions to a reference deterministic model. This characterisation provides a practical framework for quantifying uncertainty in deterministic differential equation models, with applications including oceanographic and atmospheric modelling, data assimilation and Lagrangian coherent structure extraction

4.3 Introduction

Many phenomena across geophysical, biological and socio-economic applications can be modelled using a continuous-time dynamical system, i.e., an ordinary differential equation (Brauer and Castillo-Chávez 2012, Tél et al. 2005, Wiggins 2005, e.g.). Given initial values of a multi-dimensional state variable, such equations can be solved numerically to predict the state at future times. The governing dynamics may be specified using existing phenomenological models, but in modern applications these are usually supplemented or driven by observed data. Standard examples include the modelling of weather using available data Law et al. (2015), Reich and Cotter (2015), and predicting concentrations of, for instance, temperature, pollutants or phytoplankton in the ocean using observed current velocity data Abascal et al. (2009), d'Ovidio et al. (2010).

All methods using this approach have uncertainties in the model specification arising from a variety of sources Fang et al. (2020): the model not capturing all phenomena because of the inevitable lack of a complete understanding of all processes involved, errors in measured data, information only available on spatio-temporal grids (resolution error), etc. In the absence of any other understanding of these multitudinous issues, a well-established way of tackling such uncertainties in the model is to think of these as stochastic Berner et al. (2017), Øksendal (2003). Running many realisations of stochastic perturbations to the deterministic model can generate statistics to improve predictions and estimate their associated uncertainties (Badza et al. 2023, Collins 2007, e.g.). However, in practice a very large number of simulations is necessary to generate convergent statistics Feppon and Lermusiaux (2018), Leutbecher (2019). Thus, numerically solving such stochastic systems – potentially with data-based terms – is often computationally expensive, and does not necessarily provide conceptual insight into how the model uncertainties affect predictions.

Clearly, possessing a broader theoretical understanding of how stochastic terms impact continuous dynamical systems would be valuable. Stochastic differential equations (SDEs) provide a natural framework for introducing uncertainty, as a noise process, into the continuous time evolution of a variable. Generally, in modelling situations the dynamics are highly nonlinear and one expects the noise to be multiplicative (i.e. vary with both state and time), e.g. in atmospheric Sura (2003), Sura et al. (2005) and oceanic Kamenkovich et al. (2015) systems and from experimental and observational considerations. Such SDEs are intractable to solve analytically Øksendal (2003) and computationally expensive to simulate accurately Mora et al. (2017). Having a data-based model—that is, possessing terms in the equations which are driven by data rather than by explicitly specified functions—renders additional problems in obtaining a theoretical understanding of the prediction error.

A common intuitive approach to characterising the uncertainty arising from an otherwise analytically intractable nonlinear SDE is via a multivariate Gaussian approximation, which is used across a diversity of literature. For instance, one can formally "linearise" the

SDE in some sense to obtain a Gaussian density, and this approach is used in filtering theory Jazwinski (2014). Other approaches first assume a Gaussian distribution and obtain formal computations for its mean and covariance Särkkä and Solin (2019). However, both approaches lack rigorous justification and a precise understanding of *how* the Gaussian distribution arises from the nonlinear SDE. This is particularly the case when the noise is multiplicative, which is a situation that is often ignored but necessary in practice. Sanz-Alonso and Stuart Sanz-Alonso and Stuart (2017) partially addressed these issues, by showing that the Kullback-Leibler (KL) divergence between the solutions of autonomous SDEs with additive noise and a linearised equivalent can be bounded by the scale of the noise. In this manuscript, we relax the hypotheses of Sanz-Alonso and Stuart (2017) to cater for time-dependent coefficients and for multiplicative noise. Furthermore, our result explicitly establishes the convergence rate of all moments of the deviation considered in Sanz-Alonso and Stuart (2017), which cannot be inferred from the KL divergence alone.

In this paper, we remedy this deficiency by proving rigorously that the noise-scaled deviation between the SDE solution and a reference deterministic solution converges towards a multivariate Gaussian distribution, in the limit of small noise. We consider a general class of SDEs with fully non-autonomous terms and multiplicative noise. The Gaussian distribution arises as the solution to a formal linearisation of the SDE about a deterministic trajectory (in the absence of noise). By bounding all raw moments of the difference between the SDE and the linearised solutions by the noise scale (see Theorem 4.4.1), we show that the stochastic deviation converges in distribution to a multivariate Gaussian random variable (see Theorem 4.4.2). The covariance matrix characterising this Gaussian can be explicitly written in terms of the flow map of the underlying deterministic system and the (potentially spatio-temporally varying) diffusion matrix, and is available even if the deterministic model is only available via data. The Gaussian distribution is consistent with that seen in other literature Jazwinski (2014), Sanz-Alonso and Stuart (2017), Särkkä and Solin (2019), while we additionally show convergence of *all* the moments of the deviation distribution. The results hold independently of the initial condition and for all finite times; the uncertainty evolution of any deterministic trajectory with time is therefore encapsulated in our results.

The quantification of prediction uncertainty that we present here was originally motivated by the "stochastic sensitivity" approach of Balasuriya Balasuriya (2020a). In the context of two-dimensional, unsteady fluid flow, stochastic sensitivity works with Eulerian velocity data as the underlying deterministic model, and seeks to quantify the uncertainty in an eventual Lagrangian trajectory location. This methodology was developed as a tool for determining Lagrangian coherent structures (LCS) Balasuriya et al. (2018), Hadjighasem et al. (2017) in fluid flows, in that clusters of trajectories which have small uncertainty may be thought of as more "coherent" than others. In particular, Balasuriya (2020a) derived the limiting mean and variance of the noise-scaled deviation, and provided computable expressions in terms of the deterministic flow map and velocity

field. However, this was restricted to two-dimensional systems and did not characterise the limiting distribution itself.

The contributions of this work are:

- In Section 4.4, we prove rigorously that all moments of the noise-scaled solution to a multidimensional stochastic differential equation with non-autonomous coefficients and multiplicative noise converges towards those of a linearised SDE, in the limit of small noise. The Gaussian distribution solving the linearised SDE appears in other literature and applications but often lacks justification Jazwinski (2014), Särkkä and Solin (2019). On the other hand, when the linearisation is justified, this is disregarding time-dependence in the coefficients and multiplicative noise Sanz-Alonso and Stuart (2017).

- We present characterisations of the limiting Gaussian distribution in terms of gradients of *either* the velocity field (as an ODE consistent with that arising elsewhere Jazwinski (2014), Sanz-Alonso and Stuart (2017), Särkkä and Solin (2019)) or the deterministic flow map. The latter is an alternative characterisation that allows the Gaussian distribution to be computed *entirely from the solution dynamics* of a determinstic model and specification of any multiplicative noise effects known prior.

- In Section 4.5, we generalise the two-dimensional stochastic sensitivity approach of Balasuriya (2020a) to arbitrary dimensions. Our expressions enable the computation of stochastic sensitivity in any dimension, as a scalar measure of uncertainty about any solution trajectory of the deterministic model. This also extends stochastic sensitivity as a means of Lagrangian coherent structure extraction to fluid flows of arbitrary dimension.

- In Section 4.6, we validate the results of Section 4.4 using stochastic simulations from a 2-dimensional model. In particular, we demonstrate that the first four moments of the distance between the realisations and the Gaussian limit follow the predicted bound. We also illustrate a key prediction from Section 4.5; that the computable covariance matrix of the Gaussian limit captures the time-evolution of uncertainty, even when the noise is multiplicative.

This work is relevant to the well-known "stochastic parameterisation" approach in weather and climate modelling, in which stochastic components are introduced to account for unresolved subgrid effects Berner et al. (2017), Leutbecher et al. (2017), Palmer (2019). Since this work is fundamental, in establishing a convergence result for a general class stochastic differential equations, we do not explicitly describe how to construct an appropriate stochastic parameterisation (e.g. specification of the coefficients of the SDE). Instead, we expect that this result will be useful in the analysis of stochastic parameterisations, and to convert otherwise computationally expensive schemes into an efficient

approximations, a goal explicitly identified in Leutbecher et al. (2017). We also expect that this work will find application in data assimilation Budhiraja et al. (2019), Jazwinski (2014), Law et al. (2015), Reich and Cotter (2015), as a means of characterising forecast uncertainty. The original stochastic sensitivity tools have been applied to identify Lagrangian coherent structures (LCSs) in 2-dimensional fluid flow Badza et al. (2023), Balasuriya (2020a). By extending the theory of these tools into arbitrary dimensions, our results can also be used to extract coherent structures in n-dimensional flows. These potential applications are discussed in **??**.

4.4 The Gaussian limit

Suppose we are interested in the evolution of a $\mathbb{R}^n$-valued state variable y_t over a finite time interval $[0, T]$. Our model, accounting for uncertainties arising from a range of sources, for the evolution of this variable is the Itô stochastic differential equation

$$\mathrm{d}y_t^{(\varepsilon)} = u\left(y_t^{(\varepsilon)}, t\right)\mathrm{d}t + \varepsilon\,\sigma\left(y_t^{(\varepsilon)}, t\right)\mathrm{d}W_t, \tag{4.1}$$

where $u : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ is the governing reference vector field, and can be inferred from underlying physics or available data, for instance. The canonical n-dimensional Wiener process W_t is a continuous white-noise stochastic process with independent Gaussian increments. The scale of the noise is parameterised as $0 < \varepsilon \ll 1$ and $\sigma : \mathbb{R}^n \times [0, T] \to \mathbb{R}^{n\times n}$ is a deterministic diffusion matrix. The noise in (4.1) is multiplicative, in that the diffusion matrix σ can vary with both state and time. We assume that σ is specified *a priori*, or if no such information is known, then $\sigma \equiv I_n$, the $n \times n$ identity matrix, is a default choice. We assume certain generic smoothness and boundedness conditions on the various functions outlined; these are stated explicitly in Hypothesis H in Appendix B.1 and ensure the existence of unique solutions to (4.1). The stochastic solution $y_t^{(\varepsilon)}$ to (4.1) describes the evolution of the state variable through time, accounting for ongoing uncertainty with noise-scale ε.

In the absence of any uncertainty (i.e. $\varepsilon = 0$), (4.1) reduces to the ordinary differential equation

$$\frac{\mathrm{d}w_t}{\mathrm{d}t} = u\left(w_t, t\right). \tag{4.2}$$

Let the flow map $F_0^t : \mathbb{R}^n \to \mathbb{R}^n$ be the function which evolves an initial condition from time 0 to time t according to the flow of (4.2). We refer to (4.2) as the *reference* deterministic model associated with (4.1) in that it either demonstrates the dominant physics (as would be the case if we think of the noise in (4.1) as capturing stochastic parameterisation) or is the best-available model (for example if u is available from data, and (4.1) represents the uncertainty of such data). Solutions to the reference deterministic model are more readily available, e.g. in terms of computational efficiency when solving

numerically, than those of the stochastic model, but do not account for inevitable uncertainty. Here, we establish a Gaussian characterisation and approximation of the solution to (4.1) constructed from the deterministic flow map, thereby taking advantage of the easily available solutions to (4.1) and avoiding the need for computationally expensive stochastic simulation.

To characterise the uncertainty, we fix the *identical* initial condition $x \in \mathbb{R}^n$ to *both* the stochastic model (4.1) and the reference deterministic model (4.2), and consider their evolution in time. We will show that the stochastic deviation between solutions of these can be characterised exactly in terms of a Gaussian in the limit of small noise, i.e. $\varepsilon \to 0$. This quantifies the time-evolving uncertainty of predictions from the deterministic model (4.2). We provide explicit analytical expressions for the limiting distribution, written in terms of the flow map of the deterministic system and σ, thereby providing strong theoretical insight while nullifying the need to perform expensive SDE simulations in approximating such a distribution.

To express our results, we define the noise-scaled deviation

$$z_t^{(\varepsilon)}(x) \coloneqq \frac{y_t^{(\varepsilon)} - F_0^t(x)}{\varepsilon}, \quad z_0^{(\varepsilon)}(x) = 0, \tag{4.3}$$

where $x \in \mathbb{R}^n$ is fixed and certain, and satisfies $y_0^{(\varepsilon)} = x$. We wish to understand the limiting behaviour of $z_t^{(\varepsilon)}(x)$ as ε approaches zero.

Theorem 4.4.1 (All moments are bounded) *Fix $x \in \mathbb{R}^n$ and let $z_t(x)$ be the solution to the linearised SDE*

$$\mathrm{d}z_t(x) = \nabla u\left(F_0^t(x), t\right) z_t(x)\,\mathrm{d}t + \sigma\left(F_0^t(x), t\right)\mathrm{d}W_t, \quad z_0(x) = 0, \tag{4.4}$$

where W_t is the same *Wiener process driving (4.1). Then for any $r \geq 1$ and $t \in [0, T]$, there exists a $D_r(t) \in [0, \infty)$ independent of x such that for all $\varepsilon > 0$,*

$$\mathbb{E}\left[\left\| z_t^{(\varepsilon)}(x) - z_t(x)\right\|^r\right] \leq D_r(t)\,\varepsilon^r, \tag{4.5}$$

where $\|\cdot\|$ is the Euclidean norm.

Proof. See Appendix B.2. Showing the result employs the Burkholder-Davis-Gundy inequality, Grönwall's inequality, Taylor's theorem and the bounds placed on the SDE coefficients, to explicitly construct the bounding coefficient $D_r(t)$. □

Taking the limit as ε approaches 0 in (4.5) shows that $z_t^{(\varepsilon)}(x)$ converges in rth moment to $z_t(x)$, which in turn implies convergence in probability and convergence in distribution (or weak convergence). It is important to note that the stochastic differential equation (4.1) and the linearised equation (4.4) must be defined with the *same* Wiener process W_t for Theorem 4.4.1 to hold as stated. However, by weakening the convergence we can think of $z_t^{(\varepsilon)}(x)$ as converging to a Gaussian distribution (the distribution of the linearised solution) with no reference to the driving Wiener process.

Theorem 4.4.2 (Explicit Gaussian limit) *For any $x \in \mathbb{R}^n$ and $t \in [0,T]$,*

$$z_t^{(\varepsilon)}(x) \xrightarrow{d} \mathcal{N}\big(0, \Sigma(x,t)\big) \quad \text{as} \quad \varepsilon \to 0,$$

where the covariance matrix Σ is given by

$$\Sigma(x,t) = \int_0^t L\left(x,t,\tau\right) L\left(x,t,\tau\right)^{\intercal} \mathrm{d}\tau, \tag{4.6}$$

with

$$L\left(x,t,\tau\right) \coloneqq \nabla F_0^t(x) \left[\nabla F_0^\tau(x)\right]^{-1} \sigma\left(F_0^\tau(x),\tau\right). \tag{4.7}$$

Moreover, the covariance matrix Σ is the matrix solution to the ordinary differential equation

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = \Big[\nabla u\left(F_0^t(x),t\right)\Big]\Sigma + \Sigma\Big[\nabla u\left(F_0^t(x),t\right)\Big]^{\intercal} + \sigma\left(F_0^t(x),t\right)\sigma\left(F_0^t(x),t\right)^{\intercal}, \tag{4.8}$$

subject to $\Sigma(x,0) = O$, the $n \times n$ zero matrix.

Proof. See Appendix B.3. The Gaussianity of the limiting process $z_t(x)$, and therefore the limit in distribution of $z_t^{(\varepsilon)}(x)$, is first established, and then the explicit expression for the covariance matrix Σ is derived by employing Itô's isometry and properties of the flow map. □

The covariance matrix Σ uniquely characterises the limiting Gaussian distribution in Theorem 4.4.2, and captures the impact of both the deterministic dynamics of the model (through the flow map gradients) and multiplicative noise (by evaluating the diffusion matrix σ along the deterministic trajectory). Through (4.6) and (4.7), the Gaussian distribution can be computed entirely from flow map data and specification of σ, without any reference to the governing vector field u in (4.2). Alternatively, if the velocity gradients ∇u are available, then Σ can be computed as the solution to (4.8). Solving (4.2) and (4.8) jointly describes the scheme for computing the Gaussian approximation seen elsewhere, e.g. Algorithm 9.4 of Särkkä and Solin (2019) or Equations (1.2) and (1.3) of Sanz-Alonso and Stuart (2017). Moreover, for a fixed time $t \in [0,T]$, Theorem 4.4.2 justifies the approximation

$$y_t^{(\varepsilon)} \dot{\sim} \mathcal{N}\left(F_0^t\left(x\right), \varepsilon^2 \Sigma\left(x,t\right)\right), \tag{4.9}$$

for small values of ε.

The theoretical results in this section, and the computability of the limiting distribution will be verified with numerical simulation of an example in Section 4.6. This theory has many applications and extensions which are discussed in **??**.

4.5 Extending stochastic sensitivity

The covariance matrix Σ provides a direct extension of the stochastic sensitivity tools first introduced by Balasuriya Balasuriya (2020a) for the fluid flow context. Here, the deterministic model (4.2) is seen as a "best-available" model for the evolution of trajectories, and the driving vector field u is the Eulerian velocity of the fluid. Stochastic sensitivity ascribes a scalar value to each deterministic trajectory by computing the maximum variance of the scaled deviations, when projected onto rays emanating from the origin Balasuriya (2020a). The natural restating of this original definition of stochastic sensitivity Balasuriya (2020a) in the n-dimensional setting is as follows:

Definition 4.5.1 (Stochastic sensitivity in $\mathbb{R}^n$) *The* stochastic sensitivity *is a scalar field* $S^2 : \mathbb{R}^n \times [0,T] \to [0,\infty)$ *given by*

$$S^2(x,t) := \limsup_{\varepsilon \downarrow 0} \left\{ \mathbb{V}\left[p^\intercal z_t^{(\varepsilon)}(x) \right] \ : \ p \in \mathbb{R}^n, \|p\| = 1 \right\}.$$

Definition 4.5.1 is in the spirit of principal components analysis Jolliffe (2002), performing a dimension reduction by projecting onto the direction in which the variance is maximised, thus capturing the most uncertainty in the data with a scalar value.

The anisotropic uncertainty in two-dimensions Balasuriya (2020a) is the direction-dependent projection (prior to optimising over all directions in Definition 4.5.1). Explicit theoretical expressions for both the stochastic sensitivity and the anisotropic sensitivity in two dimensions were obtained by Balasuriya Balasuriya (2020a); these allowed for quantifying certainty in eventual trajectory locations without having to perform stochastic simulations. We show here that our results in n-dimensions are a generalisation of the two-dimensional ones in Balasuriya (2020a), which moreover establish Gaussianity as well as an explicit expression for the uncertainty measure. A theoretically pleasing and computable expression for the stochastic sensitivity is obtainable:

Theorem 4.5.1 (Computation of S^2) *For any $x \in \mathbb{R}^n$ and $t \in [0,T]$,*

$$S^2(x,t) = \left\| \Sigma(x,t) \right\|, \tag{4.10}$$

where the covariance matrix Σ is defined in (4.6) and $\|\cdot\|$ denotes here the spectral norm induced by the Euclidean norm. Equivalently, $S^2(x,t)$ is given by the maximum eigenvalue of $\Sigma(x,t)$.

Proof. See Appendix B.4. This result uses Theorem 4.4.1 to establish the convergence of the covariance matrices, and then the properties of the spectral norm to establish (4.10). □

The stochastic sensitivity field can be calculated given any velocity data u, and through the explicit expression (4.6) for Σ can even be computed from only flow map data. Computation does not require knowledge of the noise scale ε, so the S^2 field is intrinsic in capturing the impact of the model dynamics on uncertainty, and any specified non-uniform diffusivity.

It has already been shown that, in the fluid flow context, stochastic sensitivity can identify coherent regions in two-dimensions Badza et al. (2023), Balasuriya (2020a). A simple approach is to define robust sets, which are those initial conditions for which the corresponding S^2 value, i.e., the uncertainty in eventual location, are below some specified threshold. This threshold can be defined precisely in terms of a spatial lengthscale of interest and the advective and diffusive characteristics of the flow, as Definition 2.9 of Balasuriya (2020a). Such a definition extends to the n-dimensional case as presented here, moreover establishing an easily computable method for determining coherent sets by using the covariance matrix Σ.

Independent of the fluid mechanics context, Theorem 4.5.1 indicates that even for general systems, the matrix norm of $\Sigma(x,t)$, i.e., the stochastic sensitivity $S^2(x,t)$, can be used as *one* number which encapsulates the uncertainty of an initial state x after t time units.

4.6 Numerical validation and applications

This section will validate the theory presented in Section 4.4. Following the example in Chapter 5 of Samelson and Wiggins (2006), we consider an unsteady meandering jet in two dimensions, which may serve as an idealised model of geophysical Rossby waves. The velocity field for $y \equiv (y_1, y_2)$ is given by Samelson and Wiggins (2006)

$$u\left(y,t\right) = \begin{bmatrix} c - A\sin\left(Ky_1\right)\cos\left(y_2\right) + \epsilon_{\mathrm{mj}} l_1 \sin\left(k_1\left(y_1 - c_1 t\right)\right)\cos\left(l_1 y_2\right) \\ AK\cos\left(Ky_1\right)\sin\left(y_2\right) + \epsilon_{\mathrm{mj}} k_1 \cos\left(k_1\left(y_1 - c_1 t\right)\right)\sin\left(l_1 y_2\right) \end{bmatrix}. \tag{4.11}$$

The velocity field describes a kinematic travelling wave with deterministic oscillatory perturbations in a co-moving frame. Here, A is the amplitude and c is the phase speed of the primary wave, and K is the wavenumber in the y_1-direction. The oscillatory perturbation has amplitude ϵ_{mj}, phase speed c_1 (in the co-moving frame), and wavenumbers k_1 and l_1 in the y_1- and y_2-directions respectively. Throughout, we take the parameter values $c = 0.5$, $A = 1$, $K = 4$, $l_1 = 2$, $k_1 = 1$, $c_1 = \pi$, and $\epsilon_{\mathrm{mj}} = 0.3$. For these values, the flow consists of a meandering jet with vortex structures within the meanders, and a chaotic zone which influences the fluid transfer between the jet and the vortices. All necessary flow map data is obtained by directly solving (4.2) numerically, with the standard Euler scheme. The flow map gradients required for computing the covariance matrix with (4.6) are calculated via a star-grid finite-difference approximation, using a spatial resolution of 0.001.

All simulations in this section were generated using the Julia programming language Bezanson et al. (2017), with implementations of the ordinary and stochastic differential equation solvers provided by the DifferentialEquations.jl package Rackauckas and Nie (2017). All figures were created using the Makie.jl package Danisch and Krumbiegel (2021). The code is available as open source[1].

4.6.1 Validation of Theorem 4.4.1

This section will validate the bound in Theorem 4.4.1 directly, and illustrate the convergence of the SDE solution towards the expected Gaussian distribution described in Theorem 4.4.2. For each value of ε considered, we use the Euler-Maruyama method Kloeden and Platen (1992) to generate $N = 10000$ independent realisations of the solutions to (4.1) and (4.4). A step size of $\delta t = 10^{-4}$ is used to ensure that numerical error does not dominate over the theoretical predictions. These solution samples are generated with the *same* realisations of the Wiener process increments $W_{t+\delta t} - W_t \sim \mathcal{N}(0, \delta t I_n)$. We consider the initial condition $x = (0, 1)$ and the prediction of the model at time $t = 1$. For each realisation of $y_t^{(\varepsilon)}$, a corresponding realisation of the scaled deviation $z_t^{(\varepsilon)}(x)$ is computed. In the following, let $\hat{z}_1^{(\varepsilon)}, \ldots, \hat{z}_N^{(\varepsilon)}$ and $\hat{z}_1, \ldots, \hat{z}_N$ denote the N realisations of $z_t^{(\varepsilon)}(x)$ and $z_t(x)$ respectively.

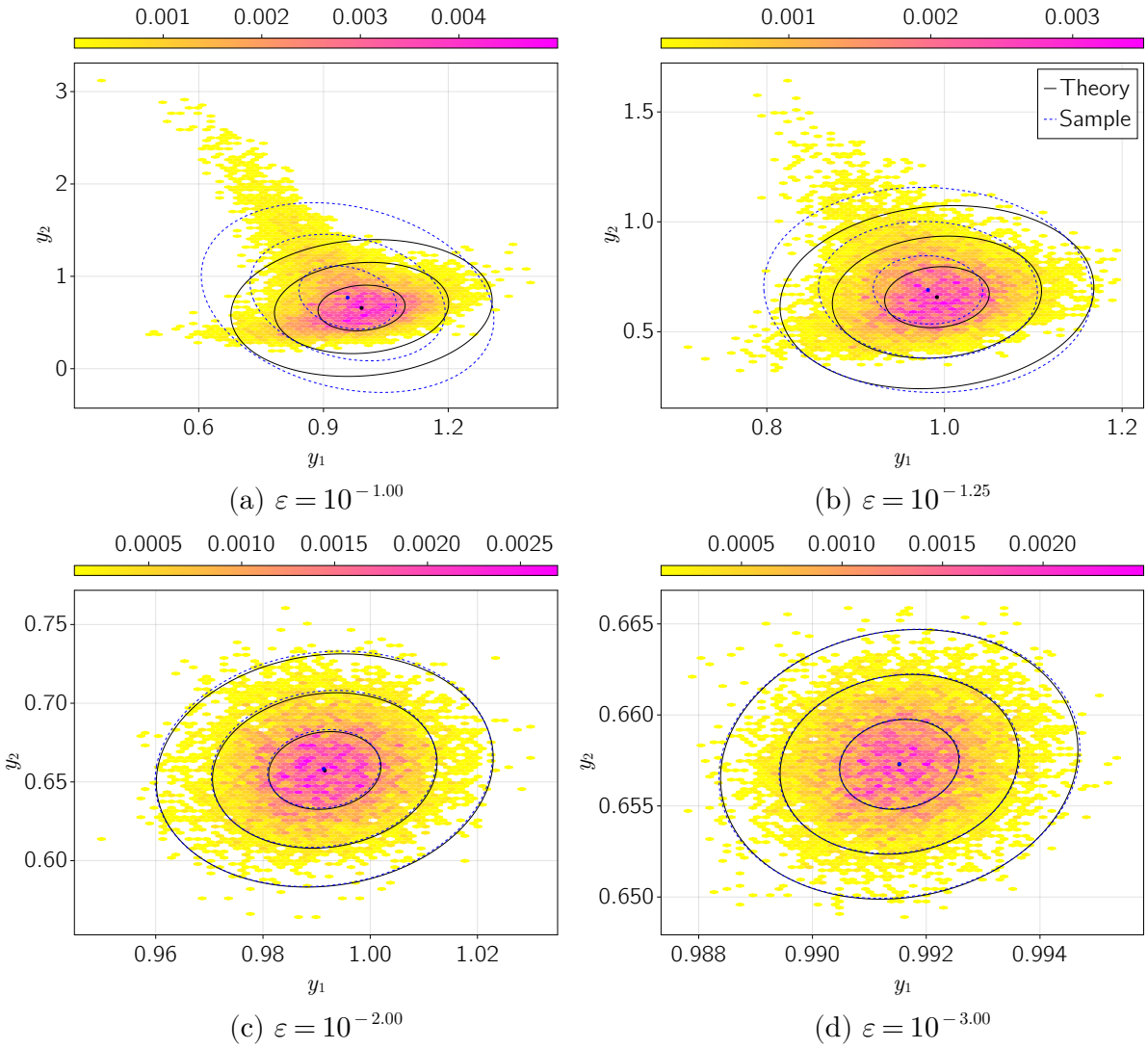
Figure 4.1 shows the resulting simulations of $y_t^{(\varepsilon)}$ for four different values of ε. The realisations are binned as a histogram and bin counts are normalised, to provide an empirical estimate of the probability density function of $y_t^{(\varepsilon)}$. Superimposed (in solid black) are the first, second and third standard-deviation contours of the probability density function of the Gaussian approximation (4.9). The first three standard-deviation levels of the 2×2 sample covariance matrix of the realisations, are also overlaid (in dashed blue). As ε decreases towards 0, the samples increasingly resemble a Gaussian distribution, and both the mean and covariance coincide with the corresponding limits.

To directly validate Theorem 4.4.1 for $r \geq 1$, define the error metric

$$\Gamma_z^{(r)}(\varepsilon) := \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{z}_i^{(\varepsilon)} - \hat{z}_i \right\|^r,$$

which is an estimator of the right-hand side of (4.5). For $r = 1, 2, 3, 4$, $\Gamma_z^{(r)}(\varepsilon)$ is shown (in a logarithmic scale) for decreasing values of ε in Figure 4.2. Theorem 4.4.1 predicts that $\log_{10}\left(\Gamma_z^{(r)}(\varepsilon)\right)$ should decay linearly with a slope greater than r as ε decreases to zero. The lines of best fit for each value of r in Figure 4.2 show this behaviour, and are therefore consistent with Theorem 4.4.1.

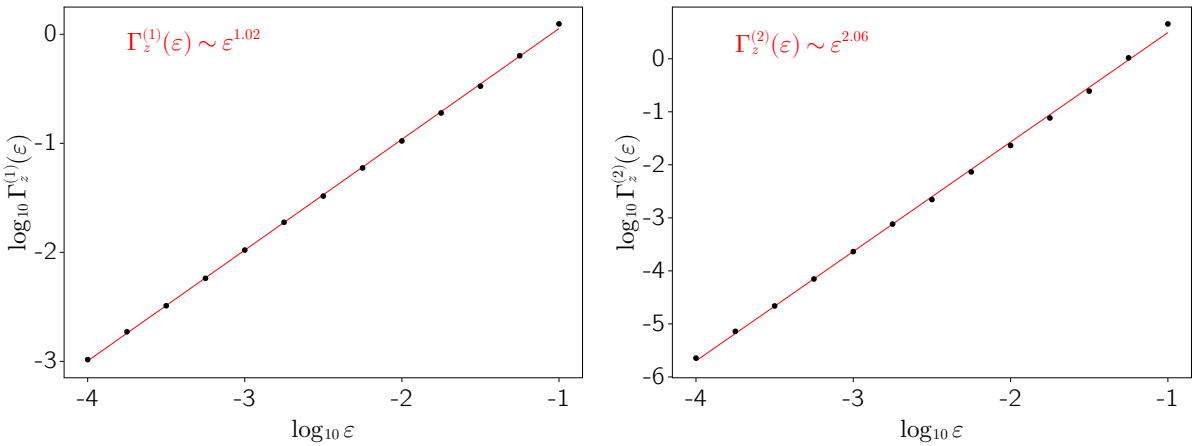
[1] at https://github.com/liamblake/explicit-gaussian-characterisation-uncertainty.



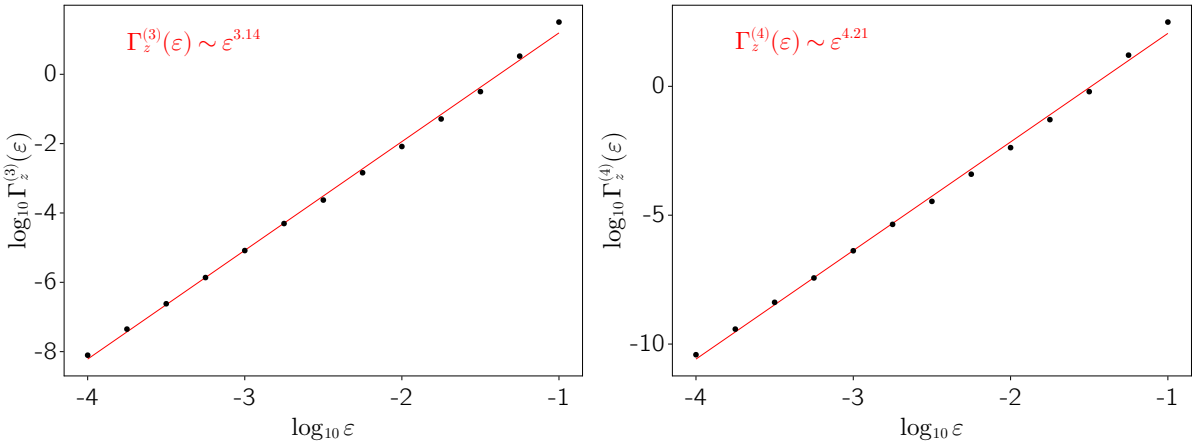
(a) $\varepsilon = 10^{-1.00}$ (b) $\varepsilon = 10^{-1.25}$

(c) $\varepsilon = 10^{-2.00}$ (d) $\varepsilon = 10^{-3.00}$

Figure 4.1: Histograms of $y_t^{(\varepsilon)}$ from direct simulation of (4.1), for four different ε values. Overlaid in black are contours of the Gaussian limit (4.9), which correspond to the first three standard deviation levels centred at the limiting mean $F_0^t(x)$. In dashed blue are corresponding contours computed from the sample covariance matrix of the realisations.



(a) $r = 1$ (mean) (b) $r = 2$ (covariance)



(c) $r = 3$ (skewness) (d) $r = 4$ (kurtosis)

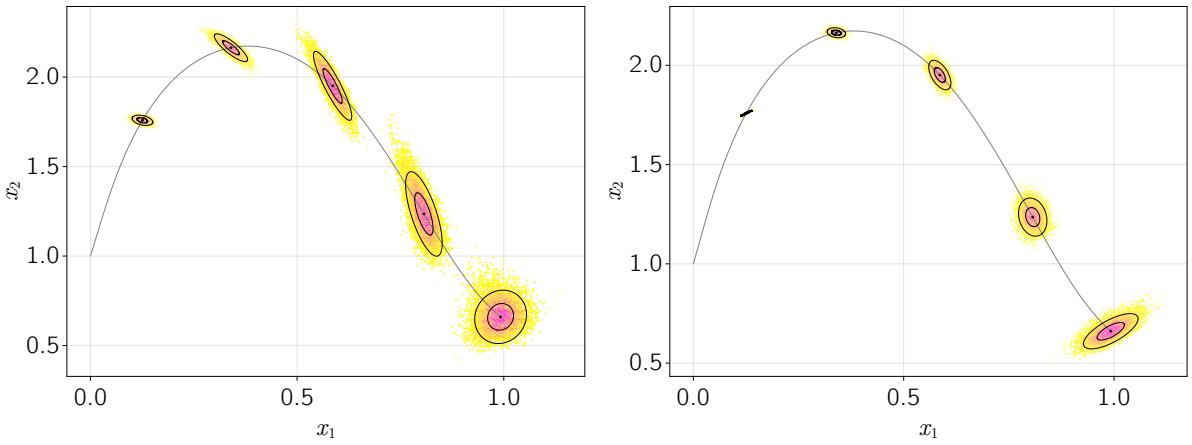
Figure 4.2: Validation of Theorem 4.4.1, by plotting the sample rth raw moment distance (the error metric $\Gamma_z^{(r)}(\varepsilon)$) between 10000 realisations of $z_t^{(\varepsilon)}(x)$ and $z_t(x)$, for decreasing values of ε. A line of best fit (in red) is placed on each, and the resulting slope indicated.

4.6.2 The evolution of $\Sigma(x,t)$ through time

Here we shall illustrate that the limiting covariance matrix Σ captures the time-evolution of model uncertainty. Consider the same meandering jet model in (4.11), with the parameter values used in the previous subsection. We fix the noise scale parameter at $\varepsilon = 0.03$ and consider the evolution of the stochastic solution to (4.1) and the limiting Gaussian distribution (4.9) for times t in the interval $[0,1]$. We also consider two different choices of the diffusion matrix σ: the identity as before, and

$$\sigma_M(x) := \begin{bmatrix} x_1 & 0.5 \\ x_1 & 0.5\,(x_1 + x_2) \end{bmatrix}, \tag{4.12}$$

to include both multiplicative and non-diagonal noise which grow for larger values of the coordinates.



(a) $\sigma(x,t) = I_2$ (b) $\sigma(x,t) = \sigma_M(x)$

Figure 4.3: Histograms of $y_t^{(0.03)}$ for (from left to right) $t = 0.2, 0.4, 0.6, 0.8, 1.0$, with the time-evolution of the deterministic trajectory in grey and contours of the limiting covariance matrix (4.6) for each time. The right figure uses the diffusion matrix $\sigma_M(x)$ as defined in (4.12).

Figure 4.3 plots histograms of realisations of the solution to (4.1) at several different times, evolving from the same initial condition $x = (0,1)$, and the time-evolution of the corresponding deterministic trajectory solving (4.2). Overlaid on each histogram are the contours of the limiting Gaussian distributions, computed entirely from covariance matrix (4.6). Although each distribution is non-Gaussian, the evolution of the uncertainty distribution through time is captured by Σ. For examples, features of the error distributions, such as stretching and rotation, are reflected in Σ. This remains the case even

when the noise is multiplicative with $\sigma = \sigma_M$. The computation of Σ circumvents the need for expensive Monte-Carlo simulation to draw conclusions about such evolution of uncertainty.

4.7 Discussion

This paper has contributed a rigorous justification, in terms of error bounds and a small-noise limit, for an easily computable linearisation approximation to the solution of nonlinear stochastic differential equations, as seen across diverse places in the literature (Jazwinski 2014, Sanz-Alonso and Stuart 2017, Särkkä and Solin 2019, e.g.). The theory applies to fully non-autonomous SDEs with multiplicative noise. This result extends the convergence bound on the Kullback-Leibler divergence by Sanz-Alonso and Stuart Sanz-Alonso and Stuart (2017) to an explicit bound on the convergence of all moments of the difference between the exact SDE solution and the approximation, and further establishes the exact Gaussian distribution in the small-noise limit. While it is known that convergence of the KL divergence leads to convergence of the moments Lu et al. (2017), this manuscript provides the exact rate of that convergence. Our bound is verified numerically by plotting the first four raw moments of the distance between the true noise-scaled solution and the linearised solution (see Figure 4.2). The results, plotted across three orders of magnitude of the small noise parameter, match our theoretical prediction exactly.

In addition, we described a framework in which uncertainty in deterministic models can be ascribed without the need for expensive stochastic simulation, and purely from the deterministic solution dynamics. We illustrated how the Gaussian limit reflects the time-evolution of uncertainty (see Figure 4.3), even when the true uncertainty distributions are themselves non-Gaussian.

A powerful advantage of this framework is that the diffusivity matrix σ is permitted to vary spatio-temporally, allowing for multiplicative noise. Multiplicative noise is often ignored in practice, due to difficulties in working with analytically (see, for instance, the prior lack of rigourous justification of linearisations when the noise is multiplicative) and generating numerical realisations efficiently (e.g. the review in Mora et al. (2017)). It has also been shown that multiplicative noise on linear dynamics can model departures from Gaussianity observed in climate statistics Sura et al. (2005), as opposed to nonlinear dynamics with only additive noise. The spatio-temporal dependence of σ can also capture experimental and observational considerations that are otherwise ignored in the deterministic model, such as cloud cover when using satellite measurements or nonuniform uncertain across the field of view of a camera. We therefore present a highly flexible framework that can capture any prior knowledge of non-uniform uncertainty that arises from modelling or experimental considerations.

This paper also supplied theoretical and computational extension to the "stochastic sensitivity" tools introduced by Balasuriya Balasuriya (2020a). Stochastic sensitivity was

hitherto derived as the variance of an unknown limiting distribution and could only be computed in two spatial dimensions: we established that stochastic sensitivity, in any number of dimensions, is computable as the operator norm of the covariance matrix of our limiting SDE. We have also established that the limiting distribution in question is Gaussian, which may provide insight into properties of stochastic sensitivity as a means of uncertainty quantification in any model (not just in the fluids context) where an n-dimensional state variable evolves according to a "best available" model.

The Gaussian approximation presented here arises as the leading order term in a power series expansion of the SDE solution in terms of the noise scale parameter ε Blagoveshchenskii (1962). A further extension would be to explore the higher-order terms in such an expansion, which could lead to a practical framework for constructing higher-order characterisations and approximations of the stochastic solution. However, the higher-order terms are known to be individually non-Markovian, and satisfy non-linear SDEs for which the solution is not expected to be analytically available Blagoveshchenskii (1962).

In this paper, we have assumed throughout that the initial condition x, from which both the stochastic differential equation and the deterministic flow map evolves from, is *certain* (i.e. not a random quantity). However, in practice there is uncertainty associated with the initial state which should also be accounted for. The bound in Theorem 4.4.1 is independent of the initial condition, suggesting that the required extension of the theoretical result is straightforward. This extension will broaden our framework, allowing for uncertainties in *both* the initial state and the time-evolution of the model to be characterised at once in a precise sense.

Similarly, we assume that the reference deterministic model (4.2) for the evolution of the state variable is "correct" and known exactly, in that in the absence of any noise (i.e. $\varepsilon = 0$), the SDE model (4.1) reduces to the deterministic (4.2). The Gaussian characterisation is computed from knowledge of either the driving vector field or the solution data itself, i.e. the flow map. However, these components of the deterministic model may not be known exactly, e.g. from solving (4.2) numerically, interpolation error, etc. There is a need to extend the theory presented here to account for this case; to, for instance, establish a bound in the error between the SDE solution and the limiting Gaussian, as in Theorem 4.4.1, if the Gaussian distribution is constructed from an "incorrect" deterministic model. Both of these theoretical extensions, to uncertain initial conditions and incorrect deterministic dynamics, are currently being pursued.

4.7.1 Applications

Here, we briefly discuss some anticipated applications of this work across a wide range of fields, including climate and ocean modelling, data assimilation and Lagrangian coherent structures.

This work fits in with recent interest in stochastic parameterisation as a means to account for unresolved subgrid effects in climate modelling Berner et al. (2017), Leutbecher

et al. (2017), Palmer (2019). In particular, the recent review Leutbecher et al. (2017) concludes, "The aim of current and future developments in stochastic representations of model uncertainty is to develop schemes that are computationally highly efficient and contribute only moderately to the overall computational cost...". This paper provides one method to convert a stochastic parameterisation (formulated as a SDE) to a computationally cheaper set of coupled ODEs for the mean and variance of an approximate Gaussian, together with a convergence proof and error estimates.

To ascribe uncertainties directly onto the deterministic model, we assume that the diffusivity matrix σ is specified *a priori*, to capture any known multiplicative noise effects. There are methods for estimating σ directly from observed data, e.g. the Bayesian inference approach of Ying et al. (2019) or via statistical estimation as in Cotter and Pavliotis (2009), which can be used in our framework. In particular, Ying et al. (2019) relies upon computationally expensive numerical approximations to compute the likelihood of each trajectory, whereas from this paper we have a potentially more efficient computation, using the analytically available Gaussian limit. Coupling these approaches with the approximation here could provide a complete and practical framework to characterise the uncertainty in the flow by efficiently estimating the (multiplicative) diffusion from observed trajectory data.

Data assimilation is a framework for improving uncertainties in predictions by combining model forecasts with observational data, accounting for error in both, and uncertainty quantification refers to the broader goal of capturing the uncertainty inherent in prediction Budhiraja et al. (2019), Jazwinski (2014), Law et al. (2015), Reich and Cotter (2015). The Gaussian limit here provides a characterisation of model uncertainty, and may therefore be useful in data assimilation and uncertainty quantification. The linearisation of the stochastic differential equation (4.1) used to construct the Gaussian approximation has been employed in data assimilation, e.g. in the continuous time continuous state-space extended Kalman filter (Jazwinski 2014, §9). The convergence analysis of this paper could contribute a new term, estimating the error due to linearisation, to the *forecast uncertainty* covariance matrix employed in these extended Kalman filters.

Stochastic sensitivity provides a novel method for extracting Lagrangian coherent structures (LCSs) Balasuriya et al. (2018), Hadjighasem et al. (2017) from fluid flow, by considering regions with uncertainty (as measured by the stochastic sensitivity field) below a prescribed threshold. Whereas the original formulation in Balasuriya (2020a) was restricted to two-dimensional flows, here we have an extension of the LCS extraction scheme to arbitrary dimensions.

Moreover, most traditional LCS measures are completely deterministic measures, not accounting for any uncertainty in the driving velocity field, and the sensitivity of these methods to such uncertainty has not been investigated in detail. The robustness of several LCS methods to stochastic noise has recently been explored in Badza et al. (2023), but via stochastic simulation and summary statistics. In this paper we have presented a

theoretical result for characterising Lagrangian trajectory uncertainty, which can be used to perform a purely theoretical analysis of such sensitivity in LCS computations. An initial study into the impact of uncertainty of one such method – the finite-time Lyapunov exponent – has already been performed using stochastic sensitivity Balasuriya (2020b), albeit in only two-dimensions and without knowledge that the limiting distribution is Gaussian.

Chapter 5

A Gaussian mixture model

First, we shall make some adjustments to the theory as presented in Chapter 4, by dropping the explicit ε notation and extending the theory to allow for Gaussian initial conditions to our stochastic differential equation.

5.1 The deterministic model versus the stochastic model

In Chapter 4, we provided a rigorous justification that the Gaussian density described in Theorem 4.4.2 provides an approximation/characterisation of the solution to a stochastic differential equation, in the sense of a small-noise limit. The scale of the noise was explicitly parameterised with a non-zero value ε, and we considered the behaviour of solutions in the limit as ε approaches zero. However, in practice there will be a prescribed value of ε, either chosen judiciously from context or informed by data. Henceforth, we shall drop the use of ε and instead consider stochastic differential equations of the form

$$\mathrm{d}y_t = u\left(y_t, t\right)\mathrm{d}t + \sigma\left(y_t, t\right)\mathrm{d}W_t,$$

where, strictly speaking, the noise scale parameter has been included in the diffusion term σ.

$$\frac{\mathrm{d}F_s^t(x)}{\mathrm{d}t} = u\left(F_s^t(x), t\right), \quad F_s^s(x) = x \tag{5.1a}$$

$$\begin{aligned}\frac{\mathrm{d}\Sigma_s^t(x;\Sigma_0)}{\mathrm{d}t} &= \nabla u\left(F_s^t(x), t\right)\Sigma_s^t(x;\Sigma_0) + \Sigma_s^t(x;\Sigma_0)\left[\nabla u\left(F_s^t(x), t\right)\right]^{\intercal} \\ &\quad + \sigma\left(F_s^t(x), t\right)\sigma\left(F_s^t(x), t\right)^{\intercal}, \quad \Sigma_s^s(x;\Sigma_0) = \Sigma_0,\end{aligned} \tag{5.1b}$$

It is also worth noting that the small noise limit can be equivalently thought of, at least heuristically, as a small time limit, using scaling properties of the Wiener process.

5.2 Motivation: The Gulf Stream

A key advantage of the Gaussian limit is the ease of computation; rather than having to generate a large number of realisations of the SDE solution to understand, either qualitatively or for the purposes of inference and estimation, the probability distribution of the solution, we can solve a smaller system of equations (4.8) for the state and covariance simultaneously.

We shall first motivate the importance of stochastic models

5.3 Propagating uncertain initial conditions

5.4 Solving for the state and covariance

To compute the Gaussian limit along a deterministic trajectory, we can solve the system of equations (5.1), providing that the Jacobian ∇u of the vector field is available, or can be approximated appropriately. Since Σ_s^t represents a covariance matrix, it must remain symmetric and positive semi-definite when solving (5.1). However, many standard numerical schemes do not take this into account, so a specialised scheme is required, as described below.

Similar equations of the form (5.1) (although often without dependence on *both* time and the state in the σ term) are solved numerically in other applications, notably when implementing the extended Kalman filter on stochastic differential equation models (Jazwinski 2014, Kulikova and Kulikov 2014). Kulikova and Kulikov (2014) identify that that the two most significant sources of numerical error when solving (5.1) are a) the estimate of the covariance matrix Σ_s^t violates the requirement of positive semi-definiteness, and b) local error propagation in the state equation without an adaptive step size. The state equation (5.1a) is the only non-linear part of (5.1), so

Mazzoni (2008) proposes an efficient hybrid method for solving (5.1) which addresses both difficulties a) and b), and takes advantage of the availability of ∇u. This method, which we shall term the Mazzoni method, combines a Taylor-Heun approximation to solve (5.1a) for the state and a Gauss-Legendre step to solve (5.1b) for the covariance.

Throughout, we use the Mazzoni method to solve (5.1)

The Taylor-Heun formula for the update of the state is then

$$F_s^{t+\Delta t}(x) \approx F_s^t(x) + \left(I - \frac{\Delta t}{2} \nabla u \left(F_s^t(x), t \right) \right)^{-1}. \tag{5.2a}$$

The Gauss-Legendre update of the covariance is

$$\Sigma_s^{t+\delta t}\left(x; \Sigma_0\right) \approx M_\tau \Sigma_s^t\left(x; \Sigma_0\right) M_\tau^\intercal + \Delta t K_\tau \sigma \left(w_\tau,\, t + \frac{\Delta t}{2} \right) \sigma \left(w_\tau,\, t + \frac{\Delta t}{2} \right)^\intercal K_\tau^\intercal, \tag{5.2b}$$

where

$$w_\tau = \frac{1}{2}\left(w_t + w_{t+\Delta t} - \frac{\Delta t^2}{4}\nabla u\left(w_t,\, t\right) u\left(w_t,\, t\right)\right) \tag{5.2c}$$

$$K_\tau = \left[I - \frac{\Delta t}{2}\nabla u\left(w_\tau,\, t + \frac{\Delta t}{2}\right)\right]^{-1} \tag{5.2d}$$

$$M_\tau = K_\tau\left[I + \frac{\Delta t}{2}\nabla u\left(w_\tau,\, t + \frac{\Delta t}{2}\right)\right]. \tag{5.2e}$$

5.5 The GMM algorithm

5.6 Analysis through exact SDE solutions

5.6.1 A linear SDE

Consider an n-dimensional linear stochastic differential equation with additive noise;

$$\mathrm{d}y_t = A(t)y_t\,\mathrm{d}t + B(t)\,\mathrm{d}W_t, \tag{5.3}$$

where $A : [0, T] \to \mathbb{R}^{n\times n}$ and $B : [0, T] \to \mathbb{R}^{n\times m}$ are specified, deterministic matrix-valued functions that are sufficiently smooth and measurable to ensure the existence of solutions, and W_t is an m-dimensional Wiener process.

The Gaussian limit described in Theorem 4.4.2

5.6.2 Benê's SDE

Chapter 6

Applications

6.1 Oceanography

6.1.1 Altimetry-derived velocity data

Suppose we have the sea surface height (SSH) $\eta = \eta(\lambda, \phi, t)$ at longitude λ and latitude ϕ (both in radians), and at time t. The SSH η is then proportional to the streamfunction for the surface flow, if we treat the surface flow as two-dimensional, where the constant of proportionality varies with latitude (Doglioni et al. 2021, Park 2004). The geostrophic zonal (east-west) and meridional (north-south) velocities u and v are then given by

$$u(\lambda, \phi, t) = -\frac{g}{f(\phi)} \frac{\partial \eta}{\partial \phi} \tag{6.1a}$$

$$v(\lambda, \phi, t) = \frac{g}{f(\phi)} \frac{\partial \eta}{\partial \lambda}, \tag{6.1b}$$

where

$$f(\phi) = 2\Omega_{\mathrm{r}} \sin \phi$$

is the Coriolis parameter, $g \approx 9.81\,\mathrm{m\,s^{-1}}$ is the standard acceleration due to gravity, and $\Omega_{\mathrm{r}} \approx 7.2921 \times 10^{-5}\,\mathrm{radians\,s^{-1}}$ is the rotation rate of the Earth.

To account for measurement error, we consider the evolution of the following stochastic model

$$\mathrm{d}\begin{bmatrix} x_t^{(\mathrm{lon})} \\ x_t^{(\mathrm{lat})} \end{bmatrix} = \begin{bmatrix} u\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right) \\ v\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right) \end{bmatrix} \mathrm{d}t + L_r \begin{bmatrix} \sqrt{u_{\mathrm{err}}\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right)} & 0 \\ 0 & \sqrt{v_{\mathrm{err}}\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right)} \end{bmatrix} \mathrm{d}W_t, \tag{6.2}$$

where t is the time in days from DATE, u and v are the interpolated zonal and meridional velocities (in degrees day^{-1}), u_{err} and v_{err} are the respective interpolated error estimates (in degrees day^{-1}), and $L_r = 0.25$ degrees is the spatial resolution of the data.

The derivatives of the deterministic velocity field in (6.2) are approximated via the finite-differences

$$\frac{\partial u\left(x^{(\text{lon})}, x^{(\text{lat})}, t\right)}{\partial x^{(\text{lon})}} \approx \frac{u\,()}{2L_r}$$

6.1.2 The Gulf Stream

A motivating example

To illustrate the importance of stochastic methods, we consider the evolution of a single particle under (6.2) as an example.

6.1.3 The Southern Ocean

6.2 Atmospheric regimes

6.2.1 Multiplicative noise regime

We consider the example consider by Sura et al. (2005), in which a stochastic differential equation linear dynamics but multiplicative noise is used to model the time-evolution of the observed streamfunction of the

The linearity of the deterministic dynamics means that the flow map F_s^t is available analytically, and so we can compute the limiting covariance exactly by using the alternative expression (4.6). However, the observed data itself displays non-Gaussianity, so the introduction of multiplicative noise is necessary to capture this.

6.3 Epidemiology

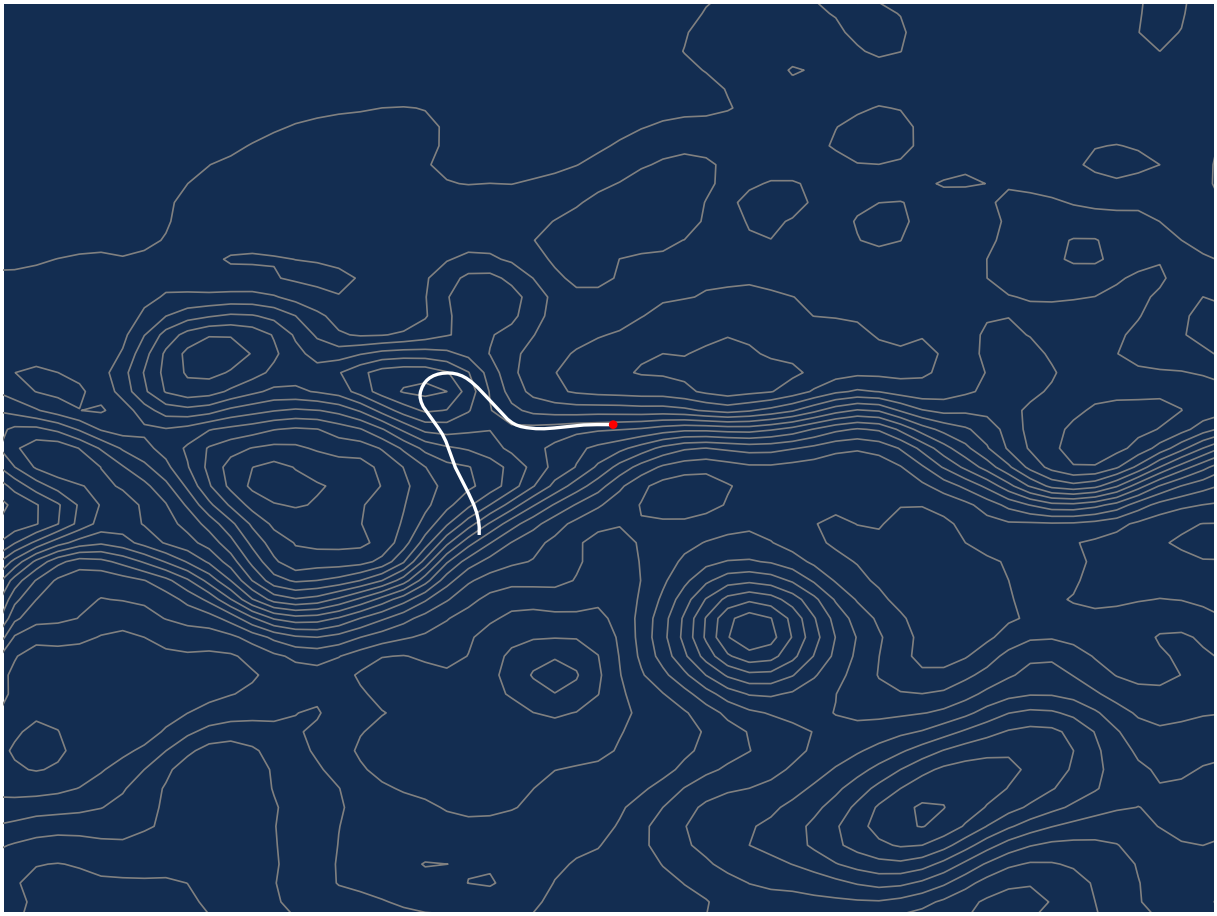
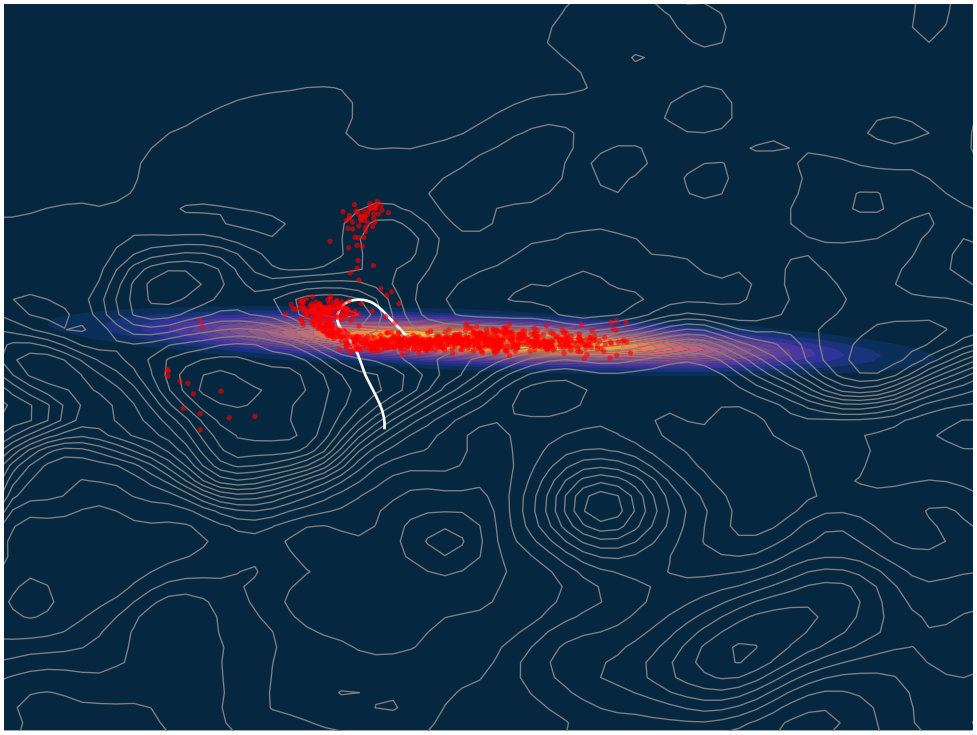
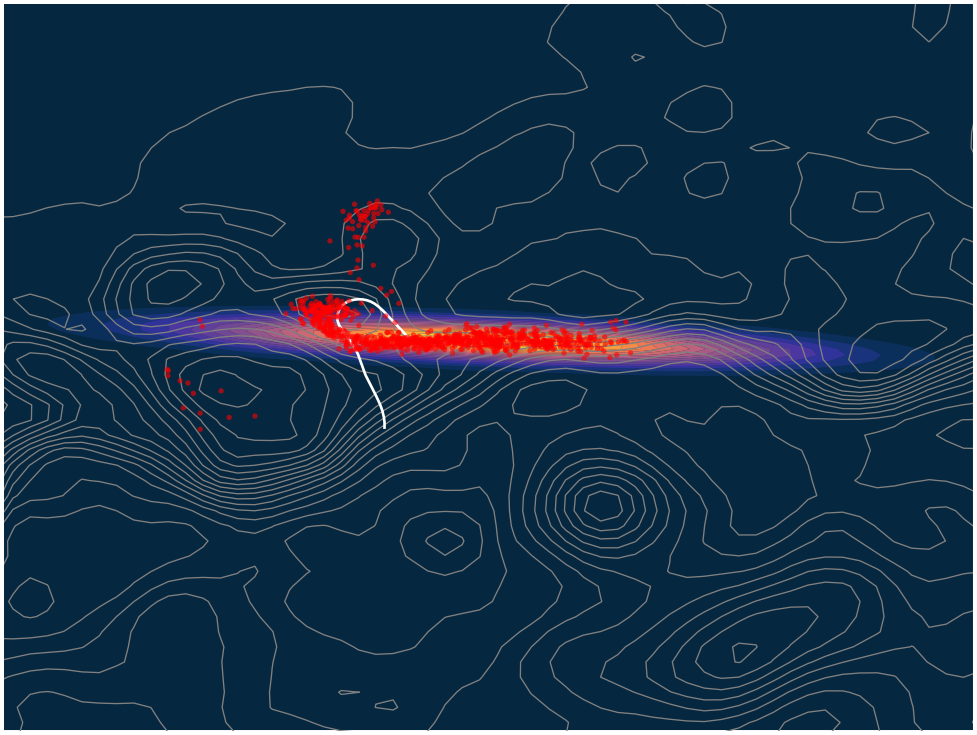


Figure 6.1: The time evolution of the solution (obtained by numerically integrating the velocity data) to the deterministic system corresponding to (6.2), from initial condition $x_0 = (-60.5, 39)$ from 01/01/2021 ($t = 0$) to 08/01/2021 ($t = 8$). Contours of the sea surface height at the final time $t = 8$ are included in grey to indicate the position of the Gulf Stream and nearby eddies in the flow.



(a) Each sample is represented by a single red marker.



(b) The samples are binned into a histogram, with contours of the Gaussian PDF overlaid.

Figure 6.2: Comparison of stochastic

Chapter 7

Future outlook and conclusions

In this chapter, we briefly discuss the implications of the work presented in this thesis, and highlight several avenues for further extensions and applications.

7.1 Further theoretical developments

7.2 Bayesian inference and data assimilation

7.3 Lagrangian coherent structures

7.4 Implications for the Fokker-Planck equation

Appendix A

Derivation of analytical SDE solutions

A.1 Linear SDEs

A.2 Benê's SDE

Consider Benê's SDE, as introduced in Example 2.4.2,

$$\mathrm{d}x_t = \tanh(x_t)\,\mathrm{d}t + \mathrm{d}W_t, \tag{A.1}$$

where x_t is a 1-dimensional stochastic process and W_t is a one-dimensional Wiener process. By employing a change of probability measure, it is shown in Section ?? of Särkkä and Solin (2019) that a weak solution to (A.1) has probability density function

$$p(x_t, t) = \frac{1}{\sqrt{2\pi t}} \frac{\cosh(x)}{\cosh(x_0)} \exp\left[-\frac{t}{2} - \frac{1}{2t}(x - x_0)^2\right] \tag{A.2}$$

at any time $t > 0$ and fixed initial condition $x_0 \in \mathbb{R}$. Here, we show that the PDF (A.2) can be expressed as the weighted sum of two Gaussian densities, which allows us to easily compute the mean and variance of the solving process x_t at any time. We can write

$$\begin{aligned} p(x_t, t) &= \frac{1}{2\sqrt{2\pi t}} \frac{\exp[x] + \exp[-x]}{\cosh(x_0)} \exp\left[-\frac{t}{2}\right] \exp\left[-\frac{1}{2t}(x - x_0)^2\right] \\ &= \frac{1}{2\sqrt{2\pi t}} \frac{\exp[x] + \exp[-x]}{\cosh(x_0)} \exp\left[-\frac{t}{2}\right] \exp\left[-\frac{1}{2t}(x - x_0)^2\right] \end{aligned}$$

Appendix B

Extended appendices of "Explicit Gaussian characterisation of model uncertainty in the limit of small noise"

B.1 Preliminaries for proofs

Throughout, we use the norm symbol $\|\cdot\|$ to denote (i) for a vector, the standard Euclidean vector norm, (ii) for a matrix, the spectral norm induced by the Euclidean norm, and (iii) for a 3rd-order tensor, the spectral norm induced by the matrix norm. The gradient symbol ∇ generically refers to derivatives with respect to the state variable. We write $W_t = \left(W_t^{(1)}, \ldots, W_t^{(n)}\right)^\intercal$ as the components of the canonical n-dimensional Wiener process, where each $W_t^{(i)}$ are mutually independent 1-dimensional Wiener processes. The flow map $F_0^t : \mathbb{R}^n \to \mathbb{R}^n$ summarises solutions of the deterministic model (4.2), given by

$$F_0^t(x) = x + \int_0^t u\left(F_0^\tau(x), \tau\right) \mathrm{d}\tau. \tag{B.1}$$

for an initial condition $x \in \mathbb{R}^n$. The spatial gradient (with respect to the initial condition) of the flow map solves the equation of variations associated with (4.2), i.e.

$$\frac{\partial}{\partial t} \nabla F_0^t(x) = \nabla u\left(F_0^t(x), t\right) \nabla F_0^t(x). \tag{B.2}$$

Hypothesis H summarises the minimum requirements of smoothness, measurability and boundedness in the coefficients of the main stochastic differential equation (4.1).

Hypothesis H *Let $u : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times [0,T] \to \mathbb{R}^{n\times n}$ be such that:*

(H.1) For each $t \in [0,T]$, the function $u(\cdot,t) : \mathbb{R}^n \to \mathbb{R}^n$ given by $u(x,t)$ is twice continuously differentiable on $\mathbb{R}^n$, and each component of the function $\sigma(\cdot,t) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ given by $\sigma(x,t)$ is differentiable on $\mathbb{R}^n$.

(H.2) For each $x \in \mathbb{R}^n$, the function $u(x,\cdot) : [0,T] \to \mathbb{R}^n$ and each component of the function $\sigma(x,\cdot) : [0,T] \to \mathbb{R}^{n\times n}$ are Borel-measurable on $[0,T]$.

(H.3) There exists a constant $K_u > 0$ such that for any $t \in [0,T]$ and $x \in \mathbb{R}^n$,

$$\max\left\{\left\|u(x,t)\right\|, \left\|\nabla u(x,t)\right\|, \left\|\nabla\nabla u(x,t)\right\|\right\} \leq K_u.$$

(H.4) There exists a constant $K_\sigma > 0$ such that for any $t \in [0,T]$ and $x \in \mathbb{R}^n$,

$$\max\left\{\left\|\sigma(x,t)\right\|, \left\|\nabla\sigma(x,t)\right\|\right\} \leq K_\sigma.$$

The conditions in Hypothesis H imply that u and σ are globally Lipschitz and guarantee that for each value of ε, there exists a unique strong solution to (4.1) Kallianpur and Sundar (2014). Specifically, for every $x, y \in \mathbb{R}^n$ and $t \in [0,T]$,

$$\left\|u\left(x,t\right) - u\left(y,t\right)\right\| \leq K_u \|x - y\|, \tag{B.3}$$

and if $\sigma_{i\cdot}$ denotes the ith row of σ, then

$$\left\|\sigma_{i\cdot}(x,t) - \sigma_{i\cdot}(y,t)\right\| \leq K_\sigma \|x-y\|. \tag{B.4}$$

The following inequalities are used several times throughout. For any real numbers $x_1, \ldots, x_m \geq 0$ and $r \geq 1$,

$$\left(\sum_{i=1}^m x_i\right)^r \leq m^{r-1} \sum_{i=1}^m x_i^r. \tag{B.5}$$

This results from an application of the finite form of Jensen's inequality. An implication of the equivalence of the L^1 and Euclidean norms and (B.5) is that for any $z \in \mathbb{R}^n$ and $r \geq 1$,

$$\|z\|^r \leq \left(\sum_{i=1}^n |z_i|\right)^r \leq n^{r-1} \sum_{i=1}^n |z_i|^r, \tag{B.6}$$

where z_i denotes the ith component of z. If each component z_i of a vector z is bounded by a constant K, then

$$\|z\| \leq \sqrt{n} K. \tag{B.7}$$

Similarly, if $f : \mathbb{R} \to \mathbb{R}^n$ is a vector-valued function such that each component of f is integrable over an interval $[0,t]$, then for all $r \geq 1$,

$$\left\|\int_0^t f(\tau)\,\mathrm{d}\tau\right\|^r \leq t^{r-1} \int_0^t \left\|f(\tau)\right\|^r \mathrm{d}\tau. \tag{B.8}$$

This inequality results from an application of Hölder's inequality.

B.2 Proof of Theorem 4.4.1

To prove the main result, we first require several lemmas. The first lemma is an extension of (Balasuriya 2020a, Lem. A.1) to the n-dimensional case.

Lemma B.2.1 *Let $y_t^{(\varepsilon)}$ denote the strong solution to (4.1) for $t \in [0,T]$, then for any $p > 0$ there exists a constant $G_p > 0$ such that*

$$\mathbb{E}\left[\left\|\int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) \mathrm{d}W_\tau\right\|^{2p}\right] \leq n^{3p} K_\sigma^{2p} G_p t^p.$$

Proof. For $i \in \{1, \ldots, n\}$, let $\sigma_{i\cdot}$ denote the ith row of σ. Define the stochastic process

$$M_\tau^{(i)} := \sigma_{i\cdot}\left(y_\tau^{(\varepsilon)}, \tau\right)$$

for $\tau \in [0,t]$, so that

$$\left[\int_0^t \sigma\left(y_\tau^{(\varepsilon)},\tau\right)\mathrm{d}W_\tau\right]_i = \int_0^t M_\tau^{(i)}\,\mathrm{d}W_\tau.$$

Then, from H.4 and using (B.7),

$$\int_0^t \left\|M_\tau^{(i)}\right\|^2 \mathrm{d}\tau \le \int_0^t nK_\sigma^2\,\mathrm{d}\tau < \infty,$$

so we can apply the Burkholder-Davis-Gundy inequality (e.g. (Kallianpur and Sundar 2014, Thm. 5.6.3)) to M_τ, which asserts that there exists a constant $G_p > 0$ depending only on p such that

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^t M_\tau^{(i)}\,\mathrm{d}W_\tau\right|^{2p}\right] &\le G_p\mathbb{E}\left[\left(\int_0^t \left\|\sigma_{i\cdot}\left(y_\tau^{(\varepsilon)},\tau\right)\right\|^2\mathrm{d}\tau\right)^p\right]\\ &\le G_p n^p K_\sigma^{2p} t^p. \end{aligned} \tag{B.9}$$

where the second inequality uses H.4 and (B.7). Then, we can write

$$\mathbb{E}\left[\left\|\int_0^t \sigma\left(y_\tau^{(\varepsilon)},\tau\right)\mathrm{d}W_\tau\right\|^{2p}\right] \le n^{3p}K_\sigma^{2p}G_p t^p,$$

using (B.6) and then (B.9). □

The next Lemma is a similar extension of (Balasuriya 2020a, Lem. 2.2).

Lemma B.2.2 *Let $q \ge 1$, then for all $\varepsilon > 0$ and $\tau \in [0,T]$*

$$\mathbb{E}\left[\int_0^t \left\|z_\tau^{(\varepsilon)}(x)\right\|^q \mathrm{d}\tau\right] \le H_q(t),$$

where

$$H_q(t) := 2^{q-1}n^{3q/2}K_\sigma^q G_{q/2}t^{q/2+1}\exp\left(2^{q-1}K_u^q t^q\right).$$

Proof. Consider the integral form of (4.1),

$$y_t^{(\varepsilon)} = x + \int_0^t u\left(y_\tau^{(\varepsilon)},\tau\right)\mathrm{d}\tau + \varepsilon\int_0^t \sigma\left(y_\tau^{(\varepsilon)},\tau\right)\mathrm{d}W_\tau.$$

Using (B.1),

$$y_t^{(\varepsilon)} - F_0^t(x) = \int_0^t \left(u\left(y_\tau^{(\varepsilon)},\tau\right) - u\left(F_0^\tau(x),\tau\right)\right)\mathrm{d}\tau + \varepsilon\int_0^t \sigma\left(y_\tau^{(\varepsilon)},\tau\right)\mathrm{d}W_\tau,$$

and so

$$\left\|y_t^{(\varepsilon)} - F_0^t(x)\right\|^q \leq 2^{q-1}\left(t^{q-1}\int_0^t \left\|u\left(y_\tau^{(\varepsilon)},\tau\right) - u\left(F_0^\tau(x),\tau\right)\right\|^q \mathrm{d}\tau + \varepsilon^q\left\|\int_0^t \sigma\left(y_\tau^{(\varepsilon)},\tau\right)\mathrm{d}W_\tau\right\|^q\right),$$

using (B.5) and then (B.8). Taking the expectation and applying Lemma B.2.1 with $p = q/2$ gives

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - F_0^t(x)\right\|^q\right] \leq 2^{q-1}t^{q-1}\mathbb{E}\left[\int_0^t \left\|u\left(y_\tau^{(\varepsilon)},\tau\right) - u\left(F_0^\tau(x),\tau\right)\right\|^q \mathrm{d}\tau\right] + 2^{q-1}\varepsilon^q n^{3q/2}K_\sigma^q G_{q/2}t^{q/2}. \tag{B.10}$$

We note that $\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - F_0^t(x)\right\|^q\right] < \infty$ from H.3, so by Tonelli's theorem (e.g. (Brémaud 2020, Thm. 2.3.9)),

$$\mathbb{E}\left[\int_0^t \left\|y_\tau^{(\varepsilon)} - F_0^\tau(x)\right\|^q \mathrm{d}\tau\right] = \int_0^t \mathbb{E}\left[\left\|y_\tau^{(\varepsilon)} - F_0^\tau(x)\right\|^q\right]\mathrm{d}\tau.$$

Now, using the Lipschitz condition (B.3) on (B.10) and interchanging the expectation and integral,

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - F_0^t(x)\right\|^q\right] \leq 2^{q-1}K_u^q t^{q-1}\int_0^t \mathbb{E}\left[\left\|y_\tau^{(\varepsilon)} - F_0^\tau(x)\right\|^q\right]\mathrm{d}\tau + 2^{q-1}\varepsilon^q n^{3q/2}K_\sigma^q G_{q/2}t^{q/2}.$$

Applying Grönwall's inequality then gives

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - F_0^t(x)\right\|^q\right] \leq 2^{q-1}\varepsilon^q n^{3q/2}K_\sigma^q G_{q/2}t^{q/2}\exp\left(2^{q-1}K_u^q t^q\right).$$

Hence, for $\tau \leq t$ we have

$$\mathbb{E}\left[\left\|z_\tau^{(\varepsilon)}(x)\right\|^q\right] \leq 2^{q-1}n^{3q/2}K_\sigma^q G_{q/2}\tau^{q/2}\exp\left(2^{q-1}K_u^q\tau^q\right) < \infty,$$

and so Tonelli's theorem allows us to again interchange the expectation and integral so that

$$\begin{aligned}\mathbb{E}\left[\int_0^t \left\|z_\tau^{(\varepsilon)}(x)\right\|^q \mathrm{d}\tau\right] &= \int_0^t \mathbb{E}\left[\left\|z_\tau^{(\varepsilon)}(x)\right\|^q\right]\mathrm{d}\tau \\ &\leq 2^{q-1}n^{3q/2}K_\sigma^q G_{q/2}t^{q/2+1}\exp\left(2^{q-1}K_u^q t^q\right),\end{aligned}$$

as desired. □

We now have all the tools to prove Theorem 4.4.1. By rearranging the integral forms of (4.1) and using (B.1) and (4.3), $z_t^{(\varepsilon)}(x)$ satisfies the integral equation

$$z_t^{(\varepsilon)}(x) = \int_0^t v^{(\varepsilon)}\left(z_\tau^{(\varepsilon)}(x), \tau\right) \mathrm{d}\tau + \int_0^t \sigma\left(\varepsilon z_\tau^{(\varepsilon)}(x) + F_0^\tau(x), \tau\right) \mathrm{d}W_\tau, \tag{B.11}$$

where

$$v^{(\varepsilon)}\left(z_\tau^{(\varepsilon)}, \tau\right) \coloneqq \frac{1}{\varepsilon}\left[u\left(\varepsilon z_\tau^{(\varepsilon)} + F_0^\tau(x), \tau\right) - u\left(F_0^\tau(x), \tau\right)\right].$$

Subtracting the integral form of (4.4) from (B.11) gives

$$\begin{aligned} z_t^{(\varepsilon)}(x) - z_t(x) &= \int_0^t \left[v^{(\varepsilon)}\left(z_\tau^{(\varepsilon)}(x), \tau\right) - \nabla u\left(F_0^\tau(x), \tau\right) z_\tau(x)\right] \mathrm{d}\tau \\ &\qquad + \int_0^t \left[\sigma\left(\varepsilon z_\tau^{(\varepsilon)} + F_0^\tau(x), \tau\right) - \sigma\left(F_0^\tau(x), \tau\right)\right] \mathrm{d}W_\tau \\ &= A(t) + B(t) + C(t), \end{aligned}$$

where

$$\begin{aligned} A(t) &\coloneqq \int_0^t \left[v^{(\varepsilon)}\left(z_\tau^{(\varepsilon)}(x), \tau\right) - \nabla u\left(F_0^\tau(x), \tau\right) z_\tau^{(\varepsilon)}(x)\right] \mathrm{d}\tau \\ B(t) &\coloneqq \int_0^t \nabla u\left(F_0^\tau(x), \tau\right)\left[z_\tau^{(\varepsilon)}(x) - z_\tau(x)\right] \mathrm{d}\tau \\ C(t) &\coloneqq \int_0^t \left[\sigma\left(\varepsilon z_\tau^{(\varepsilon)}(x) + F_0^\tau(x), \tau\right) - \sigma\left(F_0^\tau(x), \tau\right)\right] \mathrm{d}W_\tau. \end{aligned}$$

Then, using (B.5) and taking expectation,

$$\mathbb{E}\left[\left\|z_t^{(\varepsilon)}(x) - z_t(x)\right\|^r\right] \leq 3^{r-1}\left(\mathbb{E}\left[\left\|A(t)\right\|^r\right] + \mathbb{E}\left[\left\|B(t)\right\|^r\right] + \mathbb{E}\left[\left\|C(t)\right\|^r\right]\right) \tag{B.12}$$

First consider $A(t)$. Since for any $t \in [0, T]$, $u(\cdot, t)$ is twice continuously differentiable under H.1, for each $i = 1, \ldots, n$ there exists by Taylor's theorem (e.g. see (Hubbard and Hubbard 2009, Cor. A9.3.)) a function $R_i : \mathbb{R}^n \times [0, T] \to \mathbb{R}$ such that

$$u_i\left(\varepsilon z + F_0^\tau(x), \tau\right) = u_i\left(F_0^\tau(x), \tau\right) + \varepsilon\left[\nabla u_i\left(F_0^\tau(x), \tau\right)\right] z + R_i(\varepsilon z, \tau) \tag{B.13}$$

for any $z \in \mathbb{R}^n$, where u_i denotes the ith component of u. The function R_i satisfies

$$\left|R_i(\varepsilon z, \tau)\right| \leq \frac{1}{2}\left\|\nabla\nabla u_i\left(F_0^\tau(x), t\right)\right\| \|\varepsilon z\|^2 \leq \frac{\varepsilon^2 K_u}{2}\|z\|^2. \tag{B.14}$$

Rearranging (B.13),

$$v_i^{(\varepsilon)}(z,\tau) - \left[\nabla u_i\left(F_0^\tau(x),\tau\right)\right] z = \frac{1}{\varepsilon} R_i(\varepsilon z,\tau),$$

where $v_i^{(\varepsilon)}$ is the ith component of $v^{(\varepsilon)}$. Let $R(\varepsilon z,\tau) \coloneqq \left(R_1(\varepsilon z,\tau),\ldots,R_n(\varepsilon z,\tau)\right)^\intercal$, then

$$A(t) = \frac{1}{\varepsilon}\int_0^t R\left(\varepsilon z_\tau^{(\varepsilon)}(x),\tau\right)\mathrm{d}\tau,$$

and since each component of R is bounded as in (B.14), using (B.7)

$$\left\|R(\varepsilon z,\tau)\right\| \leq \frac{\sqrt{n}\varepsilon^2 K_u}{2}\|z\|^2.$$

Taking the norm and expectation then gives

$$\begin{aligned}
\mathbb{E}\left[\left\|A(t)\right\|^r\right] &= \frac{1}{\varepsilon^r}\mathbb{E}\left[\left\|\int_0^t R\left(\varepsilon z_\tau^{(\varepsilon)}(x),\tau\right)\mathrm{d}\tau\right\|^r\right] \\
&\leq \frac{t^{r-1}}{\varepsilon^r}\mathbb{E}\left[\int_0^t\left\|R\left(\varepsilon z_\tau^{(\varepsilon)}(x),\tau\right)\right\|^r\mathrm{d}\tau\right] \\
&\leq \frac{t^{r-1}n^{r/2}\varepsilon^{2r}K_u^r}{2^r\varepsilon^r}\mathbb{E}\left[\int_0^t\left\|z_\tau^{(\varepsilon)}(x)\right\|^{2r}\mathrm{d}\tau\right] \\
&\leq \frac{t^{r-1}n^{r/2}\varepsilon^r K_u^r H_{2r}(t)}{2^r},
\end{aligned} \tag{B.15}$$

where the second inequality uses (B.8), and $H_{2r}(t)$ is obtained from Lemma B.2.2 with $q = 2r$.

Next, consider $B(t)$, for which

$$\mathbb{E}\left[\left\|B(t)\right\|^r\right] \leq \int_0^t t^{r-1}K_u^r\mathbb{E}\left[\left\|z_\tau^{(\varepsilon)}(x) - z_\tau(x)\right\|^r\right]\mathrm{d}\tau. \tag{B.16}$$

using (B.8) and then H.3, and interchanging the expectation and the integral uses the fact that that $\mathbb{E}\left[\left\|z_\tau^{(\varepsilon)}(x)\right\|\right] < \infty$ and $\mathbb{E}\left[\left\|z_\tau(x)\right\|\right] < \infty$.

Finally, consider $C(t)$. For each $i \in \{1,\ldots,n\}$, define the stochastic process

$$N_t^{(i)} \coloneqq \sigma_{i\cdot}\left(\varepsilon z_\tau^{(\varepsilon)}(x) + F_0^\tau(x),\tau\right) - \sigma_{i\cdot}\left(F_0^\tau(x),\tau\right).$$

Then, the ith component of $C(t)$ is

$$\left[C(t)\right]_i = \int_0^t N_\tau^{(i)}\,\mathrm{d}W_\tau.$$

From H.4 and using (B.7),

$$\int_0^t \left\| N_\tau^{(i)} \right\|^2 \mathrm{d}\tau \leq \int_0^t 4nK_\sigma^2 \,\mathrm{d}\tau < \infty,$$

so we can apply the Burkholder-Davis-Gundy inequality on $N_\tau^{(i)}$ to write

$$\begin{aligned}
\mathbb{E}\left[\left|\left[C(t)\right]_i\right|^r\right] &\leq G_{r/2}\mathbb{E}\left[\left(\int_0^t \left\|\sigma_{i\cdot}\left(\varepsilon z_\tau^{(\varepsilon)}(x) + F_0^\tau(x), \tau\right) - \sigma_{i\cdot}\left(F_0^\tau(x), \tau\right)\right\|^2 \mathrm{d}\tau\right)^{r/2}\right] \\
&\leq G_{r/2}\mathbb{E}\left[\left(\int_0^t K_\sigma^2 \left\|\varepsilon z_\tau^{(\varepsilon)}(x)\right\|^2 \mathrm{d}\tau\right)^{r/2}\right] \\
&\leq G_{r/2}K_\sigma^r \varepsilon^r t^{r/2-1}\mathbb{E}\left[\int_0^t \left\|z_\tau^{(\varepsilon)}(x)\right\|^r \mathrm{d}\tau\right] \\
&\leq G_{r/2}K_\sigma^r \varepsilon^r t^{r/2-1} H_r(t), \qquad \text{(B.17)}
\end{aligned}$$

where the second inequality uses the Lipschitz condition on σ in (B.4), the third inequality uses (B.8), and the fourth inequality uses Lemma B.2.2 with $q = r$. Then, we have

$$\mathbb{E}\left[\left\|C(t)\right\|^r\right] \leq n^r G_{r/2}K_u^r \varepsilon^r t^{r/2-1} H_r(t), \qquad \text{(B.18)}$$

using (B.6), and then (B.17).

Combining (B.15), (B.16) and (B.18) into (B.12), we have

$$\begin{aligned}
\mathbb{E}\left[\left\|z_t^{(\varepsilon)}(x) - z_t(x)\right\|^r\right] \leq 3^{r-1}&\left(\frac{t^{r-1}n^{r/2}\varepsilon^r K_u^r H_{2r}(t)}{2^r} + n^r G_{r/2}K_u^r \varepsilon^r t^{r/2-1} H_r(t)\right) \\
&+ \int_0^t 3^{r-1}t^{r-1}K_u^r \mathbb{E}\left[\left\|z_\tau^{(\varepsilon)}(x) - z_\tau(x)\right\|^r\right] \mathrm{d}\tau.
\end{aligned}$$

Applying Grönwall's inequality, we have

$$\mathbb{E}\left[\left\|z_t^{(\varepsilon)}(x) - z_t(x)\right\|^r\right] = 3^{r-1}K_u^r\left(\frac{t^{r-1}n^{r/2}H_{2r}(t)}{2^r} + n^r G_{r/2}t^{r/2-1}H_r(t)\right)\exp\left(3^{r-1}t^r K_u^r\right)\varepsilon^r.$$

Define

$$D_r(t) \coloneqq 3^{r-1}K_u^r\left(\frac{t^{r-1}n^{r/2}H_{2r}(t)}{2^r} + n^r G_{r/2}t^{r/2-1}H_r(t)\right)\exp\left(3^{r-1}t^r K_u^r\right).$$

then we have shown (4.5), the desired result.

B.3 Proof of Theorem 4.4.2

In this appendix, we establish the exact Gaussian distribution of the solution to the linearised SDE.

Lemma B.3.1 *The strong solution (e.g.(Kallianpur and Sundar 2014, Def. 6.1.1)) to the linearised SDE (4.4) can be written as the Itô integral.*

$$z_t(x) = \int_0^t \nabla F_0^t(x) \left[\nabla F_0^\tau(x)\right]^{-1} \sigma\left(F_0^\tau(x), \tau\right) \mathrm{d}W_\tau. \tag{B.19}$$

Proof. The equation (4.4) is linear and homogeneous, so the solution method is well-known, e.g. see (Särkkä and Solin 2019, §4.3). Let $\Psi = \Psi(t)$ be the invertible fundamental matrix solution of the differential equation

$$\frac{\mathrm{d}\Psi(t)}{\mathrm{d}t} = \nabla u\left(F_0^t(x), t\right)\Psi(t), \quad \Psi(0) = I. \tag{B.20}$$

The equation (B.20) is exactly the equation of variations (B.2) for the deterministic system, so $\Psi(t) = \nabla F_0^t(x)$, noting that $\Psi(0) = \nabla F_0^0(x) = \nabla x = I$. Using Itô's Lemma (e.g. see (Kallianpur and Sundar 2014, Thm. 5.5.1)) on $M(t) \coloneqq \Psi(t)^{-1} z_t(x)$, it follows that

$$z_t(x) = \int_0^t \nabla F_0^t(x) \left[\nabla F_0^\tau(x)\right]^{-1} \sigma\left(F_0^\tau(x), \tau\right) \mathrm{d}W_\tau$$

satisfies (4.4) and is therefore a strong solution. □

We next establish that the Itô integral of a matrix-valued deterministic function with respect to a multidimensional Wiener process is a multidimensional Gaussian process.

Lemma B.3.2 *Let $a, b \in \mathbb{R}$ and let $g : [a, b] \to \mathbb{R}^{n\times n}$ be a matrix-valued deterministic function such that each element of g is Itô-integrable. Consider the Itô integral*

$$\mathcal{I}[g] \coloneqq \int_a^b g(t)\,\mathrm{d}W_t,$$

Then, the integral $\mathcal{I}[g]$ is a n-dimensional multivariate Gaussian random variable.

Proof. For $i, j \in \{1, \ldots, n\}$, let $g_{ij} : [a, b] \to \mathbb{R}$ be the (i, j)th element of g. Then, let

$$\mathcal{I}[g_{ij}] \coloneqq \int_a^b g_{ij}(t)\,\mathrm{d}W_t^{(i)},$$

so that the ith element of $\mathcal{I}[g]$ is

$$\mathcal{I}[g]_i = \sum_{j=1}^n \mathcal{I}\left[g_{ij}\right].$$

Each $\mathcal{I}[g_{ij}]$ is an Itô integral of a deterministic, scalar-valued function with respect to a one-dimensional Brownian motion, which is well-known to be a Gaussian process (e.g. see (Applebaum 2004, Lem. 4.3.11)). Moreover, each element of $\mathcal{I}[g]$ is the sum of independent Gaussian random variables and is therefore itself Gaussian. Hence, $\mathcal{I}[g]$ follows a multivariate Gaussian distribution. □

We can now prove Theorem 4.4.2. We have that

$$z_t(x) = \int_0^t \nabla F_0^t(x) \left[\nabla F_0^\tau(x)\right]^{-1} \sigma\left(F_0^\tau(x), \tau\right) \mathrm{d}W_\tau = \int_0^t L\left(x, t, \tau\right) \mathrm{d}W_\tau,$$

where L is as defined in (4.7). For any fixed $t \in [0, T]$, $z_t(x)$ is the Itô integral of a deterministic, matrix-valued function, and so $z_t(x)$ is an n-dimensional Gaussian random variable by Lemma B.3.2. Moreover Kallianpur and Sundar (2014), $\mathbb{E}\big[z_t(x)\big] = 0$. The variance of $z_t(x)$ is then

$$\mathbb{V}\big[z_t(x)\big] = \mathbb{E}\Big[\Big(z_t(x) - \mathbb{E}\big[z_t(x)\big]\Big)\Big(z_t(x) - \mathbb{E}\big[z_t(x)\big]\Big)^\intercal\Big] = \mathbb{E}\Big[z_t(x) z_t(x)^\intercal\Big].$$

Let L_{ij} denote the (i,j)th element of L for $i, j \in \{1, \ldots, n\}$. Then the (i,j)th element of $\mathbb{V}\big[z_t(x)\big]$ is

$$\begin{aligned}
\Big[\mathbb{V}\big[z_t(x)\big]\Big]_{ij} &= \mathbb{E}\left[\left(\sum_{k=1}^n \int_0^t L_{ik}\left(x, t, \tau\right) \mathrm{d}W_\tau^{(k)}\right)\left(\sum_{l=1}^n \int_0^t L_{il}\left(x, t, \tau\right) \mathrm{d}W_\tau^{(l)}\right)\right] \\
&= \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}\left[\left(\int_0^t L_{ik}\left(x, t, \tau\right) \mathrm{d}W_\tau^{(k)}\right)\left(\int_0^t L_{il}\left(x, t, \tau\right) \mathrm{d}W_\tau^{(l)}\right)\right] \\
&= \sum_{k=1}^n \int_0^t L_{ik}\left(x, t, \tau\right) L_{jk}\left(x, t, \tau\right) \mathrm{d}\tau \\
&= \big[\Sigma(x, t)\big]_{ij}
\end{aligned}$$

where the third equality uses Itô's isometry Kallianpur and Sundar (2014) and the fact that $W_t^{(k)}$ is independent of $W_t^{(l)}$ for $k \neq l$. So $\mathbb{V}\big[z_t(x)\big] = \Sigma(x, t)$ and $z_t(x) \sim \mathcal{N}\big(0, \Sigma(x, t)\big)$.

Taking the limit as $\varepsilon \to 0$ in (4.5), $z_t^{(\varepsilon)}(x)$ converges in rth moment to $z_t(x)$, which implies that $z_t^{(\varepsilon)}(x)$ converges in distribution to $z_t(x)$ also. Since $z_t(x)$ is known to follow a Gaussian distribution, we can write

$$z_t^{(\varepsilon)}(x) \xrightarrow{d} \mathcal{N}\left(0, \Sigma(x, t)\right).$$

Finally, we show that $\Sigma(x, t)$ is the solution to the matrix differential equation (4.8). We clearly have that $\Sigma(x, 0) = O$, the $n \times n$ zero matrix. Differentiating the expression

(4.6) with the Leibniz integral rule

$$\begin{aligned}\frac{\mathrm{d}\Sigma(x,t)}{\mathrm{d}t} &= L\left(x,t,t\right)L\left(x,t,t\right)^{\intercal} + \int_0^t \frac{\partial}{\partial t}\left[L\left(x,t,\tau\right)L\left(x,t,\tau\right)^{\intercal}\right]\mathrm{d}\tau \\ &= \sigma\left(F_0^t(x),t\right)\sigma\left(F_0^t(x),t\right)^{\intercal} + \int_0^t \frac{\partial L\left(x,t,\tau\right)}{\partial t}L\left(x,t,\tau\right)^{\intercal}\mathrm{d}\tau \\ &\quad + \int_0^t L\left(x,t,\tau\right)\frac{\partial L\left(x,t,\tau\right)^{\intercal}}{\partial t}\mathrm{d}\tau. \end{aligned} \tag{B.21}$$

Now, using the equation of variations (B.2),

$$\frac{\partial L\left(x,t,\tau\right)}{\partial t} = \left[\nabla u\left(F_0^t(x),t\right)\right]L\left(x,t,\tau\right). \tag{B.22}$$

Using (B.22) in (B.21),

$$\begin{aligned}\frac{\mathrm{d}\Sigma(x,t)}{\mathrm{d}t} &= \sigma\left(F_0^t(x),t\right)\sigma\left(F_0^t(x),t\right)^{\intercal} + \left[\nabla u\left(F_0^t(x),t\right)\right]\int_0^t L\left(x,t,\tau\right)L\left(x,t,\tau\right)^{\intercal}\mathrm{d}\tau \\ &\quad + \int_0^t L\left(x,t,\tau\right)L\left(x,t,\tau\right)^{\intercal}\mathrm{d}\tau\left[\nabla u\left(F_0^t(x),t\right)\right]^{\intercal} \\ &= \sigma\left(F_0^t(x),t\right)\sigma\left(F_0^t(x),t\right)^{\intercal} + \left[\nabla u\left(F_0^t(x),t\right)\right]\Sigma(x,t) + \Sigma(x,t)\left[\nabla u\left(F_0^t(x),t\right)\right]^{\intercal}. \end{aligned}$$

B.4 Proof of Theorem 4.5.1

Let $x \in \mathbb{R}^n$ and $t \in [0,T]$ be fixed. We first establish that the ε-limit and the variance operator can be exchanged. Using the fact that $\mathbb{E}\left[z_t(x)\right] = 0$ and properties of expectation, we have

$$\left\|\mathbb{E}\left[z_t^{(\varepsilon)}(x)\right]\right\| = \left\|\mathbb{E}\left[z_t^{(\varepsilon)}(x) - z_t(x)\right]\right\| \leq \mathbb{E}\left[\left\|z_t^{(\varepsilon)}(x) - z_t(x)\right\|\right] \leq D_1(t)\varepsilon,$$

so

$$\lim_{\varepsilon \downarrow 0}\left\|\mathbb{E}\left[z_t^{(\varepsilon)}(x)\right]\right\| = 0.$$

Using the fact that $\rho(z) = \left\|\mathbb{E}\left[zz^\intercal\right]\right\|^{1/2}$ defines a norm on the space of $\mathbb{R}^n$-valued random vectors and the reverse triangle inequality,

$$\begin{aligned}\left|\left\|\mathbb{V}\left[z_t^{(\varepsilon)}(x)\right]\right\|^{1/2} - \left\|\Sigma(x,t)\right\|^{1/2}\right| &= \left|\left\|\mathbb{E}\left[z_t^{(\varepsilon)}(x)z_t^{(\varepsilon)}(x)^\intercal\right] - \mathbb{E}\left[z_t^{(\varepsilon)}(x)\right]\mathbb{E}\left[z_t^{(\varepsilon)}(x)^\intercal\right]\right\|^{1/2}\right. \\ &\qquad \left. - \left\|\mathbb{E}\left[z_t(x)z_t(x)^\intercal\right]\right\|^{1/2}\right| \\ &\leq \mathbb{E}\left[\left\|z_t^{(\varepsilon)}(x) - z_t(x)\right\|^2\right]^{1/2} + \left\|\mathbb{E}\left[z_t^{(\varepsilon)}(x)\right]\mathbb{E}\left[z_t^{(\varepsilon)}(x)\right]^\intercal\right\|^{1/2} \\ &\leq D_2(t)^{1/2}\varepsilon + \left\|\mathbb{E}\left[z_t^{(\varepsilon)}(x)\right]\right\|,\end{aligned}$$

and so taking the limit as $\varepsilon \downarrow 0$ and squaring both sides, we have

$$\lim_{\varepsilon\downarrow 0}\left\|\mathbb{V}\left[z_t^{(\varepsilon)}(x)\right]\right\| = \left\|\Sigma(x,t)\right\|. \tag{B.23}$$

For $\varepsilon > 0$, define

$$S^2_{(\varepsilon)}(x,t) \coloneqq \sup\left\{\mathbb{V}\left[p^\intercal z_t^{(\varepsilon)}(x)\right] \,\middle|\, p\in\mathbb{R}^n, \|p\| = 1\right\} = \sup\left\{p^\intercal\mathbb{V}\left[z_t^{(\varepsilon)}(x)\right]p \,\middle|\, p\in\mathbb{R}^n, \|p\| = 1\right\}$$

Since $\mathbb{V}\left[z_t^{(\varepsilon)}(x)\right]$ is symmetric and positive definite, the Cholesky decomposition provides an $n\times n$ matrix $\Pi^{(\varepsilon)}(x,t)$ such that

$$\mathbb{V}\left[z_t^{(\varepsilon)}(x)\right] = \left[\Pi^{(\varepsilon)}(x,t)\right]^\intercal \Pi^{(\varepsilon)}(x,t),$$

allowing us to write

$$S^2_{(\varepsilon)}(x,t) = \left\|\Pi^{(\varepsilon)}(x,t)\right\|^2 = \left\|\mathbb{V}\left[z_t^{(\varepsilon)}(x)\right]\right\|.$$

using properties of the spectral norm. Taking the limit as ε approaches zero and using (B.23),

$$S^2(x,t) = \lim_{\varepsilon\downarrow 0} S^2_{(\varepsilon)}(x,t) = \lim_{\varepsilon\downarrow 0}\left\|\mathbb{V}\left[z_t^{(\varepsilon)}(x)\right]\right\| = \left\|\Sigma(x,t)\right\|.$$

Since $\Sigma(x,t)$ is symmetric and positive definite, the operator norm, and therefore $S^2(x,t)$, is given by the largest eigenvalue of $\Sigma(x,t)$.

Bibliography

Ana J. Abascal, Sonia Castanedo, Raul Medina, Inigo J. Losada, and Enrique Alvarez-Fanjul. Application of HF radar currents to oil spill modelling. *Marine Pollution Bulletin*, 58(2):238–248, February 2009. ISSN 0025-326X. doi: 10.1016/j.marpolbul.2008.09.020.

David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, United Kingdom, 1st edition, 2004. ISBN 978-0-511-21119-5.

Aleksandar Badza, Trent W. Mattner, and Sanjeeva Balasuriya. How sensitive are Lagrangian coherent structures to uncertainties in data? *Physica D: Nonlinear Phenomena*, 444:133580, February 2023. ISSN 0167-2789. doi: 10.1016/j.physd.2022.133580.

Sanjeeva Balasuriya. Stochastic Sensitivity: A Computable Lagrangian Uncertainty Measure for Unsteady Flows. *SIAM Review*, 62:781–816, November 2020a. doi: 10.1137/18M1222922.

Sanjeeva Balasuriya. Uncertainty in finite-time Lyapunov exponent computations. *Journal of Computational Dynamics*, 7(2):313–337, 2020b. doi: 10.3934/jcd.2020013.

Sanjeeva Balasuriya, Nicholas T. Ouellette, and Irina I. Rypina. Generalized Lagrangian coherent structures. *Physica D: Nonlinear Phenomena*, 372:31–51, June 2018. ISSN 0167-2789. doi: 10.1016/j.physd.2018.01.011.

Judith Berner, Ulrich Achatz, Lauriane Batté, Lisa Bengtsson, Alvaro de la Cámara, Hannah M. Christensen, Matteo Colangeli, Danielle R. B. Coleman, Daan Crommelin, Stamen I. Dolaptchiev, Christian L. E. Franzke, Petra Friederichs, Peter Imkeller, Heikki Järvinen, Stephan Juricke, Vassili Kitsios, François Lott, Valerio Lucarini, Salil Mahajan, Timothy N. Palmer, Cécile Penland, Mirjana Sakradzija, Jin-Song von Storch, Antje Weisheimer, Michael Weniger, Paul D. Williams, and Jun-Ichi Yano. Stochastic Parameterization: Toward a New View of Weather and Climate Models. *Bulletin of the American Meteorological Society*, 98(3):565–588, March 2017. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-15-00268.1.

Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98, January 2017. ISSN 0036-1445. doi: 10.1137/141000671.

Yu N. Blagoveshchenskii. Diffusion Processes Depending on a Small Parameter. *Theory of Probability and its Applications*, 7(2):17, 1962. ISSN 0040585X. doi: 10.1137/1107013.

Liam Blake, John Maclean, and Sanjeeva Balasuriya. Explicit Gaussian characterisation of model uncertainty in the limit of small noise. *In review*, 2023.

Michal Branicki and Kenneth Uda. Lagrangian Uncertainty Quantification and Information Inequalities for Stochastic Flows. *SIAM/ASA Journal on Uncertainty Quantification*, 9(3):1242–1313, 2021.

Michał Branicki and Kenneth Uda. Path-Based Divergence Rates and Lagrangian Uncertainty in Stochastic Flows. *SIAM Journal on Applied Dynamical Systems*, pages 419–482, March 2023. doi: 10.1137/21M1466530.

Fred Brauer and Carlos Castillo-Chávez. *Mathematical Models in Population Biology and Epidemiology*. Number 40 in Texts in Applied Mathematics. Springer, New York, 2nd ed edition, 2012. ISBN 978-1-4614-1685-2.

Pierre Brémaud. *Probability Theory and Stochastic Processes*. Universitext. Springer, first edition, 2020. ISBN 978-3-030-40182-5.

Amarjit Budhiraja, Eric Friedlander, Colin Guider, C. K. R. T. Jones, and John Maclean. Assimilating Data into Models. In *Handbook of Environmental and Ecological Statistics*. Chapman and Hall/CRC, 1st edition edition, 2019. ISBN 978-1-315-15250-9.

Mat Collins. Ensembles and probabilities: A new era in the prediction of climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857):1957–1970, June 2007. doi: 10.1098/rsta.2007.2068.

C. J. Cotter and G. A. Pavliotis. Estimating eddy diffusivities from noisy Lagrangian observations. *Communications in Mathematical Sciences*, 7(4):805–838, December 2009. ISSN 1945-0796. doi: 10.4310/CMS.2009.v7.n4.a2.

Simon Danisch and Julius Krumbiegel. Makie.jl: Flexible high-performance data visualization for Julia. *Journal of Open Source Software*, 6(65):3349, September 2021. ISSN 2475-9066. doi: 10.21105/joss.03349.

Francesca Doglioni, Robert Ricker, Benjamin Rabe, and Torsten Kanzow. Sea surface height anomaly and geostrophic velocity from altimetry measurements over the Arctic Ocean (2011–2018). Preprint, Oceanography – Physical, May 2021.

Francesco d'Ovidio, Silvia De Monte, Séverine Alvain, Yves Dandonneau, and Marina Lévy. Fluid dynamical niches of phytoplankton types. *Proceedings of the National Academy of Sciences*, 107(43):18366–18370, October 2010. doi: 10.1073/pnas.1004620107.

Lei Fang, Sanjeeva Balasuriya, and Nicholas T. Ouellette. Disentangling resolution, precision, and inherent stochasticity in nonlinear systems. *Physical Review Research*, 2(2):023343, June 2020. doi: 10.1103/PhysRevResearch.2.023343.

Florian Feppon and Pierre F. J. Lermusiaux. Dynamically Orthogonal Numerical Schemes for Efficient Stochastic Advection and Lagrangian Transport. *SIAM Review*, 60(3):595–625, January 2018. ISSN 0036-1445. doi: 10.1137/16M1109394.

Alireza Hadjighasem, Mohammad Farazmand, Daniel Blazevski, Gary Froyland, and George Haller. A critical comparison of Lagrangian methods for coherent structure detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(5):053104, May 2017. ISSN 1054-1500. doi: 10.1063/1.4982720.

John H. Hubbard and Barbara Burke Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Matrix Editions, Ithaca, NY, 4th ed edition, 2009. ISBN 978-0-9715766-5-0.

Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*, volume 64 of *Mathematics in Science and Engineering*. Elsevier Science, Burlington, 2014. ISBN 978-0-08-096090-6.

Ian T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, NY, second edition, 2002. ISBN 978-0-387-22440-4.

G. Kallianpur and P. Sundar. *Stochastic Analysis and Diffusion Processes*. Number 24 in Oxford Graduate Texts in Mathematics. Oxford University Press, Oxford, United Kingdom, first edition edition, 2014. ISBN 978-0-19-965706-3 978-0-19-965707-0.

Igor Kamenkovich, Irina I. Rypina, and Pavel Berloff. Properties and Origins of the Anisotropic Eddy-Induced Transport in the North Atlantic. *Journal of Physical Oceanography*, 45(3):778–791, March 2015. ISSN 0022-3670, 1520-0485. doi: 10.1175/JPO-D-14-0164.1.

Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics. Springer, Berlin, Heidelberg, 1st edition, 1992. ISBN 978-3-540-54062-5.

M. V. Kulikova and G. Yu. Kulikov. Adaptive ODE solvers in extended Kalman filtering algorithms. *Journal of Computational and Applied Mathematics*, 262:205–216, May 2014. ISSN 0377-0427. doi: 10.1016/j.cam.2013.09.064.

Kody Law, Andrew Stuart, and Konstantinos Zygalakis. *Data Assimilation: A Mathematical Introduction*. Number volume 62 in Texts in Applied Mathematics. Springer, Cham Heidelberg New York Dordrecht London, 2015. ISBN 978-3-319-20324-9 978-3-319-36687-6. doi: 10.1007/978-3-319-20325-6.

Martin Leutbecher. Ensemble size: How suboptimal is less than infinity? *Quarterly Journal of the Royal Meteorological Society*, 145(S1):107–128, 2019. ISSN 1477-870X. doi: 10.1002/qj.3387.

Martin Leutbecher, Sarah-Jane Lock, Pirkka Ollinaho, Simon T. K. Lang, Gianpaolo Balsamo, Peter Bechtold, Massimo Bonavita, Hannah M. Christensen, Michail Diamantakis, Emanuel Dutra, Stephen English, Michael Fisher, Richard M. Forbes, Jacqueline Goddard, Thomas Haiden, Robin J. Hogan, Stephan Juricke, Heather Lawrence, Dave MacLeod, Linus Magnusson, Sylvie Malardel, Sebastien Massart, Irina Sandu, Piotr K. Smolarkiewicz, Aneesh Subramanian, Frédéric Vitart, Nils Wedi, and Antje Weisheimer. Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143 (707):2315–2339, 2017. ISSN 1477-870X. doi: 10.1002/qj.3094.

Yulong Lu, Andrew Stuart, and Hendrik Weber. Gaussian Approximations for Probability Measures on $\mathbb{R}^d$. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1136–1165, January 2017. doi: 10.1137/16M1105384.

Thomas Mazzoni. Computational aspects of continuous–discrete extended Kalman-filtering. *Computational Statistics*, 23(4):519–539, October 2008. ISSN 1613-9658. doi: 10.1007/s00180-007-0094-4.

C. M. Mora, H. A. Mardones, J. C. Jimenez, M. Selva, and R. Biscay. A Stable Numerical Scheme for Stochastic Differential Equations with Multiplicative Noise. *SIAM Journal on Numerical Analysis*, 55(4):1614–1649, January 2017. ISSN 0036-1429. doi: 10.1137/140984488.

Bernt Øksendal. *Stochastic Differential Equations*. Universitext. Springer, Berlin, Heidelberg, sixth edition, 2003. ISBN 978-3-540-04758-2 978-3-642-14394-6. doi: 10.1007/978-3-642-14394-6.

T. N. Palmer. Stochastic weather and climate models. *Nature Reviews Physics*, 1(7): 463–471, July 2019. ISSN 2522-5820. doi: 10.1038/s42254-019-0062-2.

Young-Hyang Park. Determination of the surface geostrophic velocity field from satellite altimetry. *Journal of Geophysical Research*, 109(C5):C05006, 2004. ISSN 0148-0227. doi: 10.1029/2003JC002115.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(1):15, May 2017. ISSN 2049-9647. doi: 10.5334/jors.151.

Sebastian Reich and Colin Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, Cambridge, 2015. ISBN 978-1-107-06939-8. doi: 10.1017/CBO9781107706804.

Hannes Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Series in Synergetics. Springer, Berlin, Heidelberg, first edition, 2012. ISBN 978-3-642-96807-5.

A. J. Roberts. Modify the Improved Euler scheme to integrate stochastic differential equations. *arXiv:1210.0933 [math]*, October 2012.

Andreas Rößler. Runge-Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 48(3):922–952, 2010. ISSN 0036-1429. doi: 10.1137/09076636X.

Roger M. Samelson and Stephen Wiggins. *Lagrangian Transport in Geophysical Jets and Waves: The Dynamical Systems Approach*, volume 31 of *Interdisciplinary Applied Mathematics*. Springer, New York, NY, 2006.

Daniel Sanz-Alonso and Andrew M. Stuart. Gaussian Approximations of Small Noise Diffusions in Kullback-Leibler Divergence. *Communications in Mathematical Sciences*, 15(7):2087–2097, October 2017. ISSN 1539-6746. doi: https://dx.doi.org/10.4310/CMS.2017.v15.n7.a13.

Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, Cambridge, 2019. ISBN 978-1-316-51008-7. doi: 10.1017/9781108186735.

Philip Sura. Stochastic Analysis of Southern and Pacific Ocean Sea Surface Winds. *Journal of the Atmospheric Sciences*, 60(4):654–666, February 2003. ISSN 0022-4928, 1520-0469. doi: 10.1175/1520-0469(2003)060⟨0654:SAOSAP⟩2.0.CO;2.

Philip Sura, Matthew Newman, Cécile Penland, and Prashant Sardeshmukh. Multiplicative Noise and Non-Gaussianity: A Paradigm for Atmospheric Regimes? *Journal of the Atmospheric Sciences*, 62(5):1391–1409, May 2005. ISSN 0022-4928, 1520-0469. doi: 10.1175/JAS3408.1.

Tamás Tél, Alessandro de Moura, Celso Grebogi, and György Károlyi. Chemical and biological activity in open flows: A dynamical system approach. *Physics Reports*, 413 (2):91–196, July 2005. ISSN 0370-1573. doi: 10.1016/j.physrep.2005.01.005.

Stephen Wiggins. The Dynamical Systems Approach to Lagrangian Transport in Oceanic Flows. *Annual Review of Fluid Mechanics*, 37(1):295–328, January 2005. ISSN 0066-4189, 1545-4479. doi: 10.1146/annurev.fluid.37.061903.175815.

Y. K. Ying, J. R. Maddison, and J. Vanneste. Bayesian inference of ocean diffusivity from Lagrangian trajectory data. *Ocean Modelling*, 140:101401, August 2019. ISSN 1463-5003. doi: 10.1016/j.ocemod.2019.101401.