Computable Characterisations of Uncertainty in Differential Equations

Liam Blake

7 February, 2024

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
School of Computer and Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

List of Figures

Abstract

Differential equations are an invaluable tool for predicting and understanding phenomena across many fields. These models are inevitably subject to uncertainties, including because of measurement error, unresolved components, and resolution limitations. Accounting for these uncertainties can improve models, but this endeavour is analytically challenging and often requires computationally expensive Monte Carlo simulation. In this thesis, we consider Itô stochastic differential equations, which extend ordinary differential equations to include stochastic terms and provide a rich framework for explicitly parameterising uncertainties in the model. We look to address the need for computationally efficient characterisations of uncertainty that do not require bulk simulation.

We first build upon previous small-noise studies to provide an explicit bound for the error between stochastic differential equations and corresponding linearisations written in terms of a deterministic system. Our framework accounts for non-autonomous coefficients, multiplicative noise, and uncertain initial conditions. These linearisations are solvable and efficient to compute and so can serve as an approximate solution to the stochastic differential equation. We demonstrate the predictive power of our bound on several toy examples, providing, for the first time, a numerical validation of linearisation approximations of stochastic differential equations. In characterising this relationship, we are also able to extend stochastic sensitivity (Balasuriya, *SIAM Review*, 2020:781-816), a recently introduced tool for characterising the impact of uncertainty on differential equation solutions. Stochastic sensitivity was previously restricted to 2-dimensional flows and we overcome this limitation to empower the use of these tools on models of arbitrary dimension.

Furthering the linearisation framework, we also propose an *ad hoc* algorithm for approximating a stochastic differential equation solution with a Gaussian mixture model constructed from many different linearisations. This algorithm is computationally efficient and provides an analytic probability density function, unlike stochastic samples. Critically, the algorithm can capture non-Gaussian features in the stochastic solution that a single linearisation cannot. Our investigation into this algorithm, using a data-driven model of a drifter in the North Atlantic Ocean, yielded promising results in approximating a highly non-Gaussian distribution.

Our work provides many avenues for further development. These possibilities include establishing connections with the Fokker-Planck equation and the theoretical extension of our framework to account for different types of driving stochastic noise. We also anticipate applications of both our theoretical contributions and computational tools within the fields of data assimilation, stochastic parameterisation, and mathematical epidemiology.

Statement of Originality

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 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.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: ..... Date: 7 February, 2024

Acknowledgements

First and foremost, I give the uttermost gratitude to my two supervisors, Sanjeeva Balasuriya and Jack Maclean. Our meetings were always interesting and enjoyable and they have always had the utmost patience and reassurance for my many worries. I have learnt so much from their unique approaches to research and always have valued their advice on both academia and my broader life and career.

All the maths HDRs made working on this project a blast, but I reserve particular gratitude for the Level 7 gang: Clover Hoffman, George Savvoudis, Irulan Murphy, Kai Li, Matthew Drown, Saka Magsarjav, Steven Greenwood, and Wills Nguyen.

I also want to thank all the legends within the ATCA teams at the Adelaide University Cricket Club, who will almost surely never read this work but provided an incredibly valuable escape from it at trainings and during our matches.

To my closest friends (to follow a cliché, you all should know who you are), thank you for all the coffees, the lunches, the board games, the runs, and the drinks. All these moments went a long way to keeping me sane.

Finally, to my parents Alison and Stephen, and my grandmother Margaret, thank you always for your ongoing love and support.

Chapter 1

Introduction

Many phenomena across geophysical, biological and socio-economic applications can be modelled using ordinary differential equations (ODEs). Given an initial condition, these models can be solved, either analytically or numerically, to predict the behaviour at a future time. These models are usually specified in a purely *deterministic* fashion, in that each component is treated as being known exactly. The model consequently provides a single prediction of the future state for any given initial condition. However, in any modelling scenario, there will inevitably be uncertainty in the model specification which can arise from:

- the model being sourced from measured data, which itself includes observational error
- information only being available on a spatiotemporal grid (resolution error), either due to data availability or discretisation of the model
- a lack of a complete understanding of all the processes involved in the system
- unknown values for any parameters in the model
- imperfect knowledge of the initial conditions.

Without appropriately accounting for these many manifestations of uncertainty, any predictions and inferences from the deterministic model may be inaccurate or even misleading. This is not a trivial problem to resolve, however, as this uncertainty is inherently unknowable. Even when the uncertainty cannot be directly encoded into the model, an understanding of how reliable the predictions of the model are in the presence of this uncertainty is extremely valuable.

This problem of uncertainty plague every field, but has received particular attention in climate modelling and weather prediction. For example, resolution issues pose a significant challenge in producing accurate predictions from a high-dimensional climate model. Many

such models rely upon a spatial discretisation for tractable analysis and simulation, which often requires an extremely high resolution to produce accurate simulations that correctly capture all relevant processes (Dawson, Palmer, and Corti, 2012). Without even considering the plethora of other sources of uncertainty in climate modelling, the discretisation of the model that is necessary for tractability introduces resolution issues that require a high computational load to overcome.

One approach to account systematically for uncertainty is by replacing the deterministic model with a *stochastic* one, where the uncertainty is introduced as an inherent part of the model and the solutions are treated as random quantities. The formal introduction of stochastic terms to account for unknown and unresolved processes into an otherwise deterministic model is known as *stochastic parameterisation* in scientific circles (Berner et al., 2017; Palmer, 2019). Stochastic parameterisation has been shown to both lead to the same performance as higher-resolution models—Dawson and Palmer (2015) showed through simulation studies that the performance of a high-resolution purely deterministic model can be matched by a lower-resolution stochastic model—and to improve the predictive power of forecasts (Mitchell and Gottwald, 2012; Ha et al., 2015). More broadly, introducing stochastic terms can improve upon a deterministic model by accounting for the otherwise unknowable sources of uncertainty in the model specification. This provides a new model to generate (now probabilistic) predictions and perform analysis or even evaluate the reliability of the original deterministic model itself (Balasuriya, 2020b).

Although stochastic models can improve upon their deterministic counterparts, there is a significant compromise: stochastic models are more often intractable to solve analytically and difficult to analyse than the deterministic counterparts, instead requiring numerical solutions. However, capturing the stochasticity of the model requires taking many *realisations* of the solution. In general, a large number of samples is required for convergent statistics and accurate inference (Feppon and Lermusiaux, 2018; Leutbecher, 2019) and so generating representative realisations often reaches the limit of our computational capabilities. Nonetheless, this bulk simulation approach is the gold standard in many applications (Collins, 2007; Leutbecher et al., 2017, e.g.). Overcoming this computational expense is an area of ongoing research in *every* field in which such stochastic models are used. In the climate context, 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. In particular, the authors state that "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...". The overall aim of this thesis is to address this problem by developing computationally efficient schemes to approximate and characterise the impact of uncertainty on the solutions to differential equations. We focus our attention on one particular class of stochastic models: stochastic differential equations, which are a natural extension of ordinary differential equations.

1.1 Stochastic differential equations

Stochastic differential equations (SDEs) are a natural framework for introducing uncertainty into the continuous time evolution of a variable (Øksendal, 2003; Penland, 2003; Kallianpur and Sundar, 2014; Särkkä and Solin, 2019). The uncertainty is introduced as a (possibly time- and state-dependent) stochastic process that contributes to the vector field (rate of change) of the differential equation for the ongoing evolution of the state variable. Although an SDE allows us to account for sources of uncertainty in a differential equation model, the usage of such presents new challenges, first requiring an entirely new type of calculus to formulate (Itô, 1944, 1946). The solution is now a stochastic process evolving through time, and the analytic solution must be expressed in terms of a time-dependent probability distribution over the state space. Finding these exact solutions is almost always impossible, however, and working with the equations and their solutions analytically is more difficult than for the deterministic counterparts. Nevertheless, stochastic differential equations provide a rich and flexible modelling framework and are consequently used across a range of fields, including financial mathematics (Kabanov et al., 2006), biological modelling (Preisler et al., 2004; Bachar et al., 2013, e.g.), physics (Gardiner et al., 1992; Strauss and Effenberger, 2017, e.g.), atmospheric modelling (Wilson and Sawford, 1996), and oceanography (Berloff and McWilliams, 2002). Although analytic solutions are usually unavailable, SDEs can be solved numerically to generate ensembles of approximation solution realisations (Kloeden and Platen, 1992), which can then be used in Monte Carlo-type inferences. There are further complications to even this, however. Generally, in modelling situations, the dynamics are highly nonlinear and one expects the noise to vary with both time and state (i.e. the noise can be *multiplicative*), e.g. in atmospheric (Sura, 2003; Sura et al., 2005) and oceanic (Kamenkovich et al., 2015) systems and from experimental and observational considerations. This further complicates the usage of the model for prediction and analysis; such SDEs are intractable to solve analytically and more difficult to approximate accurately (Sancho et al., 1982; Mora et al., 2017). Data-based models—that is, models possessing terms in the equations that are driven by data rather than by explicitly specified functions—and uncertainty in the initial state renders additional problems in obtaining a theoretical understanding of the stochastic system.

Let us consider this problem from a purely mathematical perspective; we are given a possibly highly nonlinear SDE with non-trivial noise components. We cannot solve such equations analytically, and numerical simulation is computationally expensive. Motivated by the need for efficient schemes across applications, we specifically look to develop characterisations of the solutions to such SDE and algorithms for approximating the resulting probability distributions.

1.2 Linearisations of SDEs

In lieu of an exact solution, a common approach to characterising and approximating the solution of nonlinear SDEs is via a "linearisation" through time about a single deterministic trajectory. A linearised stochastic differential equation, obtained by truncating Taylor expansions of each coefficient (Blagoveshchenskii, 1962; Jazwinski, 2014, e.g.), can be solved analytically and is accordingly used across a diversity of literature and applications (Jazwinski, 2014; Sanz-Alonso and Stuart, 2017; Särkkä and Solin, 2019; Kaszás and Haller, 2020, e.g.). In filtering theory and data assimilation (Law et al., 2015; Reich and Cotter, 2015; Budhiraja et al., 2019), one looks to improve predictions from a mathematical model by including ongoing measurement. When the underlying equations are stochastic differential equations (as can be the case in a continuous-time continuous state-space scenario), linearising these SDEs provides analytically tractable expressions for the updated state, leading to tractable and efficient filtering schemes (Jazwinski, 2014). More generally, linearisations can serve as approximations of the solutions to nonlinear SDEs in any application (Särkkä and Solin, 2019). However, often these linearisations are applied without rigorous justification and a clear understanding of *how* the linearisation relates to 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. In addition, there is little understanding of the impact of the choice of initialisation of the deterministic trajectory on the validity of the linearisation. We first look to address this in this thesis by providing a fully justified linearisation framework.

Much is already known theoretically about these linearisations; classical results in the context of small-noise series expansions (Blagoveshchenskii, 1962) and large deviations theory (Freidlin and Wentzell, 1998) show that the strong error between SDE solution with a fixed initial condition and that of an appropriate linearisation is bounded. Sanz-Alonso and Stuart (2017) establish a strong result, bounding the Kullback-Leibler divergence between the solutions of autonomous SDEs with additive stationary noise and a linearised equivalent. Their result considers both an uncertain initial condition, and the evolving error due to the discrepancy between the models. We aim to build upon these studies by considering a general linearisation framework in which the original SDE we look to approximate is multidimensional, fully non-autonomous, and equipped with multiplicative noise.

When the initial state is known exactly, the solution of a linearised SDE is a Gaussian process, which is both efficient to sample from and can allow for analytically tractable expressions in later inference. The simplicity of this Gaussian process can be a drawback, however; the unimodal Gaussian density is limited in the features that it can capture. Multimodality, skewness, nonlinear correlations, and other departures from Gaussianity are commonly observed in practice (del-Castillo-Negrete, 1998; Bracco et al., 2000; Sura et al., 2005), but cannot be captured by a single Gaussian component alone. We also look to address this difficulty by proposing an *ad hoc* algorithm that combines multiple

linearisations into a Gaussian mixture model. Such an algorithm lies between the computational efficiency of the linearisation approximation and the accuracy of stochastic samples, improving upon a single Gaussian without compromising numerical speed.

1.3 Stochastic sensitivity

Even when solutions to the stochastic system are not available, one may be interested in calculating properties of the system that offer *qualitative* insight into the behaviour of these solutions. Balasuriya (2020b) introduced "stochastic sensitivity" to quantify the impact of velocity field uncertainty on Lagrangian ('following-the-flow') trajectories in a 2-dimensional fluid flow. These trajectories solve an ODE involving the velocity field, which is inevitably subject to uncertainty from all the aforementioned sources. By considering a small-noise stochastic differential equation that introduces noise to the governing ordinary differential equation, Balasuriya (2020b) defined and formally derived computable expressions for the mean and variance (in the limit of small noise) of a certain quantity measuring the deviation between the ODE and the SDE solutions. The result is a single number, termed stochastic sensitivity, that is computable for any initial condition and measures the uncertainty in the resulting trajectory. Rather than attempting to solve the SDE, stochastic sensitivity evaluates the reliability of the ODE, seen as a 'best available' deterministic model, in the presence of uncertainty. We summarise these results in Section 2.5.

By providing a scalar field across initial conditions, stochastic sensitivity was immediately applied to extract so-called "Lagrangian coherent structures" (LCSs) from a fluid flow. These coherent structures are spatial regions that persist through the time-evolution of the flow and have a significant impact on the transport properties of the fluid (Peacock and Dabiri, 2010; Hadjighasem et al., 2017; Balasuriya, Ouellette, et al., 2018). In the LCS community, there has been an emerging interest in understanding the impact of velocity field stochasticity on the resulting Lagrangian analysis (Denner et al., 2016; Balasuriya, 2017; Balasuriya and Gottwald, 2018; Haller et al., 2018; Balasuriya, 2020a; Kaszás and Haller, 2020; You and Leung, 2021), but until recently there was no extraction scheme that explicitly accounted for ongoing velocity field uncertainties. Balasuriya (2020b) addressed this deficiency by proposing that stochastic sensitivity can extract spatially coherent regions that remain robust under velocity field fluctuations.

However, the original formulation of stochastic sensitivity (Balasuriya, 2020b)—and therefore the subsequent applications (Balasuriya, 2020c; Fang, Balasuriya, et al., 2020; Fang and Ouellette, 2021; Badza et al., 2023)—is limited to 2-dimensions with no clear extension to higher-dimensions. Moreover, Balasuriya (2020b) was only able to provide expressions for the mean and variance of the limiting quantity with no understanding of the distributions involved or exactly how this work fits within broader SDE theory. We also look to address this deficiency in this thesis by providing a new definition of stochastic

sensitivity and furthering the theory to gain a clearer understanding of the distributions involved. This generalisation would have the additional advantage of extending stochastic sensitivity beyond the fluid context to any n-dimensional differential equation model, opening up a wealth of new applications.

1.4 Contributions and structure of this thesis

In Chapter 2, we provide a brief background on stochastic differential equations and the tools we will use. The remainder of the thesis, and the contributions each chapter provides, are as follows.

- In Chapter 3, we address a deficiency in the theory of SDE linearisations. The aim here is to provide a rigorously justified framework for computing linearisation approximations of nonlinear stochastic differential equations, subject to arbitrary and random initial conditions. Much of the content in this chapter and the following Chapter 4 has been submitted for review as a research article to *Communications in Mathematical Sciences* (Blake et al., 2023). This chapter is the primary theoretical contribution of this thesis.

  - We first build upon previous studies of small-noise expansions (Blagoveshchenskii, 1962; Freidlin and Wentzell, 1998; Sanz-Alonso and Stuart, 2017) to provide a direct proof of a bound on the error between the solution to an SDE and a corresponding linearisation about a deterministic trajectory. Our result holds for a multidimensional and fully nonautonomous SDE with multiplicative noise subject to an arbitrary initial condition, and the bound itself is written explicitly in terms of the scales of the uncertainty in the initial condition and the ongoing noise.
  - By finding explicit solutions to the linearised SDE, we provide in a single place a framework for efficiently computing the solution to enable efficient approximation and characterisation of the nonlinear SDE without resorting to bulk Monte Carlo simulation. Such an approximation is used across a range of literature and applications (Archambeau et al., 2007; Jazwinski, 2014; Law et al., 2015; Reich and Cotter, 2015; Sanz-Alonso and Stuart, 2017; Budhiraja et al., 2019; Särkkä and Solin, 2019; Kaszás and Haller, 2020) but is dispersed across many different sources and often stated without an explicit justification.
  - We use the linearisation framework to extend the notion of stochastic sensitivity (Balasuriya, 2020b), which seeks to quantify explicitly the impact of vector field uncertainty on the solution trajectories of dynamical systems (Balibrea-Iniesta et al., 2016; Kaszás and Haller, 2020; Branicki and Uda, 2021, 2023). Whereas previously the definition of stochastic sensitivity was restricted to 2-dimensions

(Balasuriya, 2020b), we provide a new definition in arbitrary dimensions and prove a convenient expression for computation. This is the second significant theoretical contribution of this work, enabling this tool to be applied to a much broader class of models.

- In Chapter 4, we numerically validate the results of Chapter 3 using stochastic simulations from 1- and 2-dimensional models. In particular, we show that the first four moments of the distance between the realisations and the Gaussian limit follow the predicted bound. This, to the author's knowledge, is the first time that the validity of the linearisation approximation to an SDE has been systematically verified. We also illustrate our new computation of stochastic sensitivity on the 2-dimensional example and compute the field in 3-dimensions for the first time.

- In Chapter 5, we outline an algorithm that employs the linearisation approximation to construct a Gaussian mixture model for the solution to a stochastic differential equation, which can capture non-Gaussianity while still maintaining computational efficiency. The algorithm is preliminary, with several implementation details deliberately left unspecified and warranting further development. However, we demonstrate the potential of such an approach with the example of the following chapter.

- In Chapter 6, we apply the tools developed in Chapters 3 and 5 to an example constructed from observed data: the position of a drifter on the surface of the North Atlantic Ocean. The aim here is to demonstrate the computability of the linearisation framework and stochastic sensitivity and to illustrate that even a simple implementation of the mixture model algorithm of Chapter 5 can produce a close approximation of the challenging drifter position distribution.

- In Chapter 7, we conclude this thesis by discussing the possible future extensions of our work. Since our theory is reasonably general, and we have only outlined a mixture model algorithm, there is much scope for theoretical extension, further numerical investigation, and application to a range of models. We also highlight how an analogous result in a different modelling context—stochastic models on a discrete state space—where linearised stochastic differential equations arise from large-population limits (Kurtz, 1970, 1971). To demonstrate our computations, we briefly consider two examples of discrete models for the spread of an infectious disease in a population: a simple susceptible-infected-recovered model and a 5-dimensional process modelling an outbreak of Ebola (Legrand et al., 2007). We then explain how these connections imply that the theory we develop for stochastic differential equations can apply to these models, opening up a whole new range of applications, particularly in mathematical biology and epidemiology.

Chapter 2

Background

We begin this thesis by introducing the required technical back, most notably defining the flow map of a deterministic system and building up to Itô stochastic differential equations to incorporating uncertainty into these models. Further details on the technical results and tools employed later are provided in Appendix A. The chapter is concluded by summarising stochastic sensitivity (Balasuriya, 2020b), which provided initial tools for characterising uncertainty in a computationally efficient manner and was the primary motivation for this work.

2.1 Notation

We adopt several notational conventions throughout this thesis, which are summarised here. The set of $n \times m$ matrices with real-valued entries is denoted as $\mathbb{R}^{n\times m}$. The norm symbol $\|\cdot\|$ with no 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

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

For a random variable X, we use $\mathbb{E}[X]$ to denote the expectation of X and $\mathbb{V}[X]$ to denote the variance. For a n-dimensional vector-valued random variable Y, $\mathbb{E}[Y]$ again denotes the (now vector-valued) expectation of Y, and $\mathbb{V}[Y]$ denotes the covariance matrix of Y. That is, $\mathbb{V}[Y]$ is the symmetric $n \times n$ matrix with (i, j)th component

$$\big[\mathbb{V}[Y]\big]_{ij} = \mathbb{E}\big[Y_iY_j\big] - \mathbb{E}[Y_i]\,\mathbb{E}\big[Y_j\big] = \mathrm{cov}\big(Y_i, Y_j\big) = \begin{cases} \mathbb{V}[Y_i]\,, & \text{if } i = j, \\ \mathrm{cov}\big(Y_i, Y_j\big)\,, & \text{otherwise,} \end{cases}$$

where

$$\mathrm{cov}\big(Y_i, Y_j\big) = \mathbb{E}\Big[\big(Y_i - \mathbb{E}[Y_i]\big)\,\Big(Y_j - \mathbb{E}\big[Y_j\big]\Big)\Big]$$

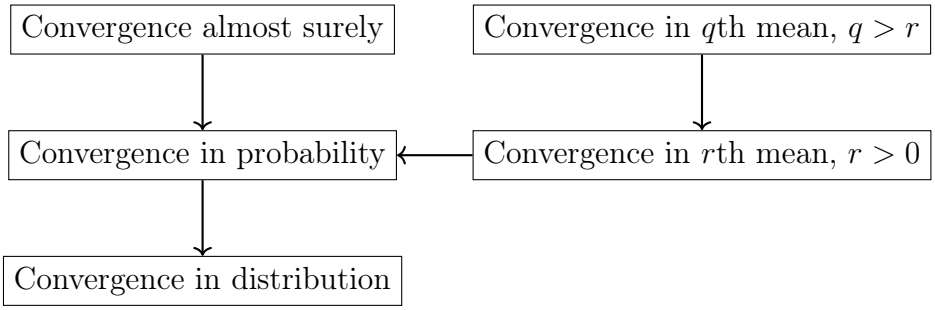


Figure 2.1: The strength of each notion of convergence for random variables, where each directed arrow corresponds to an implication. These results are stated and proven in Brémaud (2020), for instance.

denotes the (scalar) covariance between Y_i and Y_j.

There are several 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

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

  Note that the limiting random vector X need not be defined on the same probability space as the terms in the sequence $X_1, X_2, \ldots$; convergence in distribution is the only notion of convergence for which this is the case.

- 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

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

- 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

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

This type of convergence is also known as *L_r-convergence*, as it corresponds to convergence in the L_r norm on the probability space on which $X_1, \ldots, X_n$ and X are defined.

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 2.1.

We often work with stochastic processes that evolve through time, for example, the Wiener process or the solution to a stochastic differential equation. We generically use the notation y_t, for instance, to denote the stochastic process, where t refers to time and y is a label. The index set of the process is always a finite time interval, typically initialised at 0. In almost all cases where we are working with y_t mathematically, y_t refers to the state of the process at the *fixed* time t. When discussing a stochastic process in our prose, we also use y_t to refer to the *full* solution path (where the possible values of t are known in context); this is an abuse of notation, but has no impact on the rigour of our mathematical results.

2.2 The flow map

Ordinary differential equations are used to model many different phenomena across a range of fields and applications. Specifically, the continuous time evolution of a multi-dimensional state variable is governed by a system of first-order differential equations of the form

$$\frac{\mathrm{d}w_t}{\mathrm{d}t} = u(w_t, t)\,, \tag{2.1}$$

where $w_t \in \mathbb{R}^n$ is the time-evolving variable of interest and u is the vector field specified at each relevant state and time t. The vector field u may be derived from a specified model or may be driven or supplemented by observed data. Since data has a finite-time limitation, we typically consider the evolution of (2.1) over a finite time interval $[0, T]$. We can solve (3.2) analytically, or, as is often required in practice, numerically to generate trajectories,

which can inform future predictions or be used to reconstruct past behaviour. The flow map of (2.1) provides a convenient way of summarising the trajectories that solve (2.1) and working with these solutions analytically. Formally, 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 $\tau = t$. That is, the flow map is the operator mapping initial conditions at time t to their corresponding positions at time s, under the continuous-time evolution of (2.1). Equivalently, the flow map satisfies the integral form of (2.1):

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

We use the flow map to represent all solutions of the deterministic differential equation (2.1), understanding that for any relevant initial condition, the flow map at that point can be readily computed by solving the ODE either analytically or numerically. For a well-defined flow, the flow map F_s^t is reversible, in that we could have $t < s$, with the property

$$F_s^t\big(F_t^s(x)\big) = x,$$

for any times s and t. We view the flow map primarily as a function of the initial condition, so that it quantifies the impact of changes in the initial condition on future predictions. Accordingly, the gradient ∇F_s^t of the flow map (with respect to the initial condition) provides insight into the local behaviour of the dynamical system (V. I. Arnold, 1973; Truesdell and Noll, 2004). For any times $s, t \in [0, T]$, this gradient ∇F_s^t satisfies a useful property:

$$\frac{\partial \nabla F_s^t(x)}{\partial t} = \nabla u\big(F_s^t(x), t\big)\, \nabla F_s^t(x), \tag{2.3}$$

which is the equation of variations corresponding to (2.1). This result can be shown by taking the gradient of both sides in (2.2) and using the chain rule.

2.3 From ODEs to SDEs

Differential equations are well-studied and ubiquitously employed across many fields. These models are *deterministic*, in that, given an initial condition, an ODE provides a single prediction of the future state. However, as discussed in the introduction, any such model will be subject to unavoidable uncertainty, which can arise from a range of sources. By accounting for any of these in our model, we expect to improve the power and accuracy of our predictions. However, the true nature of this uncertainty is unknowable, so it is common to model it as a random process. We extend our deterministic model to a

stochastic one, where the solution (and therefore our predictions of the future state) is now a random variable with a probability distribution defined over the state space. There are several approaches to establishing such a framework, the most general of which is a random dynamical system (L. Arnold, 1998; Neckel and Rupp, 2013). When working in a continuous-time and continuous-state modelling scenario, a natural extension of an ordinary differential equation to account for this uncertainty is a stochastic differential equation (SDE) (Øksendal, 2003; Kallianpur and Sundar, 2014), where the uncertainty is modelled with an additional term. Although an SDE is an "improvement" over the deterministic ODE, in the sense that uncertainty can be encoded in the model itself, the compromise is that solutions are no longer straightforward to obtain or even analyse. In fact, the formal treatment of SDEs requires an entirely new notion of integration, as we shall discuss. Nonetheless, SDEs provide a rich framework for modelling and are used as predictive tools in a range of applications and fields. There are several distinct ways to construct stochastic differential equations from deterministic dynamics; the following motivation of stochastic differential equations loosely follows a similar one available in Øksendal (2003), but other formulations of SDEs are available, e.g. using rough path theory (Friz and Victoir, 2010) or by averaging the 'fast' dynamics in a multiscale system (Pavliotis and Stuart, 2008; Melbourne and Stuart, 2011; Gottwald and Melbourne, 2013).

Suppose that we have a deterministic ordinary differential equation, i.e. (3.2), and aim to account for uncertainty in the vector field u. In lieu of any additional understanding of this uncertainty, we model it as a stochastic noise. Ideally, we would parameterise the uncertainty with some continuous-time stochastic process ξ_t, and write

$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = u\left(x_t, t\right) + \sigma(x_t, t)\,\xi_t, \tag{2.4}$$

where σ is a specified part of the model that modulates how the magnitude and structure of the uncertainty varies with state and time. Since we are modelling with a continuum, the noise process ξ_t should be (with probability 1) continuous in time, and should, with no additional prior knowledge, evolve independently of itself through time, follow a probability distribution that does not depend on time, and have zero expectation. That is, xi_t should be a white noise process in the absence of any additional knowledge of the nature of the noise, beyond the specification of σ. However, there does not exist *any* process x_t that satisfies all of these properties simultaneously (Øksendal, 2003); continuity cannot be enforced. This complication requires an alternative formulation, which leads to an entirely new type of calculus built around the Itô integral and Itô stochastic differential equations, which we introduce in Section 2.3.2 and Section 2.3.3 respectively.

2.3.1 The Wiener process

In the absence of any additional knowledge about the noise (such as skew or heavy-tailedness), the canonical Wiener process is the standard choice as the driving stochastic

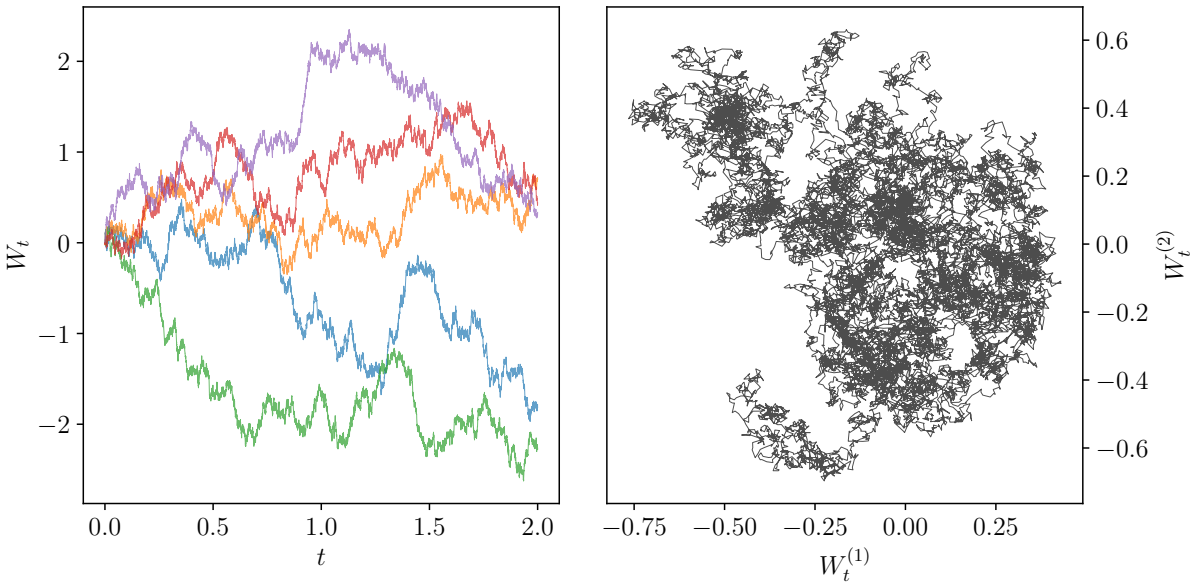


Figure 2.2: (Left) Several realisations of a 1-dimensional Wiener process W_t evolving through time, and (right) a realisation of 2-dimensional Wiener process $\left(W_t^{(1)}, W_t^{(2)}\right)^\top$.

process. The Wiener process is the definite integral of a white noise process and is, therefore, an appropriate choice to ensure that the "solution" (the result after integration through time) of (2.4) involves the idealised noise process ξ_t that we were initially after. Defined formally, the (1-dimensional) canonical *Wiener process* is a stochastic process B_t that takes values in $\mathbb{R}$ and satisfies the following properties (Kallianpur and Sundar, 2014):

(i) $B_0 = 0$ almost surely,

(ii) for every $s > 0$, the increments $B_{s+t} - B_s$ for $t \geq 0$ are independent of B_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.

Remarkably, these properties *uniquely* define the Wiener process, with the additional result that for any $t > 0$, B_t is distributed as $\mathcal{N}(0, t)$, a Gaussian distribution with mean zero and variance t. The *n-dimensional Wiener process* is a stochastic process W_t that takes values in $\mathbb{R}^n$ and is such that each component of W_t is a 1-dimensional Wiener process and the components of W_t are mutually independent. 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. Figure 2.2 plots realisations of a 1-dimensional and 2-dimensional Wiener process.

2.3.2 The Itô integral

With probability 1, the path of a Wiener process is continuous, but differentiable nowhere, so standard deterministic calculus is not sufficient to introduce continuous-time uncertainty into a differential equation. Instead, this led to an entirely new definition of the integral by Itô (1944, 1946) that allowed integration with respect to a broad class of stochastic processes, laying the foundation for the formal framework of stochastic differential equations. Here, we provide one definition of the Itô integral but do not go into technical detail. See the textbooks by Kallianpur and Sundar (2014) and Øksendal (2003), for instance, for a detailed introduction to and construction of the Itô integral and subsequent properties. We only consider integrals with respect to canonical Wiener processes, but generalisations of the driving process are possible (Applebaum, 2004). For our purposes, we can define an Itô integral as the limit in probability of a sequence of sums: for a scalar but possibly random-valued function $f\colon [a,b] \to \mathbb{R}$, the *Itô integral* of f with respect to the Wiener process W_t is the limit

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

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. The Itô integral is itself a random variable. It can be shown (Øksendal, 2003; Kallianpur and Sundar, 2014, e.g.) that this limit exists for a large class of both deterministic- and random-valued functions, by constructing appropriate approximations of the function f. There are several other definitions of the stochastic integral, the most common alternative being the Stratonovich integral (Stratonovich, 1966), which results from a different interpretation of the noise term in (2.4) and is often used in physics. The Stratonovich integral in particular can be re-interpreted as an Itô integral with an appropriate transformation of the integrand, so we focus our attention in this thesis solely on the Itô formulation of stochastic calculus.

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 the n-dimensional vector

$$\int_a^b g(t)\,\mathrm{d}W_t \coloneqq \left(\mathcal{I}_1, \ldots, \mathcal{I}_n\right)^\top,$$

where

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

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

The Itô integral behaves similarly to classical notions of the integral, including acting as a linear operator. For any Itô-integrable functions $f, g\colon [a,b] \to \mathbb{R}$ and values $\alpha, \beta \in \mathbb{R}$, which may be random but are constant with respect to t:

1. Linearity:
$$\int_a^b \big[\alpha f(t) + \beta g(t)\big]\,\mathrm{d}W_t = \alpha \int_a^b f(t)\,\mathrm{d}W_t + \beta \int_a^b g(t)\,\mathrm{d}W_t.$$
2. Zero expectation:
$$\mathbb{E}\left[\int_a^b f(t)\,\mathrm{d}W_t\right] = 0.$$
3. The Itô isometry:
$$\mathbb{E}\left[\left(\int_a^b f(t)\,\mathrm{d}W_t\right)^2\right] = \int_a^b \mathbb{E}\Big[f(t)^2\Big]\,\mathrm{d}t.$$

The first two of these properties immediately extend to Itô integrals of vector- and matrix-valued functions. The third property, the Itô isometry, is a fundamental result that enables the calculation of the variance of an Itô integral.

2.3.3 Itô stochastic differential equations

Equipped with the Itô integral as a formal definition of an integral with respect to a stochastic process, we can now extend the notion of an ordinary differential equation to include stochasticity. 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 (or sometimes diffusion) matrix. The driving process W_t is the m-dimensional Wiener process. There is a heuristic interpretation of (2.6): over a small time interval $(t, t+\delta t)$, the value of y_t changes by a Gaussian increment with expected value $u(y_t,t)\,\delta t$ and variance $\sigma(y_t,t)\,\sigma(y_t,t)^\top\,\delta t$. The product $\sigma\sigma^\top$ can therefore be informally seen as the variance of the noise term. In the most general case, the drift u and diffusivity σ are permitted to be random functions (Kallianpur and Sundar, 2014), but in this thesis, we assume that both are deterministic. The notation in (2.6) is not rigorously defined, but rather taken as equivalent to the integral form
$$y_t = y_0 + \int_0^t u\left(y_\tau, \tau\right)\mathrm{d}\tau + \int_0^t \sigma\left(y_\tau, \tau\right)\mathrm{d}W_\tau. \tag{2.7}$$

where y_0 is the possibly random initial condition. The integral form (2.7) provides the rigorous foundation of the stochastic differential equation, overcoming the difficulties of working with Wiener processes by introducing the Itô integral to compute the ongoing contributions from the stochastic terms in the equation. As with ordinary differential equations, under certain conditions on the drift u and diffusivity σ there exist solutions to the SDE (2.6) that are unique in some sense—see Theorem 6.2.1 of Kallianpur and Sundar (2014), for instance.

There are many analytic tools available for working with stochastic differential equations, such as Itô's Lemma, an analogy of the chain rule. The results that are used in this thesis (most notably in the proofs presented in Chapter 3) are summarised in Appendix A.2.

The diffusivity matrix σ characterises the spatiotemporal structure of the noise. When σ does not depend on the solution y_t, the noise is termed *additive*, whereas if σ depends on the solution, then the noise is *multiplicative*. SDEs with additive noise are typically easier to solve and analyse than those with multiplicative noise (Sancho et al., 1982), but multiplicative noise is often required in practice to capture uncertainty that varies with state.

As an example, in Figure 2.3 we show 10 realisations of the solution to the SDE $\mathrm{d}x_t = \sin(x_t)\,\mathrm{d}t + \mathrm{d}W_t$. At any fixed time t, the solution x_t follows a probability distribution over $\mathbb{R}$, the (numerically estimated—see Section 2.3.4) probability density function of which is shown at $t = \pi$ on the right-hand side of Figure 2.3. Although the deterministic dynamics can give a loose indication of the behaviour of the stochastic samples, even additive noise can result in complicated behaviour in the stochastic system and departures from the deterministic solutions when the drift is nonlinear, evidenced by the trajectory (in orange) that deviates from the deterministic one.

2.3.4 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 involves random sampling (typically of the driving noise process) and produces approximate *realisations* of the solution. With a large number of these Monte Carlo realisations, one can estimate statistical properties and approximate the distribution of the solution. The stochastic sampling approach is the gold standard in many applications, most notably climate and weather modelling (Collins, 2007).

The simplest scheme for numerically solving SDEs is the Euler-Maruyama (EM) method, which is analogous to the Euler method for ODEs (Kloeden and Platen, 1992). The update step of the EM scheme, with step size δt, is

$$\hat{x}_{t+\delta t} = \hat{x}_t + \delta t u(\hat{x}_t, t) + \sqrt{\delta t}\sigma(\hat{x}_t, t)\, Z_t, \tag{2.8}$$

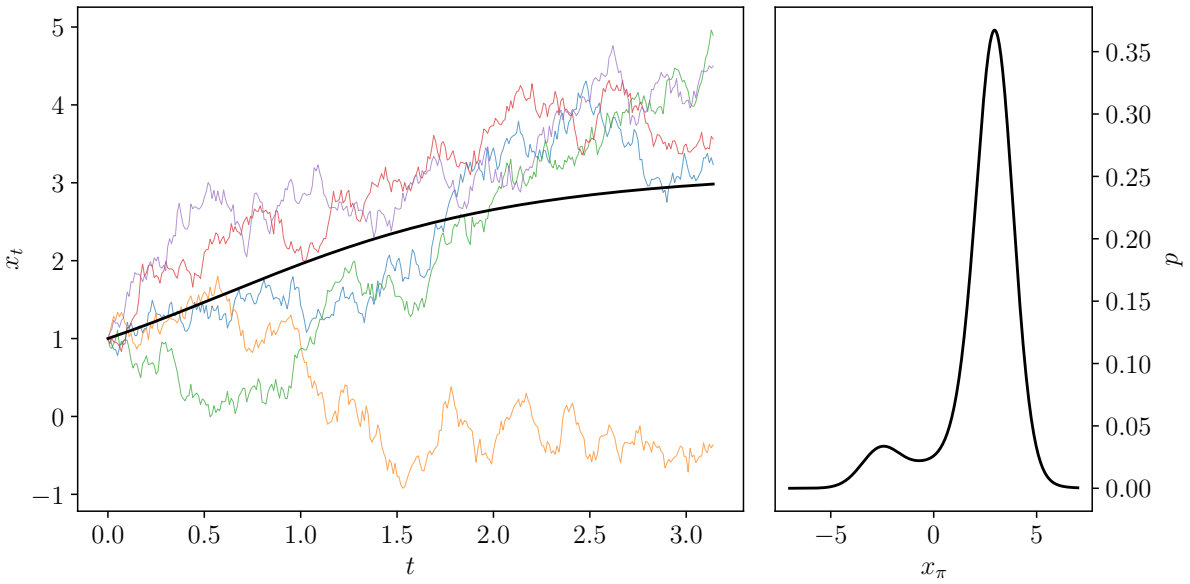


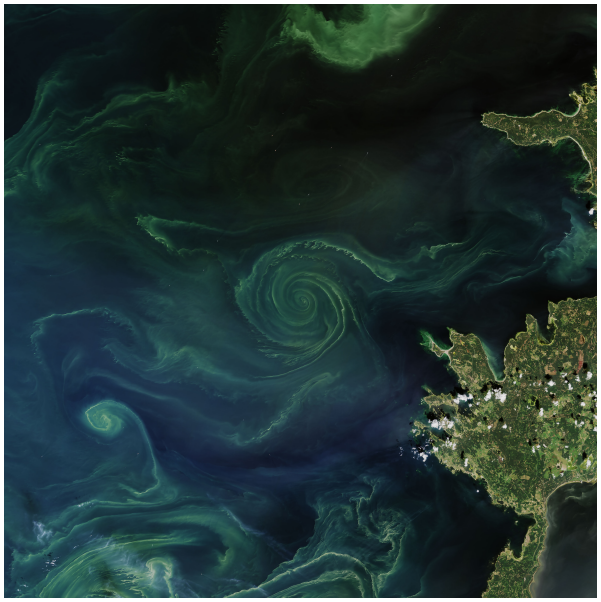
Figure 2.3: (Left) Sample paths of the solution to the stochastic differential equation $\mathrm{d}x_t = \sin(x_t)\,\mathrm{d}t + \mathrm{d}W_t$, from the initial condition $x_0 = 1$ and over the time interval $(0, \pi)$. The solution to the corresponding deterministic system $\frac{\mathrm{d}w_t}{\mathrm{d}t} = \sin(w_t)$ with the same initial condition is in black. (Right) The numerically estimated probability density function of the solution x_π, using 10000 samples.

where Z_t is sampled from the standard Gaussian $\mathcal{N}(0, I)$, and the scheme is initialised as $\hat{x}_0 = x_0$. When the initial condition is random, one samples $\hat{x}_0$ from that distribution. There are many other schemes for generating approximate samples of a stochastic differential equation, of varying precision and computational complexity, many of which are given in Kloeden and Platen (1992). Numerical schemes give us access to approximate solutions to otherwise intractable SDEs. However, this comes at a computational cost: a large number of sample paths is often required to generate convergent statistics and make accurate inferences. One of the primary aims of this thesis is to overcome this expense by devising alternative ways of approximating and characterising SDE solutions that are computationally cheaper.

2.4 Lagrangian coherent structures

In this section, we take a brief sojourn into the field of Lagrangian coherent structures (LCSs), which provide qualitative insight into the behaviour of a dynamical system, particularly in the context of fluid flows. Although LCSs are not the primary focus of this work, the field provides potential applications for our uncertainty quantification, that

(a) An oil slick from the *Deepwater Horizon* oil spill in 2010 (NASA Image of the Day, May 19, 2010). The southwards extension of the slick appeared unexpectedly.



(b) Phytoplankton blooms in the Baltic Sea (NASA Earth Observatory, July 18, 2018). The structure of these blooms reflect the underlying flow of the Sea.

(c) Jupiter's Great Red Spot, as photographed by the Voyager 1 probe (NASA/JPL-Caltech and processed by Björn Jónsson, March 5, 1979). The storm is an example of a vortex or eddy within the atmospheric flow of the planet.

Figure 2.4: Examples of coherent patterns emerging in fluid flows observed in nature.

builds upon preliminary work by Balasuriya (2020b,c) and Badza et al. (2023). There is no universally accepted definition of a coherent structure, but typically these are structures within a flow that remain together over the time-evolution of the system and separate the spatial domain into regions with qualitatively different behaviour (Balasuriya, Ouellette, et al., 2018). Figure 2.4 shows examples of coherent structures in observed fluid flows, which can be considered LCSs. These are structures such as vortices, eddies, and jets that influence the transport of material within the fluid. For instance, in Figures 2.4a and 2.4b, the observed patterns of the oil slick and phytoplankton blooms respectively reflect the behaviour of the underlying ocean flow and are influenced by jet- and eddy-like structures.

In a steady system (that is, the vector field u in (2.1) is independent of time), we can gain this insight by using classical methods in dynamical systems, such as phase portrait analysis and identifying unstable and stable manifolds. For example, solution trajectories cannot intersect unstable and stable manifolds, and so these manifolds can act as barriers for the transport of material within a flow. However, when the system is non-autonomous (that is, the vector field explicitly depends on time t), these structures can themselves vary with time and the problem of identifying them is far more non-trivial. Another complication is that in practice, the data driving a system is only available over a finite timeframe, whereas classical dynamical systems techniques often tell us about the long-term (in the infinite time limit) behaviour of solutions. Lagrangian coherent structure theory provides a mathematical framework for defining and identifying such structures within a flow (Balasuriya, Ouellette, et al., 2018). There are many procedures and heuristics for extracting these regions from a given flow, which draw upon different mathematical techniques, including classical dynamical system theory, variational calculus, transfer operators, and statistical clustering. Detailed reviews of approaches to Lagrangian coherent structure extraction are provided by Peacock and Dabiri (2010), Hadjighasem et al. (2017), and Balasuriya, Ouellette, et al. (2018).

Coherent structures can provide valuable qualitative insight into the behaviour of a flow, by providing an outline of the transport properties and underlying dynamics. An example is in the study of the spread of the oil slick resulting from the 2010 Deepwater Horizon disaster in the Gulf of Mexico (see Figure 2.4a). Investigations showed that the behaviour of the slick, including a sudden and unexpected extension of the slick, could be understood and therefore predicted in future cases by using Lagrangian coherent structures and the insight they provide (Mezić et al., 2010; Olascoaga and Haller, 2012; Olascoaga, Beron-Vera, et al., 2013). This approach explained dynamics that were otherwise poorly understood because of the time-varying and complex nature of the flow.

One of the most well-studied and frequently used procedures for extracting Lagrangian coherent structures is via the finite-time Lyapunov exponent (FTLE), which is a measure quantifying the stretching of infinitesimal regions of the flow over a time period. The FTLE can be computed as a scalar field over a set of initial conditions, from which the maximising ridges can correspond to flow barriers (Shadden et al., 2005). Importantly,

the finite-time Lyapunov exponent can be computed only using the gradients of the flow map, which ensures a highly practical and flexible procedure that can be used across many different contexts. The FTLE is an example of a common class of LCS methods that first compute a field over a set of initial conditions and then extract coherent structures based on that field.

Most well-established LCS frameworks and extraction procedures are purely deterministic, in that they are defined and computed solely in terms of the behaviour of the underlying ordinary differential equation. However, uncertainty in such systems is inevitable in practice and these methods fail to account explicitly for this. Accordingly, there is an emerging interest (Balasuriya, 2020a) in extending LCS theory to stochastic settings. There are two primary ways in which stochastic has been recently investigated in the LCS community:

(i) in creating novel procedures that explicitly account for such ongoing uncertainty, such as by using properties of corresponding stochastic systems. For example, see (in Section 2.5) stochastic sensitivity introduced by Balasuriya (2020b), model sensitivity introduced by Kaszás and Haller (2020), and the finite-time divergence rate by Branicki and Uda (2023). These are scalar fields defined on initial conditions that measure the certainty in the corresponding deterministic trajectories and can distinguish coherent regions in similar ways to deterministic methods. As another example, Denner et al. (2016) directly compute coherent sets by working with a discretised Fokker-Planck equation, which is a partial differential equation that governs the probability density function of an SDE solution (an overview is provided in Section 7.3). This Fokker-Planck approach extends the transfer operator method of Froyland (2013), which is a popular method for LCS detection and extraction by encoding how densities are pushed forward by the flow. Until recently, the transfer operator, at least in the context of Lagrangian analysis, has been viewed as purely deterministic but provides a framework for naturally including velocity field uncertainties (Balasuriya, 2020a).

(ii) in understanding the direct impact of velocity uncertainty on well-established deterministic LCS measures. Badza et al. (2023) provide a systematic analysis, using Monte Carlo simulation and summary statistics to evaluate the robustness of several common LCS extraction schemes to velocity uncertainty. The finite-time Lyapunov exponent has received particular attention, with recent studies aiming to quantify the impact of velocity field uncertainty on the FTLE computation: Guo et al. (2016) use stochastic simulation and statistical analysis, Balasuriya (2020c) provides theoretical error bounds on the FTLE computation, and You and Leung (2021) propose an approach for computing the (statistically) expected FTLE field.

In this thesis, we are primarily interested in point (i), by exploring how our characterisations of uncertainty can be applied to extract coherent structures. This is directly extending

the stochastic sensitivity of Balasuriya (2020b), which is summarised in Section 2.5. We will also discuss (in Section 7.6) how we anticipate our work could be applied to point (ii), as a quantification of uncertainty in computations involving the flow map in other LCS schemes.

2.5 Stochastic sensitivity

To conclude our background and motivation, this section summarises stochastic sensitivity, a measure of uncertainty in differential equations introduced by Balasuriya (2020b). These tools are *computable* given only velocity data, which enables an efficient quantification of uncertainty in a stochastic system with no need for bulk simulation. Stochastic sensitivity (also termed S^2 in both notation and prose) was originally provided for 2-dimensional systems only, and the primary motivation was to understand the impact of velocity field uncertainty on fluid flows. Given possibly time-dependent velocity data $u\colon \mathbb{R}^2\times[0,T]\to\mathbb{R}^2$, Balasuriya (2020b) considers the evolution of solutions to the ordinary differential equation

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = u(x_t,t)\,, \tag{2.9}$$

subject to some fixed initial condition x_0. The velocity field u is *Eulerian*, in that it describes the fluid velocity at a given point in space and time. The trajectories that solve (2.9) are *Lagrangian* and correspond to the movement of idealised infinitesimal particles within the flow. These Lagrangian trajectories are summarised by the flow map F_s^t of (2.9). As we have continued to emphasise, the velocity field u is in practice subject to unavoidable uncertainties. Balasuriya (2020b) aims to quantify the impact of this Eulerian uncertainty, directly attributed to u, on the Lagrangian trajectories arising from integrating the velocity field.

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

$$\mathrm{d}y_t = u(y_t,t)\,\mathrm{d}t + \varepsilon\sigma(y_t,t)\,\mathrm{d}W_t, \quad y_0 = x_0, \tag{2.10}$$

where $0<\varepsilon\ll 1$ is a parameter quantifying the scale of the noise, $\sigma\colon \mathbb{R}^2\times[0,T]\to\mathbb{R}^{2\times 2}$ is the 2×2 diffusion matrix, and W_t is the canonical 2-dimensional Wiener process. In the original formulation (Balasuriya, 2020b), 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 Balasuriya (2020c), Fang, Balasuriya, et al. (2020), and Badza et al. (2023) for example). Since σ can vary by both space and time, the noise is permitted to be 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, physics-informed models, etc. If no such prior information is known, then $\sigma\equiv I$, the 2×2 identity matrix, is the default choice.

Next, to quantify uncertainty at a time t, Balasuriya (2020b) defined the random variable $z_\varepsilon(x_0)$ as

$$z_\varepsilon(x_0) \coloneqq \frac{y_t - F_0^t(x_0)}{\varepsilon},$$

which captures the random deviation between the "true" stochastic trajectories and the deterministic flow map. The aim was to compute certain statistics of z_ε. To derive such quantities that can be computed in practice, Balasuriya (2020b) considers the signed projection of $z_\varepsilon(x_0)$ onto a ray emanating from the deterministic position $F_0^t(x_0)$ in the direction θ, defining

$$P_\varepsilon(x_0, \theta) \coloneqq \hat{n}^\top(\theta)\, z_\varepsilon(x_0), \quad \hat{n}(\theta) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.$$

where $\theta \in \big[-\pi/2, \pi/2\big)$. The statistics of $z_\varepsilon(x_0)$ and $P_\varepsilon(x_0, \theta)$ are considered in the limit as $\varepsilon \downarrow 0$, which provides a characterisation of the uncertainty of the model that is *independent* of the scale of the noise. Balasuriya (2020b) defines two measures of uncertainty from the variance of P_ε in this limit:

Definition 2.1 (Balasuriya 2020b) *a) The **anisotropic uncertainty** is a scalar field $A : \mathbb{R}^2 \times \big[-\pi/2, \pi/2\big) \to [0, \infty)$ defined by*

$$A(x_0, \theta) \coloneqq \sqrt{\lim_{\varepsilon \downarrow 0} \mathbb{V}\big[P_\varepsilon(x_0, \theta)\big]}.$$

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

$$S^2(x_0) \coloneqq \lim_{\varepsilon \downarrow 0} \sup_\theta \mathbb{V}\big[P_\varepsilon(x_0, \theta)\big].$$

The anisotropic uncertainty is a measure of the uncertainty in a specified direction θ, whereas stochastic sensitivity is a scalar field which, for a given initial condition measures the uncertainty in the corresponding Lagrangian trajectory. Figure 2.5 shows a pictorial representation of the set-up in two dimensions: the deterministic flow (2.9) (in black) takes the initial condition and provides a computable prediction of the state at time t. Simultaneously, the solution to the stochastic system (2.10) (in green) gives a different, random value y_t for the true position at time t. We take the difference (in blue) between the deterministic position and the stochastic and project (in red) this vector onto a ray of angle θ. The anisotropic uncertainty in the direction of θ is then calculated by computing the variance of $P_\varepsilon(x, \theta)$ and taking the $\varepsilon \downarrow 0$ limit. By maximising this limiting variance across all angles θ, we get the stochastic sensitivity value, a single scalar number associated with the initial condition x. Using techniques from both deterministic and stochastic calculus, Balasuriya (2020b) further established expressions for both the anisotropic uncertainty and the stochastic sensitivity that are computable given only the flow map and velocity data.

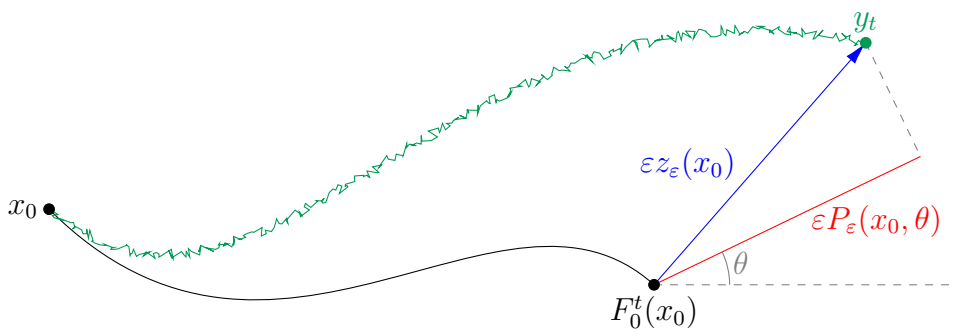


Figure 2.5: The entities used in the definition of stochastic sensitivity, including the mapping (in black) from the deterministic flow map solving (2.9) and the 'true', but random, trajectory that solves the stochastic equation (2.10) (in green).

Theorem 2.1 (Balasuriya 2020b) *For $x_0 \in \mathbb{R}^2$, set $w := F_0^t(x_0)$ and fix $t \in [0, T]$. Then, for any $\theta \in \big[-\pi/2, \pi/2\big)$,*

$$A(x_0, \theta) = \left(\int_0^t \left\| \Lambda(x_0, \tau)\, J\hat{n}(\theta) \right\| \mathrm{d}\tau \right)^{1/2},$$

where

$$\Lambda(x_0, \tau) := e^{\int_\tau^t [\nabla \cdot u]\left(F_\tau^\xi(x_0), \xi\right) \mathrm{d}\xi} \sigma \left(F_0^\tau(x), \tau\right)^\top J \nabla_w F_t^\tau(w),$$

with the gradient ∇_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

$$S^2(x_0) = P(x_0) + N(x_0),$$

with

$$
\begin{aligned}
L(x_0) &\coloneqq \frac{1}{2}\sum_{i=1}^{2}\int_0^T \left[\Lambda_{i2}(x_0,\tau)^2 - \Lambda_{i1}(x_0,\tau)^2\right] \mathrm{d}\tau \\
M(x_0) &\coloneqq \sum_{i=1}^{2}\int_0^T \Lambda_{i1}(x_0,\tau)\,\Lambda_{i2}(x_0,\tau)\,\mathrm{d}t \\
N(x_0) &\coloneqq \sqrt{L^2(x_0) + M^2(x)} \\
P(x_0) &\coloneqq \left|\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\int_0^T \Lambda_{ij}(x_0,\tau)^2\,\mathrm{d}t\right|,
\end{aligned}
$$

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

Proof. See the appendices of Balasuriya (2020b). □

Balasuriya (2020b) also provides nonlinear scalings of stochastic sensitivity that are informed by the spatial resolution and diffusivity scale in the specific model. However, stochastic sensitivity provides a theoretical field that needs no reference to any of these physical considerations, and in this thesis, we will only consider the unscaled field. We do note that taking the logarithm of the stochastic sensitivity field has often proved useful for visualisation (Badza et al., 2023, e.g.).

Since stochastic sensitivity provides a scalar field on the set of initial conditions to (2.9), Balasuriya (2020b) also proposes that stochastic sensitivity can distinguish spatial regions of interest within the flow in the spirit of Lagrangian coherent structure analysis. A *robust set* is a subset of initial conditions such that the stochastic sensitivity is below a specified threshold. That is, a robust set comprises those initial conditions for which the uncertainty in the corresponding flow trajectory, as measured by stochastic sensitivity, is small enough. Given a threshold R, a robust set is defined as

$$
\mathrm{RS}(R) \coloneqq \left\{x \in \Omega_0 \,\middle|\, S^2(x) < R\right\},
$$

where $\Omega_0 \subset \mathbb{R}^2$ is the domain of initial conditions of interest. Numerical experiments by Balasuriya (2020b), and Badza et al. (2023) suggest that such sets can correspond to coherent structures within the flow. Unlike a majority of previous methods for identifying LCSs which treat the dynamical system as fully deterministic and ignore any uncertainty, stochastic sensitivity, and therefore the extracted robust sets, explicitly accounts for stochasticity and is therefore highly novel within the LCS community. The threshold R can be informed by lengthscales from the resolution (e.g. the resolution of the underlying velocity data, if the noise in (2.10) is thought of as accounting for subgrid effects) and advective and diffusivity properties of the flow (such as the Péclet number)—see Balasuriya (2020b) for details on this. However, when we later use a similar idea to extract robust sets,

we select R with no prior knowledge and instead choose a value that results in coherent structures that match our qualitative expectations.

Since stochastic sensitivity is only a recent development, it has only been applied in a limited number of places so far:

- Balasuriya (2020c) uses stochastic sensitivity to compute an error bound for the finite-time Lyapunov (FTLE) computation. The computable S^2 value is used as an estimate of the standard deviation of the deviation between the deterministic and "true" (stochastic) trajectories, leading to a computable error bound on the FTLE value.

- Fang, Balasuriya, et al. (2020) use stochastic sensitivity in an investigation of the interplay between three different sources of error in dynamical systems: errors from unresolved processes because of limited resolution, errors from limited precision, and the inherent stochasticity in the dynamical behaviour of the system.

- Fang and Ouellette (2021), extending previous work by Fang, Balasuriya, et al. (2019), use stochastic sensitivity to construct spatial objects that assess the information content (in an information-theoretic sense) of a dynamical system. Stochastic sensitivity is used to provide a state-dependent lengthscale to assess whether the inherent stochasticity in the system dominates over resolution and linearisation errors.

- Badza et al. (2023) investigate the impact of velocity uncertainty on Lagrangian coherent structures extracted as robust sets with stochastic sensitivity. The stochastic model (2.10) is used to generate realisations of Lagrangian trajectories subject to noise on the velocity field, and summary statistics are used to evaluate the stability of a range of coherent structure techniques to this noise. By directly capturing such uncertainty as a means of coherent set Badza et al. (2023) showed that robust sets extracted with stochastic sensitivity typically remain stable under perturbations, since the framework directly accounts for this stochasticity unlike other LCS methods.

However, as stochastic sensitivity has only been defined in 2-dimensions, these studies are also all restricted to 2-dimensional systems. This is a significant limitation of the original work, and one we will overcome; in Chapter 3, our primary theoretical chapter, we will present a new definition of stochastic sensitivity in n-dimensions that maintains the computability of the measure. 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, and accordingly, the extension of all the aforementioned studies and applications. Moreover, Balasuriya (2020b) only computes the expectation and variance of the projections $P_\varepsilon(x, \theta)$, which does not describe the full distribution under the $\varepsilon \downarrow 0$ limit. The computational formula for the anisotropic uncertainty and stochastic

sensitivity in Theorem 2.1 requires knowledge of the divergence $\nabla \cdot u$ of the velocity field, and the computation of several integrals. While these expressions are computable given only velocity data, they are cumbersome. We also address these difficulties by directly relating stochastic sensitivity to the theory of stochastic differential equation linearisations and provide a computation for the value that only requires taking the operator norm of the matrix solution to an ordinary differential equation.

Chapter 3

Characterising SDE linearisations: the theory

In general, stochastic differential equations cannot be solved analytically and instead require computationally expensive numerical simulation. An alternative approach is to approximate the SDE by a simplified one, which can be solved analytically. A linearisation procedure is one such approach when the noise is small, in which the coefficients of the SDE are replaced with first-order Taylor expansions. As we highlighted in the introduction (Chapter 1), this linearisation scheme provides a practical approach to using SDEs and is accordingly used across a range of literature and applications (LeGland and Wang, 2002; Archambeau et al., 2007; Jazwinski, 2014; Law et al., 2015; Reich and Cotter, 2015; Sanz-Alonso and Stuart, 2017; Budhiraja et al., 2019; Särkkä and Solin, 2019; Kaszás and Haller, 2020).

In this chapter, we build upon previous small-noise studies of SDE linearisations (Blagoveshchenskii, 1962; Freidlin and Wentzell, 1998; Sanz-Alonso and Stuart, 2017) to provide an explicit bound for the error between a general class of stochastic differential equations and linearisations thereof about trajectories of a corresponding deterministic system. Our framework accounts for non-autonomous coefficients, multiplicative noise, and uncertain initial conditions in a stochastic differential equation of arbitrary dimension—a more general situation than that considered by any of Blagoveshchenskii (1962), Freidlin and Wentzell (1998), or Sanz-Alonso and Stuart (2017). In Section 3.1, we state the bound, written in terms of scales of the initial and the ongoing uncertainty, and provide an explicit characterisation of the solution to the linearised SDE including computations for the first two moments. We directly compare our newly derived bound in Section 3.1.1 to that on the KL-divergence by Sanz-Alonso and Stuart (2017) and postulate that our bound is tighter.

The second contribution of this chapter is to extend the original formulation of stochastic sensitivity by Balasuriya (2020b), in Section 3.2. We provide a new definition of stochastic sensitivity for n-dimensions and establish that the value can be computed from the solution

of a linearised SDE.

Much of the content in this chapter and the following (Chapter 4) has been submitted as a research article to *Communications in Mathematical Sciences* (Blake et al., 2023), and is currently under review. The preprint is available on arXiv at arXiv:2309.16334. Section 3.2.1, which discusses the connections between stochastic sensitivity and the finite-time Lyapunov exponent, does not appear in the submitted article and is instead a new contribution in this thesis.

3.1 Convergence of a SDE to a linearisation

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

$$\mathrm{d}y_t^{(\varepsilon)} = u\left(y_t^{(\varepsilon)}, t\right)\mathrm{d}t + \varepsilon\,\sigma\left(y_t^{(\varepsilon)}, t\right)\mathrm{d}W_t, \quad y_0^{(\varepsilon)} = x \tag{3.1}$$

where $u\colon \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ is the governing reference vector field. The canonical m-dimensional Wiener process W_t is a continuous white-noise stochastic process with independent Gaussian increments. The scale of the ongoing noise is assumed to be small and is parameterised as $0 < \varepsilon \ll 1$. The noise in (3.1) is multiplicative, in that the diffusion matrix $\sigma\colon \mathbb{R}^n \times [0,T] \to \mathbb{R}^{n\times m}$ can vary with state x, as well as with time t. We assume that σ is specified *a priori*, or if no such information is known, then $\sigma \equiv I$, the $n \times m$ identity matrix, is a default modelling choice. We consider (3.1) subject to the *general* uncertain initial condition $y_0^{(\varepsilon)} = x$, where x is an n-dimensional random vector with some given distribution. The two sources of randomness, x and W_t, are assumed independent from each other.

In the absence of any uncertainty (i.e. $\varepsilon = 0$ and the initial condition is a known deterministic quantity), (3.1) reduces to the ordinary differential equation

$$\frac{\mathrm{d}y_t^{(0)}}{\mathrm{d}t} = u\left(y_t^{(0)}, t\right), \quad y_0^{(0)} = x_0. \tag{3.2}$$

where the initial condition $x_0 \in \mathbb{R}^n$ is fixed. The formal convergence of the stochastic solution $y_t^{(\varepsilon)}$ (under certain conditions on the initial condition) to the deterministic $y_t^{(0)}$ in the limit as $\varepsilon \to 0$ is well-established using the large deviations principle (Freidlin and Wentzell, 1998, e.g). We refer to (3.2) as the *reference* deterministic model associated with (3.1). 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. Let the flow map $F_0^t\colon \mathbb{R}^n \to \mathbb{R}^n$ be the function which evolves an initial condition from time 0 to time t according to the flow of (3.2), i.e. $F_0^t(x_0) = y_t^{(0)}$.

We assume certain smoothness and boundedness conditions on the various terms outlined, which are stated explicitly in Hypothesis H. Throughout this article, 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.

Hypothesis H *Let the deterministic functions $u\colon \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ and $\sigma\colon \mathbb{R}^n \times [0,T] \to \mathbb{R}^{n\times m}$, and the random initial condition x be such that:*

(H.1) For all $t \in [0,T]$ flow map $F_0^t\colon \mathbb{R}^n \to \mathbb{R}^n$ is well-defined, and continuously differentiable (with respect to the initial condition) with invertible derivative.

(H.2) 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 $\sigma(\cdot,t) : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ given by $\sigma(x,t)$ is differentiable on $\mathbb{R}^n$.

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

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

Equivalently, for all $t \in [0,T]$, the function $u(\cdot,t)$ is Lipschitz continuous with Lipschitz constant $K_{\nabla u}$.

(H.4) For each $x \in \mathbb{R}^n$, the function $u(x,\cdot)\colon [0,T] \to \mathbb{R}^n$ and each component of the function $\sigma(x,\cdot)\colon [0,T] \to \mathbb{R}^{n\times m}$ are Borel-measurable on $[0,T]$.

(H.5) There exists a constant K_L such that for any $t \in [0,T]$ and $x \in \mathbb{R}^n$,

$$\left\|u\left(x,t\right)\right\| + \left\|\sigma\left(x,t\right)\right\| \leq K_L\left(1 + \|x\|\right).$$

(H.6) There exists a constant $K_{\nabla\sigma} \geq 0$ such that for any $t \in [0,T]$ and $x \in \mathbb{R}^n$,

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

and we take $K_{\nabla\sigma} = 0$ if there is no spatial dependence in σ. Equivalently, for all $t \in [0,T]$, the function $\sigma(x,\cdot)$ is Lipschitz continuous with Lipschitz constant $K_{\nabla\sigma}$.

(H.7) The initial condition x is defined on the same probability space as W_t, and is independent of W_t for all $t \in [0,T]$.

(H.8) There exists a constant $K_{\nabla\nabla u} \geq 0$ such that for any $t \in [0,T]$ and $x \in \mathbb{R}^n$,

$$\left\|\nabla\nabla u(x,t)\right\| \leq K_{\nabla\nabla u},$$

and we take $K_{\nabla\nabla} = 0$ if the second spatial derivatives of u are all zero.

(H.9) There exists a constant $K_\sigma \geq 0$ such that for any $t \in [0,T]$ and $x \in \mathbb{R}^n$,

$$\left\|\sigma(x,t)\right\| \leq K_\sigma.$$

The conditions H.2 to H.7 guarantee that (3.1) with the initial condition $y_0 = x$ has a unique strong solution (Kallianpur and Sundar, 2014). The bound $K_{\nabla\nabla u}$ placed on the second derivatives of u in H.8 quantifies exactly when the deterministic dynamics (that is, u) of (3.1) are linear. Similarly, the bound $K_{\nabla\sigma}$ on the spatial derivatives of σ in H.6 allows us to distinguish when the noise in (3.1) is multiplicative.

Our aim is to construct and formally justify a computable linearisation of (3.1) about a trajectory solving the deterministic system (3.2). To that end, we take a *fixed* initial condition $x_0 \in \mathbb{R}^n$ to the reference deterministic model (3.2) and consider linearising the SDE (3.1) about the corresponding trajectory $F_0^t(x_0)$. We consider the following linearisation of (3.1):

$$\mathrm{d}l_t^{(\varepsilon)} = \left[u\big(F_0^t(x_0)\,,t\big) + \nabla u\big(F_0^t(x_0)\,,t\big)\left(l_t^{(\varepsilon)} - F_0^t(x_0)\right)\right]\mathrm{d}t + \varepsilon\sigma\big(F_0^t(x_0)\,,t\big)\mathrm{d}W_t, \quad l_0^{(\varepsilon)} = x, \tag{3.3}$$

where the initial condition x is still permitted to be random. Informally, we can arrive at (3.3) by performing a Taylor expansion of the coefficient u up to first-order and σ to zeroth-order about the time-varying trajectory $F_0^t(x_0)$. Such a linearisation is advantageous over the nonlinear SDE (3.1), since (3.3) can be solved analytically. We will later (see Corollary 3.1) provide explicit expressions for computing the distribution of the solution $l_t^{(\varepsilon)}$ solely in terms of the solution dynamics of the deterministic system (3.2), the specified diffusion matrix σ, and the distribution of x.

In order to quantify the error arising from the choice of reference point x_0, we define

$$\delta_r := \mathbb{E}\left[\left\|x - x_0\right\|^r\right]^{1/r},$$

i.e. δ_r is the L_r distance between x and the deterministic point x_0. We can think of δ_r as a scalar measure of the uncertainty in the initial condition, relative to the choice of reference point x_0. Alternatively, the limit as δ_r approaches zero is equivalent to convergence in rth mean of x to the fixed point x_0. We can therefore distinguish two sources of uncertainty in our model; that arising from the initial condition, quantified by δ_r, and the ongoing uncertainty driven by the Wiener process W_t as measured by ε.

Our first and primary result, Theorem 3.1, provides an explicit bound on the rth moment of the error between the SDE solution $y_t^{(\varepsilon)}$ and the linearised solution $l_t^{(\varepsilon)}$.

Theorem 3.1 (Linearisation error is bounded) *Let $y_t^{(\varepsilon)}$ be the strong solution to the SDE (3.1) and $l_t^{(\varepsilon)}$ be the strong solution to the corresponding linearisation (3.3), both*

driven by the same Wiener process W_t and subject to the same random initial condition $y_0^{(\varepsilon)} = l_0^{(\varepsilon)} = x$. Then, for any $r \geq 1$ such that $\delta_{2r} < \infty$ and $t \in [0, T]$, there exist constants

$$D_1(r, t, K_{\nabla u}, K_\sigma)\,,\ D_2(r, t, K_{\nabla u})\,,\ D_3(r, t, K_{\nabla u}) \in [0, \infty)$$

independent of x and x_0 such that for all $\varepsilon > 0$,

$$\begin{aligned}\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq\ & (K_{\nabla\nabla u}^r + K_{\nabla\sigma}^r)\, D_1(r, t, K_{\nabla u}, K_\sigma)\ \varepsilon^{2r} \\ & + K_{\nabla\nabla u}^r D_2(r, t, K_{\nabla u})\, \delta_{2r}^{2r} + K_{\nabla\sigma}^r D_3(r, t, K_{\nabla u})\, \delta_r^r \varepsilon^r.\end{aligned} \tag{3.4}$$

Proof. See Section 3.3.2. Our proof employs the Burkholder-Davis-Gundy inequality, Grönwall's inequality, and Taylor's theorem to explicitly construct the bounding coefficients in terms of the conditions on the SDE coefficients set out in Hypothesis H. The bounding coefficients D_1, D_2, and D_3 are given explicitly in (3.33). □

In (3.4), we have an explicit scaling of the error in terms of ε and δ_r. The three terms can be informally considered as: a contribution purely from the ongoing linearisation error, a contribution purely from the initial uncertainty, and a term resulting from the interaction between the initial and ongoing uncertainties. By explicitly identifying the dependence of the bound on $K_{\nabla\nabla u}$ and $K_{\nabla\sigma}$, we note three special cases that are summarised by Remarks 3.1 to 3.3.

Remark 3.1 (Linear drift) *When the deterministic dynamics are linear, we can set $K_{\nabla\nabla u} = 0$ and (3.4) becomes*

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq K_{\nabla\sigma}^r D_1(r, t, K_{\nabla u}, K_\sigma)\ \varepsilon^{2r} + K_{\nabla\sigma}^r D_3(r, t, K_{\nabla u})\, \delta_r^r \varepsilon^r.$$

The linearisation of the drift term u is exact, so the error is purely due to the spatial dependency of the diffusion term σ.

Remark 3.2 (Additive noise) *When the noise in (3.1) is additive, we can set $K_{\nabla\sigma} = 0$ and (3.4) becomes*

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq K_{\nabla\nabla u}^r D_1(r, t, K_{\nabla u}, K_\sigma)\ \varepsilon^{2r} + K_{\nabla\nabla u}^r D_2(r, t, K_{\nabla u})\, \delta_{2r}^{2r}.$$

The error is then purely due to the linearisation of the drift term u, and as expected is of second order in both the initial condition uncertainty δ_{2r} and the ongoing uncertainty ε.

Remark 3.3 (Exact linearisation) *When the deterministic dynamics are linear and the noise in (3.1) is additive (non-multiplicative), the linearisation (3.3) should be exact. Accordingly, we can set $K_{\nabla\nabla u} = K_{\nabla\sigma} = 0$ and (3.4) becomes*

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] = 0.$$

In turn, this implies that $y_t^{(\varepsilon)} = l_t^{(\varepsilon)}$ almost surely, for any choice of reference point x_0.

We postpone a discussion of an additional special case—where the initial condition is fixed or Gaussian—to a later section. For the general situation, we next explicitly establish the solution to the linearisation (3.3), in terms of the initial condition and the deterministic evolution of (3.2).

Theorem 3.2 (Solution of the linearised SDE) *The linearised SDE (3.3) has the strong solution*

$$l_t^{(\varepsilon)} = \nabla F_0^t(x_0)\,(x - x_0) + F_0^t(x_0) + \varepsilon \nabla F_0^t(x_0) \int_0^t L(x_0, \tau)\,\mathrm{d}W_\tau. \tag{3.5}$$

where the term involving the uncertain initial condition x and the Itô integral are independent, and

$$L(x_0, \tau) \coloneqq \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\left(F_0^\tau(x_0), \tau\right). \tag{3.6}$$

Proof. See Section 3.3.3. □

The representation of the linearised solution as an independent sum in (3.5) can be seen as a decomposition into contributions from the initial uncertainty (the transformation of initial condition x), a deterministic prediction (the flow map $F_0^t(x_0)$) and the ongoing uncertainty in u (the remaining Itô integral term).

We can further show that the Itô integral term follows a Gaussian random variable, which ensures that the independent sum in (3.5) is a convenient expression for both theoretical analysis and numerical computation. We also provide explicit expressions for the mean and covariance matrix of the linearised solution, written in terms of the deterministic dynamics and σ.

Corollary 3.1 (Distribution of the linearisation solution) *The Itô integral term in the linearised solution (3.5) follows a Gaussian distribution independently of x, namely*

$$\int_0^t L(x_0, \tau)\,\mathrm{d}W_\tau \sim \mathcal{N}\left(0, \int_0^t L(x_0, \tau)\,L(x_0, \tau)^\top\,\mathrm{d}\tau\right).$$

The mean of the linearisation solution is

$$\mathbb{E}\left[l_t^{(\varepsilon)}\right] = F_0^t(x_0) + \nabla F_0^t(x_0)\,\mathbb{E}[x - x_0]\,. \tag{3.7}$$

The $n \times n$ covariance matrix of the linearisation solution is given explicitly by

$$\mathbb{V}\left[l_t^{(\varepsilon)}\right] = \nabla F_0^t(x_0)\left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0, \tau)\,L(x_0, \tau)^\top\,\mathrm{d}\tau\right)\left[\nabla F_0^t(x_0)\right]^\top \tag{3.8}$$

where $L(x_0, \tau)$ is as defined in (3.6) and the integral is taken in the elementwise sense.

Proof. See Section 3.3.4. The expressions follow from the representation of the linearised solution as an independent sum in (3.5). □

In Theorem 3.2, we have provided expressions for the distribution of the solution $l_t^{(\varepsilon)}$ to the linearised SDE (3.3) written solely in terms of the behaviour of the deterministic system (3.2), the specified diffusion matrix σ, and the distribution of the initial condition x. This describes a method for approximating the solution to the nonlinear SDE (3.1), or for characterising the impact of uncertainty in a dynamical system (3.2), that circumvents the need for expensive stochastic simulation.

Thus far, we have stated our results in terms of a general initial condition x, and provided expressions for the linearised solution in terms of this otherwise arbitrary distribution. However, we later consider two special cases for the initial condition x, a fixed (deterministic) initial condition in Section 3.1.3, and a Gaussian initial condition in Section 3.1.2. In both these cases, the linearised solution also follows a Gaussian distribution which is characterised entirely by the mean and covariance described in Corollary 3.1, allowing for easy computation. We also relate these results directly to other literature (Blagoveshchenskii, 1962; Freidlin and Wentzell, 1998; Jazwinski, 2014; Sanz-Alonso and Stuart, 2017; Särkkä and Solin, 2019; Balasuriya, 2020b) which uses linearisation procedures and Gaussian process approximations for nonlinear SDEs in these situations.

Next, we establish the ordinary differential equation satisfied by the covariance matrix, which is an expression consistent with linearisations schemes described elsewhere (Archambeau et al., 2007; Jazwinski, 2014; Sanz-Alonso and Stuart, 2017; Särkkä and Solin, 2019). This ODE enables rapid computation of the mean and covariance of the linearised solutions by solving a system of ODEs, i.e. (3.2) and (3.9).

Remark 3.4 *The $n \times n$ covariance matrix $\mathbb{V}\left[l_t^{(\varepsilon)}\right]$ of the linearised solution is the symmetric positive-semidefinite $n \times n$ matrix solution to the ordinary differential equation*

$$\frac{\mathrm{d}\Pi(t)}{\mathrm{d}t} = \nabla u\left(F_0^t(x_0)\,,t\right)\Pi(t) + \Pi(t)\left[\nabla u\left(F_0^t(x_0)\,,t\right)\right]^\top + \varepsilon^2\sigma\left(F_0^t(x_0)\,,t\right)\sigma\left(F_0^t(x_0)\,,t\right)^\top, \tag{3.9}$$

subject to the initial condition $\Pi(0) = \mathbb{V}[x]$. We show that the variance satisfies (3.9) in Section 3.3.4.

3.1.1 Comparison to existing results

In this section we connect our work to the cognate bound derived by Sanz-Alonso and Stuart (2017). That paper considers the following SDE:

$$\mathrm{d}y_t^{(\varepsilon)} = u\left(y_t^{(\varepsilon)}\right)\mathrm{d}t + \varepsilon\,\tilde{\sigma}\,\mathrm{d}W_t, \tag{3.10}$$

where the diffusion coefficient $\tilde{\sigma}$ is a constant matrix, which is a special case of (3.1). In this section, we apply our results to (3.10) to enable both bounds to be compared. Note that ε in our article is written as $\sqrt{\varepsilon}$ in Sanz-Alonso and Stuart (2017); we will translate results from Sanz-Alonso and Stuart (2017) to use our notation, so that all results in this article are directly comparable. In the following, c denotes an arbitrary finite and non-negative constant that can vary between inequalities.

Theorem 2.2 of Sanz-Alonso and Stuart (2017), summarised, is as follows. Let $\xi_t^{(\varepsilon)}$ be the probability measure associated with $y_t^{(\varepsilon)}$ (as defined in (3.10)), and let $\nu_t^{(\varepsilon)}$ be the probability measure associated with the corresponding linearisation $l_t^{(\varepsilon)}$ (as defined in Section 3.1). Then there exists a constant c such that the Kullback–Leibler (KL) divergence D_{KL} between $\xi_t^{(\varepsilon)}$ and $\nu_t^{(\varepsilon)}$ is bounded;

$$D_{\mathrm{KL}}\left(\xi_t^{(\varepsilon)} \,\middle\|\, \nu_t^{(\varepsilon)}\right) \leq D_{\mathrm{KL}}\left(\xi_0^{(\varepsilon)} \,\middle\|\, \nu_0^{(\varepsilon)}\right) + c\,\varepsilon^2 \,. \tag{3.11}$$

To focus on the scaling with ε, assume a fixed initial condition with $D_{\mathrm{KL}}\left(\xi_0^{(\varepsilon)} \,\middle\|\, \nu_0^{(\varepsilon)}\right) = 0$ (and $\delta_r = 0$ in our bound (3.4)). Then, employing the Hellinger distance D_{H}, (3.11) implies

$$\left\|\mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right]\right\| \leq c D_{\mathrm{H}}\left(\xi_t^{(\varepsilon)}, \nu_t^{(\varepsilon)}\right) \leq c\sqrt{D_{\mathrm{KL}}\left(\xi_t^{(\varepsilon)} \,\middle\|\, \nu_t^{(\varepsilon)}\right)} \leq c\,\varepsilon \,, \tag{3.12}$$

while our result (3.4) and Jensen's inequality imply

$$\left\|\mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right]\right\| \leq \mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|\right] \leq c\,\varepsilon^2 \,.$$

Thus, our bound on the moments is quadratic in ε rather than linear. If our conversion in (3.12) was optimal, then our approach in this article provides a sharper bound on $\left\|\mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right]\right\|$ that the results of Sanz-Alonso and Stuart (2017) imply, and do so for a more general σ. The results in Sanz-Alonso and Stuart (2017) on the KL divergence would be more natural in information-theoretic contexts, and our hope is that our explicit bound on the moments would be similarly preferred in other contexts.

3.1.2 Gaussian initial condition

We now briefly consider the case when the initial condition follows a Gaussian distribution, i.e. $x \sim \mathcal{N}(\mu_0, \Sigma_0)$, where $\mu_0 \in \mathbb{R}^n$ and $\Sigma_0 \in \mathbb{R}^{n\times n}$ are fixed and specified. The linearisation then follows a Gaussian distribution itself, which is entirely characterised by the mean and covariance matrix described in Corollary 3.1. Alternatively, these moments can be conveniently computed by simultaneously solving (3.2) for the state variable and (3.9) for the linearised covariance.

A natural choice of reference point x_0 is the mean of the initial Gaussian density, i.e. $x_0 = \mu_0$. The L_r distance between x and the mean μ_0 can be bounded by the trace of Σ_0; for example, one such bound is

$$\delta_r^r \leq n^{3r/2-1} M_r \mathrm{tr}\left(\Sigma_0\right)^{r/2}, \quad M_r \coloneqq \frac{2^{r/2}\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}, \tag{3.13}$$

where Γ denotes the Gamma function, with equality when $n = 1$. The initial covariance Σ_0 directly measures the uncertainty in the initial condition, and we see through (3.13) that as the components of Σ_0 approach zero, the contribution of the initial uncertainty to the linearisation error in (3.4) approaches zero also. The linearised solution is then

$$l_t^{(\varepsilon)} \sim \mathcal{N}\left(F_0^t(x_0)\,,\, \nabla F_0^t(x_0)\,\Sigma_0\left[\nabla F_0^t(x_0)\right]^\top + \varepsilon^2 \Sigma_0^t(x_0)\right),$$

where $\Sigma_0^t(x_0)$ is given explicitly by

$$\Sigma_0^t(x_0) = \nabla F_0^t(x_0)\left(\int_0^t L(x_0,\tau)\,L(x_0,\tau)^\top \,\mathrm{d}\tau\right)\left[\nabla F_0^t(x_0)\right]^\top, \tag{3.14}$$

and is the solution to the matrix differential equation (3.9) in Remark 3.4, subject to $\Sigma_0^0(x_0) = O$, the $n \times n$ zero matrix. The covariance matrix $\Sigma_0^t(x_0)$ characterises the contribution of the ongoing uncertainty in the stochastic system. The full covariance matrix $\mathbb{V}\left[l_t^{(\varepsilon)}\right]$ is also the solution to (3.9) subject to the initial condition $\Pi(0) = \mathbb{V}[x]$. By jointly solving (3.2) for the deterministic trajectory (the mean of $l_t^{(\varepsilon)}$) and (3.9) for the covariance matrix, one can easily compute the linearised solution, describing exactly the assumed Gaussian approximation presented in Särkkä and Solin (2019), and the dynamics linearisation used in the extended Kalman filter (Jazwinski, 2014).

3.1.3 Fixed initial condition

Consider when the initial condition x is itself a fixed and known deterministic value, in which case we take $x = x_0$ and $\delta_r = 0$ for all r. In this situation, the bound (3.4) on the linearisation error reduces to

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq \left(K_{\nabla\nabla u}^r + K_{\nabla\sigma}^r\right) D_1(r,t,K_{\nabla u},K_\sigma)\,\varepsilon^{2r}. \tag{3.15}$$

We can consider the linearisation as equivalently arising from a first-order power series expansion of $y_t^{(\varepsilon)}$ in the noise-scale parameter ε, i.e.

$$y_t^{(\varepsilon)} = F_0^t(x_0) + \varepsilon z_t^{(\varepsilon)} + R_2\left(x,t,\varepsilon\right).$$

where $z_\varepsilon := \left(l_t^{(\varepsilon)} - F_0^t(x_0)\right)/\varepsilon$ is the first order term and R_2 is a random quantity capturing the remaining deviation between $y_t^{(\varepsilon)}$ and the linearisation. By rearranging and taking $r = 1$ in (3.15), we therefore have the explicit Taylor-like bound

$$\frac{\mathbb{E}\left[\left\|R_2\left(x,t,\varepsilon\right)\right\|\right]}{\varepsilon^2} \leq \left(K_{\nabla\nabla u} + K_{\nabla\sigma}\right) D_1(1,t,K_{\nabla u},K_\sigma)\,,$$

This result is consistent in formulation with the error bounds of Blagoveshchenskii (1962) and Freidlin and Wentzell (1998), for instance. Moreover, the distribution of the linearisation solution (3.5) is Gaussian, which through Corollary 3.1 we can again explicitly characterise in terms of the deterministic system, namely

$$l_t^{(\varepsilon)} \sim \mathcal{N}\left(F_0^t(x_0)\,, \varepsilon^2\Sigma_0^t(x_0)\right), \tag{3.16}$$

where $\Sigma_0^t(x_0)$ is defined in (3.14). The distribution can be computed *entirely* from the solution behaviour of the deterministic equation (3.2) and prior specification of σ. In Section 3.2, we demonstrate an application of these results to extend stochastic sensitivity (Balasuriya, 2020b) to arbitrary dimension.

3.2 Extending stochastic sensitivity

The results of Section 3.1.3 for a fixed initial condition provide a direct extension of the stochastic sensitivity tools first introduced by Balasuriya (2020b) for the fluid flow context. Here, the deterministic model (3.2) is seen as a "best-available" model for the evolution of Lagrangian 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 a maximum variance of projected deviation (Balasuriya, 2020b). The aim is to provide a *single* computable number for each deterministic trajectory quantifying the impact of uncertainty in the velocity, independent of the scale (ε) of the noise. The natural restating of this original definition of stochastic sensitivity (Balasuriya, 2020b) in the n-dimensional setting is as follows:

Definition 3.1 (Stochastic sensitivity in $\mathbb{R}^n$) *The* stochastic sensitivity *is the scalar field* $S^2\colon \mathbb{R}^n \times [0,T] \to [0,\infty)$ *given by*

$$S^2(x_0,t) := \limsup_{\varepsilon\downarrow 0} \left\{ \mathbb{V}\left[\frac{1}{\varepsilon} p^\top \left(y_t^{(\varepsilon)} - F_0^t(x_0)\right)\right] \,:\, p\in\mathbb{R}^n, \|p\| = 1 \right\}.$$

Figure 3.1 illustrates the quantities involved in Definition 3.1, to be compared directly to Figure 2.5 which represented the original definition of stochastic sensitivity. The ray

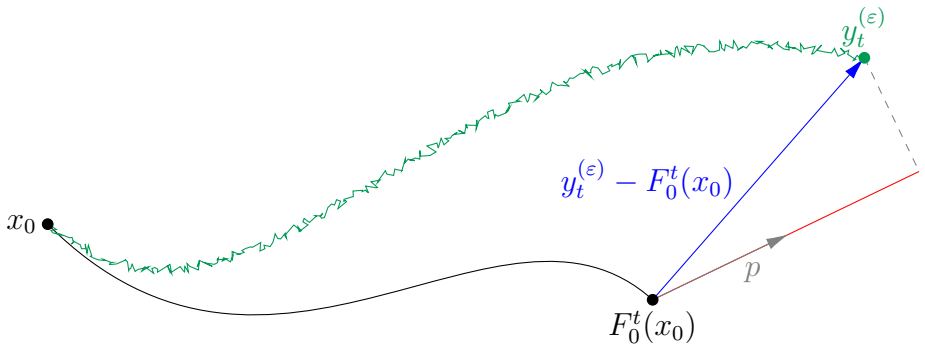


Figure 3.1: The entities used in Definition 3.1 of stochastic sensitivity in arbitrary dimensions. This diagram should be compared to Figure 2.5.

of angle θ in the original definition has been replaced with an arbitrary unit vector p. We compute the projection (in red) of the deviation (in blue) between the deterministic (in black) and stochastic (in green) predictions and then maximise over all possible unit vectors p to obtain the stochastic sensitivity value. Definition 3.1 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 2-dimensions (Balasuriya, 2020b) is the direction-dependent projection (before optimising over all directions in Definition 3.1). Explicit theoretical expressions for both the stochastic sensitivity and the anisotropic sensitivity in two dimensions were obtained by Balasuriya (2020b); 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 2-dimensional ones by Balasuriya (2020b), which moreover establishes Gaussianity as well as an explicit expression for the uncertainty measure. A theoretically pleasing and computable expression for the stochastic sensitivity is obtainable:

Theorem 3.3 (Computation of S^2) *For any $x_0 \in \mathbb{R}^n$ and $t \in [0, T]$,*

$$S^2(x_0, t) = \left\| \Sigma_0^t(x_0) \right\|, \tag{3.17}$$

where the covariance matrix Σ_0^t is defined in (3.14). Equivalently, $S^2(x_0, t)$ is given by the maximum eigenvalue of $\Sigma_0^t(x_0)$.

Proof. See Section 3.3.5. This result uses Theorem 3.1 to establish the convergence of the covariance matrices, and then the properties of the spectral norm to establish (3.17). □

Independent of the fluid mechanics context, Theorem 3.3 indicates that even for general systems, the matrix norm of $\Sigma_0^t(x)$, 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. The significance of this result is that the stochastic sensitivity has here been recovered as the maximal eigenvalue of a covariance matrix that is ubiquitous in the literature. Stochastic sensitivity was formerly defined by Balasuriya (2020b) as the maximal value of the anisotropic uncertainty of a particular stochastic flow, and the connection to a linearisation was not apparent.

The stochastic sensitivity field can be calculated given any velocity data u, and through the explicit expression (3.14) for Σ_0^t 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 2-dimensions (Balasuriya, 2020b; Badza et al., 2023). 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 (2020b). Such a definition extends to the n-dimensional case as presented here, moreover establishing an easily computable method for determining coherent sets from the covariance matrix Σ_0^t.

3.2.1 Connections to the finite-time Lyapunov exponent

We briefly consider the implications of our results on uncertain initial conditions from Section 3.1 on stochastic sensitivity, which suggests a connection between this measure and the finite-time Lyapunov exponent. The finite-time Lyapunov exponent (FTLE) is a measure of the local stretching in a dynamical system and is one of the most commonly used tools for identifying Lagrangian coherent structures in fluid flow (Shadden et al., 2005). In a particular direction p from the initial condition x_0, the stretching due to the evolution of the ODE (2.1) is measured as

$$s(x_0,p) = \frac{\left\|F_0^t(x_0+p) - F_0^t(x_0)\right\|}{\|x_0\|} \approx \frac{\left\|F_0^t(x_0)\,p\right\|}{\|p\|}$$

assuming the perturbation p is small so that $\mathcal{O}\left(\|p\|^2\right)$ terms can be neglected. The FTLE is then computed by finding the direction in which the stretching is maximised, leading to the operator norm computation

$$\sup_{p\in\mathbb{R}^n,\,p\neq 0} s(x,p) = \left\|\nabla F_0^t(x_0)\right\| = \sqrt{\left\|\left[\nabla F_0^t(x_0)\right]^\top \nabla F_0^t(x_0)\right\|}.$$

The stochastic sensitivity definition in Definition 3.1 is analogous to this formation of the FTLE. We consider a small stochastic perturbation (the solution to the SDE (3.1)) to a deterministic model (the corresponding deterministic system (2.1)). The size of the perturbation is parameterised by ε, and the notion that this perturbation is 'small' is formalised by considering the limit as ε approaches zero. Since we are now dealing with stochastic quantities, we take the direction in which the variance of the mapped perturbation is maximised. Theorem 3.3 further establishes that stochastic sensitivity can be computed by taking the operator norm of the covariance matrix of the linearised SDE with a fixed initial condition. That is, stochastic sensitivity is the operator norm of

$$\frac{1}{\varepsilon^2}\mathbb{V}\left[l_t^{(\varepsilon)}\right] = \Sigma_0^t(x_0)\,.$$

This matrix captures (up to leading order in ε) the impact of the ongoing uncertainty. However, the theory of Section 3.1 allows for uncertain initial conditions, so suppose that we instead consider the perturbation SDE (3.1) initial condition x with expectation $\mathbb{E}[x] = x_0$ and variance $\mathbb{V}[x] = \varepsilon^2 I$. If there was no ongoing uncertainty ($\sigma \equiv O$), then any uncertainty in the system arises purely from the initial condition x. Then from (3.8), the variance of the small noise linearisation is

$$\frac{1}{\varepsilon^2}\mathbb{V}\left[l_t^{(\varepsilon)}\right] = \nabla F_0^t(x_0)\left[\nabla F_0^t(x_0)\right]^\top,$$

resulting only from the first term in (3.8). Thus, if we compute stochastic sensitivity in the same way as in Definition 3.1 and Theorem 3.3 we get

$$\tilde{S}^2(x_0,t) = \left\|\nabla F_0^t(x_0)\left[\nabla F_0^t(x_0)\right]^\top\right\| = \sup_{p\in\mathbb{R}^n,\,p\neq 0} s(x_0,p)^2$$

By maximising the projection of this variance over all directions, as in (3.17) to compute stochastic sensitivity, we perform exactly the computation to determine the maximal stretching rate in the computation of the finite-time Lyapunov exponent. The finite-time Lyapunov exponent (FTLE) quantifies the sensitivity of a dynamical system to initial conditions (Shadden et al., 2005), and can be equivalently considered a measure of the impact of an uncertain initial condition on the *deterministic* evolution of trajectories. The construction of $\tilde{S}^2$ is completely consistent with this idea. Thus, while the original formulation of stochastic sensitivity with a fixed initial condition captures only the impact of ongoing uncertainty, by including an uncertain initial condition in the computation we can also account for the information on the impact of initial condition uncertainty provided by the FTLE. There has been recent interest in extending the FTLE for systems with ongoing uncertainty (Guo et al., 2016; Balasuriya, 2020c; You and Leung, 2021), but no established approach as of yet. A framework that computes stochastic sensitivity with uncertain initial conditions can be seen as such an extension of the FTLE, in the sense

that the measure would characterise the sensitivity of a dynamical system to *both* initial conditions and ongoing uncertainty. Formalising this connection (e.g. by considering the appropriate limits and justifying the choice of initial variance) is a matter for future work.

3.3 Proofs of results

In this section, we provide proofs of all the results in this chapter. We use several results from deterministic and Itô calculus, which are stated in Appendix A for reference.

3.3.1 Preliminaries for proofs

There are several generic results and inequalities that we use several times throughout our proofs, which we state here for completeness. We write $W_t = \left(W_t^{(1)}, \ldots, W_t^{(m)}\right)^\top$ as the components of the canonical m-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 (3.2) and is given by

$$F_0^t(x) = x + \int_0^t u\left(F_0^\tau(x), \tau\right) \mathrm{d}\tau, \tag{3.18}$$

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 (3.2), i.e. recall (2.3):

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

For any real numbers $x_1, \ldots, x_p \geq 0$ and $r \geq 1$,

$$\left(\sum_{i=1}^p x_i\right)^r \leq p^{r-1} \sum_{i=1}^p x_i^r. \tag{3.19}$$

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 (3.19) is that for any $z \in \mathbb{R}^n$ and $r \geq 1$,

$$\|z\|^r \leq \left(\sum_{i=1}^n |z_i|\right)^r \leq n^{r-1} \sum_{i=1}^n |z_i|^r, \tag{3.20}$$

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

$$\|z\| \leq \sqrt{n} K. \tag{3.21}$$

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 $r \geq 1$,

$$\left\| \int_0^t f(\tau)\, \mathrm{d}\tau \right\|^r \leq t^{r-1} \int_0^t \left\| f(\tau) \right\|^r \mathrm{d}\tau. \tag{3.22}$$

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

3.3.2 Proof of Theorem 3.1

To prove the main result, we first require a lemma establishing a bound on the time integral of the expectation of the distance between the SDE solution and the reference deterministic trajectory.

Lemma 3.1 *Let $q \geq 1$ be such that $\delta_q < \infty$, then for all $\varepsilon > 0$ and $\tau \in [0, T]$*

$$\mathbb{E}\left[\int_0^t \left\| y_\tau^{(\varepsilon)} - F_0^t(x_0) \right\|^q \mathrm{d}\tau \right] \leq H_1(q, t, K_{\nabla u}, K_\sigma)\, \varepsilon^q + H_2(q, t, K_{\nabla u})\, \delta_q^q,$$

where

$$\begin{aligned} H_1(q, t, K_{\nabla u}, K_\sigma) &:= 3^{q-1} n^{3q/2} K_\sigma^{q/2} G_{q/2} t^{q/2+1} \exp\left(3^{q-1} K_{\nabla u}^q t^q\right), \\ H_2(q, t, K_{\nabla u}) &:= 3^{q-1} t \exp\left(3^{q-1} K_{\nabla u}^q t^q\right). \end{aligned}$$

Proof. Consider the integral form of (3.1),

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

Using (3.18),

$$y_t^{(\varepsilon)} - F_0^t(x_0) = x - x_0 + \int_0^t \left(u\left(y_\tau^{(\varepsilon)}, \tau\right) - u\left(F_0^\tau(x_0), \tau\right) \right) \mathrm{d}\tau + \varepsilon \int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) \mathrm{d}W_\tau,$$

and so

$$\begin{aligned} \mathbb{E}\left[\left\| y_t^{(\varepsilon)} - F_0^t(x_0) \right\|^q \right] \leq\ & 3^{q-1} \mathbb{E}\left[\|x - x_0\|^q \right] \\ & + 3^{q-1} t^{q-1} \mathbb{E}\left[\int_0^t \left\| u\left(y_\tau^{(\varepsilon)}, \tau\right) - u\left(F_0^\tau(x_0), \tau\right) \right\|^q \mathrm{d}\tau \right] \\ & + 3^{q-1} \varepsilon^q \mathbb{E}\left[\left\| \int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) \mathrm{d}W_\tau \right\|^q \right], \end{aligned} \tag{3.23}$$

using (3.19) followed by (3.22), and taking the expectation on both sides.

Next, we establish a bound on the Itô integral term in (3.23). For $i \in \{1, \dots, n\}$, let $\sigma_{i\cdot}$ denote the ith row of σ. Define the stochastic process

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

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

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

Since $y_t^{(\varepsilon)}$ is a strong solution to (3.1), we have that (e.g. see Definition 6.1.1 of Kallianpur and Sundar (2014))

$$\int_0^t \left\| M_\tau^{(i)} \right\|^2 \mathrm{d}\tau \leq \int_0^t n K_\sigma^2 \, \mathrm{d}\tau < \infty, \quad \text{almost surely,}$$

so we can apply the Burkholder-Davis-Gundy inequality (see Theorem A.4) to M_τ, which asserts that there exists a constant $G_{q/2} > 0$ depending only on q such that

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^t M_\tau^{(i)} \, \mathrm{d}W_\tau\right|^q\right] &\leq G_{q/2}\mathbb{E}\left[\left(\int_0^t \left\|\sigma_{i\cdot}\left(y_\tau^{(\varepsilon)}, \tau\right)\right\|^2 \mathrm{d}\tau\right)^{q/2}\right] \\ &\leq G_{q/2} n^p K_\sigma^{q/2} t^{q/2},\end{aligned}$$

where the second inequality uses H.9. Then,

$$\mathbb{E}\left[\left\|\int_0^t \sigma\left(y_\tau^{(\varepsilon)}, \tau\right) \mathrm{d}W_\tau\right\|^q\right] \leq n^{3q/2} K_\sigma^{q/2} G_{q/2} t^{q/2}, \tag{3.24}$$

using (3.20).

Applying the bound (3.24) to (3.23), we have

$$\begin{aligned}\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - F_0^t(x_0)\right\|^q\right] &\leq 3^{q-1}\delta_q^q + 3^{q-1}\varepsilon^q n^{3q/2} K_\sigma^{q/2} G_{q/2} t^{q/2} \\ &\quad + 3^{q-1} t^{q-1} \mathbb{E}\left[\int_0^t \left\|u\left(y_\tau^{(\varepsilon)}, \tau\right) - u\left(F_0^\tau(x_0), \tau\right)\right\|^q \mathrm{d}\tau\right].\end{aligned} \tag{3.25}$$

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)),

$$\mathbb{E}\left[\int_0^t \left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^q \mathrm{d}\tau\right] = \int_0^t \mathbb{E}\left[\left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^q\right] \mathrm{d}\tau.$$

Now, using the Lipschitz continuity of u from H.6 on (3.25) and interchanging the expectation and integral,

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - F_0^t(x_0)\right\|^q\right] \leq 3^{q-1}\delta_q^q + 3^{q-1}K_{\nabla u}^q t^{q-1}\int_0^t \mathbb{E}\left[\left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^q\right]\mathrm{d}\tau + 3^{q-1}\varepsilon^q n^{3q/2}K_\sigma^{q/2}G_{q/2}t^{q/2}.$$

Applying Grönwall's inequality and using the monotonicity of the resulting bound in t, we have that for any $\tau \in [0,t]$,

$$\mathbb{E}\left[\left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^q\right] \leq 3^{q-1}\varepsilon^q n^{3q/2}S^{q/2}G_{q/2}t^{q/2}\exp\left(3^{q-1}K_{\nabla u}^q t^q\right) + 3^{q-1}\exp\left(3^{q-1}K_{\nabla u}^q t^q\right)\delta_q^q.$$

Integrating both sides with respect to time and again using Tonelli's theorem, we have

$$\mathbb{E}\left[\int_0^t \left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^q \mathrm{d}\tau\right] \leq 3^{q-1}n^{3q/2}S^{q/2}G_{q/2}t^{q/2+1}\exp\left(3^{q-1}K_{\nabla u}^q t^q\right)\varepsilon^q + 3^{q-1}t\exp\left(3^{q-1}K_{\nabla u}^q t^q\right)\delta_q^q,$$

as desired. □

With these bounds established, we can now prove Theorem 3.1. Subtracting the integral forms of (3.3) and (3.1) gives

$$\begin{aligned}
y_t^{(\varepsilon)} - l_t^{(\varepsilon)} &= \int_0^t \left[u\left(y_\tau^{(\varepsilon)},\tau\right) - u\left(F_0^\tau(x_0),\tau\right) - \nabla u\left(F_0^\tau(x_0)\right)\left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right)\right]\mathrm{d}\tau \\
&\qquad + \int_0^t \left[\varepsilon\sigma\left(y_\tau^{(\varepsilon)},\tau\right) - \varepsilon\sigma\left(F_0^\tau(x_0),\tau\right)\right]\mathrm{d}W_\tau \\
&= \int_0^t \left[u\left(y_\tau^{(\varepsilon)},\tau\right) - \left(u\left(F_0^\tau(x_0),\tau\right) + \nabla u\left(F_0^\tau(x_0)\right)\left(y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right)\right)\right]\mathrm{d}\tau \\
&\qquad + \int_0^t \nabla u\left(F_0^\tau(x_0),\tau\right)\left[y_\tau^{(\varepsilon)} - l_\tau^{(\varepsilon)}\right]\mathrm{d}\tau \\
&\qquad + \varepsilon\int_0^t \left[\sigma\left(y_\tau^{(\varepsilon)},\tau\right) - \sigma\left(F_0^\tau(x_0),\tau\right)\right]\mathrm{d}W_\tau \\
&= A(t) + B(t) + \varepsilon C(t),
\end{aligned}$$

where

$$\begin{aligned}
A(t) &\coloneqq \int_0^t \left[u\left(y_\tau^{(\varepsilon)},\tau\right) - \left(u\left(F_0^\tau(x_0),\tau\right) + \nabla u\left(F_0^\tau(x_0)\right)\left(y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right)\right)\right]\mathrm{d}\tau \\
B(t) &\coloneqq \int_0^t \nabla u\left(F_0^\tau(x_0),\tau\right)\left[y_\tau^{(\varepsilon)} - l_\tau^{(\varepsilon)}\right]\mathrm{d}\tau \\
C(t) &\coloneqq \int_0^t \left[\sigma\left(y_\tau^{(\varepsilon)},\tau\right) - \sigma\left(F_0^\tau(x_0),\tau\right)\right]\mathrm{d}W_\tau.
\end{aligned}$$

Then, using (3.19) and taking expectation,

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq 3^{r-1}\left(\mathbb{E}\left[\left\|A(t)\right\|^r\right] + \mathbb{E}\left[\left\|B(t)\right\|^r\right] + \varepsilon^r \mathbb{E}\left[\left\|C(t)\right\|^r\right]\right). \tag{3.26}$$

First consider $A(t)$, for which the integrand is the expected difference between the drift u evaluated along SDE solution and the first-order Taylor expansion of u about the reference deterministic trajectory. Since for any $t \in [0,T]$, $u(\cdot,t)$ is twice continuously differentiable under H.2, for each $i = 1,\ldots,n$ there exists by Taylor's theorem (e.g. see J. H. Hubbard and B. B. Hubbard (2009, Cor. A9.3.)) a function $R_i : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ such that

$$u_i(z,\tau) = u_i\left(F_0^\tau(x_0),\tau\right) + \left[\nabla u_i\left(F_0^\tau(x_0),\tau\right)\right]\left(z - F_0^\tau(x_0)\right) + R_i(z,\tau) \tag{3.27}$$

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

$$\left|R_i(z,\tau)\right| \leq \frac{1}{2}\left\|\nabla\nabla u_i\left(F_0^\tau(x),t\right)\right\|\left\|z - F_0^\tau(x_0)\right\|^2 \leq \frac{K_{\nabla\nabla u}}{2}\left\|z - F_0^\tau(x_0)\right\|^2. \tag{3.28}$$

Let $R(z,\tau) \coloneqq \left(R_1(z,\tau),\ldots,R_n(z,\tau)\right)^\top$, then

$$A(t) = \int_0^t R\left(y_t^{(\varepsilon)},\tau\right)\mathrm{d}\tau,$$

and since each component of R is bounded as in (3.28), using (3.21)

$$\left\|R\left(y_t^{(\varepsilon)},\tau\right)\right\| \leq \frac{\sqrt{n}K_{\nabla\nabla u}}{2}\left\|y_t^{(\varepsilon)} - F_0^\tau(x_0)\right\|^2.$$

Taking the norm and expectation then gives

$$\begin{aligned}
\mathbb{E}\left[\left\|A(t)\right\|^r\right] &= \mathbb{E}\left[\left\|\int_0^t R\left(y_t^{(\varepsilon)},\tau\right)\mathrm{d}\tau\right\|^r\right] \\
&\leq t^{r-1}\mathbb{E}\left[\int_0^t \left\|R\left(y_\tau^{(\varepsilon)},\tau\right)\right\|^r \mathrm{d}\tau\right] \\
&\leq \frac{t^{r-1}n^{r/2}K_{\nabla\nabla u}^r}{2^r}\mathbb{E}\left[\int_0^t \left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^{2r}\mathrm{d}\tau\right] \\
&\leq \frac{t^{r-1}n^{r/2}K_{\nabla\nabla u}^r H_1(2r,t,K_{\nabla u},K_\sigma)}{2^r}\varepsilon^{2r} \\
&\quad + \frac{t^{r-1}n^{r/2}K_{\nabla\nabla u}^r H_2(2r,t,K_{\nabla u})}{2^r}\delta_{2r}^{2r},
\end{aligned} \tag{3.29}$$

where the first inequality uses (3.22), and H_1 and H_2 are obtained from Lemma 3.1.

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

$$\mathbb{E}\left[\left\|B(t)\right\|^r\right] \leq \int_0^t t^{r-1} K_{\nabla u}^r \mathbb{E}\left[\left\|y_\tau^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \mathrm{d}\tau. \tag{3.30}$$

using (3.22) and then H.3, and interchanging the expectation and the integral uses the fact that that $\mathbb{E}\left[\left\|y_\tau^{(\varepsilon)}\right\|\right] < \infty$ and $\mathbb{E}\left[\left\|l_\tau^{(\varepsilon)}\right\|\right] < \infty$.

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

$$N_\tau^{(i)} := \sigma_{i\cdot}\left(y_\tau^{(\varepsilon)}, \tau\right) - \sigma_{i\cdot}\left(F_0^\tau(x_0), \tau\right).$$

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.9 and using (3.21),

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

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

$$\begin{aligned}
\mathbb{E}\left[\left|\left[C(t)\right]_i\right|^r\right] &\leq G_{r/2}\mathbb{E}\left[\left(\int_0^t \left\|\sigma_{i\cdot}\left(y_\tau^{(\varepsilon)}, \tau\right) - \sigma_{i\cdot}\left(F_0^\tau(x_0), \tau\right)\right\|^2 \mathrm{d}\tau\right)^{r/2}\right] \\
&\leq G_{r/2}\mathbb{E}\left[\left(\int_0^t K_{\nabla\sigma}^2 \left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^2 \mathrm{d}\tau\right)^{r/2}\right] \\
&\leq G_{r/2}K_{\nabla\sigma}^r t^{r/2-1}\mathbb{E}\left[\int_0^t \left\|y_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right\|^r \mathrm{d}\tau\right] \\
&\leq G_{r/2}K_{\nabla\sigma}^r t^{r/2-1} H_1(r, t, K_{\nabla u}, K_\sigma)\, \varepsilon^r \\
&\quad + G_{r/2}K_{\nabla\sigma}^r t^{r/2-1} H_2(r, t, K_{\nabla u})\, \delta_r^r,
\end{aligned} \tag{3.31}$$

where the second inequality uses the Lipschitz condition on σ in H.6, the third inequality uses (3.22), and the fourth inequality uses Lemma 3.1 with $q = r$. Then, we have

$$\begin{aligned}
\mathbb{E}\left[\left\|C(t)\right\|^r\right] &\leq n^r G_{r/2}K_{\nabla\sigma}^r t^{r/2-1} H_1(r, t, K_{\nabla u}, K_\sigma)\, \varepsilon^r \\
&\quad + n^r G_{r/2}K_{\nabla\sigma}^r t^{r/2-1} H_2(r, t, K_{\nabla u})\, \delta_r^r,
\end{aligned} \tag{3.32}$$

using (3.20), and then (3.31).

Combining (3.29), (3.30) and (3.32) into (3.26), we have

$$\begin{aligned}\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq{}& \frac{3^{r-1}t^{r-1}n^{r/2}K_{\nabla\nabla u}^r H_1(2r,t,K_{\nabla u},K_\sigma)}{2^r}\varepsilon^{2r} \\ &+ \frac{3^{r-1}t^{r-1}n^{r/2}K_{\nabla\nabla u}^r H_2(2r,t,K_{\nabla u})}{2^r}\delta_{2r}^{2r} \\ &+ \int_0^t 3^{r-1}t^{r-1}K_{\nabla u}^r \mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right]\mathrm{d}\tau \\ &+ 3^{r-1}G_{r/2}K_{\nabla\sigma}^r t^{r/2-1}H_1(r,t,K_{\nabla u},K_\sigma)\,\varepsilon^{2r} \\ &+ 3^{r-1}G_{r/2}K_{\nabla\sigma}^r t^{r/2-1}H_2(r,t,K_{\nabla u})\,\varepsilon^r\delta_r^r.\end{aligned}$$

Applying Grönwall's inequality, noting that H_1 and H_2 are non-decreasing in t, we have

$$\begin{aligned}\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq{}& \frac{3^{r-1}t^{r-1}n^{r/2}K_{\nabla\nabla u}^r H_1(2r,t,K_{\nabla u},K_\sigma)}{2^r}\exp\left(3^{r-1}t^r K_{\nabla u}^r\right)\varepsilon^{2r} \\ &+ \frac{3^{r-1}t^{r-1}n^{r/2}K_{\nabla\nabla u}^r H_2(2r,t,K_{\nabla u})}{2^r}\exp\left(3^{r-1}t^r K_{\nabla u}^r\right)\delta_{2r}^{2r} \\ &+ 3^{r-1}G_{r/2}K_{\nabla\sigma}^r t^{r/2-1}\exp\left(3^{r-1}t^r K_{\nabla u}^r\right)H_1(r,t,K_{\nabla u},K_\sigma)\,\varepsilon^{2r} \\ &+ 3^{r-1}G_{r/2}K_{\nabla\sigma}^r t^{r/2-1}\exp\left(3^{r-1}t^r K_{\nabla u}^r\right)H_2(r,t,K_{\nabla u})\,\varepsilon^r\delta_r^r.\end{aligned}$$

Set

$$D_1(r,t,K_{\nabla u},K_\sigma) \coloneqq 3^{r-1}\exp\left(3^{r-1}t^r K_{\nabla u}^r\right)K_M(r,t,K_{\nabla u},K_\sigma) \tag{3.33a}$$

$$D_2(r,t,K_{\nabla u}) \coloneqq 3^{r-1}t^{r-1}n^{r/2}H_2(2r,t,K_{\nabla u})\exp\left(3^{r-1}t^r K_{\nabla u}^r\right) \tag{3.33b}$$

$$D_3(r,t,K_{\nabla u}) \coloneqq 3^{r-1}G_{r/2}t^{r/2-1}H_2(r,t,K_{\nabla u})\exp\left(3^{r-1}t^r K_{\nabla u}^r\right), \tag{3.33c}$$

where

$$K_M(r,t,K_{\nabla u},K_\sigma) \coloneqq \max\left\{\frac{t^{r-1}n^{r/2}H_1(2r,t,K_{\nabla u},K_\sigma)}{2^r},\, G_{r/2}t^{r/2-1}H_1(r,t,K_{\nabla u},K_\sigma)\right\},$$

then we have shown the desired result.

3.3.3 Proof of Theorem 3.2

Next, we show that the strong solution to the linearised SDE (3.3) can be written as the independent sum (3.5). Let

$$M_t = h\left(l_t^{(\varepsilon)}, t\right) \coloneqq \frac{1}{\varepsilon}\left[\nabla F_0^t(x_0)\right]^{-1}\left(l_t^{(\varepsilon)} - F_0^t(x_0)\right),$$

where $l_t^{(\varepsilon)}$ is the strong solution to (3.3). Then, the required derivatives for applying Itô's Lemma (see Theorem A.3) are

$$\begin{aligned}
M_0 &= h\Big(l_0^{(\varepsilon)}, 0\Big) = \frac{1}{\varepsilon}\left(x - x_0\right) \\
\frac{\partial h}{\partial t} &= -\frac{1}{\varepsilon}\left[\nabla F_0^t(x_0)\right]^{-1} \frac{\partial \nabla F_0^t(x_0)}{\partial t} \left[\nabla F_0^t(x_0)\right]^{-1} \left(l_t^{(\varepsilon)} - F_0^t(x_0)\right) \\
&\qquad - \frac{1}{\varepsilon}\left[\nabla F_0^t(x_0)\right]^{-1} u\big(F_0^t(x_0), t\big) \\
\nabla h\Big(l_t^{(\varepsilon)}, t\Big) &= \frac{1}{\varepsilon}\left[\nabla F_0^t(x_0)\right]^{-1} \\
\nabla\nabla h\Big(l_t^{(\varepsilon)}, t\Big) &= O,
\end{aligned}$$

where in computing the t derivative we have used the fact that $F_0^t(x_0)$ solves the deterministic ODE (3.2). Thus,

$$\begin{aligned}
M_t = \frac{1}{\varepsilon}\left(x - x_0\right) + \frac{1}{\varepsilon}\int_0^t \Bigg(& -\left[\nabla F_0^\tau(x_0)\right]^{-1} \frac{\partial \nabla F_0^\tau(x_0)}{\partial \tau} \left[\nabla F_0^\tau(x_0)\right]^{-1} \left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right) \\
& - \left[\nabla F_0^\tau(x_0)\right]^{-1} u\big(F_0^\tau(x_0), \tau\big) \\
& + \left[\nabla F_0^\tau(x_0)\right]^{-1} \left[u\big(F_0^\tau(x_0), \tau\big) + \nabla u\big(F_0^\tau(x_0), \tau\big)\left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right)\right]\Bigg) \mathrm{d}\tau \\
& + \int_0^t \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\big(F_0^\tau(x_0), \tau\big)\, \mathrm{d}W_\tau \\
= \frac{1}{\varepsilon}\left(x - x_0\right) + \frac{1}{\varepsilon}\int_0^t \Bigg(& -\left[\nabla F_0^\tau(x_0)\right]^{-1} \nabla u\big(F_0^\tau(x_0), \tau\big)\left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right) \\
& + \left[\nabla F_0^\tau(x_0)\right]^{-1} \nabla u\big(F_0^\tau(x_0), \tau\big)\left(l_\tau^{(\varepsilon)} - F_0^\tau(x_0)\right)\Bigg) \mathrm{d}\tau \\
& + \int_0^t \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\big(F_0^\tau(x_0), \tau\big)\, \mathrm{d}W_\tau \\
= \frac{1}{\varepsilon}\left(x - x_0\right) + \int_0^t & \left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\big(F_0^\tau(x_0), \tau\big)\, \mathrm{d}W_\tau,
\end{aligned}$$

where we reach the second line by using the equation of variations (2.3) satisfied by $\nabla F_0^\tau(x_0)$. It follows that the expression (3.5), i.e.

$$l_t^{(\varepsilon)} = \nabla F_0^t(x_0)\left(x - x_0\right) + F_0^t(x_0) + \varepsilon \int_0^t \nabla F_0^t(x_0)\left[\nabla F_0^\tau(x_0)\right]^{-1} \sigma\big(F_0^\tau(x_0), \tau\big)\, \mathrm{d}\tau,$$

is a strong solution to (3.3).

Finally, under H.7, x is independent of the Wiener process W_t, and since independence is preserved under limits and linear transformations, it follows that the two random terms in (3.5) are independent.

3.3.4 Proof of Corollary 3.1

We first establish that the Itô integral of a matrix-valued deterministic function with respect to a multidimensional Wiener process is a multidimensional Gaussian process. This is a well-known result in the scalar case, and the extension to our case is straightforward.

Lemma 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 1-dimensional Brownian motion, which is well-known to be a Gaussian process (e.g. see Lemma 4.3 of Applebaum (2004)). 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. □

Now, we move onto showing Corollary 3.1. Consider the Itô integral

$$\mathcal{I}[L] = \int_0^t L(x_0, \tau)\,\mathrm{d}W_\tau.$$

For any fixed $t \in [0, T]$, the integrand is a deterministic, matrix-valued function, and is therefore follows a n-dimensional Gaussian distribution. Moreover (Kallianpur and Sundar, 2014),

$$\mathbb{E}\big[\mathcal{I}[L]\big] = 0,$$

and

$$\mathbb{V}\big[\mathcal{I}[L]\big] = \mathbb{E}\left[\left(\int_0^t L(x_0, \tau)\,\mathrm{d}W_\tau\right)\left(\int_0^t L(x_0, \tau)\,\mathrm{d}W_\tau\right)^{\top}\right].$$

Let L_{ij} denote the (i,j)th element of L, then the (i,j)th element of the variance is

$$\begin{aligned}\Big[\mathbb{V}\big[\mathcal{I}[L]\big]\Big]_{ij} &= \sum_{k=1}^{m}\sum_{l=1}^{m}\mathbb{E}\left[\left(\int_0^t L_{ik}(x_0,\tau)\,\mathrm{d}W_\tau^{(k)}\right)\left(\int_0^t L_{jl}(x_0,\tau)\,\mathrm{d}W_\tau^{(l)}\right)\right] \\ &= \sum_{k=1}^{m}\mathbb{E}\left[\left(\int_0^t L_{ik}(x_0,\tau)\,\mathrm{d}W_\tau^{(k)}\right)\left(\int_0^t L_{jk}(x_0,\tau)\,\mathrm{d}W_\tau^{(k)}\right)\right] \\ &\quad + \sum_{k=1}^{m}\sum_{\substack{l=1\\ l\neq k}}^{m}\mathbb{E}\left[\left(\int_0^t L_{ik}(x_0,\tau)\,\mathrm{d}W_\tau^{(k)}\right)\right]\mathbb{E}\left[\left(\int_0^t L_{jl}(x_0,\tau)\,\mathrm{d}W_\tau^{(l)}\right)\right] \\ &= \sum_{k=1}^{m}\int_0^t L_{ik}(x_0,\tau)\,L_{jk}(x_0,\tau)\,\mathrm{d}\tau,\end{aligned}$$

where the second equality the fact that $W_t^{(k)}$ is independent of $W_t^{(l)}$ for $k \neq l$ and the third equality uses Itô's isometry (Kallianpur and Sundar, 2014). Hence, we have that

$$\int_0^t L(x_0,\tau)\,\mathrm{d}\tau \sim \mathcal{N}\left(0, \int_0^t L(x_0,\tau)\,L(x_0,\tau)^\top\,\mathrm{d}\tau\right),$$

completing the proof of Theorem 3.2. Next, we show that the mean and covariance of $l_t^{(\varepsilon)}$ are given explicitly by (3.7) and (3.8) respectively. It follows immediately from (3.5) that the mean of $l_t^{(\varepsilon)}$ is

$$\begin{aligned}\mathbb{E}\left[l_t^{(\varepsilon)}\right] &= \mathbb{E}\big[\nabla F_0^t(x_0)\,(x-x_0)\big] + F_0^t(x_0) + \varepsilon\nabla F_0^t(x_0)\,\mathbb{E}\left[\int_0^t L\,(x_0,\tau)\,\mathrm{d}\tau\right] \\ &= \nabla F_0^t(x)\,\big(\mathbb{E}[x]-x_0\big) + F_0^t(x_0)\,,\end{aligned}$$

thus showing (3.7).

Since the two summands in (3.5) are independent, the variance of $l_t^{(\varepsilon)}$ is

$$\begin{aligned}\mathbb{V}\left[l_t^{(\varepsilon)}\right] &= \mathbb{V}\big[\nabla F_0^t(x_0)\,(x-x_0)\big] + \mathbb{V}\left[\varepsilon\nabla F_0^t(x_0)\int_0^t L(x_0,\tau)\,\mathrm{d}W_\tau\right] \\ &= \nabla F_0^t(x_0)\left(\mathbb{V}[x] + \varepsilon^2\int_0^t L(x_0,\tau)\,L(x_0,\tau)^\top\,\mathrm{d}\tau\right)\big[\nabla F_0^t(x_0)\big]^\top\end{aligned}$$

where the variance of the Itô integral was established in Section 3.3.3.

Finally, we show Remark 3.4, i.e. that $\mathbb{V}\left[l_t^{(\varepsilon)}\right]$ is the solution to the matrix differential equation (3.9). Directly differentiating the expression (3.8)

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{V}\left[l_t^{(\varepsilon)}\right]}{\mathrm{d}t} &= \frac{\partial \nabla F_0^t(x_0)}{\partial t}\left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0,\tau)\, L(x_0,\tau)^\top \,\mathrm{d}\tau\right)\left[\nabla F_0^t(x_0)\right]^\top \\
&\quad + \nabla F_0^t(x_0)\left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0,\tau)\, L(x_0,\tau)^\top \,\mathrm{d}\tau\right)\left[\frac{\partial \nabla F_0^t(x_0)}{\partial t}\right]^\top \\
&\qquad + \varepsilon^2 \nabla F_0^t(x_0)\, L(x_0,t)\, L(x_0,t)^\top \left[\nabla F_0^t(x_0)\right]^\top \\
&= \nabla u\big(F_0^t(x_0),t\big)\, \nabla F_0^t(x_0)\left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0,\tau)\, L(x_0,\tau)^\top \,\mathrm{d}\tau\right)\left[\nabla F_0^t(x_0)\right]^\top \\
&\quad + \nabla F_0^t(x_0)\left(\mathbb{V}[x] + \varepsilon^2 \int_0^t L(x_0,\tau)\, L(x_0,\tau)^\top \,\mathrm{d}\tau\right)\left[\nabla F_0^t(x_0)\right]^\top \left[\nabla u\big(F_0^t(x_0),t\big)\right]^\top \\
&\qquad + \varepsilon^2 \sigma\big(F_0^t(x_0),t\big)\sigma\big(F_0^t(x_0),t\big)^\top \\
&= \nabla u\big(F_0^t(x_0),t\big)\, \mathbb{V}\left[l_t^{(\varepsilon)}\right] + \mathbb{V}\left[l_t^{(\varepsilon)}\right]\left[\nabla u\big(F_0^t(x_0),t\big)\right]^\top \\
&\qquad + \varepsilon^2 \sigma\big(F_0^t(x_0),t\big)\,\sigma\big(F_0^t(x_0),t\big)^\top,
\end{aligned}$$

where the second inequality has used the equation of variations (2.3).

3.3.5 Proof of Theorem 3.3

Let $t \in [0,T]$ and consider the solutions $y_t^{(\varepsilon)}$ to (3.1) and $l_t^{(\varepsilon)}$ to (3.3) subject to the fixed initial condition $x_0 \in \mathbb{R}^n$. On the vector space of n-dimensional random vectors with each component having finite expectation and variance, define the function ρ as

$$\rho(z) \coloneqq \left\|\mathbb{V}[z]\right\|^{\frac{1}{2}}.$$

Then, ρ is a semi-norm, which can be verified using properties of the spectral norm and the Cauchy-Schwarz inequality. This proof is provided in the supplementary materials.

Then,

$$\begin{aligned}
\left|\left\|\mathbb{V}\left[y_t^{(\varepsilon)}\right]\right\|^{\frac{1}{2}} - \left\|\mathbb{V}\left[l_t^{(\varepsilon)}\right]\right\|^{\frac{1}{2}}\right| &= \left|\rho\left(y_t^{(\varepsilon)}\right) - \rho\left(l_t^{(\varepsilon)}\right)\right| \\
&\leq \rho\left(y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right) \\
&= \left\|\mathbb{E}\left[\left(y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right)\left(y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right)^\top\right] - \mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right]\mathbb{E}\left[y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right]^\top\right\|^{\frac{1}{2}} \\
&\leq \left(\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^2\right] + \mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|\right]^2\right)^{\frac{1}{2}} \\
&\leq \left(\left(K_{\nabla\nabla u} + K_{\nabla\sigma}\right) D_1(2,t) + \left(K_{\nabla\nabla u} + K_{\nabla\sigma}\right)^2 D_1(1,t)^2\right)^{1/2} \varepsilon^2
\end{aligned}$$

where the first inequality uses the reverse triangle inequality, the second inequality uses the Jensen's inequality and properties of the spectral norm, and the third inequality results from Theorem 3.1. Thus,

$$\begin{aligned}
\left|\left\|\frac{1}{\varepsilon^2}\mathbb{V}\left[y_t^{(\varepsilon)}\right]\right\|^{\frac{1}{2}} - \left\|\frac{1}{\varepsilon^2}\mathbb{V}\left[l_t^{(\varepsilon)}\right]\right\|^{\frac{1}{2}}\right| \leq\ & \Big(\left(K_{\nabla\nabla u} + K_{\nabla\sigma}\right) D_1(2,t) \\
& + \left(K_{\nabla\nabla u} + K_{\nabla\sigma}\right)^2 D_1(1,t)^2\Big)^{\frac{1}{2}} \varepsilon,
\end{aligned}$$

and so taking the limit of ε to zero and squaring both sides,

$$\lim_{\varepsilon\downarrow 0}\left\|\frac{1}{\varepsilon^2}\mathbb{V}\left[y_t^{(\varepsilon)}\right]\right\| = \lim_{\varepsilon\downarrow 0}\left\|\frac{1}{\varepsilon^2}\mathbb{V}\left[l_t^{(\varepsilon)}\right]\right\|. \tag{3.34}$$

Now, for $\varepsilon > 0$, define

$$\begin{aligned}
S^2_{(\varepsilon)}(x_0,t) &:= \sup\left\{\mathbb{V}\left[\frac{1}{\varepsilon}p^\top\left(y_t^{(\varepsilon)} - F_0^t(x_0)\right)\right] \middle| p \in \mathbb{R}^n, \|p\| = 1\right\} \\
&= \frac{1}{\varepsilon^2}\sup\left\{p^\top\mathbb{V}\left[y_t^{(\varepsilon)}\right]p \middle| p \in \mathbb{R}^n, \|p\| = 1\right\}
\end{aligned}$$

Since $\mathbb{V}\left[y_t^{(\varepsilon)}\right]$ is symmetric and positive definite, the Cholesky decomposition provides a lower triangular $n \times n$ matrix $\Pi^{(\varepsilon)}$ such that $\mathbb{V}\left[y_t^{(\varepsilon)}\right] = \Pi^{(\varepsilon)}\left[\Pi^{(\varepsilon)}\right]^\top$, allowing us to write

$$\begin{aligned} S^2_{(\varepsilon)}(x_0, t) &= \frac{1}{\varepsilon^2} \sup\left\{ \left\| \Pi^{(\varepsilon)} p \right\|^2 \,\middle|\, p \in \mathbb{R}^n, \|p\| = 1 \right\} \\ &= \frac{1}{\varepsilon^2} \left\| \Pi^{(\varepsilon)} \right\|^2 \\ &= \frac{1}{\varepsilon^2} \left\| \mathbb{V}\left[y_t^{(\varepsilon)}(x) \right] \right\|, \end{aligned}$$

using properties of the spectral norm. Taking the limit as ε approaches zero and using (3.34),

$$S^2(x_0, t) = \lim_{\varepsilon \downarrow 0} S^2_{(\varepsilon)}(x_0, t) = \lim_{\varepsilon \downarrow 0} \left\| \frac{1}{\varepsilon^2} \mathbb{V}\left[y_t^{(\varepsilon)} \right] \right\| = \left\| \Sigma_0^t(x_0) \right\|,$$

where Σ_0^t is defined in (3.14). Since $\Sigma_0^t(x_0)$ is symmetric and positive definite, the operator norm, and therefore $S^2(x_0, t)$, is given by the largest eigenvalue of $\Sigma_0^t(x_0)$.

Chapter 4

Characterising SDE linearisations: the numerics

We now provide numerical validation of the results in Chapter 3 on four different example stochastic differential equations in 1-, 2-, and 3-dimensions. A majority of the content in this chapter again appears in the submitted article (Blake et al., 2023). However, the overview in Section 4.1 of an algorithm (Mazzoni, 2008) for efficiently solving the mean-covariance equations for the linearisation approximation and Section 4.3.2, where we provide an example of stochastic sensitivity computed for a 3-dimensional system, do not appear in the submitted article.

4.1 The Mazzoni method

In Corollary 3.1, we established expressions for the first two moments of the solution $l_t^{(\varepsilon)}$ of the solution $l_t^{(\varepsilon)}$ to the linearisation (3.3) of the SDE (3.1). When the initial condition of the SDE is fixed or Gaussian—a case often used in practice and one which we restrict ourselves to in this chapter—the linearisation solution $l_t^{(\varepsilon)}$ is also Gaussian and characterised entirely by those two moments. The mean is given by the deterministic trajectory $F_0^t(x_0)$ about which (3.1) was linearised and the variance can be computed by solving the matrix differential equation (3.9) in Remark 3.4. The deterministic trajectory $F_0^t(x_0)$ is obtained by solving the deterministic equation (2.1), so we can re-frame the computation of $l_t^{(\varepsilon)}$ as the joint solving of a pair of differential equations. Suppose that the initial condition $y_0^{(\varepsilon)} = l_t^{(\varepsilon)}$ is Gaussian with mean x_0 and covariance matrix Σ_0. For notational brevity, set $w_t \coloneqq F_0^t(x_0)$ and $\Pi_t \coloneqq \mathbb{V}[l_t]$, so that at any time t

$$l_t^{(\varepsilon)} \sim \mathcal{N}(w_t, \Pi_t)\,.$$

We can then compute w_t and Π_t by solving the system of ordinary differential equation

$$\frac{\mathrm{d}w_t}{\mathrm{d}t} = u(w_t, t)\,, \quad w_0 = x_0 \tag{4.1a}$$

$$\frac{\mathrm{d}\Pi_t}{\mathrm{d}t} = \nabla u(w_t, t)\,\Pi_t + \Pi_t \left[\nabla u\,(w_t, t)\right]^\top + \varepsilon\sigma\,(w_t, t)\,\sigma\,(w_t, t)^\top\,, \quad \Pi_0 = \Sigma_0. \tag{4.1b}$$

where (4.1a) is (2.1) and (4.1b) comes from (3.9). Solving (4.1) jointly can provide a more convenient computation to obtain the distribution of the linearisation solution in practice than evaluating (3.7) and (3.8) directly.

An important consideration when solving (4.1) numerically is that Π_t represents a covariance matrix and must remain symmetric and positive semi-definite. However, many standard numerical ODE schemes do not take this into account, so a specialised scheme is required. Similar equations of the form (4.1) (although often without explicit dependence on both time and the state in the σ term) are solved numerically in other applications, notably when implementing the extended Kalman filter (Jazwinski, 2014; Kulikova and Kulikov, 2014). Kulikova and Kulikov (2014) identify that that the two most significant sources of numerical error when solving equations of this form are a) the estimate of the covariance matrix Σ_s^t violating the requirement of positive semi-definiteness, and b) local error propagation in the state equation. Moreover, a computationally efficient algorithm is critical to ensure that the linearisation approximation has an advantage over bulk Monte Carlo simulation.

Mazzoni (2008) proposes an efficient hybrid method for solving (4.1) which addresses both difficulties a) and b) and takes advantage of the availability of the Jacobian ∇u. This method, which we shall term the Mazzoni method, combines a Taylor-Heun approximation for (4.1a) and a Gauss-Legendre step for (4.1b). With a step size of $\delta\tau$, integration for both the state variable and the covariance matrix are convergent with order $\mathcal{O}\big(\delta\tau^2\big)$. Moreover, Mazzoni (2008) shows through numerical simulations that the algorithm is computationally efficient when compared to alternatives with moderate precision, and since the algorithm relies on matrix operations, it scales well with the dimension of the model. We therefore employ the Mazzoni method for all subsequent computations of the linearisation solution.

We summarise the algorithm in the following and provide a diagram of the full implementation in Figure 4.1. Further details on the derivation of these equations are available in the original article (Mazzoni, 2008). The Taylor-Heun formula for the update of the state over the interval $[\tau, \tau + \delta\tau]$ is

$$w_{\tau+\delta\tau} \approx w_\tau + \left(I - \frac{\delta\tau}{2}\nabla u(w_\tau, \tau)\right)^{-1}. \tag{4.2a}$$

The Gauss-Legendre update of the covariance matrix is

$$\Pi_{\tau+\delta\tau} \approx M_\iota \Pi_\tau M_\iota^\top + \delta\tau K_\iota \sigma\left(w_\iota,\, \tau + \frac{\delta\tau}{2}\right)\sigma\left(w_\iota,\, \tau + \frac{\delta\tau}{2}\right)^\top K_\iota^\top, \tag{4.2b}$$

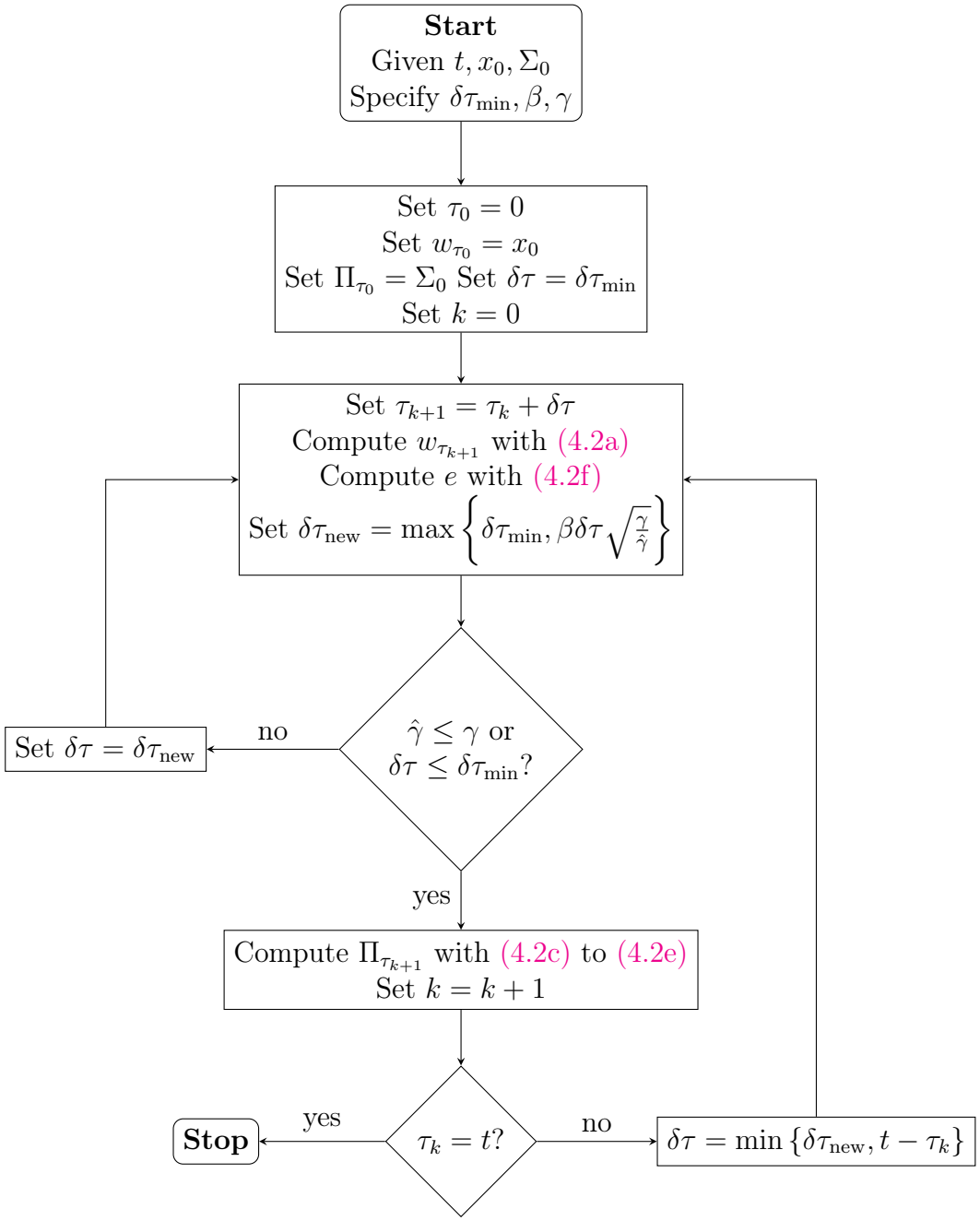


Figure 4.1: A flowchart of the Mazzoni algorithm with adaptive time stepping to solve (4.1) over the time interval $[0, t]$. The minimum step size $\delta\tau_{\min}$ is specified and enforced to ensure that the step size does not become too small to compromise computationally efficiency. Recreated from Figure 2 of Mazzoni (2008).

where

$$w_\iota = \frac{1}{2}\left(w_\tau + w_{\tau+\delta\tau} - \frac{\delta\tau^2}{4}\nabla u(w_\tau,\,\tau)\,u(w_\tau,\,\tau)\right) \tag{4.2c}$$

$$K_\iota = \left[I - \frac{\delta\tau}{2}\nabla u\left(w_\iota,\,\tau+\frac{\delta\tau}{2}\right)\right]^{-1} \tag{4.2d}$$

$$M_\iota = K_\iota\left[I + \frac{\delta\tau}{2}\nabla u\left(w_\iota,\,\tau+\frac{\delta\tau}{2}\right)\right]. \tag{4.2e}$$

The vector w_ι serves as an interpolation between w_τ and $w_{t+\delta\tau}$ for the state at time $\tau+\delta\tau/2$, which is used to provide a more accurate approximation of the covariance matrix. Mazzoni (2008) also provides the option of an adaptive time step to fix the numerical precision while ensuring a computationally efficient algorithm. The step size is adjusted by monitoring the error in the estimation of the state variable, aiming to maintain a specified tolerance level $\gamma > 0$. Given the error vector $\mathcal{E}$ of the state approximation, the maximum total-relative error is

$$\hat{\gamma} = \max_{i=1,\dots,n} \frac{\left|\mathcal{E}^{(i)}\right|}{\left|w^{(i)}_{\tau+\delta\tau}\right| + 1} \tag{4.2f}$$

where $\mathcal{E}^{(i)}$ and $w^{(i)}_{\tau+\delta\tau}$ denote the respective ith elements of $\mathcal{E}$ and $w_{\tau+\delta\tau}$. The resulting adjustment factor for the time step is

$$\delta\tau_{\text{new}} = \beta\delta\tau\sqrt{\frac{\gamma}{\hat{\gamma}}}. \tag{4.2g}$$

The factor β is a control parameter inserted to avoid frequent recalculations of the step size and is specified prior. Mazzoni (2008) suggests setting $\beta = 0.8$. By again taking advantage of the availability of the Jacobian ∇u, the error vector is approximated as

$$\mathcal{E} \approx \frac{\delta\tau^2}{2}\left[\frac{1}{3\delta t}\left(\nabla u(w_{\tau+\delta\tau},\tau+\delta\tau) - \nabla u(w_\tau,\tau)\right) - \frac{1}{6}\nabla u(w_\tau,\tau)^2\right] u(w_\tau,\tau)\,. \tag{4.2h}$$

The set of equations (4.2) describe the Mazzoni algorithm with an adaptive step size.

4.2 Numerical validation & examples

This section will validate the theory presented in Sections 3.1 and 3.2, by considering three example SDEs, each leading to a different form of the strong error bound (3.4). For each example, we first demonstrate heuristically that the solution converges to the

limiting distribution described by Theorem 3.2. We then verify the error bound in Theorem 3.1 directly by considering a range of values for the noise scale ε and initial condition uncertainty δ_r. In doing so, we demonstrate numerically that the form of the bound on the linearisation error predicted by Theorem 3.1 is sharp, in the sense that estimates of the error scale exactly with the initial uncertainty δ_r and ongoing uncertainty ε as predicted.

All simulations in this section were generated using the Julia programming language (Bezanson et al., 2017), with the implementations of numerical ODE and SDE solvers provided by the DifferentialEquations.jl package (Rackauckas and Nie, 2017b). The code is available at github.com/liamblake/explicit-characterisation-sde-linearisation.

4.2.1 Nonlinear dynamics, additive noise

Consider the following SDE in 1D;

$$\mathrm{d}y_t^{(\varepsilon)} = \sin\left(y_t^{(\varepsilon)}\right)\mathrm{d}t + \varepsilon\,\mathrm{d}W_t. \tag{4.3}$$

The deterministic system corresponding to (4.3) has solution

$$F_0^t(x_0) = 2\arctan\left(e^{-t}\tan\left(\frac{x_0}{2}\right)\right).$$

Further details of this example, including computation of the derivatives required in the linearisation, are provided in the supplementary material.

To explore the impact of initial condition uncertainty, we consider the univariate Gaussian initial condition $y_0 = x \sim \mathcal{N}(\mu, \rho^2)$, where the mean μ is specified and the standard deviation ρ is a non-negative scaling parameter. We linearise (4.3) about the deterministic trajectory $F_0^t(\mu)$ originating from the mean, that is, μ is the chosen reference point. This ensures that for any $r \geq 0$

$$\delta_r^r = \mathbb{E}\left[|x-\mu|^r\right] = M_r\rho^r. \tag{4.4}$$

where M_r is as defined in (3.13). This property of the univariate Gaussian distribution allows us to easily control the uncertainty in the initial condition and verify the bounds; by sending the parameter ρ to zero we ensure that δ_r approaches zero also.

The linearised equation is then

$$\mathrm{d}l_t^{(\varepsilon)} = \left[F_0^t(\mu) + \cos\left(F_0^t(\mu)\right)\left(l_t^{(\varepsilon)} - F_0^t(\mu)\right)\right]\mathrm{d}t + \varepsilon\,\mathrm{d}W_t, \quad l_0^{(\varepsilon)} \sim \mathcal{N}\left(\mu, \rho^2\right). \tag{4.5}$$

and the solution follows a Gaussian distribution, specifically

$$l_t^{(\varepsilon)} \sim \mathcal{N}\left(F_0^t(\mu)\,,\ \rho^2\nabla F_0^t(\mu)^2 + \varepsilon^2\Sigma_0^t(\mu)\right). \tag{4.6}$$

where $\Sigma_0^t(\mu)$ is computed by solving (3.9) numerically subject to a zero initial condition.

In this example, we take $\mu_0 = 0.5$ and consider the solutions of (4.3) and (4.5) at time $t = 1.5$. We generate accurate samples of (4.3) and (4.5) jointly (i.e. using the same numerical realisations of the Wiener process W_t) using the stochastic Runge-Kutta scheme SRI (Rößler, 2010) with an adaptive step size (Rackauckas and Nie, 2017a).

In Figure 4.2, we show histograms of $N = 10000$ samples of the solution to nonlinear SDE (4.3) and the corresponding probability density function of the linearised solution (4.6), for different combinations of ε and ρ. Even when the ongoing noise is small, the nonlinearity of the drift term means that a large initial uncertainty results in a non-Gaussian distribution. However, in situations where both the initial and ongoing uncertainties are small, the Gaussian solution to the linearised equation provides a reasonable approximation. In the limit of both small initial ($\rho \to 0$) and small ongoing ($\varepsilon \to 0$) uncertainty (towards the bottom right), we see that the distribution of the samples approach the Gaussian density of the linearisation solution, matching the understanding that the linearisation approximation is "reasonable" for small noise regimes.

Since the drift term is nonlinear and the noise is additive in (4.3), the bound predicted by Theorem 3.1 has the form

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq D_1(r, t, K_{\nabla u}, K_\sigma)\,\varepsilon^{2r} + M_{2r} D_2(r, t, K_{\nabla u})\,\rho^{2r}.$$

where we have taken $K_{\nabla\nabla u} = 1$ and $K_{\nabla\sigma} = 0$. To numerically validate this bound under the Gaussian initial condition (4.4), define for $r \geq 1$ the error measure

$$E_r(\varepsilon, \rho) \coloneqq \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{y}_i^{(\varepsilon)} - \hat{l}_i^{(\varepsilon)}\right\|^r, \tag{4.7}$$

which is a Monte Carlo estimator of the right-hand side of (3.4), where $\hat{y}_1^{(\varepsilon)}, \ldots, \hat{y}_N^{(\varepsilon)}$ and $\hat{l}_1^{(\varepsilon)}, \ldots, \hat{l}_N^{(\varepsilon)}$ are N numerical samples of the solutions to SDE (3.1) and the linearisation (3.3) respectively.

We directly validate the *form* of the error bound (as a function of ε and ρ) in Figure 4.3, by computing E_1 using samples for each pair of ε and ρ values. In Figure 4.3a, we demonstrate the relationship between E_1 and the ongoing uncertainty ε for several different fixed values of ρ, each corresponding to a different colour. A least squares estimate of a line of best fit of the form $E_1 = \beta_0 + \beta_1\varepsilon^2$, for fixed coefficients β_0 and β_1, is fitted to the observed errors (in untransformed space) to verify the scaling of our bound in Theorem 3.1. We see that the line of best fit accurately matches the observed values of E_1, verifying that E_1 is in fact scaling with ε^2 as predicted. Figure 4.3b provides a similar demonstration between E_1 and the initial uncertainty ρ, where now each colour corresponds to a different fixed value of ε. We again fit lines of the form $E_1 = \beta_0 + \beta_1\rho^2$ to verify the scaling of the bound, and see that the lines match the observed values of E_2. Thus, we have also validated that E_1 scales with ρ^2, as expected from Theorem 3.1.

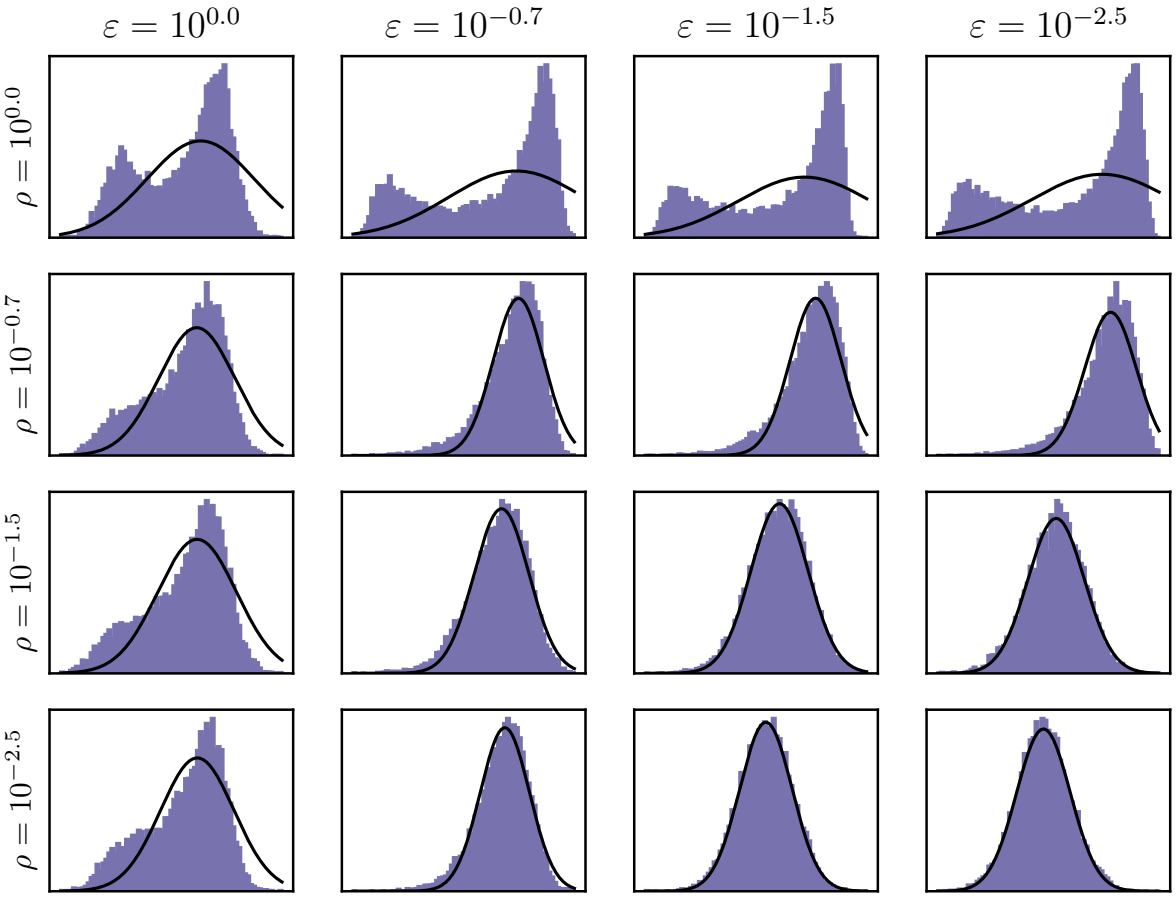
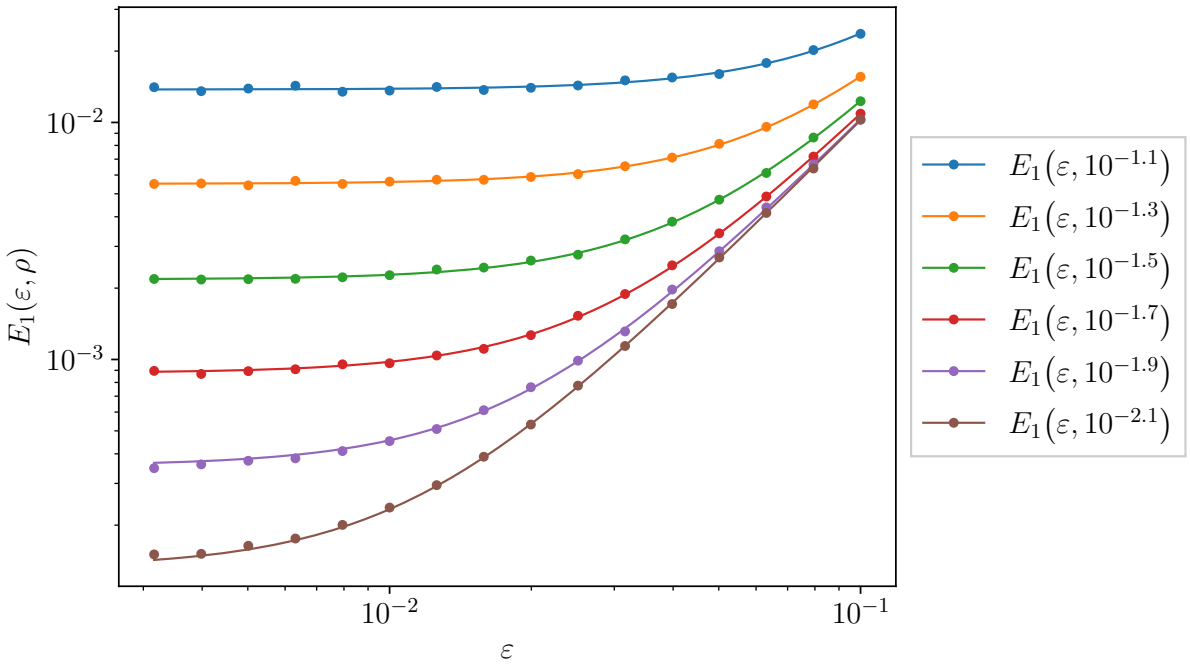
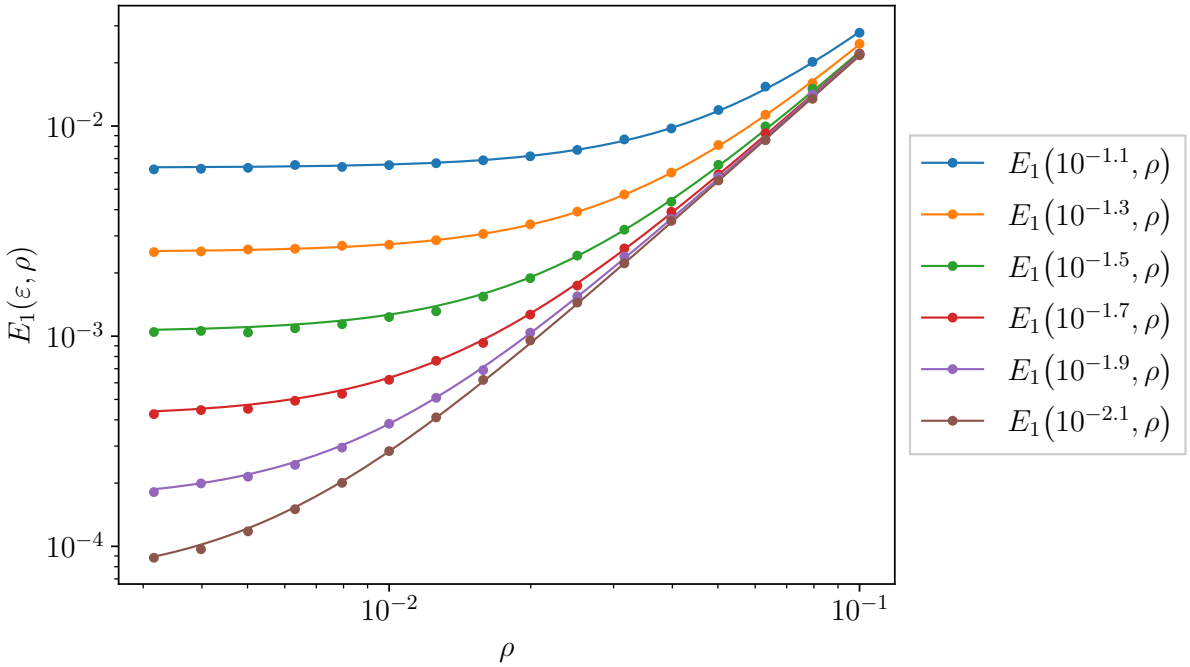


Figure 4.2: Histograms of stochastic samples of (4.3), subject to the Gaussian initial condition (4.4), for varying initial uncertainty scale ρ and ongoing uncertainty scale ε. The distribution of the corresponding solution (4.6) to the linearised equation is overlaid in black.



(a) Estimates of the strong error (with $r = 1$) in linearising (4.3) with (4.5), for varying ongoing uncertainty parameter ε. Each colour corresponds to a different value of the initial uncertainty parameter ρ. A (least squares) line of best fit of the form $\beta_0 + \beta_1 \varepsilon^2$ is included in the corresponding colour.



(b) Estimates of the strong error (with $r = 1$) for varying initial uncertainty parameter ρ. Each colour corresponds to a different value of the ongoing uncertainty parameter ε. A (least squares) line of best fit of the form $\beta_0 + \beta_1 \rho^2$ is included in the corresponding colour.

Figure 4.3: Validation of the theoretical bound predicted by Theorem 3.1, when $r = 1$, on numerical realisations of the solution to the 1D example (4.3).

4.2.2 Linear dynamics, multiplicative noise

Now consider the following SDE with multiplicative noise in 1D;

$$\mathrm{d}y_t^{(\varepsilon)} = \frac{1}{2} y_t^{(\varepsilon)} \,\mathrm{d}t + \varepsilon \cos\left(y_t^{(\varepsilon)}\right) \mathrm{d}W_t. \tag{4.8}$$

The corresponding deterministic system is linear and has solution

$$F_0^t(x_0) = \exp\left(\frac{t}{2}\right) x_0, \tag{4.9}$$

with additional details provided in the supplementary materials. As with the previous example in Section 4.2.1, we take the Gaussian initial condition (4.4) with variance ρ^2 and linearised (4.8) about the initial mean μ. The linearised equation is then

$$\mathrm{d}l_t^{(\varepsilon)} = \frac{1}{2} l_t^{(\varepsilon)} \,\mathrm{d}t + \varepsilon \cos\left(\exp\left(\frac{t}{2}\right)\mu\right) \mathrm{d}W_t, \quad l_0^{(\varepsilon)} \sim \mathcal{N}\left(\mu, \rho^2\right), \tag{4.10}$$

with Gaussian solution (4.6). We take the initial point $\mu = 2$ and consider the solutions at time $t = 1$. To generate numerical realisations of the solutions to (4.8) and (4.10) with the same realisations of W_t, we use the same set-up as in the previous example.

In Figure 4.4, we show histograms of $N = 10000$ samples of the multiplicative noise SDE (4.8) and the corresponding probability density function of the linearised solution, for different combinations of ε and ρ. We again see that in the limit of both small initial and small ongoing uncertainty (towards the bottom right), we see that the distribution of the samples approach the Gaussian density of the linearisation solution.

Since the drift term is linear and the noise multiplicative in (4.8), the bound predicted by Theorem 3.1 has the form

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq D_1(r, t, K_{\nabla u}, K_\sigma)\, \varepsilon^{2r} + M_r D_3(r, t, K_{\nabla u})\, \varepsilon^r \rho^r,$$

where we have $K_{\nabla\nabla u} = 0$ and $K_{\nabla\sigma} = 1$. In Figure 4.5, we again validate the form of this bound (for $r = 1$; results for additional values of r are provided in the supplementary material) by approximating the left-hand side with E_1 computed from realisations of the solution to (4.8) and the linearisation (4.10). For each fixed value of the initial uncertainty ρ, in Figure 4.5a, we fit a line of best fit of the form $\beta_1\varepsilon + \beta_2\varepsilon^2$ to validate that the strong error scales as predicted. Similarly, in Figure 4.5a we fit a line of best fit of the form $\beta_0 + \beta_1\rho$ and confirm that the linearisation error follows this scaling.

4.2.3 Fixed initial condition

In this example, we consider a 2-dimensional model and a fixed initial condition, to validate the results presented in Section 3.1.3. Following the example in Chapter 5 of Samelson

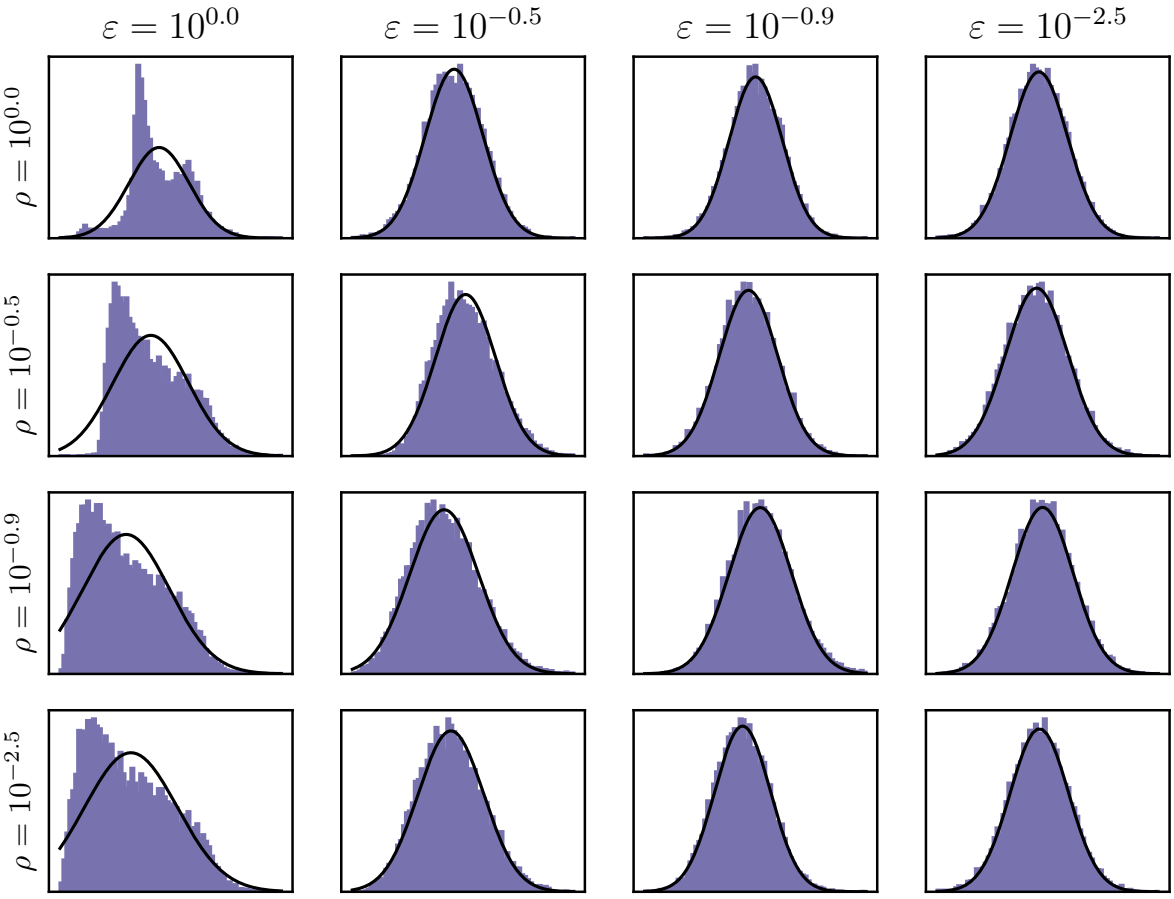
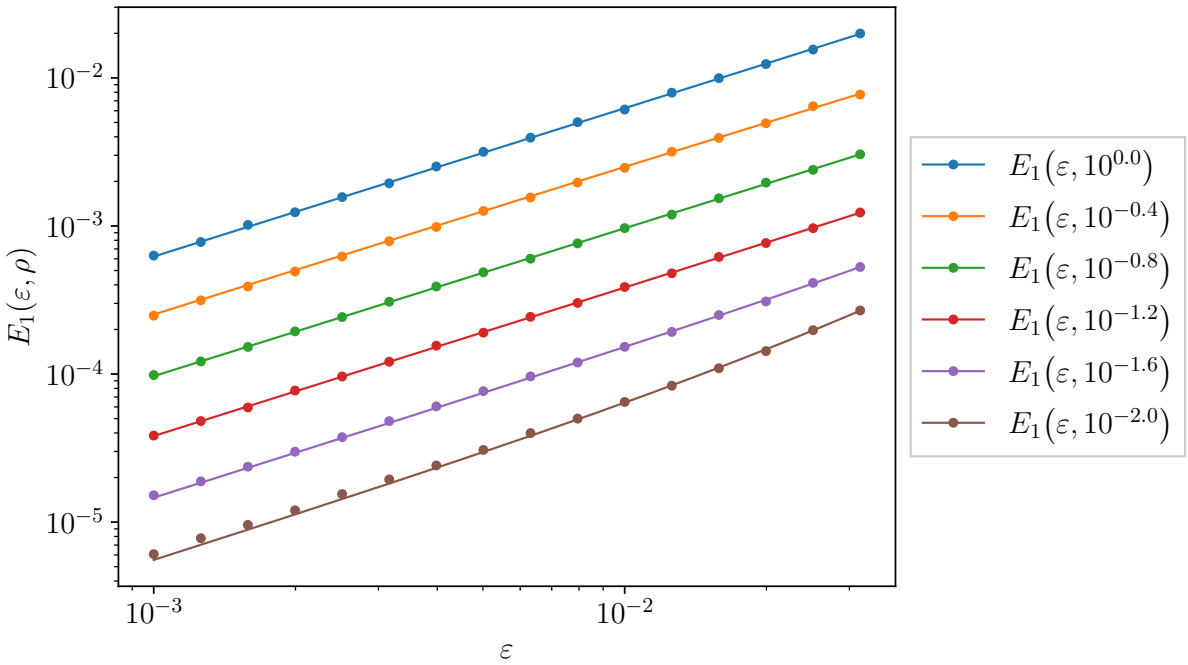
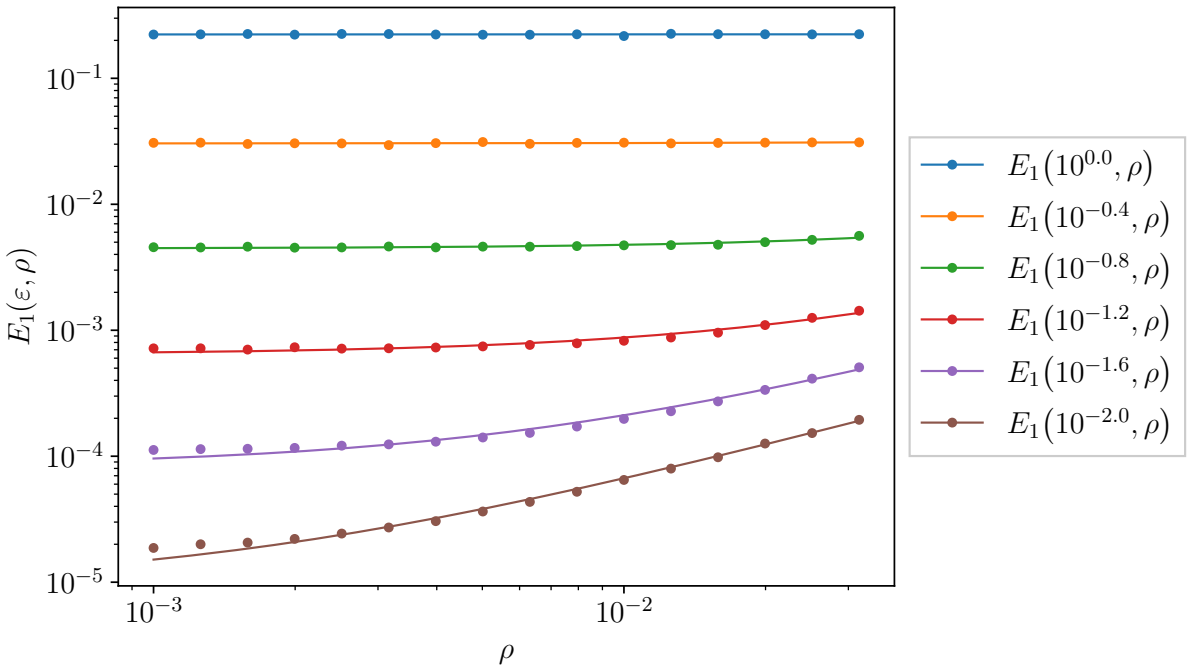


Figure 4.4: The same arrangement as Figure 4.2, but for the 1D multiplicative noise SDE (4.8).



(a) Estimates of the strong order (with $r = 1$) for varying ongoing uncertainty parameter ε. Each colour corresponds to a different value of the initial uncertainty parameter ρ. A (least squares) line of best fit of the form $\beta_1\varepsilon + \beta_2\varepsilon^2$ is included in the corresponding colour.



(b) Estimates of the strong order (with $r = 1$) for varying initial uncertainty parameter ρ. Each colour corresponds to a different value of the ongoing uncertainty parameter ε. A (least squares) line of best fit of the form $\beta_0 + \beta_1\rho$ is included in the corresponding colour.

Figure 4.5: Validation of the theoretical bound predicted by Theorem 3.1, when $r = 1$, on numerical realisations of the solution to the 1D example (4.8).

and Wiggins (2006), we consider an unsteady meandering jet in two dimensions, which may serve as an idealised model of geophysical Rossby waves (Pierrehumbert, 1991). The velocity field for $y \equiv (y_1, y_2)^\top$ is given by

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

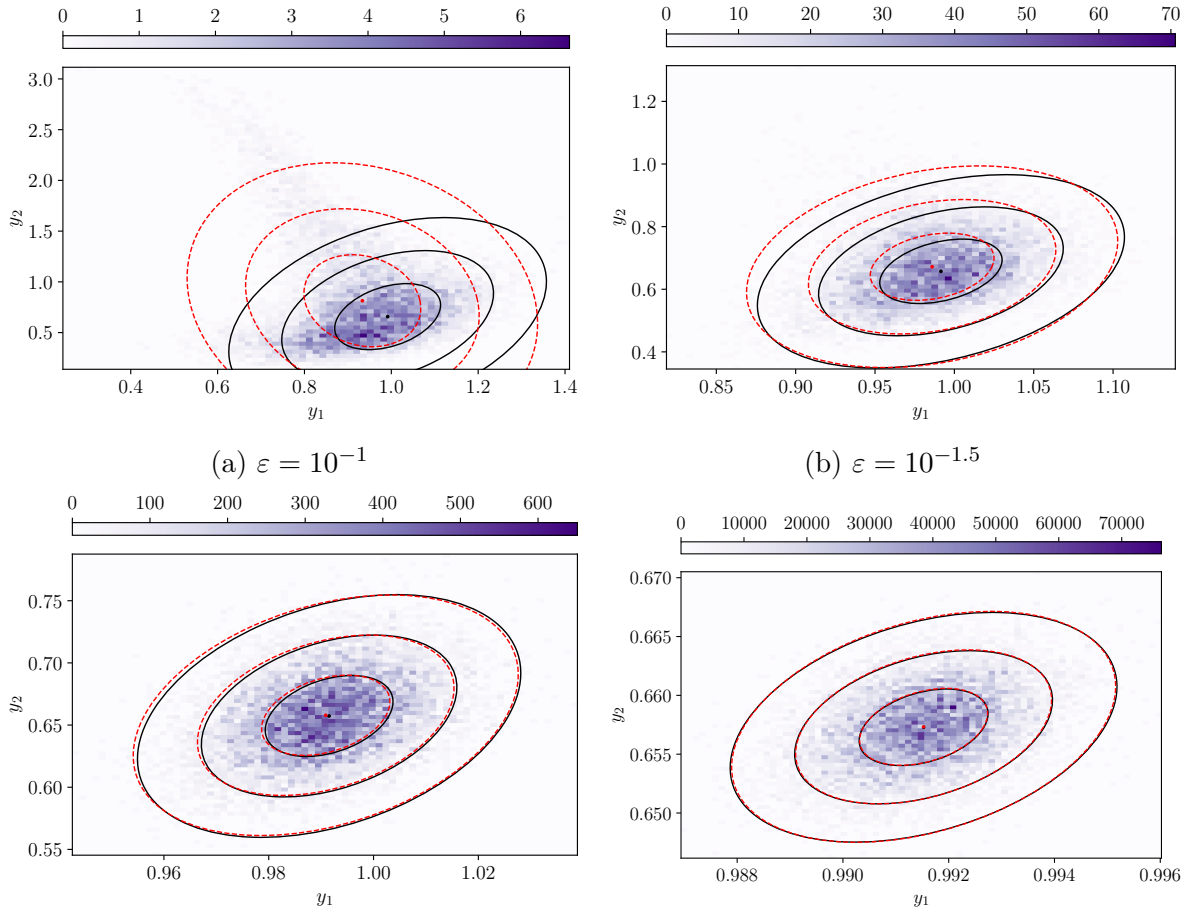
The velocity field describes a kinematic travelling wave with deterministic oscillatory perturbations in a co-moving frame. Here, A is the amplitude and c is the phase speed of the primary wave, and K is the wavenumber in the y_1-direction. The oscillatory perturbation has amplitude ϵ_{mj}, phase speed 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.

We introduce multiplicative noise by considering stochastic perturbations to the phase speed c and the primary amplitude A, which we model with the respective components of a 2-dimensional Wiener process $W_t = \left(W_t^{(1)}, W_t^{(2)}\right)^\top$. Then, we specify the diffusion term as

$$\sigma(y,t) = \begin{bmatrix} 1 & \sin(Ky_1)\cos(y_2) \\ 0 & K\cos(Ky_1)\sin(y_2) \end{bmatrix}. \tag{4.12}$$

We consider the fixed initial condition $x_0 = (0,1)$ and the prediction of the model at time $t = 1$. We then consider a linearisation of the SDE about the deterministic trajectory $F_0^t(x_0)$, where F_0^t is the deterministic flow map corresponding to the vector field (4.11). To compute the Gaussian distribution (3.16) of the linearised solution, we again solve (3.9) numerically with initial condition $\Sigma_0^t(x_0) = O$. Specifically, (3.9) is solved jointly with the deterministic state equation (3.2) using the hybrid method proposed by Mazzoni (2008). This hybrid method combines a Taylor-Heun approximation with a Gauss-Legendre one and ensures that the numerical solution of the covariance equation is symmetric and positive semi-definite while maintaining both accuracy and computational efficiency.

Figure 4.6 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 distribution that solves the linearised equation. The first three standard-deviation levels of the 2×2 sample covariance matrix of the realisations of $y_t^{(\varepsilon)}$, 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.



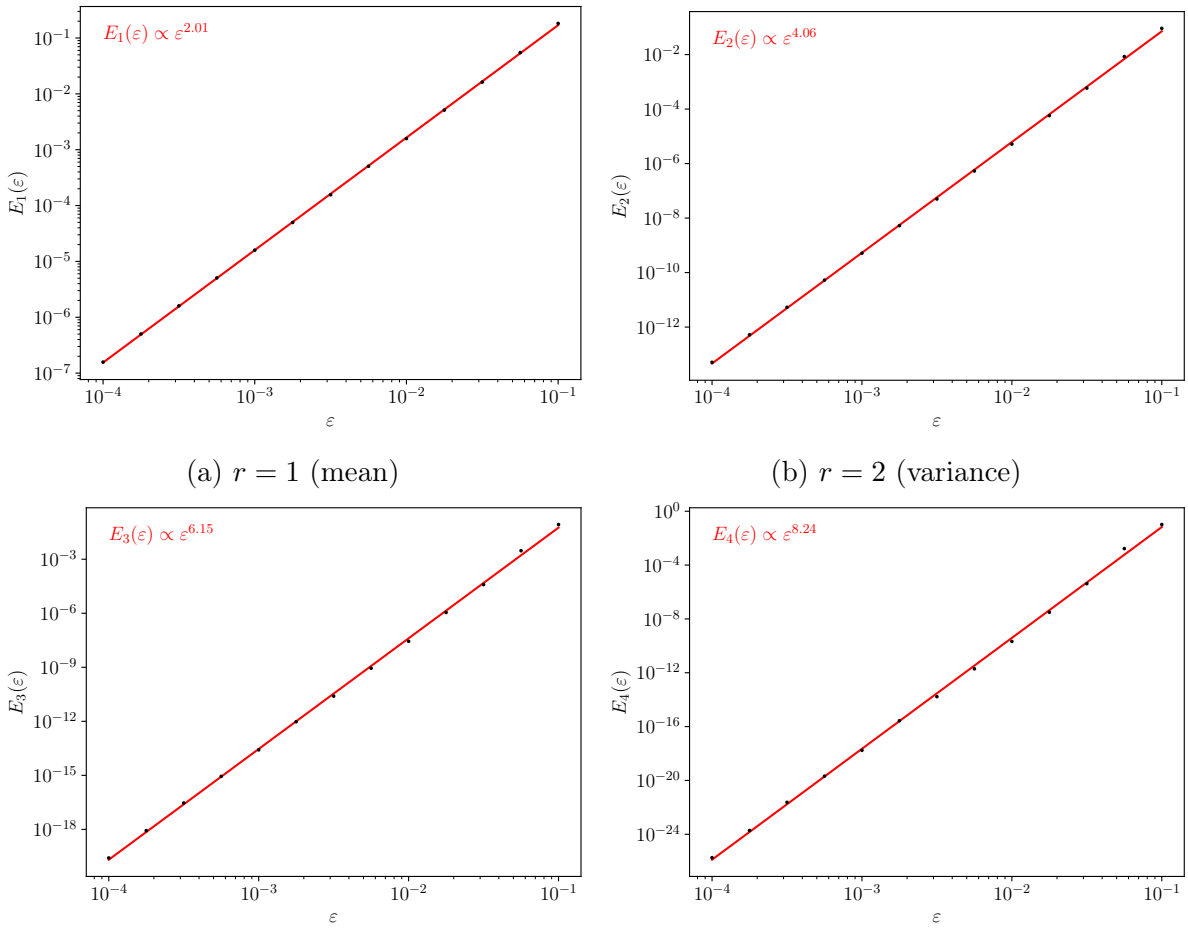
(a) $\varepsilon = 10^{-1}$

(b) $\varepsilon = 10^{-1.5}$

(c) $\varepsilon = 10^{-2}$

(d) $\varepsilon = 10^{-3}$

Figure 4.6: Histograms of $y_t^{(\varepsilon)}$ from direct simulation of the SDE with drift (4.11) and diffusivity (4.12) subject to the fixed initial condition, for four different ε values. Overlaid in black are contours of the Gaussian solution (3.16) of the linearised SDE (3.3), which correspond to the first three standard deviation levels centred at the 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) $r = 2$ (variance)

(c) $r = 3$ (skewness)

(d) $r = 4$ (kurtosis)

Figure 4.7: Validation of Theorem 3.1, by plotting the sample rth raw moment distance (the error metric $E_r(\varepsilon)$) between 10000 realisations of the meandering jet SDE and a corresponding linearisation, for decreasing values of ε. A line of best fit (in red) is placed on each, and the resulting slope indicated.

For a fixed initial condition, (3.4) predicts that the expected distance between the original SDE solution and that of a linearisation satisfies

$$\mathbb{E}\left[\left\|y_t^{(\varepsilon)} - l_t^{(\varepsilon)}\right\|^r\right] \leq \left(K_{\nabla\nabla u} + K_{\nabla\sigma}\right) D_1(r, t, K_{\nabla u}, K_\sigma)\, \varepsilon^{2r}.$$

To numerically estimate the left-hand side of (3.4), we again use a Monte Carlo estimator;

$$E_r(\varepsilon) \coloneqq \frac{1}{N} \sum_{i=1}^{N} \left\|\hat{y}_i^{(\varepsilon)} - \hat{l}_i^{(\varepsilon)}\right\|^r.$$

For $r = 1, 2, 3, 4$, $E_r(\varepsilon)$ is shown (in a logarithmic scale) for decreasing values of ε in Figure 4.7. Theorem 3.1 predicts that $\log_{10}\left(E_r(\varepsilon)\right)$ should decay linearly with a slope greater than $2r$ as ε decreases to zero. The least squares lines of best fit for each value of r in Figure 4.7 show this behaviour, and are therefore consistent with Theorem 3.1.

4.3 Computing stochastic sensitivity

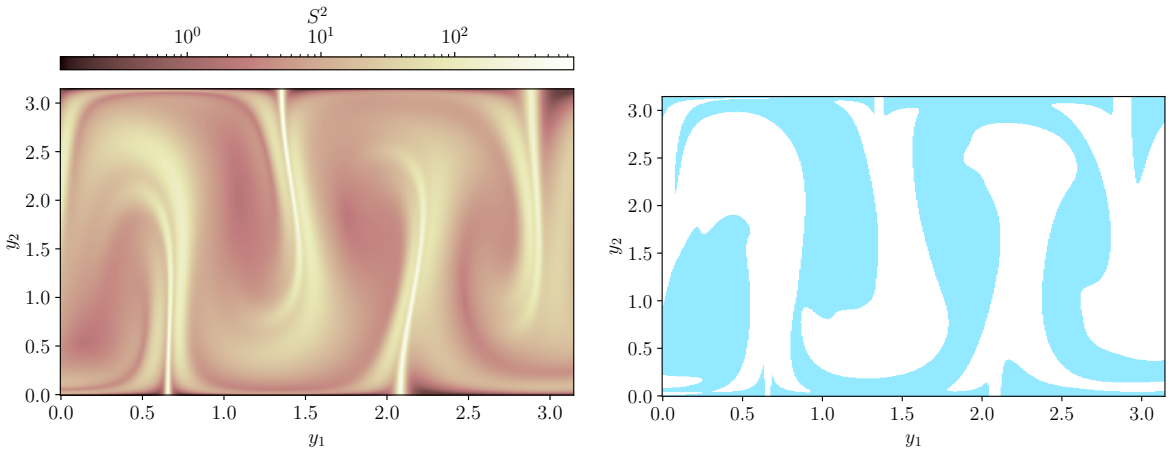
In this section, we illustrate the computability of stochastic sensitivity as described in Theorem 3.3. These computations are primarily demonstrative, so we keep the analysis of the results to a minimum and instead refer the reader to the works of Balasuriya (2020b), Fang, Balasuriya, et al. (2020), and Badza et al. (2023), where stochastic sensitivity has been calculated and interpreted in context.

4.3.1 In 2-dimensions

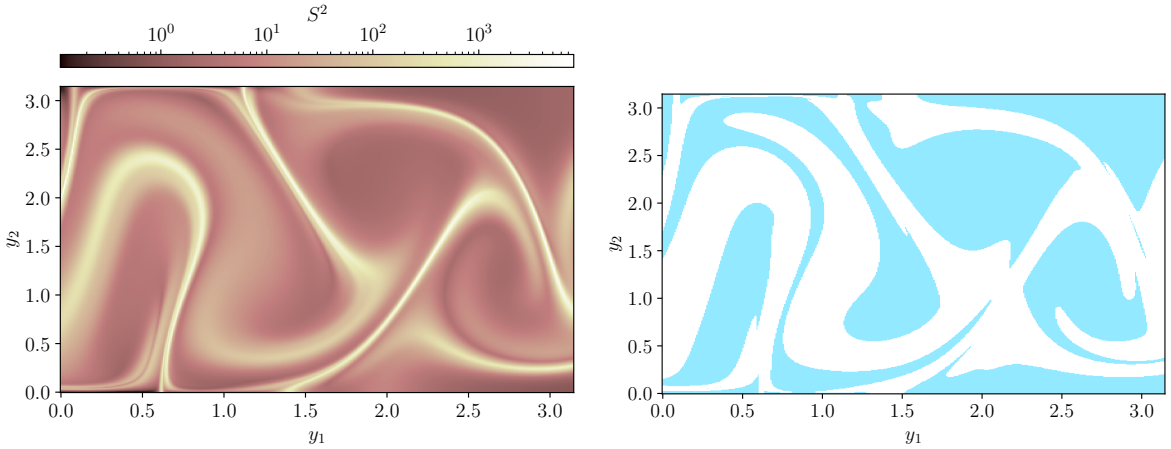
We again consider the meandering jet (4.11) with multiplicative noise described by (4.12) and take the same choice of parameters as in Section 4.2.3, except for the perturbation amplitude ϵ_{mj}, which is varied to obtain qualitatively different behaviour in the system. For each initial condition x_0 in a 400×400 uniform grid on $[0, \pi] \times [0, \pi]$, the S^2 value is calculated by first computing (by solving (4.1b)) the covariance matrix $\Sigma_0^t(x_0)$ of the fixed initial condition linearisation and then taking the operator norm per (3.17). Figure 4.8 shows the resulting S^2 field from time 0 to $t = 1$, for two different values of ϵ_{mj}. We also extract robust sets from each field, by highlighting in cyan on the right side of Figure 4.8 those initial conditions with a stochastic sensitivity value less than the specified threshold of 10. That is, cyan corresponds to the set

$$\text{RS}(10) = \left\{x_0 \in [0, \pi] \times [0, \pi] \,\middle|\, S^2(x_0, t) < 10\right\}.$$

When $\epsilon_{\text{mj}} = 0.3$ (Figure 4.8a), the S^2 field is largest in the elongated gyre regions outside of the meandering jet, where the flow exhibits chaotic behaviour (Pierrehumbert, 1991)



(a) $\epsilon_{\mathrm{mj}} = 0.3$



(b) $\epsilon_{\mathrm{mj}} = 1.0$

Figure 4.8: (Left) The S^2 field of the meandering jet flow (4.11) over the time interval $[0, 1]$, for two different sets of parameters with qualitatively different behaviour. (Right) Robust sets RS(10) extracted from the stochastic sensitivity fields, by taking the initial conditions with a stochastic sensitivity value about a threshold of 10.

and we accordingly expect larger uncertainty due to the model dynamics. As a region of small S^2 value, the meandering jet emerges as a robust set, consisting of initial points whose eventual fate is significantly more certain than in other regions. When $\epsilon_{\text{mj}} = 1.0$ (Figure 4.8b), the deterministic flow is dominated by oscillatory perturbation, which further increases the chaotic nature of the solving trajectories and the boundaries between the gyres and the meandering jet are no longer distinguishable (Crocker, 2021). Rotational eddies begin to dominate the flow and we see this reflected in the stochastic sensitivity field in Figure 4.8b: the eddies exhibit a smaller stochastic sensitivity, as even under stochasticity trajectories, are inclined to remain within them. The resulting robust sets consequently also highlight these eddies.

4.3.2 In 3-dimensions

In Definition 3.1, we have provided a new definition for stochastic sensitivity in arbitrary dimensions, whereas previously, the definition and computation (Balasuriya, 2020b) were limited to only two. This is the second main contribution of this thesis and so we shall demonstrate this extension on an example toy model in 3-dimensions. Consider the Gromeka-Arnold-Beltrami-Childress flow (Dombre et al., 1986):

$$u(y) = \begin{bmatrix} A\sin(y_3) + C\cos(y_2) \\ B\sin(y_1) + A\cos(y_3) \\ C\sin(y_2) + B\cos(y_1) \end{bmatrix}, \tag{4.13}$$

where $A, B, C > 0$ are constants and $y \equiv (y_1, y_2, y_3)^\top$. The flow arises as an exact solution to Euler's equation and is often used as a testbed for Lagrangian analysis in 3-dimensions (Haller, 2001; Brunton and Rowley, 2010; Sulman et al., 2013; Nelson and Jacobs, 2016, e.g.). We take the parameter values $A = \sqrt{3}$, $B = \sqrt{2}$, and $C = 1$, for which the flow is spatially periodic in the cube $[0, 2\pi] \times [0, 2\pi] \times [0, 2\pi]$. The flow then comprises a network of vortex structures and homoclinic orbits that cause distinct regions of 'order' and chaos to emerge (Dombre et al., 1986).

To introduce 3-dimensional noise to the system, we set $\sigma = I$, the 3×3 identity matrix, and compute stochastic sensitivity for a $200 \times 200 \times 200$ grid of initial conditions within the cube $[0, 2\pi] \times [0, 2\pi] \times [0, 2\pi]$. As with the 2-dimensional example, for each initial condition, we compute the 3×3 covariance matrix using the Mazzoni method. The stochastic sensitivity value is then computed by taking the operator norm of this covariance matrix. To visualise the three-dimensional structure of the field, in Figure 4.10 we plot the stochastic sensitivity field on slices of the cube $[0, 2\pi] \times [0, 2\pi] \times [0, 2\pi]$ in the arrangement given in Figure 4.9.

The stochastic sensitivity field highlights the 3-dimensional structure of the vortices—compare Figure 4.10 to the Poincarè sections in Figure 7 of Dombre et al. (1986), which show the structure of the vortices because of the *deterministic* dynamics. Smaller stochastic

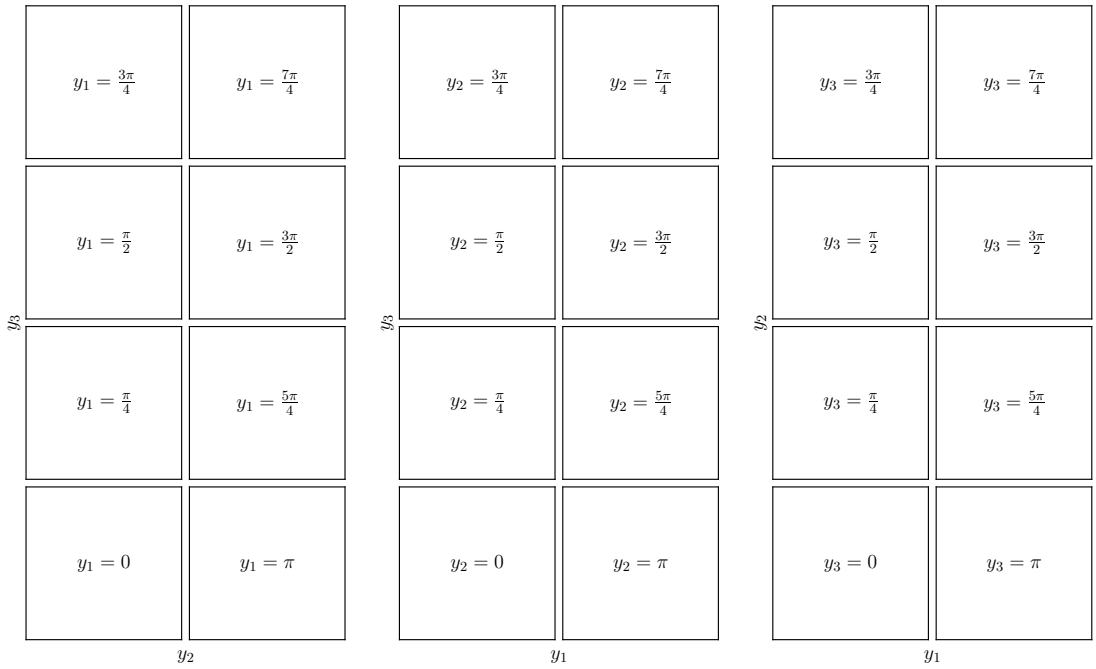


Figure 4.9: The arrangement of figures in Figure 4.10.

sensitivity values distinguish the interiors of most vortices, suggesting that even under small stochasticity, these structures remain coherent. This matches conclusions from the deterministic flow, where there interiors of the vortices are 'ordered' rather than chaotic; the stochastic sensitivity field provides a similar qualitative insight. The boundaries of these vortices are also highlighted by sharp ridges of large stochastic sensitivity values. These narrow ridge-like structures correspond to the unstable manifolds within the flow (Dombre et al., 1986); the repelling nature of these regions results in a 'larger' stochasticity.

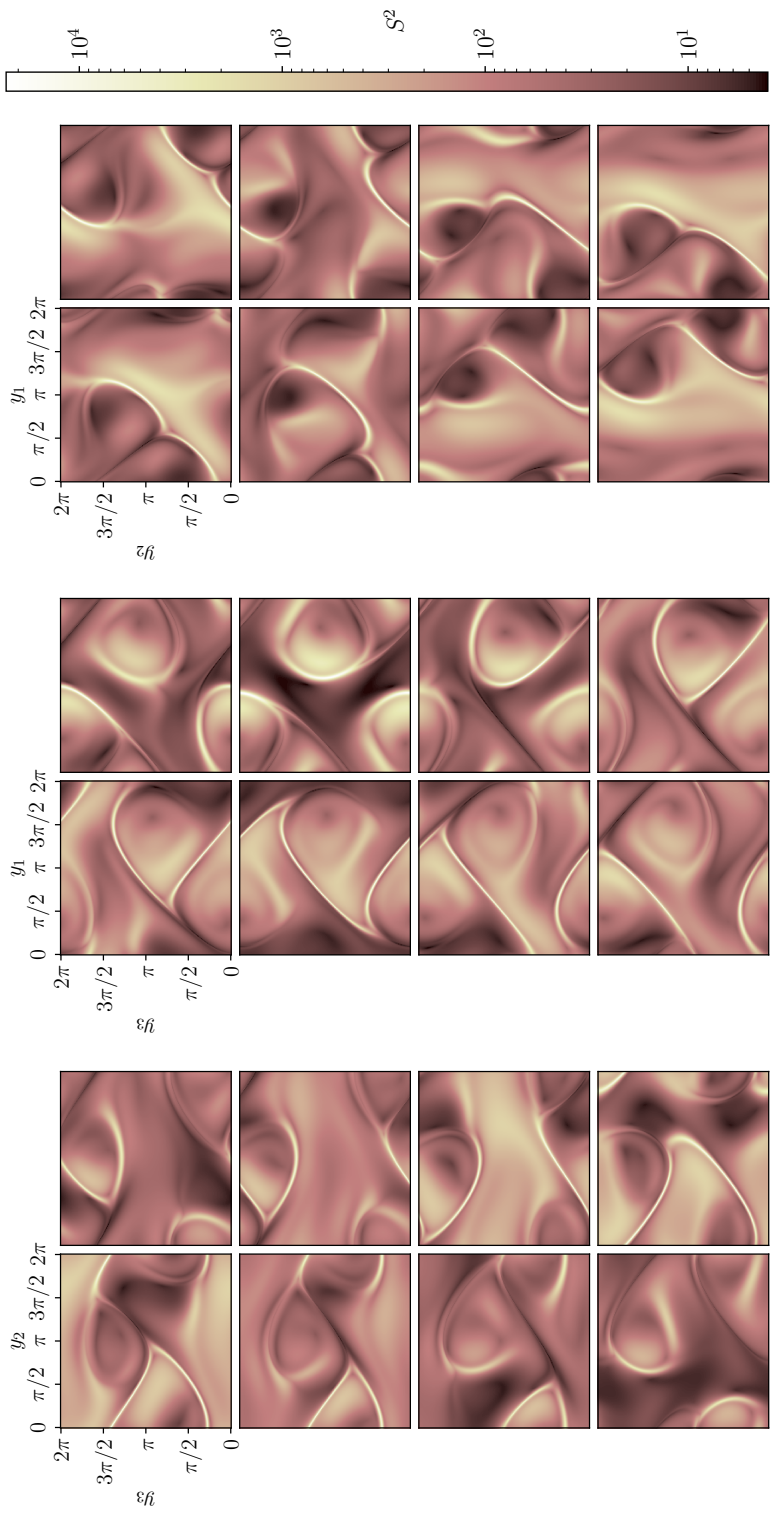


Figure 4.10: The stochastic sensitivity field for slices of the GABC flow (4.13) with identity diffusivity matrix, following the arrangement specified in Figure 4.9. The colour bar applies to every individual figure.

Chapter 5

A Gaussian mixture model

In Chapter 3, we described and justified a framework for computing approximations to the solutions of nonlinear SDEs via linearisations about solutions to the corresponding deterministic system. A key advantage of this approximation is the efficiency of computation; rather than having to generate a large number of realisations of the SDE solution we can solve a smaller system of differential equations to obtain the first two moments. When the initial condition is fixed or Gaussian, the resulting linearisation solution is a Gaussian process, providing an approximation characterised entirely by these two moments and an analytically available probability density function. A Gaussian approximation has further practical advantages, leading to its use across many different applications (Archambeau et al., 2007; Jazwinski, 2014; Sanz-Alonso and Stuart, 2017; Särkkä and Solin, 2019; Kaszás and Haller, 2020). For example, even in inference requiring Monte Carlo simulations, a Gaussian distribution with a specified mean and covariance is faster to sample from than generating numerical solutions of the nonlinear SDE.

Recall that we are interested in approximating the solution at a time t to the nonlinear stochastic differential equation

$$\mathrm{d}y_t = u(y_t, t)\,\mathrm{d}t + \varepsilon\sigma(y_t, t)\,\mathrm{d}W_t, \tag{5.1}$$

subject to some random initial condition $y_s = x$, by using linearisations about solutions of the corresponding deterministic system

$$\frac{\mathrm{d}F_s^t(x_0)}{\mathrm{d}t} = u\big(F_s^t(x_0)\,, t\big)\,, \quad F_s^s(x_0) = x_0. \tag{5.2}$$

The behaviour of the solution (5.1) over the time interval $[s, t]$ for small ε can be approximated by the linearised SDE

$$\mathrm{d}l_t = \Big(u\big(F_s^t(x)\,, t\big) + \nabla u\big(F_s^t(x)\,, t\big)\,\big[l_t - F_s^t(x)\big]\Big)\,\mathrm{d}t + \varepsilon\sigma\big(F_s^t(x)\,, t\big)\,\mathrm{d}W_t, \tag{5.3}$$

subject to the initial condition $l_s = x$. Chapter 3 considered the evolution of (5.1) over a time interval $[0, t]$ and subject to some initial condition at time 0. However, by using

a simple time shift our theory can be applied to any finite time interval $[s, t]$, where the solution is specified at time s instead of 0, without loss of generality. Note that we have now also dropped the dependence of ε in the notation y_t and l_t, as ε is now treated as a fixed value specified as part of the model. Suppose that the random initial condition x follows a Gaussian distribution with mean x_s and a specified covariance matrix Σ_s. This also permits the initial condition to be fixed as x_s: we then set $\Sigma_s = O$, the $n \times n$ zero matrix, and follow the convention that the resulting zero-variance Gaussian is a Dirac delta centred at the mean x_s. We will focus our attention on this case for the remainder of this thesis, having already provided a general framework in Chapter 3 for other initial conditions. By taking the mean x_s as the initial condition to (5.2), we ensure that the initial uncertainty (measured by the L_r-norm as in Chapter 3) scales with the trace of Σ_s (recall (3.13)). The solution to (5.3) is then a Gaussian process characterised by the mean $F_s^t(x_0)$ and covariance matrix $\mathbb{V}[l_t]$, for which explicit expressions are given in Corollary 3.1. To recall the notation used in Section 4.1, set $w_t \equiv F_s^t(x_s)$ and $\Pi_t \equiv \mathbb{V}[l_t]$. When the Jacobian ∇u of the vector field u is available or can be approximated appropriately, the moments of the Gaussian solution can be obtained by the system of ordinary differential equations

$$\frac{\mathrm{d}w_t}{\mathrm{d}t} = u(w_t, t)\,, \quad w_s = x_0 \tag{5.4a}$$

$$\frac{\mathrm{d}\Pi_t}{\mathrm{d}t} = \nabla u(w_t, t)\,\Pi_t + \Pi_t \left[\nabla u\,(w_t, t)\right]^\top + \sigma\,(w_t, t)\,\sigma\,(w_t, t)^\top\,, \quad \Pi_s = \Sigma_s. \tag{5.4b}$$

In practice, (5.4) must be solved numerically, but can be more computationally efficient than the alternative of generating many realisations of (5.1). Thus, we can efficiently compute a Gaussian approximation $\mathcal{N}(w_t, \Sigma_t)$ to the solution to the nonlinear SDE (5.1) at time t by solving (5.4) only. The Mazzoni (2008) method, outlined in Section 4.1, provides a computationally efficient algorithm for solving (5.4) jointly.

Despite the computational appeal, this Gaussian approximation is only 'reasonable' in the limit of small initial and ongoing uncertainty, which we quantified precisely with the bound in Theorem 3.1, and demonstrated heuristically with examples in Figures 4.2, 4.4 and 4.6. In practice, the scale of the noise is an inherent part of a model and there is no guarantee that it is sufficiently small for this Gaussian approximation to be reasonable. When the noise scale is larger, we see non-Gaussianity emerge—the numerical solutions in Figures 4.2, 4.4 and 4.6 exhibit skewness, bi-modality, and other departures from Gaussianity. Non-Gaussianity is also often observed in statistical measurements in practice, such as in atmospheric regimes (Sura et al., 2005), experimental fluid flows (del-Castillo-Negrete, 1998), and oceanic flows (Bracco et al., 2000). Although the Gaussian approximation can still provide qualitative insight into the behaviour of these solutions (see stochastic sensitivity, for instance), to produce a reasonable approximation we need a method that can capture these departures from Gaussianity.

To further illustrate this point, we consider another example of a 1-dimensional

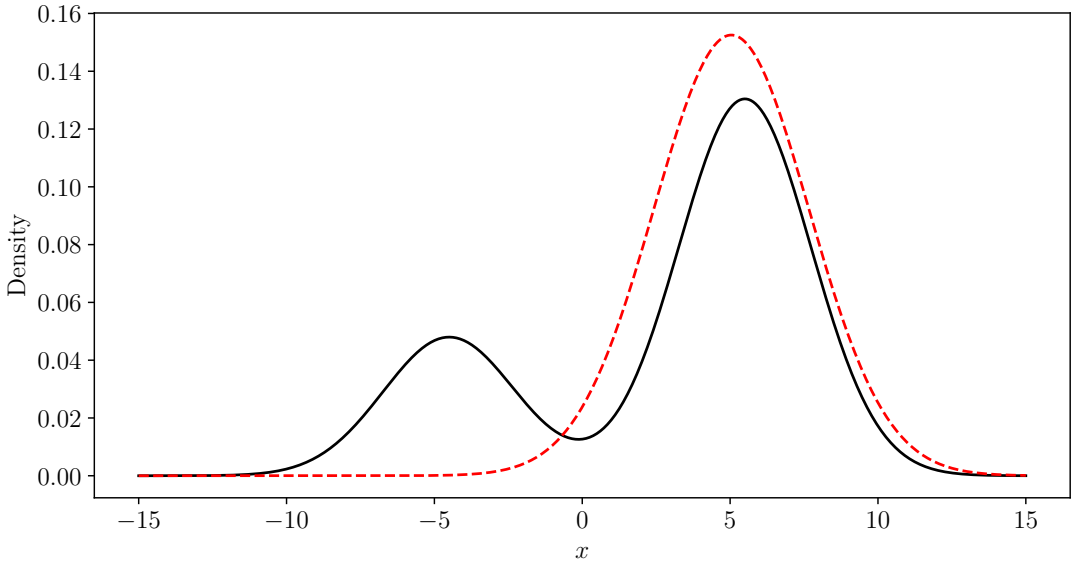


Figure 5.1: The probability density function (5.8) of the solution x_5 (in black) to the Beneš SDE (5.5), with fixed initial condition $x_0 = 1/2$. The density function of the Gaussian solution to the corresponding linearisation is overlaid in dashed red.

stochastic differential equation. The Beneš SDE (Särkkä and Solin, 2019) is

$$\mathrm{d}x_t = \tanh(x_t)\,\mathrm{d}t + \mathrm{d}W_t. \tag{5.5}$$

We are implicitly taking $\varepsilon = 1$, which is a larger noise scale than we typically expect to apply our results to. However, this example is purely demonstrative to highlight a limitation of Gaussian approximations and show the potential of the algorithm we are about to propose; later examples will involve smaller values of ε. The deterministic system corresponding to (5.5) is

$$\frac{\mathrm{d}F_s^t(x)}{\mathrm{d}t} = \tanh\big(F_s^t(x)\big)\,, \quad F_s^s(x) = x, \tag{5.6}$$

which includes an unstable fixed point at 0. The solution to (5.6) over the time interval (s,t) is

$$F_s^t(x) = \operatorname{arcsinh}\big(e^{t-s}\sinh(x)\big)\,. \tag{5.7}$$

The probability density function of the weak solution to (5.5) can be derived using an appropriate change of measure with Girsanov's theorem (see Section 7.3 of Särkkä and Solin (2019)). The solution x_t at time t has the probability density function

$$p(x,t) = \frac{1}{\sqrt{2\pi t}}\frac{\cosh(x)}{\cosh(x_0)}\exp\left[-\frac{t}{2} - \frac{1}{2t}\,(x - x_0)^2\right], \tag{5.8}$$

where x_0 is the fixed initial condition. In this example, we consider the solution to (5.5) subject to the fixed initial condition $x_0 = 1/2$ and at time $t = 5$ and plot the resulting PDF of the solution (5.8) in black in Figure 5.1. The density is not symmetric with two distinct modes that result from the unstable fixed point of (5.6) at $x = 0$. Many stochastic trajectories are driven away from zero, resulting in the predominant mode centred at $x = 11/2$. However, when the stochastic perturbations force a trajectory through the fixed point and into negative values, they are pushed further in the negative direction, leading to the second mode at $x = -9/2$. Nonetheless, we can linearise (5.5) about the deterministic trajectory $F_0^t(1/2)$ solving (5.6) to obtain a Gaussian approximation. In general, if we linearise (5.5) over the time interval $[s, t]$ about the trajectory $F_s^t(x_s)$, we are considering the equation

$$\mathrm{d}l_t = \Big(\tanh\big(F_s^t(x_s)\big) + \operatorname{arcsech}\big(F_0^t(x_s)\big)\,\big[l_t - F_s^t(x_s)\big]\Big)\,\mathrm{d}t + \mathrm{d}W_t,$$

where the deterministic flow map F_s^t given by (5.7). Since the deterministic flow map is available analytically in (5.7), we can determine the variance of the solution to the linearised SDE exactly by evaluating (3.8). The solution to (5.5) at time t is then approximated by the Gaussian distribution

$$l_t \sim \mathcal{N}\left(\operatorname{arcsinh}\big(e^{t-s}\sinh(x_s)\big)\,,\ \frac{2\sinh^2(x_s)\,t + e^{2t-2s} - 1}{2\sinh^2(x_s) + 2e^{-2s-2t}}\right).$$

Returning to our specific example ($x_0 = 1/2$, $s = 0$, and $t = 5$), we plot the density of the Gaussian approximation l_5 in dashed red in Figure 5.1. The Gaussian approximation cannot capture the bimodality of the true solution, and so only a single mode is captured. This is a significant limitation of using a single linearisation approximation.

To improve upon the single linearisation approximation, we seek a scheme that can capture departures from Gaussianity in the SDE solution, while still taking advantage of the efficient computation of the linearisation solution. An alternative perspective is that the linearised SDE captures the behaviour of stochastic fluctuations close to a deterministic trajectory, in a similar sense to how a Taylor polynomial captures the local behaviour of a nonlinear function. By 'piecing' together several of these approximations together we can capture the stochastic behaviour in different regions of the state space. A Gaussian mixture model (GMM) provides a framework for combining multiple Gaussian densities into a single distribution and is thus the obvious choice for constructing non-Gaussian densities out of our Gaussian approximations. In this chapter, we outline an algorithm that uses *multiple* linearisation approximations to construct a Gaussian mixture model that approximates the nonlinear SDE solution.

In general, a Gaussian mixture model with K components is a probability density function of the form

$$p(z) = \sum_{k=1}^{K} \omega_k \mathcal{N}\Big(z;\, \mu^{(k)}, \Sigma^{(k)}\Big),$$

(a) Skewness (b) Bimodality

Figure 5.2: The probability density functions of two Gaussian mixture models in 1-dimension both using two equally weighted components. When individual components (dashed) are combined to produce non-Gaussian mixture densities (solid), they can exhibit both (a) skewness and (b) bimodality.

consisting of K Gaussian components $\mathcal{N}\left(\mu^{(k)}, \Sigma^{(k)}\right)$ with respective weights $\omega_1, \ldots, \omega_K \geq 0$ that satisfy $\sum_{k=1}^{K} \omega_k = 1$. With sufficient components, a Gaussian mixture model can recreate any probability distribution in $\mathbb{R}^n$ while having many of the properties that make Gaussian distributions appealing in practice (McLachlan et al., 2019). Figure 5.2 shows two examples of Gaussian mixture models in 1-dimension, where departures from Gaussianity such as multimodality and skewness can be captured with an appropriate combination of components.

5.1 The GMM algorithm

We will now outline our algorithm to approximate the SDE solution with a Gaussian mixture model while taking advantage of the computational efficiency of the Gaussian linearisation approximation. Given a Gaussian component at a time s, we can 'propagate' the component forward to a later time t by solving (5.4) initialised with the component mean and covariance matrix. Given a mixture model with multiple Gaussian components, we can propagate each component *separately* to update the full model through time in a computationally efficient manner. Each component is propagated by approximating the original SDE (5.1) with a different linearisation, about the deterministic trajectory resulting from the component mean.

Our proposed method is *ad hoc* and based on an intuition: the Gaussian solution provides a reasonable approximation for the local behaviour of the true SDE solution over a short timeframe, but once this is no longer the case, we can capture departures from Gaussianity by introducing more components in the mixture model. We expect heuristically that as the number of components increases, the full mixture model should provide a closer approximation of the SDE solution density, provided that the components are appropriately placed. However, our goal is to provide a numerically efficient algorithm, so we wish to minimise the number of components and use the Gaussian approximation wherever possible. We therefore propose propagating Gaussian components forward through the linearisation model until are no longer reasonable approximations of the local solution

behaviour. Then, we replace the violating component with several smaller judiciously chosen components and propagate each new component individually. We term this the *splitting* step, where a Gaussian component is split into several smaller ones. The new components should be chosen in such a way as to 'preserve' the original component, which can be achieved in one sense as follows. Let Ξ follow a Gaussian mixture density with N components, with respective weights $w^{(1)}, \dots, w^{(N)}$, means $\mu^{(1)}, \dots, \mu^{(N)}$, and covariance matrices $\Sigma^{(1)}, \dots, \Sigma^{(N)}$. The variance of the mixture model is then

$$\begin{aligned}\mathbb{V}[\Xi] &= \sum_{i=1}^{N} w^{(i)}\Sigma^{(i)} + \sum_{i=1}^{N} w^{(i)}\left(\mu^{(i)} - \bar{\mu}\right)\left(\mu^{(i)} - \bar{\mu}\right)^{\top} \\ &= \text{Mean of covariances} + \text{Covariance of means},\end{aligned}$$

where $\bar{\mu} = \mathbb{E}[\Xi] = \sum_{i=1}^{N} w^{(i)}\mu^{(i)}$ is the overall mean. This decomposition suggests, at least heuristically, that we can include additional uncertainty (in the form of contributions to the overall variance) within the component means themselves. By replacing a single component with points, we can preserve the mean and covariance of the component while introducing additional components that can closely match the non-Gaussian target distribution. This leads to the following condition: the K splitting points (the new component means) $x^{(1)}, \dots, x^{(K)}$ that replace x should be chosen so that

$$\sum_{i=1}^{K} \hat{w}_k x^{(k)} = x, \quad \sum_{i=1}^{K} \hat{w}_k \left(x^{(k)} - x\right)\left(x^{(k)} - x\right)^{\top} = \Sigma_0. \tag{5.9}$$

with weights $\hat{w}_1, \dots, \hat{w}_K > 0$ satisfying $\sum_{k=1}^{K} \hat{w}_k = 1$. Each new component is assigned a zero variance in this formulation, but with an appropriate adjustment of (5.9) each can have a specified variance. This ensures that the mean and covariance of the original Gaussian are preserved within the (sample) mean and covariance of the new points themselves. Note that at least $K = n + 1$ points are necessary for (5.9) to be satisfied. The selection of splitting points is similar to the notion of 'sigma points', employed in the unscented transform to encode an initial mean and covariance (Uhlmann, 1995; Julier et al., 2000). Any such sigma points satisfy (5.9)—Table I in Menegaz et al. (2015) provides a list of sigma points and their weightings used across other literature (and reviewed in the context of Kalman filtering)—and therefore can be used in our algorithm.

The canonical set of sigma points originally proposed by Uhlmann (1995), which we use to apply the GMM algorithm in Chapter 6, are

$$x^{(1)} = x, \tag{5.10a}$$

$$x^{(1+i)} = x + \sqrt{n + \frac{1}{2}}\left[\sqrt{\Sigma}\right]_{\cdot i}, \tag{5.10b}$$

$$x^{(1+n+i)} = x - \sqrt{n + \frac{1}{2}}\left[\sqrt{\Sigma}\right]_{\cdot i}, \tag{5.10c}$$

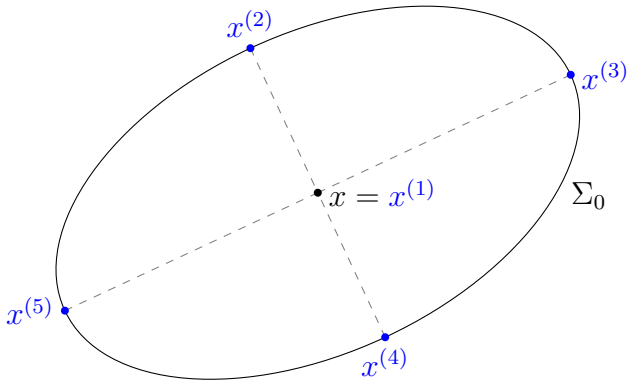


Figure 5.3: The splitting of a 2-dimensional mean x and 2×2 covariance matrix Σ (in black) into 5 sigma points $x^{(1)}, \ldots, x^{(5)}$ (in blue), using the canonical set (5.10). The mean x is preserved as one of the points, and the four others are placed at the vertices and co-vertices of the first standard deviation ellipse of Σ.

for $i = 1, \ldots, n$, where $\sqrt{\Sigma}$ denotes the symmetric square root of Σ, and $\left[\sqrt{\Sigma}\right]_{\cdot i}$ denotes the ith column of $\sqrt{\Sigma}$. The points are uniformly weighted, i.e. $\hat{w}_k = 1/(2n+1)$. Figure 5.3 depicts the splitting of a mean and covariance matrix pair into 5 sigma points in 2-dimensions, using the canonical set (5.10). By perturbing the mean by the columns of the square root of the covariance matrix, the sigma points are placed at the vertices of the ellipse representing the matrix. However, we do not enforce a particular approach for selecting these points and leave this to be investigated further as future work.

The two critical steps of the algorithm are *when* to split, a criterion that should decide when the Gaussian approximation is no longer a reasonable representation of the nearby solution behaviour, and *how* to split, with the appropriate method that ensures the new components satisfy (5.9). We do not provide specific choices for either step and save a thorough investigation for future work. In Section 5.2, we do however include a brief discussion of some options for the splitting criterion.

The mixture model algorithm is as follows:

1. Initialise a Gaussian mixture model with N components, setting $x^{(1)}, \ldots, x^{(N)}$ to be the component means, $\Sigma^{(1)}, \ldots, \Sigma^{(N)}$ to be the component covariance matrices, and $w^{(1)}, \ldots, w^{(N)}$ to be the component weights. For a fixed initial condition x_0, set $N = 1$, $x^{(1)} = x_0$, $\Sigma^{(1)} = O$ and $\omega^{(1)}$. Otherwise, the components and weights are represent the random initial distribution. Set $\tau^{(1)} = \cdots = \tau^{(N)}$.

2. While $\tau^{(i)} < t$ for any $i = 1, \ldots, N$, iterate the following;

   (a) Set j to be any i for which $\tau^{(i)} < T$.

(b) Update $x^{(j)}$ and $\Sigma^{(j)}$ by solving the joint system (5.4) with initial state $x^{(j)}$ and covariance $\Sigma^{(j)}$, terminating when a split condition is met or the final time T is reached. Set $\tau^{(j)}$ to the time at which this solution terminates.

(c) If $\tau^{(j)} = t$, then complete this branch of the algorithm.

(d) Otherwise, if $\tau^{(j)} < t$, construct K points $\hat{x}^{(1)}, \dots, \hat{x}^{(K)}$ with corresponding weights $\hat{w}^{(1)}, \dots, \hat{w}^{(K)}$ that preserve the propagated mean and covariance (i.e. satisfying (5.9)). Set $x^{(j)} = \hat{x}^{(1)}$ and $\Sigma^{(j)} = O$, and for each $k = 2, \dots, K$:

$$\begin{aligned} \Sigma^{(N+k-1)} &= O \\ w^{(N+k-1)} &= \hat{w}^{(k)} w^{(j)} \\ \tau^{(N+k-1)} &= \tau^{(j)}. \end{aligned}$$

Update the weight of the first sigma point as $w^{(j)} = \hat{w}^{(1)} w^{(j)}$, and set $N = N + K - 1$.

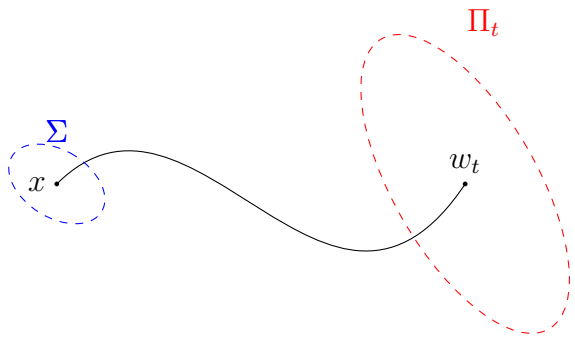
3. Construct the final mixture model with density function

$$G(x) = \sum_{i=1}^{N} w^{(i)} \mathcal{N}\left(x;\, x^{(i)}, \Sigma^{(i)}\right).$$

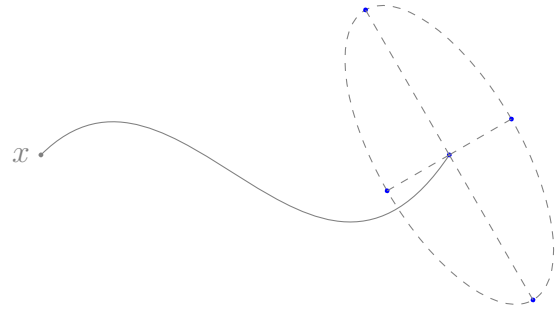
Figure 5.4 provides a pictorial representation of the propagation and splitting of a single component in the mixture model. The algorithm is fully deterministic, up to the choices for criterion and splitting, and since each component evolves and is monitored independently of the others can be implemented in parallel. Thus, the algorithm would present a significant computational improvement over bulk stochastic simulation.

There are several options for initialising the mixture model, depending on the initial condition at time 0. For a fixed initial condition, we can take the degenerate mixture model with a single component and zero variance; this implementation is summarised with an algorithm flow chart in Figure 5.5, which also describes the main propagation-splitting loop (indicated by red). If the initial condition is Gaussian, this can be used as a single component mixture model and propagated forward with (5.4) by including the initial covariance matrix as the initial condition to (5.4b). For a non-Gaussian initial condition, if this distribution can be represented, at least approximately, with a Gaussian mixture model, then this can be used immediately in the algorithm. The components and weights of the initial condition are updated in Step 2 of the algorithm.

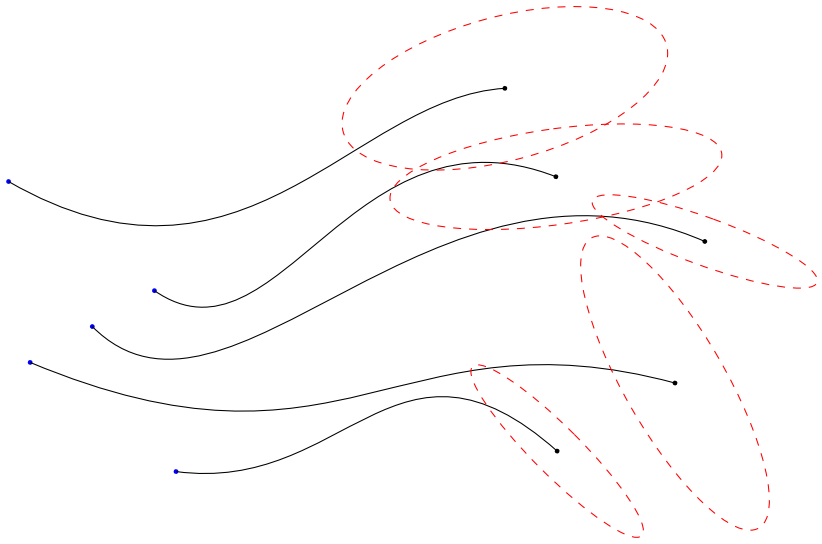
The number of components in the final mixture is determined by the splitting procedure and in general may not be known in advance. This may lead to computational difficulties: many splits can result in an unbounded increase in the number of components. Each component requires the computation of a mean and covariance matrix to propagate forward in time, and so it would be ideal to keep the number of components low. One can easily



(a) The mean x and covariance matrix Σ of the component is propagated forward by solving (5.4) to obtain w_t and Π_t respectively.



(b) Once the splitting criterion is met, the mapped component is split into K new components with corresponding means (in blue) satisfying (5.9).



(c) Each new component is propagated forward with (5.4) and the process is repeated for each component *individually* until the final time is reached.

Figure 5.4: The propagation and splitting of a component in the Gaussian mixture model

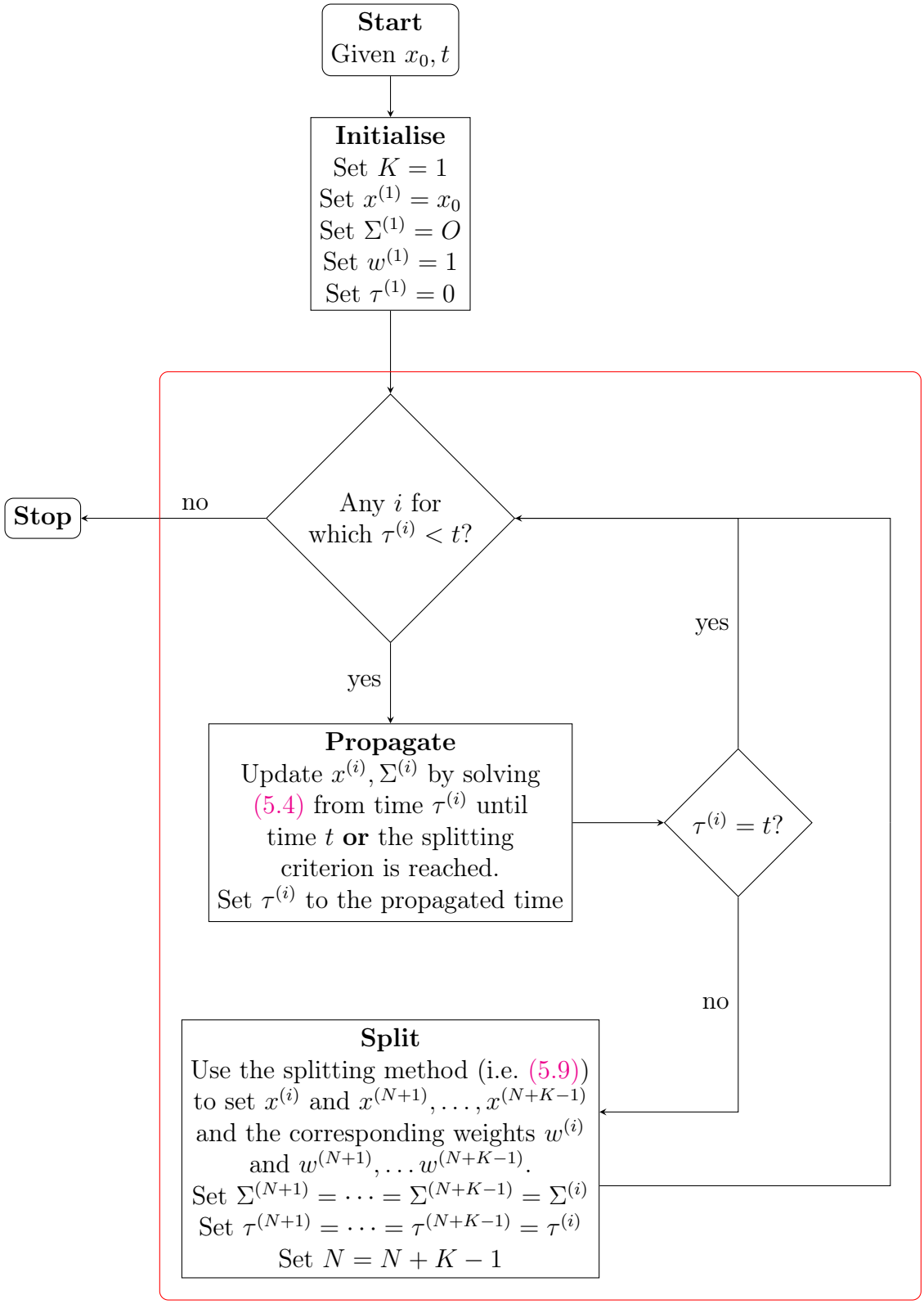


Figure 5.5: An algorithm flow chart of one particular implementation of the Gaussian mixture model with a certain initial condition. The main propagation-splitting loop is highlighted by the outlined area.

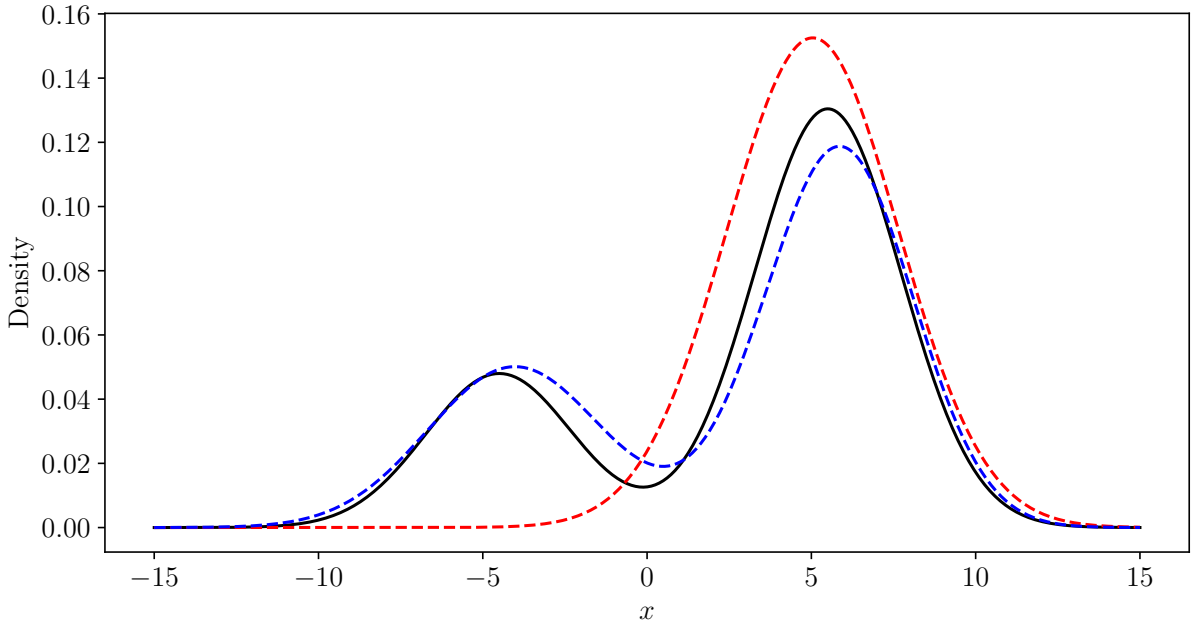


Figure 5.6: The mixture model algorithm implemented on the Beneš SDE (5.5) with fixed initial condition $x_0 = 0.5$ and over the time interval $(0, 5)$. The probability density function of the true solution is shown in solid black and the corresponding linearisation solution in dashed red (as in Figure 5.1). The Gaussian mixture model density is overlaid in blue, constructed using a single split into three canonical sigma points (using (5.10)) at time $\tau = 0.8$).

enforce an upper limit on the number of components, by preventing any further splitting once this threshold is reached.

As a simple example, we apply the algorithm to the Beneš SDE (5.5), subject again to the initial condition $x_0 = 1/2$ and considered at the final time of $t = 5$. Recall from (5.8) and Figure 5.1 that the true solution has a bimodal density, which the Gaussian distribution resulting from a single linearisation is unable to capture. We implement the Gaussian mixture model with a single split, manually chosen to be at time $t = 0.8$, with the resulting mixture density shown in Figure 5.6. The algorithm is initialised with $x^{(1)} = 1/2$, $\Sigma^{(1)} = 0$ and $w^{(1)}$. After propagating the initial condition by solving (4.1) to the splitting time $t = 0.8$, we use the canonical sigma points ((5.10)) to replace the Gaussian density with three points $x^{(1)}, x^{(2)}, x^{(3)}$, preserving the mean and covariance. We then propagate these three points forward separately, in that for each one we take the deterministic trajectory $F^5_{0.8}\left(x^{(i)}\right)$ and compute the solution to the SDE (5.5) linearised about this trajectory. The result is three Gaussian distributions at $t = 5$, which we then combine as an equally weighted sum to produce a Gaussian mixture model approximating the

SDE solution. In Figure 5.6, we compare the probability density function of the resulting Gaussian mixture model to the true solution density and the Gaussian approximation from a single linearisation. Importantly, the mixture model density includes two modes that resemble those of the true solution, which is an important feature of the solution that was not captured by the single linearisation. Further configuration of the algorithm may result in a better fit; this was a simple and contrived example implemented to demonstrate the *potential* of the mixture model algorithm in overcoming the limitations of a single Gaussian approximation. Computing the mixture model only required the propagation of 6 values—the three means and covariance values—by solving the pair (5.4) of differential equations three times. We can capture the bimodality, an important feature of the solution, with only a small number of calculations. This would be particularly advantageous in a situation where the true solution is not analytically available (as in a majority of practical scenarios) and we would have to otherwise rely on bulk simulation to observe these features.

5.2 The splitting criterion

In the implementation of our algorithm on the Beneš SDE, we manually chose a single split time. However, in general, the algorithm requires a choice of criterion for when to split a Gaussian component, for which there are several choices. A given Gaussian component should be split when the linearised SDE is no longer a reasonable approximation for the original SDE about the deterministic trajectory corresponding to the component. This requires an ongoing evaluation of the error in using the approximation. Some possibilities include:

- Alongside the propagation of the mixture model, one can also compute a small number of stochastic samples that solve the original SDE (5.1). These stochastic samples can be compared to the ongoing Gaussian evolutions, for instance using a probabilistic distance measure. However, this would increase the computational load of the algorithm and may require many samples to accurately quantify departures from Gaussianity. The Gaussian mixture model does provide an analytic probability density function that can lend itself to further inference, as opposed to solely stochastic samples that require an additional step to compute a density function. There may be a trade-off between the number of samples and the desired computational efficiency.

- A diagnostic based purely on the deterministic model will be computationally efficient, not requiring any stochastic simulation. Stochastic sensitivity (introduced by Balasuriya (2020b) and extended in Section 3.2) is a scalar value that is computed directly from the covariance matrix of the Gaussian approximation and quantifies the magnitude and direction of maximum uncertainty. In highly nonlinear systems, stochastic sensitivity may provide a measure for evaluating non-Gaussianity. This can

be computed from each Gaussian component with minimal additional computational cost, by simply taking the operator norm of the component covariance matrix. However, in regions of linearity and non-multiplicative noise, the solution to the SDE can be Gaussian, but stochastic sensitivity, being the magnitude of the noise, can increase. A large value of stochastic sensitivity needs to imply non-Gaussianity in these cases.

- DeMars et al. (2013) propose an algorithm for propagating an initial uncertainty through a nonlinear time-varying mapping, by using a Gaussian mixture model and a similar splitting algorithm. The principle is similar to ours; the uncertainty is propagated forward using a linearisation of the model until this is no longer a reasonable approximation. A split occurs when nonlinearity in the mapping, which would result in non-Gaussianity, is detected via an entropy-based measure. This measure uses the sigma point method to evaluate the mapping of a covariance matrix, which is compared to the ongoing propagation of a Gaussian component. The sigma point method provides an *exact* computation for the covariance matrix of a deterministic nonlinear transformation of a random variable, provided that the transformation can be evaluated exactly. In our case, however, there is ongoing uncertainty from the multiplicative noise and the SDE cannot be solved exactly, so we cannot evaluate the mapping required for the sigma point method. Nonetheless, this previous work may suggest a direction for further tuning the implementation of our method, including the selection of points when splitting a component.

We leave further development of this step of the algorithm for future work.

Chapter 6

An application: drifter in the Gulf Stream

We now apply the theoretical developments of Chapter 3 and the mixture model algorithm of Chapter 5 to a data-driven model. Using altimetry-derived velocity data, we model the motion of a drifter on the surface of the Gulf Stream, a climatically important part of the North Atlantic Ocean. We construct a pair of 2-dimensional models: an ordinary differential equation treating the measurements as known exactly and an 'improved' stochastic differential equation that accounts for measurement error and unresolved effects. Our tools then provide both qualitative and quantitative insight into the behaviour of the stochastic model and, consequently, the impact of uncertainty on the predictions of the deterministic one.

6.1 The Hellinger distance

Before moving on to the example, we first provide a measure that quantitatively evaluates the performance of the Gaussian approximation and our mixture model as estimators of SDE solutions. We will use the Hellinger distance, which measures the distance between probability distributions. For two continuous probability distributions in $\mathbb{R}^n$ with respective probability density functions p and q, the Hellinger distance between the two is (Le Cam and Lo Yang, 2000)

$$D_H(p,q) = \sqrt{\frac{1}{2}\int_{\mathbb{R}^n}\left(\sqrt{p(x)}-\sqrt{q(x)}\right)^2 \mathrm{d}x} = \sqrt{1-\int_{\mathbb{R}^n}\sqrt{p(x)q(x)}\,\mathrm{d}x}. \tag{6.1}$$

We previously discussed (in Section 3.1.1) the Kullback-Leibler divergence—Sanz-Alonso and Stuart (2017) used this measure to quantify the performance of Gaussian approximations to SDE solutions—but opt now to use the Hellinger distance. There are attractive

properties of the Hellinger distance: the Hellinger distance is a metric on the space of probability measures and always takes values in the interval $[0, 1]$, unlike the KL-divergence, which is not a formal metric and can become infinite.

We cannot solve the stochastic model analytically, so we will use stochastic samples as our ground 'truth', obtained by numerically solving the SDE. That is, without access to the true solutions to our model, we instead take many stochastic samples and compare our methods to those. This fits in with the philosophy of our approach; stochastic sampling is the standard in practical settings, and we seek approximations that give the same conclusions without the computational expense.

However, using stochastic samples presents a complication when using the Hellinger distance: the calculation of (6.1) requires evaluating the probability density function from which these realisations were sampled. This would require a choice for how to approximate the density function and then compute the integral in (6.1). One option is to construct a kernel density estimator, where a certain known distribution (usually a Gaussian with a small variance) is placed at each realisation and then combined into a mixture density (Silverman, 2017). To avoid the need to make such a choice here, we will instead use an *empirical* estimator that approximates (6.1) only using samples and without having to evaluate either probability density function directly. Although we will compare the numerical realisations to analytically available probability density functions (from either a single Gaussian approximation of the mixture model algorithm), we will generate samples from both distributions to use a purely empirical estimator. We will employ the empirical estimator recently proposed by Ding and Mullhaupt (2023), which extends a similar estimator for the KL-divergence by Perez-Cruz (2008). Let $\{\hat{x}_1, \dots, \hat{x}_N\}$ and $\{\hat{y}_1, \dots, \hat{y}_N\}$ denote the two sets of N realisations sampled from the probability density functions p and q respectively. The estimator is first computed as

$$\hat{H}_a^2(p, q) = 1 - \frac{\sqrt{N-1}\,\Gamma(k)^2}{N^{3/2}\,\Gamma\left(k - \frac{1}{2}\right)\Gamma\left(k + \frac{1}{2}\right)} \sum_{i=1}^{N} \frac{r_k(x_i)^{n/2}}{s_k(\hat{x}_i)^{n/2}}, \tag{6.2}$$

where Γ denotes the Gamma function, and $r_k(\hat{x}_i)$ and $s_k(\hat{x}_i)$ denote the Euclidean distance to the kth nearest neighbour of $\hat{x}_i$ in $\{\hat{x}_1, \dots, \hat{x}_N\} \setminus \{\hat{x}_i\}$ and $\{\hat{y}_1, \dots, \hat{y}_N\}$ respectively. The pre-specified parameter k is the number of neighbouring points used to construct a k-nearest-neighbour density estimate as an intermediate step in the calculation. We set $k = 5$ throughout to ensure a reasonable computation time. This provides a consistent (in the sense of almost sure convergence as $N \to \infty$) statistical estimator of the squared Hellinger distance (Ding and Mullhaupt, 2023). However, (6.2) is not symmetric in p and q, i.e. $\hat{H}_a^2(p, q) \neq \hat{H}_a^2(q, p)$ with the asymmetry arising in the distances r_k and s_k. In practice, one direction can provide a more accurate estimator depending on the underlying distributions. To ensure symmetry in the final estimate and find a 'middle ground' between

the two asymmetric estimators, Ding and Mullhaupt (2023) propose taking the average:

$$\hat{H}(p,q) = \sqrt{\frac{\hat{H}_a^2(p,q) + \hat{H}_a^2(q,p)}{2}}, \tag{6.3}$$

which still exhibits the same convergence properties as $\hat{H}_a$. Note that this estimator is random, as it relies upon samples from both densities. We therefore take multiple realisations of $\hat{H}$, each using a new set of samples, and use the average to reduce the variance in the estimator.

6.2 Model setup

The Gulf Stream is a warm water current that originates in the Gulf of Mexico, travels through the North Atlantic Ocean, and plays a vital role in the climate patterns of the Northern and Western hemispheres (Palter, 2015). The stream itself consists of a rapidly moving jet which varies dramatically with time, and small eddies of warm and cold water that are formed and shed from the stream (Kang and Curchitser, 2013). The Gulf Stream is a well-studied region of the ocean, due to this climatic importance and the interesting dynamical behaviour exhibited by the flow.

Consider tracking the longitudinal and latitudinal position of a drifter moving on the surface of the ocean. We first construct an ordinary differential equation for the time evolution of the drifter position from geostrophic velocity data inferred from altimetry (sea surface height) observations by satellite. The dataset is supplied by the E.U. Copernicus Marine Service (CMEMS) (2020) and has been processed by the Data Unification and Altimeter Combination System (DUACS). The sea surface height is proportional to the streamfunction of the surface flow, provided that the flow is treated as 2-dimensional, with a constant of proportionality that varies with latitude (Park, 2004; Doglioni et al., 2021). This enables approximation of the zonal (eastwards) and meridional (northwards) geostrophic velocities, which are originally given in metres per second. These measurements are taken hourly and available on a $0.25° \times 0.25°$ spatial grid. We convert the measured velocities to degrees per day via the following transformation: at longitude λ and latitude ϕ (both in degrees), the components u_m and v_m in metres per second are transformed as (Capderou, 2014)

$$u(\lambda,\phi,t) = \frac{1-\left(2f-f^2\right)^2}{a}\left(1-\left(2f-f^2\right)^2\sin^2(\phi)\right)^{3/2} u_\mathrm{m}(\lambda,\phi,t) \tag{6.4a}$$

$$v(\lambda,\phi,t) = \frac{1}{a\cos(\phi)}\left(1-\left(2f-f^2\right)^2\sin^2(\phi)\right)^{1/2} v_\mathrm{m}(\lambda,\phi,t)\,, \tag{6.4b}$$

where a is the semi-major axis and f is the flattening of an ellipsoid model of the Earth.

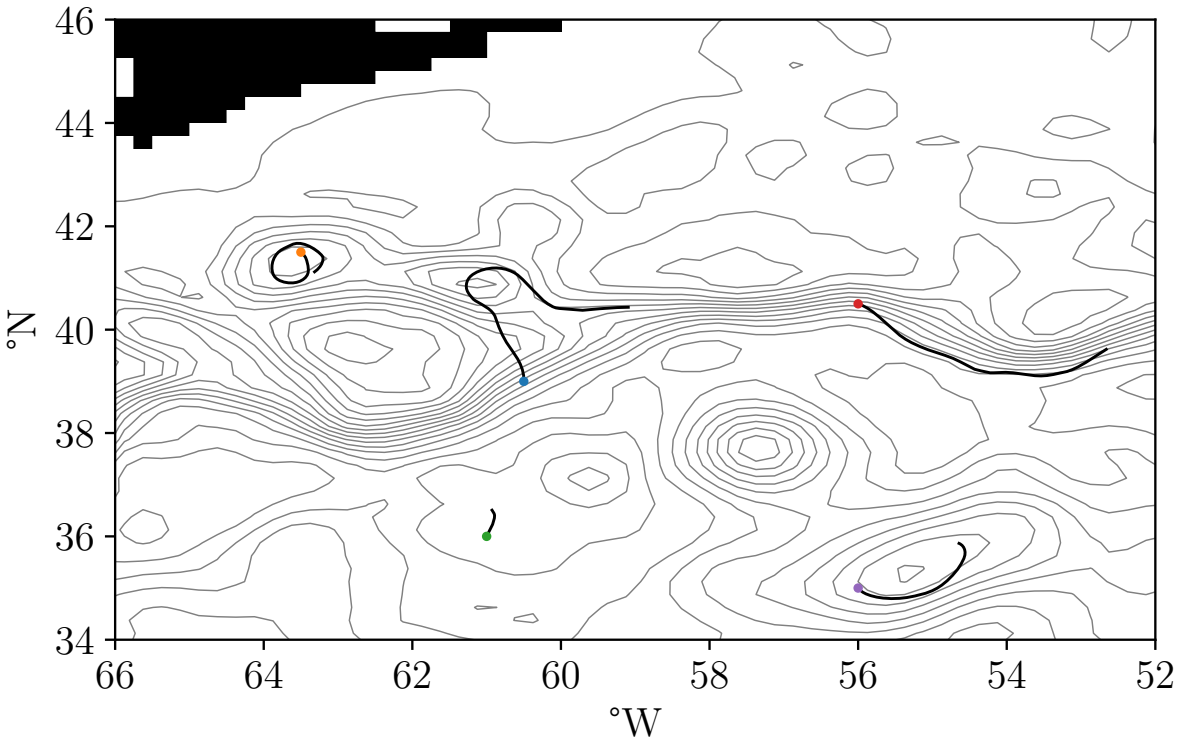


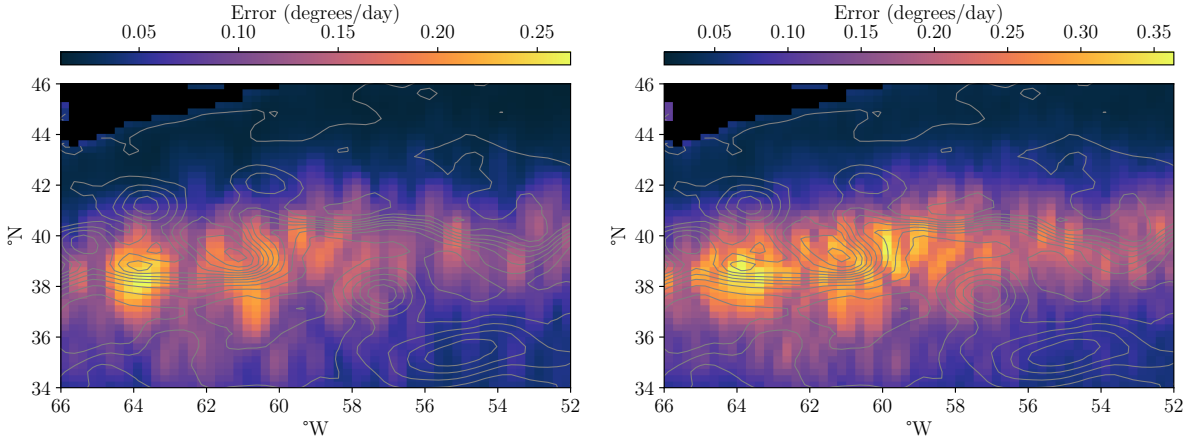
Figure 6.1: Several trajectories (in heavy black) that solve (6.5) and describe the motion of drifters on the surface of the Gulf Stream. Each trajectory is initialised at a coloured point at $t = 0$ (midnight 01/01/2020), and evolved by numerically solving (6.5) up to midnight $t = 7$ (midnight 08/01/2020). Instantaneous streamlines at the final time $t = 7$ are displayed in grey.

We use the World Geodesic System 84, which gives the values (Capderou, 2014)

$$a = 6378137\,\mathrm{m}, \quad f = 1/(298.257223563).$$

Although the dataset provides global coverage, we focus our attention on longitudes between $-66°$E and $-52°$E and latitudes between 34°N and 46°N, and measurements starting from midnight 01/01/2020. We use a larger spatial subset for calculations, however, to prevent issues at the boundaries of the spatial domain. There is a region of land—a small part of the southeast coast of Canada—in the dataset, where velocity data is missing. We set the velocity to zero in this region, as we know from physical considerations that the drifter could not travel on land, and all the following figures indicate this region with black.

The velocity data is Eulerian and unsteady (varying with time), so to predict the position of a drifter we must solve for a Lagrangian trajectory. This requires interpolating



(a) Zonal component

(b) Meridional component

Figure 6.2: The mapping error in the (a) zonal (longitudinal) and (b) meridional (latitudinal) velocity components at midnight 01/01/2020 ($t = 0$).

the velocity data between the spatial and temporal gridpoints, for which we use a linear interpolate. Let

$$x_t = \left(x_t^{(\text{lon})}, x_t^{(\text{lat})}\right)^\top$$

denote the longitudinal (in degrees east) and latitudinal (in degrees north) position of the drifter t days after midnight 01/01/2020. The deterministic model for x_t, constructed purely from the interpolated velocity data, is

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x_t^{(\text{lon})} \\ x_t^{(\text{lat})} \end{bmatrix} = \begin{bmatrix} \tilde{u}\left(x_t^{(\text{lon})}, x_t^{(\text{lat})}, t\right) \\ \tilde{v}\left(x_t^{(\text{lon})}, x_t^{(\text{lat})}, t\right) \end{bmatrix}, \tag{6.5}$$

where $\tilde{u}$ and $\tilde{v}$ are the interpolated zonal and meridional velocities respectively. Figure 6.1 shows some example trajectories that solve (6.5), evolving over the span of a week. Contours of the sea surface height at $t = 7$ are also shown in grey, which corresponds to those of the instantaneous streamfunction and provides an indication of the flow structure. The stream itself is fast-moving and a prominent feature of the flow, and drifters starting within the stream (i.e. the blue and red points) are accordingly transported along it. Along the stream, eddies form and break off (roughly indicated by the circular regions in the sea-surface height contour map), and trajectories that start within an eddy (i.e. orange and purple points) typically remain there. Away from the stream and the eddies, the flow is relatively isotropic and slow-moving, resulting in the short and unremarkable trajectory starting from the green point.

We may think of (6.5) as our "best available" deterministic model for the time evolution of the drifter position; if the data (and the subsequent interpolation) were correct, then (6.5) would provide accurate predictions. However, this is certainly not the case and there are many sources of uncertainty not accounted for by (6.5), including measurement error and the unresolved flow behaviour between gridpoints. The dataset additionally includes estimates for the mapping error (an estimate of the variance) in the zonal and meridional velocity measurements, which indicates measurement error. The mapping error in each velocity component at $t = 0$ is shown in Figure 6.2, as an example of the structure of the field, along with contours of the sea surface height. The error varies with time but retains the main structure where it is largest along the Gulf Stream and features a stratified appearance because satellite measurements are more accurate directly under the tracks followed by the instruments, which is not unexpected as this region contains the most complicated part of the flow. We can capture the spatial and temporal variation in this mapping error by adding multiplicative noise to the deterministic model (6.5). We use the following stochastic model:

$$\begin{aligned} \mathrm{d}\begin{bmatrix} x_t^{(\mathrm{lon})} \\ x_t^{(\mathrm{lat})} \end{bmatrix} &= \begin{bmatrix} \tilde{u}\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right) \\ \tilde{v}\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right) \end{bmatrix} \mathrm{d}t \\ &\quad + \sqrt{L_r} \begin{bmatrix} \sqrt{\tilde{u}_{\mathrm{err}}\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right)} & 0 \\ 0 & \sqrt{\tilde{v}_{\mathrm{err}}\left(x_t^{(\mathrm{lon})}, x_t^{(\mathrm{lat})}, t\right)} \end{bmatrix} \mathrm{d}W_t, \end{aligned} \tag{6.6}$$

where $\tilde{u}_{\mathrm{err}}$ and $\tilde{v}_{\mathrm{err}}$ are the linearly interpolated error estimates for the zonal and meridional velocities (converted to degrees per day using the same transformation (6.4) as was done on the velocity data), and $L_r = 0.25°$ is the spatial resolution of the data. Our choice for the diffusion matrix σ comes from the intuition that $L_r \sigma \sigma^\top$ represents the instantaneous variance of the noise in our stochastic differential equation and to ensure that the units in (6.6) are consistent with the convention that $\mathrm{d}W_t \sim \sqrt{\mathrm{days}}$ (Øksendal, 2003). Both coefficients depend on time and the noise is multiplicative, making (6.6) the most general type of SDE that our theory is equipped to handle. In the framework of Chapter 3, the noise scale is $\varepsilon = \sqrt{L_r} = \sqrt{0.25°}$ and we treat this value as being fixed by the data and model. The introduction of noise in (6.6) can improve over (6.5) by accounting for both the data-informed measurement error and unresolved behaviours between gridpoints, in the fashion of stochastic parameterisation. This is, however, only one choice for such a model; the following analysis could be applied to a better-informed stochastic model.

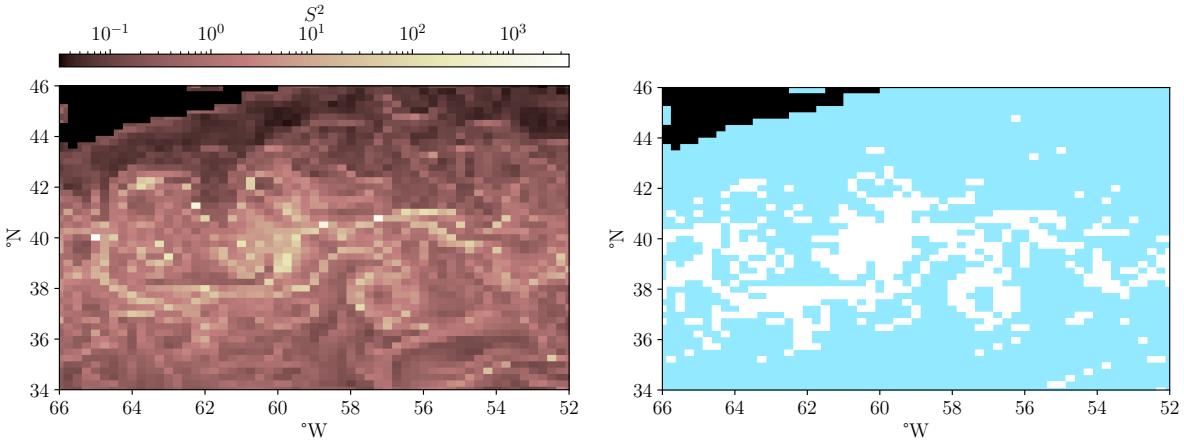
6.3 Stochastic sensitivity

We first compute the stochastic sensitivity field to evaluate the impact of uncertainty (formulated by the SDE (6.6)) on the predictions of the deterministic model (6.5). As in Section 4.3, this example demonstrates the computation of stochastic sensitivity from the covariance matrix, rather than provide any new insight. Badza et al. (2023) provide a similar example of computing the stochastic sensitivity field for a similar sea-surface dataset of the Gulf Stream, albeit using the original method of calculation (Balasuriya, 2020b).

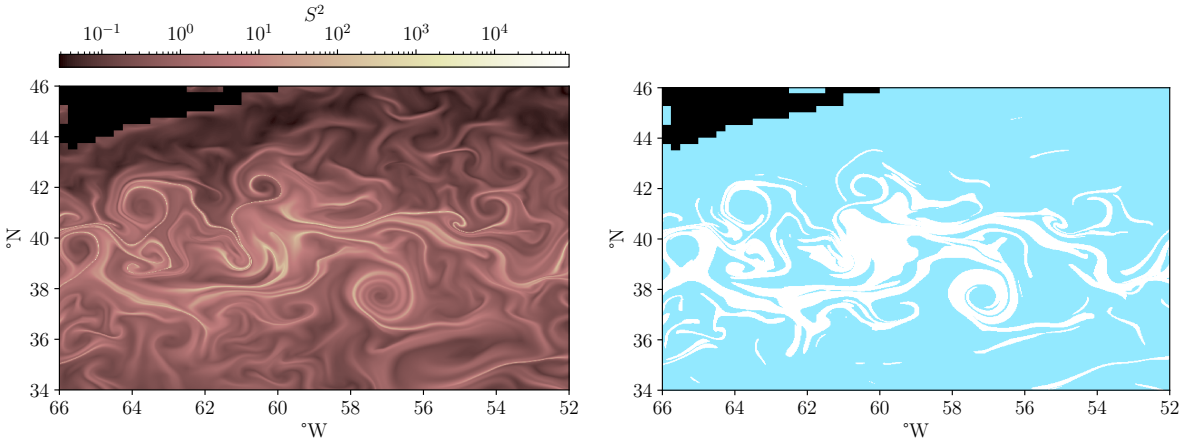
Figure 6.3a plots the stochastic sensitivity field for a grid of initial conditions at the $0.25° \times 0.25°$ resolution of the velocity data, over the timespan of a week, from midnight 01/01/2020 ($t = 0$) to midnight 08/01/2020 ($t = 7$). Each initial condition corresponds to a point at which the surface velocity data was available. Although the resolution of the velocity data is low, we can still distinguish the stream itself as a region of high uncertainty. Away from the stream (either north or south of it), the flow is reasonably isotropic and the level of noise from measurement error is small, so there are relatively small S^2 values. The S^2 field also highlights an eddy-like structure, centred around 57°W and 38°N, with a larger uncertainty on the boundary and a smaller within.

Balasuriya (2020b) provided a simple method for extracting possibly coherent sets from the stochastic sensitivity field, by taking the initial conditions for which the S^2 value is under a certain threshold. Using a threshold of $R = 2°$, the right-hand side of Figure 6.3a shows in cyan the initial conditions which correspond to a robust set. We see that the regions away from the stream emerge as robust sets. These sets do not include parts of the stream, because of the larger uncertainty. However, the low resolution of the data means we cannot make any further conclusions, as the finer structure is unresolved.

The resolution of the data is a significant limitation of the deterministic model; since the velocity data has been interpolated between gridpoints, any conclusions relying on the deterministic model alone cannot be trusted at higher resolutions. Although the stochastic model also uses this interpolated data, the model is in some sense accounting for the uncertainty introduced by the interpolation. We are resolving the behaviour of the system between the gridpoints by introducing stochastic noise. Since stochastic sensitivity is a property of the *stochastic* system and not just the deterministic one, we are permitted to investigate and make conclusions from this field at a higher resolution than that prescribed by the data. This is an advantage of stochastic sensitivity over other deterministic measures, such as the finite-time Lyapunov exponent, which do not account for unresolved effects. Figure 6.3b show the stochastic sensitivity field computed on a $0.025° \times 0.025$ grid of initial conditions, at a higher resolution than the velocity data. We see several eddy-like structures highlighted above and below the stream, where uncertainty is large on the boundaries of the eddies and smaller within. These eddies were not all apparent in the lower-resolution S^2 field or in the corresponding robust sets. The complicated structure of the stream is better resolved in the higher-resolution field



(a) At the $0.25° \times 0.25°$ resolution of the velocity data.



(b) At the higher resolution of $0.025° \times 0.025°$.

Figure 6.3: (Left) The stochastic sensitivity fields computed from the linearisation covariance matrix on a grid of initial conditions, at two different resolutions. (Right) From each field, robust sets are extracted with a threshold of $R = 2°$ and shown in cyan.

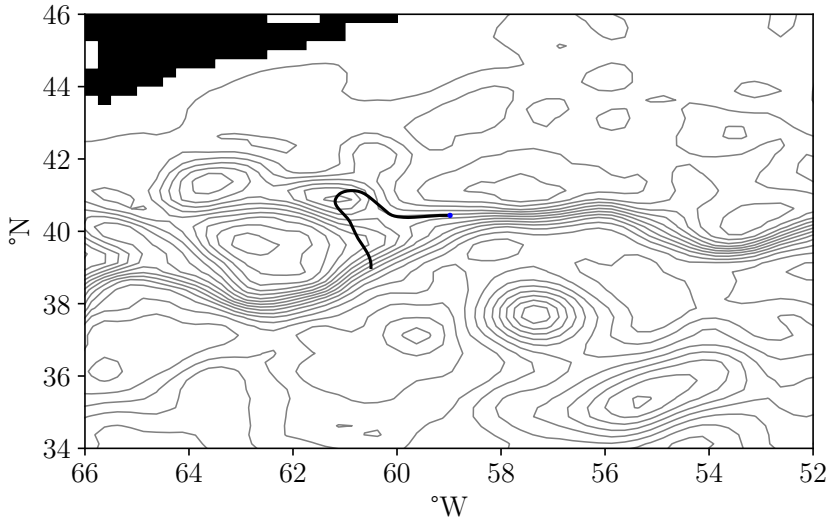
and emerges as regions of non-robustness (that is, not contained in the robust sets). The centres of the eddies, where we expect coherency in the deterministic system, are also included in the robust set. By increasing the resolution, we can therefore identify regions of low uncertainty and distinguish finer structures within the flow.

6.4 Exploring a single trajectory

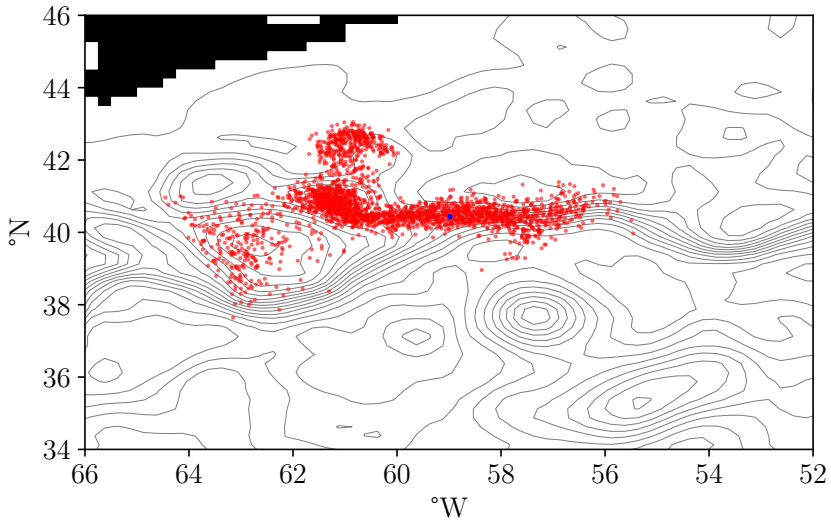
We now focus our attention on a fixed initial condition for the drifter. Suppose that, at midnight 01/01/2021, the drifter is located within the stream at longitude 60.5°W and latitude 39°N. We accordingly consider the evolution of the two models (6.5) and (6.6) with the initial condition $x_0 = (-60.5, 39)^\top$. By solving the deterministic model (6.5) numerically, we predict the position of the drifter after $t = 7$ days (at midnight 08/01/2021) and show the time evolution of this solution trajectory in Figure 6.4a. The jet stream transports the drifter. This is the *only* prediction of the deterministic model (6.5), but when accounting for uncertainty with the stochastic model (6.6), we get a more complicated picture. Figure 6.4b shows the result of 2500 numerical realisations (obtained via Euler-Maruyama integration with a step size of 1/24 days) of the stochastic model (6.6), starting from the fixed initial condition x_0 and evolved over the same week-long timeframe. Each red point corresponds to a single realisation of the stochastic solution at midnight 08/01/2021 ($t = 7$). Although many of the realisations are transported along the stream, many break off in eddies, resulting in several clusters of realisations. Thus, with uncertainty in the model, the drifter may end up following distinctly different qualitative behaviour. This example highlights the importance of stochastic models; the deterministic model provides a single prediction, but the dynamic behaviour of the flow means that any uncertainty can lead to vastly different predictions of the future state of the system.

To further understand the impact of the deterministic flow dynamics on the behaviour of the stochastic solution, Figure 6.5 plots the evolution, as histograms, of now 10000 EM samples up to midnight 08/01/2021 ($t = 7$). The figure also includes contours of the sea surface height at each time to give some indication of the deterministic dynamics. The stream propagates a majority of the stochastic realisations, but we see several trajectories leave the stream early and move westwards, creating an arc-like structure of smaller density in the empirical distribution. A second cluster of realisations departs the stream starting around $t = 4$ and moves northwards, likely due to the formation and break off of an eddy from the stream in the region. The final distribution of realisations is therefore complicated, comprising numerous realisations that remained in the stream, a second cluster that is transported northwards by an eddy, and a small number that are scattered westwards.

The SDE (6.6) is highly nonlinear and data-driven, so analytical solutions are not available. However, we can compute a Gaussian approximation by linearising (6.6) about the corresponding deterministic trajectory solving (6.5) (with the same fixed initial

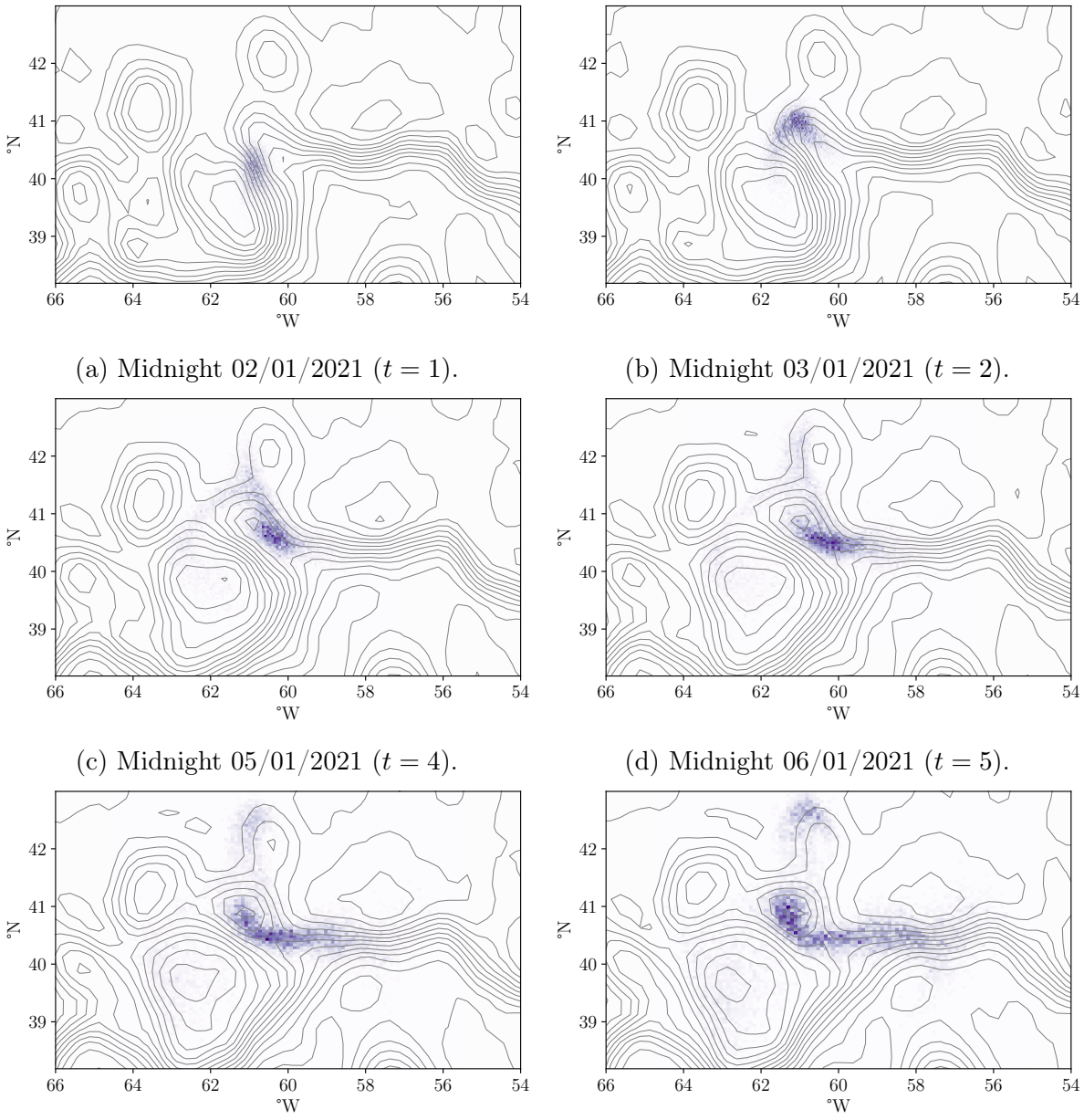


(a) The solution to the deterministic model (6.5).



(b) Realisations of the solution to the stochastic model (6.6).

Figure 6.4: Solutions to the (a) deterministic model (6.5) and (b) stochastic model (6.6) for fixed initial condition $(-60.5, 39)^\top$ from midnight 01/01/2021 ($t = 0$) to midnight 08/01/2021 ($t = 7$), corresponding to a drifter on the surface of the Gulf Stream. The deterministic prediction at $t = 7$ is indicated in blue in both figures, with the time evolution of the deterministic trajectory in black in (a). In (b), each red point corresponds to one of 2500 EM realisations of the stochastic solution at midnight 08/01/2021. Contours of the sea surface height at $t = 7$ are included in grey.



(a) Midnight 02/01/2021 ($t = 1$).

(b) Midnight 03/01/2021 ($t = 2$).

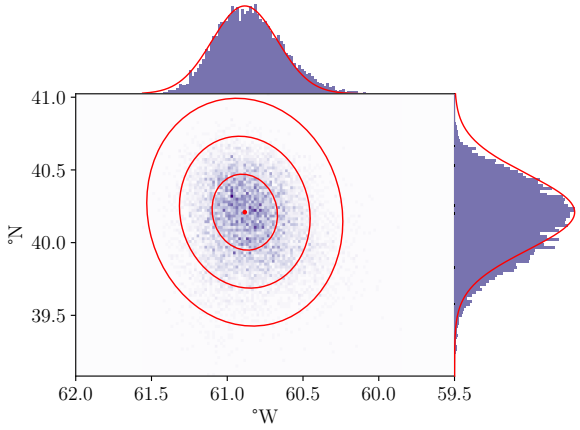
(c) Midnight 05/01/2021 ($t = 4$).

(d) Midnight 06/01/2021 ($t = 5$).

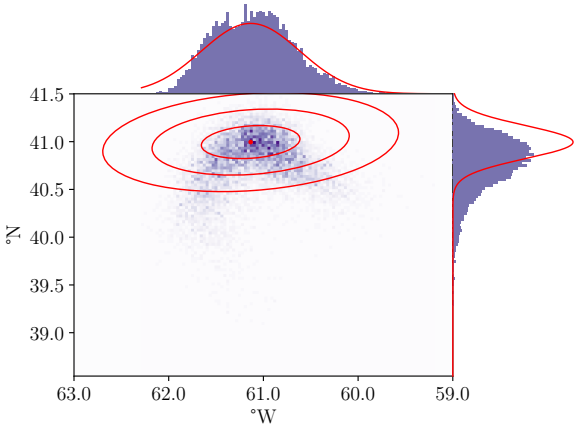
(e) Midnight 07/01/2021 ($t = 6$).

(f) Midnight 08/01/2021 ($t = 7$).

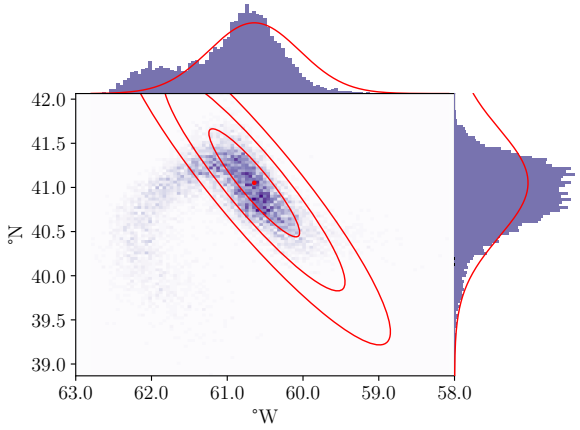
Figure 6.5: Histograms representing the empirical distribution of the solution to (6.6), constructed from 10000 Euler-Maruyama samples, with darker colours indicating higher density. In grey are contours of the sea surface height which correspond to the instantaneous streamlines at time t.



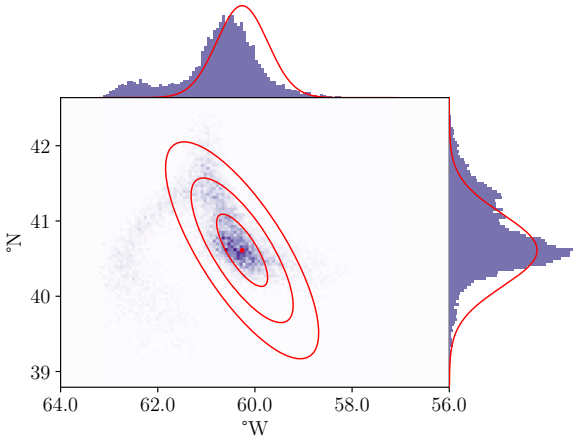
(a) $t = 1$ (midnight 01/01/2021)



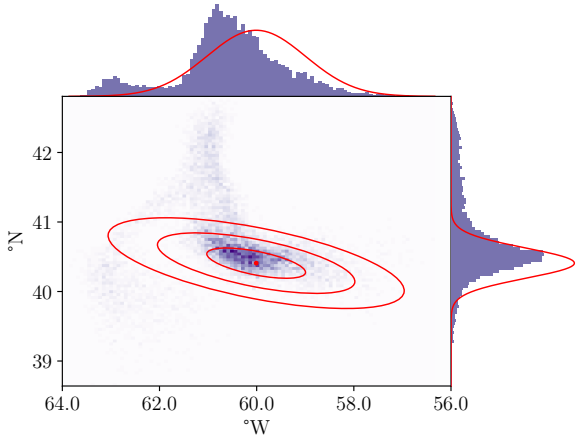
(b) $t = 2$ (midnight 02/01/2021)



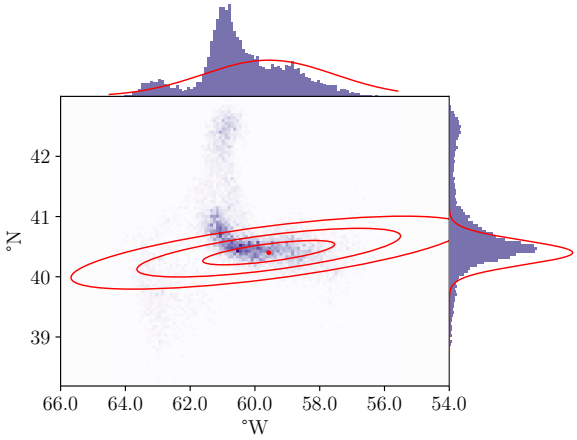
(c) $t = 3$ (midnight 03/01/2021)



(d) $t = 4$ (midnight 04/01/2021)



(e) $t = 5$ (midnight 05/01/2021)



(f) $t = 6$ (midnight 06/01/2021)

Figure 6.6: The time evolution of 10000 Euler-Maruyama realisations of the solution to (6.6), with the Gaussian density arising from a linearisation of (6.6) about the deterministic trajectory overlaid in red. The bivariate plot shows contours of the probability density function (or equivalently standard deviation bounds) of the Gaussian approximation, whereas each marginal plot on the longitudinal and latitudinal axis show the PDFs themselves of the Gaussian marginals.

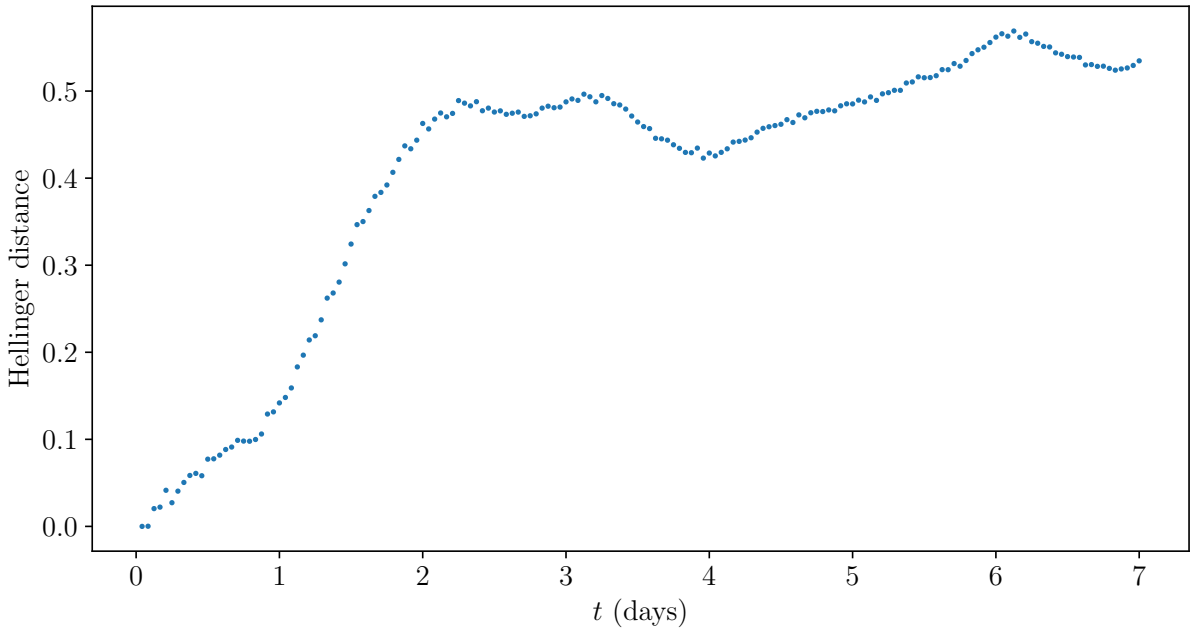


Figure 6.7: The estimated Hellinger distance between 10000 Euler-Maruyama samples of (6.6) and the Gaussian process solution to the linearisation, t days after midnight 01/01/2021.

condition), and employing the theory and computations of Chapter 3. In Figure 6.6, we again plot histograms of the 10000 Euler-Maruyama realisations of the solution, and include marginal distributions in the longitudinal and latitudinal directions on each axis. Overlaid in red is the Gaussian solution to the linearisation of (6.6). The deterministic trajectory and covariance matrix are computed simultaneously using the Mazzoni method outlined in Section 4.1. This computation requires the gradient of the velocity field, which we approximate with a centred finite difference. In Figure 6.6, on each joint histogram, we have plotted contours of the bivariate Gaussian solution. On each marginal histogram, we have plotted the probability density function of the marginal Gaussian corresponding to that component. After a single day ($t = 1$), the numerical realisations match the Gaussian approximation, but the distribution of the realisations quickly becomes non-Gaussian due to both the nonlinearity of the flow and the multiplicative noise of the stochastic model. While the large number of realisations that remain in the stream maintain a Gaussian-like shape with a single model, the scatter of trajectories leaving the stream results in a highly non-Gaussian distribution. To account for this, the variance of the Gaussian approximation grows larger over time, particularly in the longitudinal direction.

To quantitatively evaluate the quality of the Gaussian approximation over time, Figure 6.7 plots the estimated Hellinger distance between the Gaussian approximation and Euler-Maruyama samples solving (6.6). We use 100 realisations of the Hellinger distance—

that is, for each of those 100 calculations, we take 10000 new EM samples and 10000 samples of the Gaussian approximation and compute the empirical Hellinger distance with (6.3)—and plot the average of these in Figure 6.7. The sample variance in this calculation is approximately $\mathcal{O}(10^{-4})$, so we do not include error bars in Figure 6.7 as they would be negligible. As expected, the distance increases over time, as the numerical solution to (6.6) quickly becomes non-Gaussian. However, there are two times where the Gaussian approximation *improves*: at $t \approx 3.3$ and $t \approx 6$. This improvement may be because of contracting dynamics within the Gulf Stream, such as when the trajectories pass through the 'bend' in the stream (evident around $t = 3$ and $t = 4$ in Figure 6.5) which advect some realisations back towards each other.

The Gaussian solution to the linearisation can provide a reasonable representation of the drifter position distribution over short timeframes. This linearisation solution still provides valuable insight into the stochastic system over longer periods of time, as we saw when computing the stochastic sensitivity field. However, the highly non-linear dynamics of the system quickly drive the distribution away from the Gaussian distribution, and the aforementioned limitations of a Gaussian approximation, such as the inability to capture multiple modes and skew, become apparent. The Gaussian approximation is still efficient to compute, however, so we will look to employ the Gaussian mixture model algorithm we described in Chapter 5.

6.5 The mixture model

We will explore a simple implementation of the mixture model algorithm described in Chapter 5 to predict the distribution of stochastic solutions about the single trajectory in Section 6.4. As discussed in Chapter 5, we expect that the mixture model approach will work best in moderate noise models and over reasonably short timeframes. We take the same fixed initial condition $x_0 = (-60.5, 39)^\top$ but consider the position of the drifter at $t = 3$, so at midnight 04/01/2021. The distribution of numerical samples at this time is given in Figure 6.6c; the aim here is to use the mixture model algorithm to approximate this non-Gaussian distribution with a small number of Gaussian components, each constructed as solutions to the linearised SDE. The two marginal distributions (in the longitudinal and latitudinal directions) in Figure 6.6c demonstrate the qualitative departures from Gaussianity mentioned previously: in the longitudinal direction, the distribution of samples is bimodal, whereas in the latitudinal direction we see skew in the southern direction. The distribution is 2-dimensional, however, so there are aspects of the joint density that are not reflected by the two marginal distributions, such as the curved structure of the density function. This is a highly non-Gaussian distribution with distinct features that we would like to capture with a mixture model approximation.

First, we shall implement the GMM algorithm with a single manually specified split, for which we use the canonical sigma points (5.10) to split the covariance matrix at the

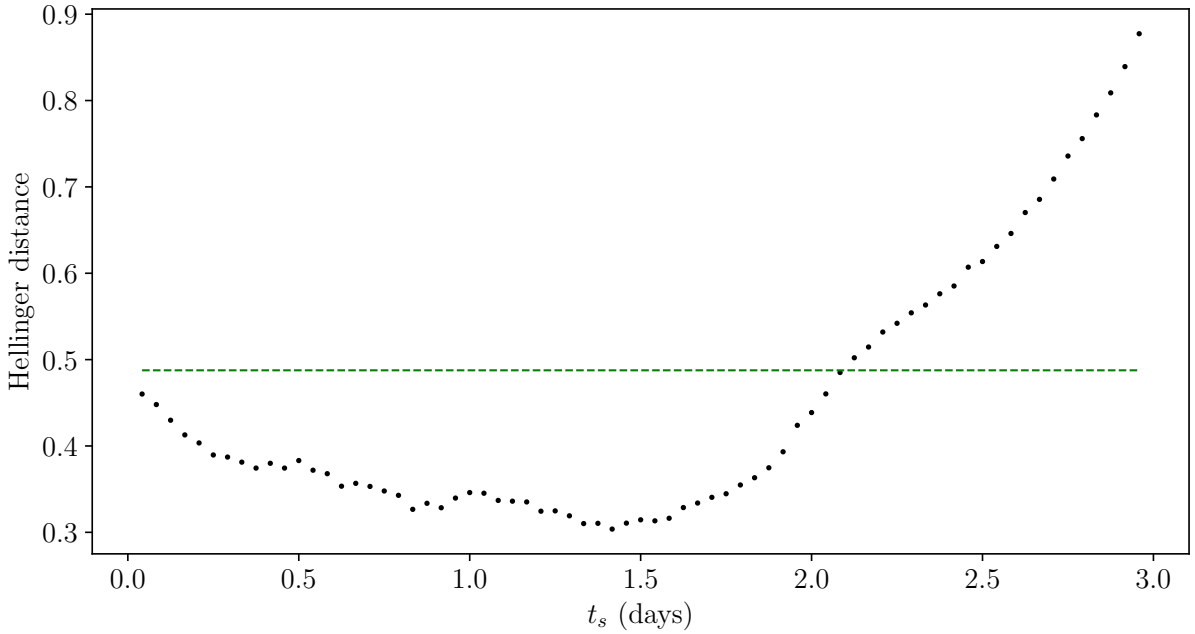


Figure 6.8: The Hellinger distance between the mixture model implemented with a single split at time t_s and 10000 Euler-Maruyama samples of the solution to (6.6) at $t = 3$. The green line indicates the Hellinger distance between the single Gaussian component and the samples.

specified time into 4 additional points. With a single split, the mixture model comprises 5 equally weighted components: the original Gaussian component evolved from the initial condition, and the 4 additional components resulting from the splitting step. For each splitting time, we can use the Hellinger distance between the resulting mixture model and the 10000 Euler-Maruyama realisations of the true solution to evaluate that choice of time. This allows us to 'tune' the choice of splitting time by finding that which results in the smallest distance. This method, of course, relies upon us having already obtained the numerical samples over which we are trying to improve computational efficiency. However, the purpose of this example is to demonstrate that such an optimal time can be found and to suggest that, once equipped with an appropriate online splitting criterion, the mixture model algorithm provides an efficient *ad hoc* method for capturing key features of the distribution. As a benchmark, the single Gaussian component (shown in Figure 6.6c overlaid on the histograms of samples) gave a Hellinger distance of approximately 0.48761. Figure 6.8 plots, against the split time t_s, the Hellinger distance (averaged over 100 calculations) between each mixture model and a set of 10000 EM samples. We can see that a split time earlier than $t_s = 2$ results in a mixture model that improves over the single Gaussian approximation. When the split occurs later, the quality of the resulting

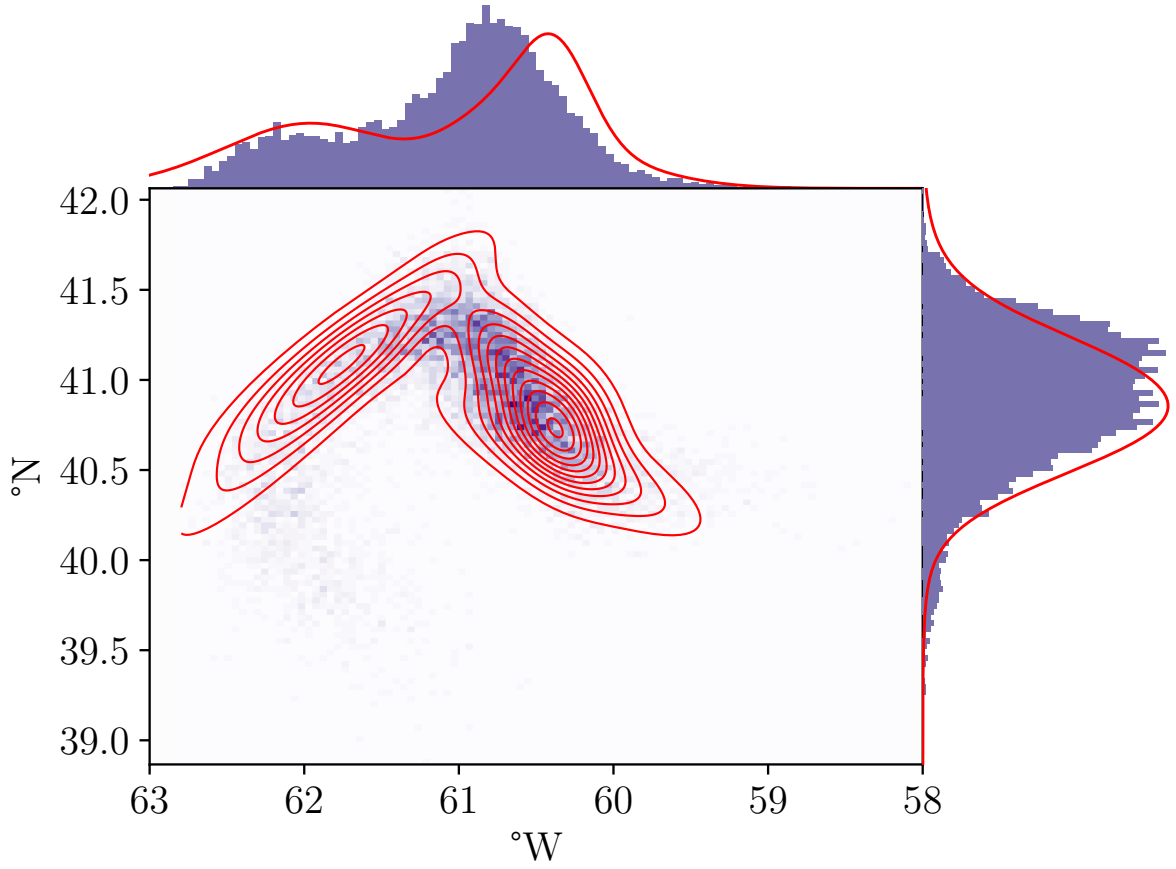


Figure 6.9: The best-fitting mixture model with a single split at $t_s = 33/24$ (9am 02/01/2021). The joint histogram (centre) includes contours of the mixture probability density function. Each axis provides marginal histograms of the numerical samples, with the corresponding marginal PDFs of the mixture density in red.

mixture model worsens. The minimum Hellinger distance of approximately 0.30191 occurs with a split at $t_s = 33/24$ days, or at 9am 02/01/2021, which we take as our 'best' fit.

We compare the resulting mixture model against the numerical samples in Figure 6.9, in the same fashion as Figure 6.6 where contours of the mixture probability density function are shown on the joint histogram and the marginal PDFs on each marginal histogram. The mixture model captures several important qualitative features of the empirical distribution, including the bimodality in the longitudinal marginal and part of the 'curved' shape of the joint distribution. Comparing directly to the single Gaussian in Figure 6.6c, the latitudinal marginal, with a smaller variance, more closely matches the spread of the samples. However, the mixture fails to capture the full scatter of samples in the southern direction, even though this only corresponds to a small proportion of the

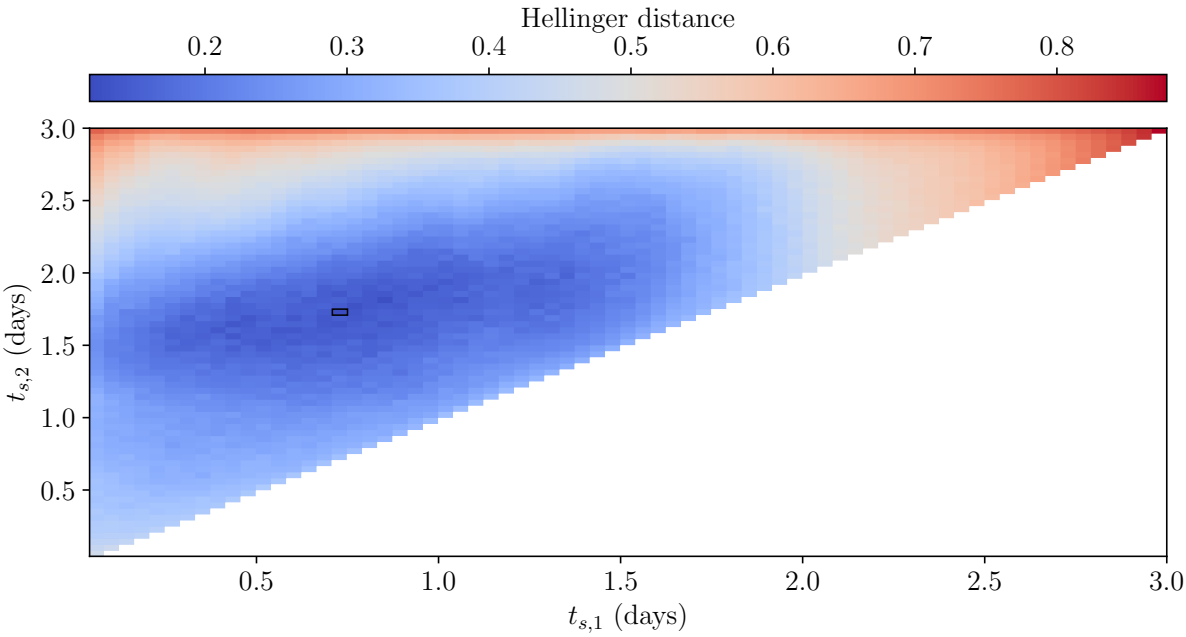


Figure 6.10: The Hellinger distance between the mixture model implemented with two splits—the first at $t_{s,1}$ days, and the second at $t_{s,2} > t_{s,1}$ days—and 10000 Euler-Maruyama samples of the solution to (6.6) at $t = 3$. When $t_{s,1} = t_{s,2}$, the result for a single split at that time (from Figure 6.8) is shown for comparison. The minimising value is indicated by the black box.

samples. Regardless, this result is promising, as with only a single split and four more components, we have improved over the single Gaussian approximation.

Five components may not be sufficient to fully capture the shape and spread of the empirical distribution, so we now consider adding an additional splitting step to each component. After the first split at time $t_{s,1}$, we will split each new trajectory at a later time $t_{s,2}$. For simplicity, each of the five trajectories is split at the *same* time $t_{s,2}$, but the algorithm allows for each trajectory to be split at different times as some regions may exhibit more Gaussianity than others. Each component pair is split at time $t_{s,2}$ into four additional points, again using the canonical sigma points (5.10) to preserve the mean and covariance. We consider a range of times for $t_{s,1}$ and $t_{s,2}$, construct the mixture model using each pair of times, and compute the empirical Hellinger distance (again averaged over 100 calculations) between the mixture density and the EM samples to find the best configuration. The resulting Hellinger distances are shown in Figure 6.10. The optimal configuration with two splits has the first at $t_{s,1} = 3/4$ (at 6pm 01/01/2021) and the second at $t_{s,2} = 43/24$ (at 7pm 02/01/2021), and results in a Hellinger distance of approximately 0.14684. This distance is a substantial improvement over both the single

Gaussian component and the mixture model with a single split.

In Figure 6.11, we compare this mixture density against the stochastic samples. The mixture model captures the shape of the empirical distribution, including the scatter of points leaving the stream in the westward direction. The bimodality in the longitudinal marginal and the skew in the latitudinal are captured by the mixture model. The 25-component mixture model provides a close representation of the empirical distribution, both heuristically (by capturing qualitative departures from Gaussianity) and supported by the small Hellinger distance of 0.14684. Whereas we propagated 10000 Euler-Maruyama samples to construct the empirical distribution, the mixture model comprises only 25 mean-covariance pairs, the computation of which can be thought of as requiring the propagation of only 125 values (the two components of each mean and the three components of each symmetric covariance matrix). A mixture model with fewer components (by having fewer splits in total) may also provide the same quality of approximation, which would be even more computationally efficient. Estimating the true solution distribution is a difficult problem, as the deterministic dynamics are complicated and highly non-linear, and the non-uniformity of error in the observations used to construct it means that the noise is multiplicative. With a simple implementation of the mixture model, we can capture key features and provide a close approximation of the distribution of the solution at $t = 3$. Evaluating the mixture model and finding the best splitting times required a set of numerical realisations of the SDE solution when our overall aim is to avoid this computational cost. However, we continue to emphasise that this example is demonstrative and the mixture model has only been proposed as an outline of an algorithm. Nonetheless, the results of this section are encouraging and motivate further development and investigation of the algorithm. Once equipped with an appropriate splitting criterion, the mixture model algorithm could become a highly effective method for circumventing bulk simulation in stochastic systems. Even without this splitting criterion, the mixture model provides an explicit form for the probability density function, which may lend itself to further applications.

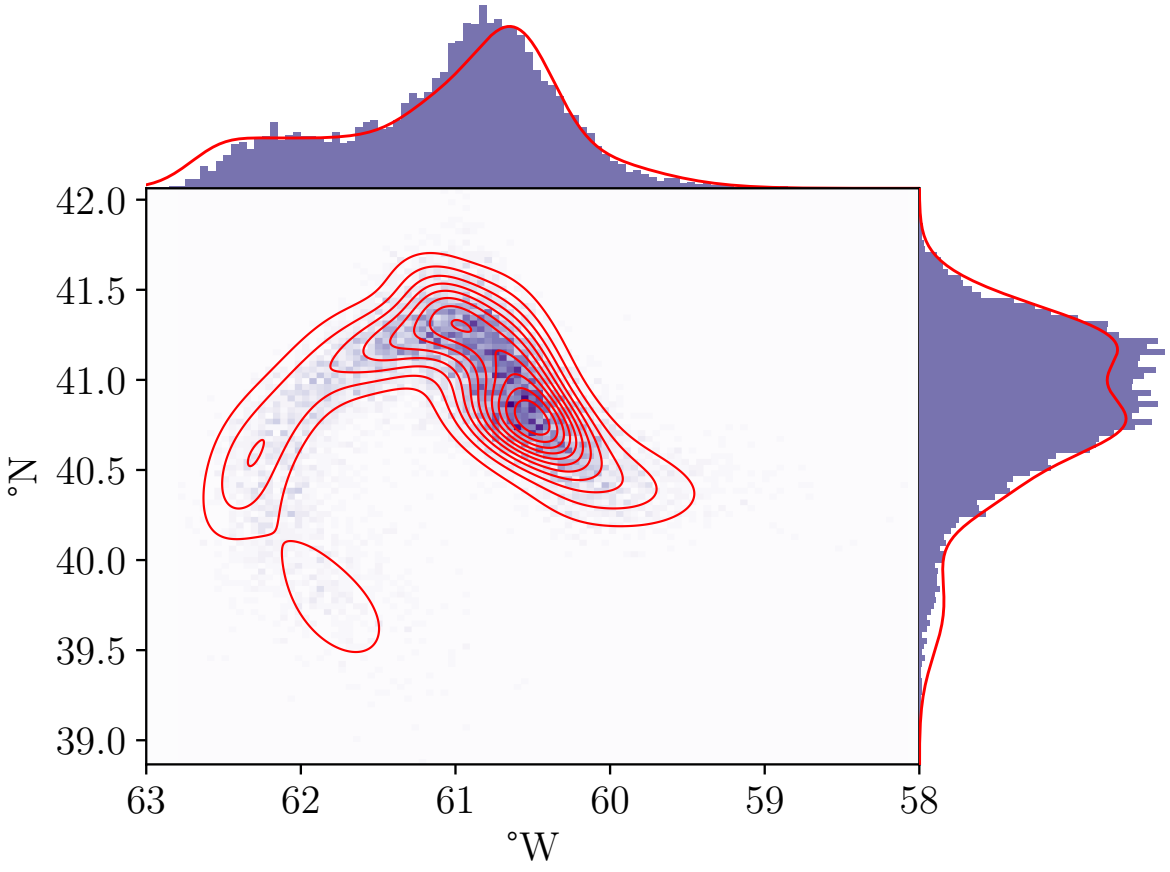


Figure 6.11: The best-fitting mixture model (in red) with two splits on each component, over histograms of 10000 Euler-Maruyama samples. The central joint histogram shows contours of the mixture probability density function. On the longitudinal and latitudinal marginals on each axis, the probability density functions of the corresponding marginals of the mixture model are shown.

Chapter 7

Discussion and future outlook

To conclude this thesis, we shall summarise the results we have presented and raise several discussion points on the implications and extensions of our work. This is by no means an exhaustive list; since we have provided results for a general class of stochastic differential equations, our results have the potential to be applied across a wide range of fields, well beyond the scope of this thesis. Broadly speaking, our theoretical work (namely the error bound in Theorem 3.1) may be useful to investigate the validity of SDE linearisations in application domains such as Stochastic Parameterisation (Berner et al., 2017; Leutbecher et al., 2017; Palmer, 2019) or Data Assimilation (Law et al., 2015; Reich and Cotter, 2015; Budhiraja et al., 2019), or to improve such linearisations by monitoring the skew or kurtosis.

The later sections of this chapter cover:

- In Section 7.1, we discuss approaches to selecting the diffusivity matrix σ.
- In Section 7.2, we discuss the difficulties that boundary conditions can present to both the formulation of an SDE and our methods.
- In Section 7.3, we introduce the Fokker-Planck equation, which provides an alternative view of the solution of a stochastic differential equation, and discuss how this perspective can further our theoretical understanding.
- In Section 7.4, we discuss the possibility of replacing the Wiener process in the driving stochastic differential equation with a more general Lévy process.
- In Section 7.5, we discuss the extension of our theory in Chapter 3 to include higher-order terms in the small-noise expansion of the SDE.
- In Section 7.6, we discuss the implications of our extension to stochastic sensitivity and the applications thereof.

- In Section 7.7, we draw analogies between our theoretical work and results for the limits of population processes, a class of stochastic models evolving on a discrete state space. This suggests that our tools can be applied to a broader class of stochastic models. The section culminates in a 5-dimensional implementation of the Gaussian computation outlined in Chapter 4 on a data-informed model for the spread of Ebola.

We began the main contribution of this thesis in Chapter 3, where we provided and justified a framework for computing linearisations of stochastic differential equations. We provided in Theorem 3.1 an explicit bound on the error between the solution of a small-noise nonlinear stochastic differential equation and an easily computable linearisation approximation, building upon previous studies (Blagoveshchenskii, 1962; Freidlin and Wentzell, 1998; Sanz-Alonso and Stuart, 2017). The linearisation approximation is used across many applications and contexts (Archambeau et al., 2007; Jazwinski, 2014; Sanz-Alonso and Stuart, 2017; Kaszás and Haller, 2020, e.g.), but often without a clear mathematical justification. The theory applies to fully non-autonomous SDEs with multiplicative noise and a random initial condition. Our bound is written in terms of a scaling of the diffusivity matrix and a measure of the uncertainty in the initial condition using the L_r-norm. A comparison in Section 3.1.1 suggests that our bound on the moments is tighter than implied by the gold standard in the literature (a comparable bound on the Kullback-Leibler divergence by Sanz-Alonso and Stuart (2017)). We also provided in Theorem 3.2 and Corollary 3.1 an explicit characterisation of the distribution of the solution to the linearised SDE, enabling efficient approximation of the original nonlinear SDE using solutions to the corresponding deterministic equation, whereas previously special cases of these computations were dispersed across other literature (Jazwinski, 2014; Sanz-Alonso and Stuart, 2017; Särkkä and Solin, 2019, e.g.). In Section 3.1.3 and Section 3.1.2 we highlighted two application-relevant special cases: when the initial condition is fixed and when it is Gaussian, respectively. Coupled with the efficient Mazzoni (2008) method (summarised in Section 4.1) to compute the moments of the linearisation solution, we have provided a rigorous and practical framework for approximating small-noise SDEs with linearisations about the solutions of their corresponding deterministic systems.

Our next contribution was to extend the stochastic sensitivity tools introduced by Balasuriya (2020b). Stochastic sensitivity was hitherto derived as the variance of an unknown limiting distribution and could only be computed in two spatial dimensions. We provided a new definition of stochastic sensitivity that extends the original to any number of dimensions and is computable as the operator norm of the covariance matrix of the linearised SDE (for which we outlined the computation in Theorem 3.2 and Corollary 3.1). We have also established that the limiting distribution in question is Gaussian, as the computation is related directly to a linearised SDE with a fixed initial condition. This 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" deterministic model.

With three example SDEs in 1- and 2-dimensions in Chapter 4, we validated the form of our theoretical error bound and showed heuristically that the solutions approach that of their respective linearisations in the limit of small noise. In particular, we found that the strong error scales with the initial uncertainty and ongoing uncertainty exactly as predicted. We also demonstrated the new computation of stochastic sensitivity in two dimensions (in Section 4.3.1) and, for the first time, in 3-dimensions (in Section 4.3.2).

Although linearisation approximations of SDEs have proven to be useful in both applications and as a theoretical tool (e.g. stochastic sensitivity), in Chapter 5 we highlighted that the limitations of using such approximations, particularly when the linearisation solution is a Gaussian process. A Gaussian approximation cannot capture features such as skew and multimodality, despite these distinctly non-Gaussian attributes arising both in the solutions of nonlinear and multiplicative noise SDEs and observed statistics (del-Castillo-Negrete, 1998; Bracco et al., 2000; Sura et al., 2005). To overcome these limitations while still taking advantage of the computationally efficient linearisation approximation, we proposed an *ad hoc* algorithm that uses a splitting process to approximate SDE solutions with a Gaussian mixture model. Each component of the mixture model arises from a solution of the SDE linearised about a *different* deterministic trajectory, which can capture uncertainty in different regions of the state space. With a simple propagation-splitting procedure, the algorithm can capture multimodality, skewness, and other departures from Gaussianity to improve upon a single Gaussian component. Another advantage of this algorithm is that it provides an analytically available probability density function, as opposed to stochastic samples, which require an additional step to estimate the density function. Our aim throughout was to outline the algorithm and demonstrate its potential on several simple implementations, rather than explore the specifics of each step or perform a mathematical analysis. Instead, we briefly discussed some suggestions for these specifics throughout Chapter 5 while leaving a thorough investigation further as future work.

In Chapter 6, we applied the developments of the previous chapters to a data-driven model. Using satellite-measured velocity data, we constructed two 2-dimensional models for the motion of a drifter on the surface of the Gulf Stream in the North Atlantic Ocean. The first was a purely deterministic ordinary differential equation, using only the interpolated velocity data. The second was a stochastic differential equation that 'improved' upon the ODE by accounting for measurement errors in the velocity data and unresolved effects due to the spatial resolution of the data. By computing stochastic sensitivity, we used the stochastic model to diagnose the reliability of the deterministic model over the spatial domain of the data. We then focussed our attention on a single solution trajectory, corresponding to the motion of the drifter starting from a fixed initial condition. The stochastic model produced a highly non-Gaussian distribution (estimated from Euler-Maruyama samples) for the position of the drifter at a later time. We compared to this empirical distribution the Gaussian approximation arising from a linearisation of the stochastic model about the deterministic trajectory but found that the Gaussian

density was limited in representing the key features of this distribution. Motivated by this, we explored two simple implementations of the mixture model algorithm, using the Hellinger distance to 'tune' the splitting times with reference to the stochastic samples. The mixture model showed highly promising results on this example trajectory; with only 2 splits and 25 components, the algorithm could closely recreate the highly non-Gaussian distribution (in Figure 6.11).

7.1 Selecting the diffusivity matrix

A powerful advantage of our framework is that the diffusivity matrix σ may vary spatiotemporally, allowing for multiplicative noise. Multiplicative noise is often ignored in practice, because of difficulties in analytic work (Sancho et al., 1982) and generating numerical realisations efficiently (Mora et al., 2017). Despite these challenges, multiplicative noise is often necessary in practice (Sura, 2003; Kamenkovich et al., 2015, e.g.). Such noise can also arise from experimental and observational considerations that are otherwise ignored in the deterministic model, such as cloud cover when using satellite measurements or nonuniformity across the field of view of a camera. This raises the question of exactly *how* to specify σ, particularly when the aim is to measure uncertainty in a model that is only initially specified deterministically. If there is no *a priori* knowledge about the spatiotemporal structure of the noise, then the default choice of $\sigma = I$ addresses a generic situation in which noise is uniformly ascribed to each component of the state. In Chapter 6, we took a naïve approach to include measurement error estimates in our model by selecting σ so that $\sigma\sigma^\top$ corresponded to a variance. However, there is potential for a more informed construction of σ that accounts for the many limitations of the deterministic model. For instance, the velocity field in the deterministic ODE was constructed by interpolating the observed data, which was only available on a finite spatiotemporal grid. One could encode in σ the error introduced by using this interpolant, for example, reducing the magnitude of the diffusivity at gridpoints and increasing it further away. The impact of interpolation, quantified by a multiplicative noise SDE, warrants a full investigation, in which stochastic sensitivity could provide a computable quantification of the impact of this introduced uncertainty (Fang, Balasuriya, et al., 2020; Fang and Ouellette, 2021). It is also pertinent to note that the driving Wiener process in the SDE model need not have the same dimension as the state variable. The model can thus account for different sources of uncertainty that do not need to be attributed directly to the components of the velocity field. This was an implicit limitation of the original stochastic sensitivity work (Balasuriya, 2020b), where the Wiener process could only be 1- or 2-dimensional. For instance, in a model containing several unknown or estimated parameters, we could encode the uncertainty in these parameters with the components of a Wiener process. This would lead to a certain construction of σ from the relationship between those parameters and the vector field of the ODE model—we provided a simple example of this in our 2-dimensional

toy model in Chapter 4, where we devised a spatially varying diffusion matrix (4.12) by attributing the uncertainty to two of the model parameters.

There are many application-specific methods available for estimating σ directly from observed data, e.g. via statistical estimation (Cotter and Pavliotis, 2009) or the Bayesian inference approach of Ying et al. (2019) in the context of ocean modelling. For a review of the statistical estimation of the diffusion matrix from noisy data, see Nielsen et al. (2000). Coupling these approaches with our methods could provide a complete and practical framework for characterising uncertainty using only observed trajectory data, without the need to specify directly the diffusion matrix as part of the model. In oceanography, a common approach is to specify the diffusion term as part of a physics-informed model (Berloff and McWilliams, 2002). For example (van Sebille et al., 2018; L. Li et al., 2023), in modelling Lagrangian particles, the movement of material via unresolved subgrid processes is quantified as diffusive transport. This then leads to the specification of an advection-diffusion equation with a spatially varying diffusivity that specifies the time-evolution of the spatial distribution of Lagrangian particles. Via the Fokker-Planck equation (to be introduced in Section 7.3), this purely deterministic formulation is equivalent to a stochastic differential equation with a drift and diffusion matrix relating to the terms of the advection-diffusion equation. A similar approach is employed in stochastic parameterisations of atmospheric convection (Wilson and Sawford, 1996). We must emphasise, however, that the specification of the diffusion term is an open question and depends on the specifics of the situation and modelling choices. Our theory is general enough to cover all of these aforementioned cases.

7.2 Boundary conditions

In many practical situations, we can enforce boundary conditions on the evolution of our state variable. These conditions are typically informed by physical aspects of the model, such as expecting that a variable of interest is non-negative, the presence of impermeable boundaries, or limitations of the underlying data. The Gulf Stream example in Chapter 6 provided one such situation. Firstly, the subset of altimetry data that we used was only available on a finite spatial region, which is almost always the case in data-driven models. Outside of this region, we have no data and so there is an implicit boundary: once a trajectory leaves the data domain we can no longer propagate it forward or make further inferences. We also expected, from physical considerations, that drifter trajectories cannot move onto the regions of land within the dataset, creating impermeable boundaries. For an ordinary differential equation, boundary conditions usually do not pose a significant problem. However, the treatment of these boundaries in stochastic differential equations can be more complicated, requiring adjustments to both the drift and diffusion. There are two primary types of boundary conditions: absorbing, which acts as a set of fixed points, and reflecting, where the trajectories are 'pushed away' from the boundary. Reflecting

boundary conditions can be enforced with a modification of the drift—see Pilipenko (2014) for an introduction. An absorbing boundary condition presents a more difficult proposition, such as requiring specialised numerical schemes (Mannella, 1999), causing a breakdown in the Markovian behaviour of the solution (Muñoz, 1998), and even causing ambiguity in the interpretation of the driving noise (Correales and Escudero, 2019).

Even without considering these technical modifications, there are difficulties in using our tools in the presence of boundaries. An n-dimensional Gaussian density has support on the entirety of $\mathbb{R}^n$, meaning that for any sensible non-empty region in $\mathbb{R}^n$, there is a non-zero probability that a random variable following a non-degenerate Gaussian distribution takes a value in that region. This property means that a Gaussian approximation, such as that arising from the linearisation, cannot capture boundary behaviour. Our Gaussian mixture model consequently suffers from the same limitation. A further difficulty in the mixture model algorithm is the selection of new components when splitting; without adjustment, the new components may be placed across boundaries and violate conditions on the state. This can be avoided by appropriately adjusting the new components and their weights, for instance, but is another practical issue that needs further consideration. We avoided complications in our Gulf Stream example by ensuring that the target densities and the evolution of solution trajectories were sufficiently far from the boundaries. The probability that a solution or the resulting approximations were close to the boundaries was then negligible. Without these considerations, however, the boundaries pose a significant problem—any inferences that use the SDE linearisation, notably stochastic sensitivity, cannot be trusted near those boundaries. When generating realisations, a naïve approach to handling these boundaries is to discard any trajectories that leave the spatial domain. In a similar vein, one may adjust the approximate probability density functions obtained by our techniques by discarding probability mass that violates the boundary conditions and renormalising the function. This approach is not supported with a mathematical justification, however, and would not account for the behaviour of the solution at the boundary (e.g. absorbing versus reflecting). A more intriguing prospect is to adjust the linearised SDEs themselves to include these boundary conditions and attempt to find an appropriate solution. This is not easy, however, due to the aforementioned theoretical modifications that must be made to the formulation and solution approach to a stochastic differential equation in the presence of boundaries. Any headway in this direction could overcome one of the most significant limitations of Gaussian/linearisation approximations of stochastic differential equations.

7.3 The Fokker-Planck equation

The Fokker-Planck equation is a partial differential equation that describes the time evolution of the probability density function of the solution to a stochastic differential

equation. Given an SDE

$$\mathrm{d}x_t = u(x_t, t)\,\mathrm{d}t + \sigma(x_t, t)\,\mathrm{d}W_t, \tag{7.1}$$

the probability density function ρ for the solution to (7.1) at time t solves the corresponding Fokker-Planck (FP) equation (Risken, 2012)

$$\frac{\partial \rho(x,t)}{\partial t} = \frac{1}{2}\nabla \cdot \nabla \cdot \Big(\rho(x,t)\,\sigma(x,t)\,\sigma(x,t)^\top\Big) - \nabla \cdot \big(\rho(x,t)\,u(x,t)\big) \tag{7.2}$$

subject to some initial density $\rho(x, 0)$ given by the distribution of the initial condition to (7.1). Solving the Fokker-Planck equation provides an alternative method for finding the solution to a stochastic differential equation; rather than working directly with the stochastic trajectories that solve (7.1), we instead seek solutions to a partial differential equation. Our work could both contribute to understanding the Fokker-Planck equation and conversely, make use of this alternative view on SDEs. We could extend our theoretical work on SDE linearisations to consider the corresponding Fokker-Planck equations. We presented the solution to the linearised SDE in Theorem 3.2, for which we can derive an expression for the probability density function—the PDF is a Gaussian density or can otherwise be computed as the convolution between a Gaussian and the initial condition PDF. Therefore, we have solutions available to the Fokker-Planck equation corresponding to the linearised SDE. If we could understand the relationship between the Fokker-Planck equations of the original SDE and the linearisation, this would accordingly offer insight into the relationship between the probability density functions of the respective SDE solutions. Such understanding cannot be inferred from our bound on the expectation in Theorem 3.1 alone.

Secondly, the Fokker-Planck equation cannot be solved analytically except for simple cases and so is typically approximated. There are two approaches to numerically solving the Fokker-Planck equation: either by generating many stochastic realisations by solving the corresponding SDE (2.6) numerically and using a density estimation method (Silverman, 2017), or by directly solving the Fokker-Planck equation with a finite-difference or finite-element scheme (Pichler et al., 2013). Both approaches have practical difficulties, including poor scaling with dimensionality and the enforcement of boundary conditions. These difficulties mean that the computational cost of solving the Fokker-Planck equation is considered impractical in 3- or higher dimensions (Allawala and Marston, 2016; Y. Li, 2019; Zhai et al., 2022; W. Anderson and Farazmand, 2024). The linearisation approximation and our Gaussian mixture model algorithm both provide analytic probability density functions and are both efficient methods that scale reasonably with dimension. These methods may therefore be viewed as approximate solution methods to the Fokker-Planck equation that overcome the computational limitations of numerical solutions. Understanding the relationship between SDE linearisations and the Fokker-Planck equation may provide an avenue for further developing the mixture model algorithm. The aim of the mixture

model is to approximate the probability density function of the solution to the SDE, so the Fokker-Planck equation may be a more appropriate framework for investigating this algorithm, as opposed to the solution-trajectory focus of the theory in Chapter 3.

7.4 Non-Gaussian noise processes

Throughout this thesis, we have considered Itô stochastic differential equations driven by the canonical Wiener process, and so we assume that the noise in our system is white (zero temporal correlation) and Gaussian. However, when modelling physical systems there is observational evidence to suggest that in some modelling scenarios, the underlying noise process may be better formulated as a more general Lévy process (Viecelli, 1998; Ditlevsen, 1999). A Lévy process satisfies the same properties as the Wiener process but without Gaussian increments and can therefore capture skew and heavy-tailedness in the ongoing noise process. Many of the theoretical results for stochastic differential equations hold for when the Wiener process is replaced by a Lévy process (Applebaum, 2004).

There is scope to determine whether the theory presented in Chapter 3, and importantly the computability of the linearisation solution, can be generalised to stochastic differential equations driven by arbitrary Lévy processes. If we replace the Wiener process in Theorem 3.1, the proof of Theorem 3.1 will still hold as it relies upon results for general Itô integrals and square-integrable stochastic processes. This would show that such a solution, for small noise, can be approximated by an expression involving an Itô integral of a deterministic function (i.e. (3.5)) with respect to the driving process. The applicability of this approximation then depends upon how this integral can be evaluated without having to solve numerically the original stochastic differential equation. When the Wiener process is replaced, Itô's isometry does not hold and we will need a different characterisation of the linearisation solution—the mean and variance that suffice to describe a Gaussian process may not be enough. Regardless, a theoretical understanding of the linearisation of SDEs driven by Lévy processes would be a significant contribution to the literature.

7.5 Pursuing higher-order terms

The linearisation approximation that underpins the work in this thesis equivalently arises by considering a formal power-series expansion of the SDE solution. Given the solution $y_t^{(\varepsilon)}$ to a small-noise SDE

$$\mathrm{d}y_t^{(\varepsilon)} = u\left(y_t^{(\varepsilon)}, t\right)\mathrm{d}t + \varepsilon\sigma\left(y_t^{(\varepsilon)}, t\right)\mathrm{d}W_t, \tag{7.3}$$

one can consider a power-series expansion in ε

$$y_t^{(\varepsilon)} = z_t^{(0)} + \varepsilon z_t^{(1)} + \varepsilon^2 z_t^{(2)} + \cdots + \varepsilon^m z_t^{(m)} + \cdots, \tag{7.4}$$

where the coefficients $z_t^{(0)}, z_t^{(1)}, z_t^{(2)}, \ldots$ are themselves stochastic processes independent of ε. Blagoveshchenskii (1962) considers such expansions and found generic bounds for the expected distance between the solution to (7.3) and truncations of (7.4). The linearised SDE we consider corresponds to truncating (7.4) to the $m = 1$ term. In finding the linearisation solution $l_t^{(\varepsilon)}$ we implicitly established that $z_t^{(0)} = F_0^t(x_0)$, the deterministic flow map and $z_t^{(1)}$ can be expressed as the Itô integral of a deterministic function—compare $z_t^{(0)} + \varepsilon z_t^{(1)}$ with (3.5). We expect that including more terms when truncating (7.4) will result in a 'better' approximation of the original SDE solution, in the same fashion as a Taylor expansion of a deterministic function. An obvious extension to our theoretical results would therefore be to higher-order expansion, where we look to continue building upon the work of Blagoveshchenskii (1962) to provide explicit bounds on the distance between the SDE solution and truncations of (7.4). From a practical perspective, this also raises the question of whether the higher-order terms can be computed to provide a more accurate approximation of nonlinear SDE solutions. However, these higher-order coefficients obey stochastic differential equations (Blagoveshchenskii, 1962), which cannot in general be solved analytically. For instance, by performing Taylor expansions of the coefficients of (7.3) and rearranging, the second-order term $z_t^{(2)}$ is shown to solve

$$\mathrm{d}z_t^{(2)} = \left[\frac{1}{2}\left[z_t^{(1)}\right]^\top \nabla\nabla u\left(z_t^{(0)}, t\right) z_t^{(1)} + \nabla u\left(z_t^{(0)}, t\right) z_t^{(2)}\right] \mathrm{d}t + \nabla\sigma\left(z_t^{(0)}, t\right) z_t^{(1)}\, \mathrm{d}W_t$$

Although linear in $z_t^{(2)}$, this equation must be viewed jointly with that satisfied by $z_t^{(1)}$, cannot be solved analytically and is even difficult to solve numerically due to the dependence on the second derivatives of u. Thus, while there is potential to extend our *theoretical* results to higher-order expansions, we expect that there would not be the same *practicality* of computation that we relied upon later in the thesis. It remains to be seen, however, if qualitative measures can be computed from the higher order terms, in the fashion in which stochastic sensitivity was shown to be the variance of the first-order term $z_t^{(1)}$.

7.6 Applying stochastic sensitivity

As a recent development, stochastic sensitivity has only seen limited application (Balasuriya, 2020c; Fang, Balasuriya, et al., 2020; Fang and Ouellette, 2021; Badza et al., 2023), which we reviewed in Section 2.5. We have presented a complete generalisation of these tools to arbitrary dimensions and established the connection to SDE linearisation that empowers a rapid computation. Equipped with this extension, one can now investigate stochastic sensitivity as a Lagrangian diagnostic and a procedure to extract coherent structures via robust sets in higher-dimensional flows, and in particular 3-dimensions. More broadly, stochastic sensitivity can now be used as a theoretical and computational tool in any differential equation model. There are consequently many possible applications of stochastic

sensitivity, but in this short discussion, we focus on those related to Lagrangian coherent structures. The work by Fang, Balasuriya, et al. (2020) and Fang and Ouellette (2021), which used stochastic sensitivity as a dynamic lengthscale, can now immediately be extended to dynamical systems of arbitrary dimension.

Balasuriya, Ouellette, et al. (2018) provide a framework that extends the notion of a Lagrangian coherent structure to capture spatial "regions of interest" in the time-evolution of other quantities that are transported by, but not fully locked to, a flow. Examples of these quantities include the concentration of pollutants, the density of an organism such as phytoplankton, temperature, and salinity in the fluid context. The Fokker-Planck equation is a generalisation of the classical advection-diffusion equation that allows for a spatiotemporally varying diffusivity tensor. The spread of a tracer under the advection-diffusion equation can, through the Fokker-Planck formulation, be equivalently captured by the solution to a stochastic differential equation. The spread of "uncertainty" in the stochastic formulation corresponds to the movement and dispersion of the tracer, in the sense that the probability density function of the stochastic solution is the normalised tracer concentration. In describing the behaviour of the stochastic system and enabling the extraction of coherent regions, stochastic sensitivity can offer insight into the spread of such tracers and potentially be used to extract generalised Lagrangian coherent structures in a broader framework. Hence, stochastic sensitivity is a far more general framework that can apply to systems beyond the tracking of particles, particularly with the developments into arbitrary dimensions provided by this thesis. Establishing a theoretical connection between stochastic sensitivity and the Fokker-Planck equation would also be valuable (Balasuriya, 2020a) and many now be viable as we have established that stochastic sensitivity is associated with a density function (the Gaussian solution to the linearised SDE) that itself solves a Fokker-Planck equation.

Stochastic sensitivity also has the potential as a *theoretical* quantity. Most traditional Lagrangian coherent structure procedures are completely deterministic, not accounting for any uncertainty in the driving velocity field. The robustness of several popular LCS methods to stochastic noise has recently been explored by Badza et al. (2023) but via stochastic simulation and summary statistics. Stochastic sensitivity can be used to perform a purely theoretical analysis of such sensitivity in LCS computations. More generally, the linearised covariance matrix and stochastic sensitivity values are computable using the flow map gradient ∇F_0^t, which is a quantity often used in LCS computation, and so can enable a rapid computation to *supplement* an LCS extraction scheme. Balasuriya (2020c) provides a preliminary derivation (using the previously 2-dimensional formulation of stochastic sensitivity) of error bounds for the finite-time Lyapunov exponent, as an example of how stochastic sensitivity can be used as a theoretical tool in LCS analysis. In this work, stochastic sensitivity provided an estimate of the standard deviation in the flow map, but with the extension of stochastic sensitivity to arbitrary dimensions and the additional knowledge of Gaussianity in the relevant underlying distribution, this initial

Transition	Event		Rate
1	Infection	$(S, I) \to (S-1, I+1)$	$\beta SI/M$
2	Recovery	$(S, I) \to (S, I-1)$	γI

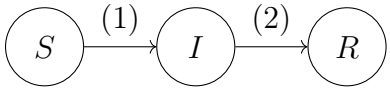


Figure 7.1: A description of the SIR model formulated as a continuous-time Markov chain, including (top) the transition rates and (bottom) the corresponding transitions between states for an individual in a population.

work could be further formalised to provide a leading-order estimate for the error in FTLE computations.

7.7 Applications to population processes

To conclude this thesis, we discuss results for stochastic systems evolving over *discrete* state space that are analogous to ours. This suggests that the tools we have developed can apply to a far broader range of modelling scenarios. A population process is a stochastic process where each component of the process typically evolves over some subset of the positive integers. As the name suggests, such a process is used to model the number of individuals in a population (Kendall, 1949). These processes are used extensively across many fields but most notably within epidemiological, biological, and ecological modelling (Brauer, 2008). A classic application of population processes is in the modelling of the spread of an infectious disease in a population, where the process counts the number of individuals in each stage of the disease. For a well-mixed and homogeneous population, a continuous time Markov chain (CTMC) is a natural choice to formulate mathematically the evolution of a population process. A CTMC is a stochastic process evolving over a discrete (but not necessarily finite) state space in continuous time, where the process moves from one state to another. The process dynamics are characterised by these transition rates, which give the instantaneous probability per unit time of the process transitioning from one state to another at a specified rate. The term "Markov" refers to the assumption that the behaviour of the process at a given time only depends on the current state and not any of the preceding history. For a detailed introduction to CTMCs, see W. J. Anderson (1991)——the key properties for our purposes are that a CTMC is fully described by the state space and transition rates and can be simulated numerically to generate realisations.

As a simple example, consider the susceptible-infected-recovered (SIR) model (Allen, 2017), which represents the spread of an infectious disease through a population where each individual is one of susceptible (able to be infected), infected (infectious to susceptible individuals), or recovered (no longer infectious and unable to be re-infected). The transition

rates are listed in Figure 7.1, with the diagram showing how an individual in the population moves through the three stages. In the table, S denotes the number of susceptible individuals and I denotes the number of infected individuals. For a fixed population size M, the number of recovered individuals R is given by $R = M - S - I$, so we can formulate the stochastic model as 2-dimensional. Let $X_t = (S_t, I_t)^\top$ be the stochastic process, where S_t is the number of susceptible individuals and I_t is the number of infected at time t. The state space of the process is then $\mathcal{S} = \left\{(s,i) \,\middle|\, s, i \in \{0, 1, \ldots, M\}\,,\ s + i \leq M\right\}$. Two possible events lead to a change in the state of the process: an infection of a susceptible individual, which occurs with a rate of $\beta S_t I_t$ with S_t and I_t denoting the number of susceptible and infected individuals respectively at the time t, and a recovery of an infected individual, which occurs with rate γI_t. This is sufficient to fully describe the CTMC model.

As with stochastic differential equations, CTMC models cannot generally be solved (in the sense of finding distributions across the state space at certain times, determining the long-term behaviour, etc.) analytically, but can be simulated (Gillespie, 1977). We summarise a method for simulating from a CTMC in Appendix B.2, which is used in the two examples in this section. The need for stochastic simulation leads to the same practical issues as with SDEs; for complicated models, this is numerically inefficient and a large number of samples are needed for accurate inference. However, these are approaches to approximating the behaviour of the CTMC that circumvent these computational difficulties. Seminal work by Kurtz (1970, 1971) showed that the stochastic evolution of certain CTMCs converges to the solutions to an ordinary and a stochastic differential equation in the limit of infinite population size. These results hold when the process is modelling a fixed population size and is density dependent. A CTMC population process X_t with a fixed population size M is density dependent when the transition rates only depend on $\frac{1}{M}X_t$ and not X_t directly. Similar results for relaxations of these conditions exist (Pollett, 1990), however. We briefly summarise these results here—Appendix B.1 provides further details—and can draw analogies with our work. Let $X_t^{(M)}$ denote an n-dimensional population process, with each component taking values on the discrete set $\{0, 1, \ldots, M\}$. The density process is then $Y_t^{(M)} = \frac{1}{M}X_t^{(M)}$, which describes the proportion of the population represented by each component. Kurtz (1970) showed that in the limit of large population size, $M \to \infty$, the density process $Y_t^{(M)}$ converges in probability to the deterministic trajectory $Y_t^{(\infty)}$ that solves the ODE (Kurtz, 1970)

$$\frac{\mathrm{d}Y_t^{(\infty)}}{\mathrm{d}t} = Q\left(Y_t^{(\infty)}\right), \quad Y_0^{(\infty)} = \lim_{M\to\infty} \frac{1}{M} X_0^{(M)}, \tag{7.5}$$

where Q is determined by the transition rates of the CTMC and captures the average (macroscopic) behaviour of the density process. The deterministic ODE (7.5) is known as the *fluid limit* of the population process. Although the initial condition to (7.5) is formally written as a limit, in practice $Y_0^{(M)}$ can be specified directly and independently of

M. Kurtz (1971) then further establishes a stronger result; that the stochastic variation of $Y_t^{(M)}$ for large M can be described by the solution to a stochastic differential equation. The scaled process

$$Z_t^{(M)} = \sqrt{M}\left(Y_t^{(M)} - Y_t^{(\infty)}\right),$$

which captures the deviation between the density process and the fluid limit, converges in distribution to the solution to the linear stochastic differential equation (Kurtz, 1971)

$$\mathrm{d}Z_t^{(\infty)} = \nabla Q\left(Y_t^{(\infty)}\right) Z_t^{(\infty)}\,\mathrm{d}t + G\left(Y_t^{(\infty)}\right)\mathrm{d}W_t, \quad Z_0^{(M)} = 0, \tag{7.6}$$

which is termed the *diffusion limit*. The $n \times n$ diffusion matrix G is also determined by the transition rates of the Markov chain, and W_t is a canonical n-dimensional Wiener process. In the initialisation of (7.6), it is assumed that the initial condition $Y_0^{(\infty)}$ is fixed. The diffusion term captures the microscopic behaviour resulting from individual transition events, and the uncertainty in this is parameterised with W_t. The diffusion limit (7.6) is equivalent to the unscaled SDE

$$\mathrm{d}L_t^{(N)} = \left[Q\left(Y_t^{(\infty)}\right) + \frac{1}{\sqrt{N}}\nabla Q\left(Y_t^{(\infty)}\right)\left(L_t^{(N)} - Y_t^{(\infty)}\right)\right]\mathrm{d}t + \frac{1}{\sqrt{N}}G\left(Y_t^{(\infty)}\right)\mathrm{d}W_t, \tag{7.7}$$

which is then a linearisation of the nonlinear SDE

$$\mathrm{d}\hat{Y}_t^{(N)} = Q\left(\hat{Y}_t^{(N)}\right)\mathrm{d}t + \frac{1}{\sqrt{N}}G\left(\hat{Y}_t^{(N)}\right)\mathrm{d}W_t \tag{7.8}$$

about the deterministic limit $Y_t^{(\infty)}$. The diffusion limit (7.6) can be solved analytically and follows a Gaussian distribution at any fixed time. For large populations, the diffusion limit is used as an approximation to the population process that avoids the need for bulk simulation (Pollett et al., 2010).

For example, the SIR model is a density dependent process, with the fluid limit

$$\frac{\mathrm{d}Y_t^{(\infty)}}{\mathrm{d}t} = \begin{bmatrix} -\beta Y_t^{(\infty,1)} Y_t^{(\infty,2)} \\ \beta Y_t^{(\infty,1)} Y_t^{(\infty,2)} - \gamma Y_t^{(\infty,2)} \end{bmatrix}, \tag{7.9}$$

where $Y_t^{(\infty)} = \left(Y_t^{(\infty,1)}, Y_t^{(\infty,2)}\right)^\top$ is the density process. Here, the components of the density process are the proportion of the population that are susceptible and infected respectively. The diffusion limit is the linear SDE

$$\mathrm{d}Z_t^{(\infty)} = \begin{bmatrix} -\beta X_t^{(\infty,2)} & -\beta X_t^{(\infty,1)} \\ \beta X_t^{(\infty,2)} & \beta X_t^{(\infty,1)} - \gamma \end{bmatrix} Z_t^{(\infty)}\,\mathrm{d}t + \begin{bmatrix} \sqrt{\beta X_t^{(\infty,1)} X_t^{(\infty,2)}} & 0 \\ -\sqrt{\beta X_t^{(\infty,1)} X_t^{(\infty,2)}} & \sqrt{\gamma X_t^{(\infty,2)}} \end{bmatrix} \mathrm{d}W_t \tag{7.10}$$

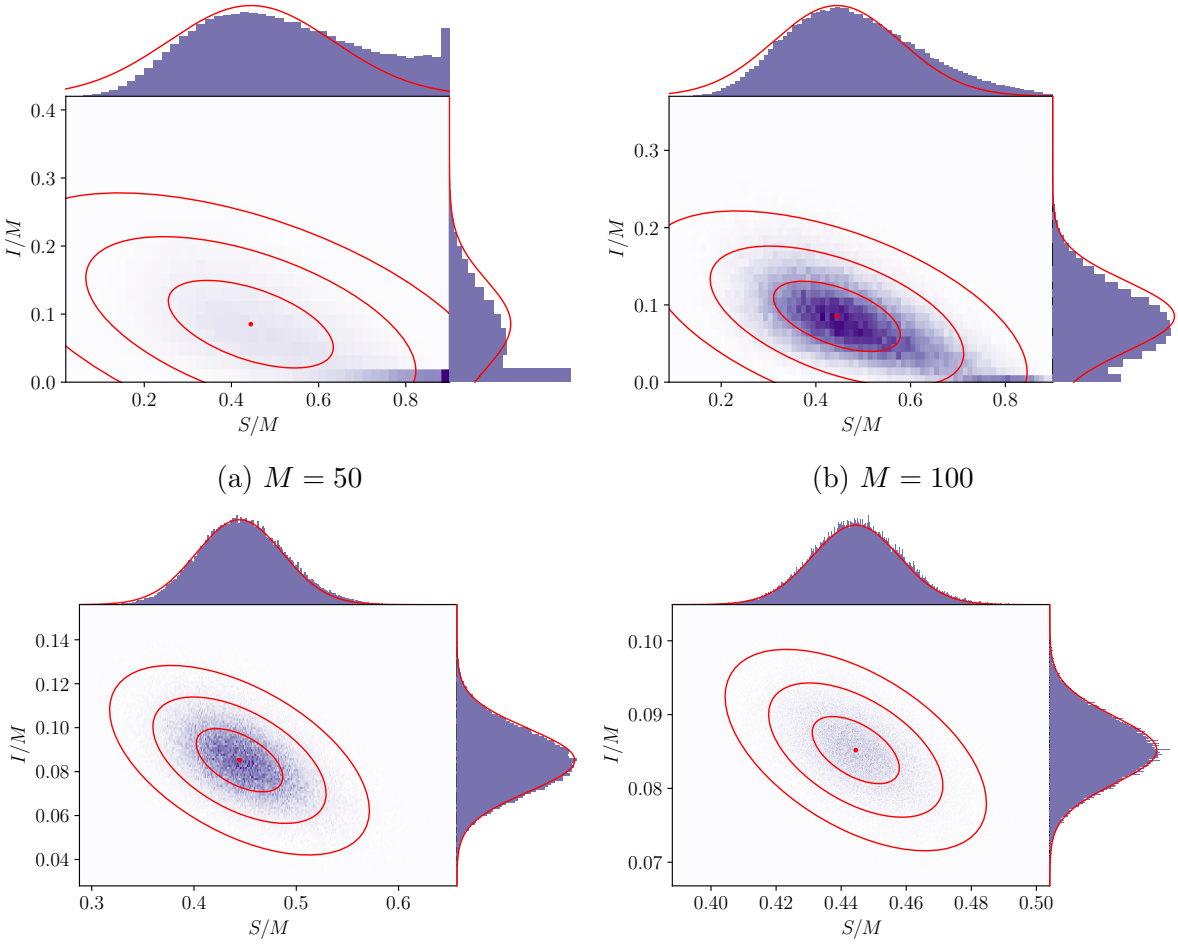
Solving (7.9) and (7.10) provides the approximation

$$Y_t^{(M)} \approx Y_t^{(\infty)} + \frac{1}{\sqrt{M}} Z_t^{(\infty)},$$

that is valid in the sense of a limit for large M.

We can draw analogies between these results and the SDE linearisation procedure in Chapter 3. The fluid limit (7.5) is parallel to the convergence of a small-noise SDE solution to the corresponding deterministic ODE solution, i.e. $y_t^{(\varepsilon)} \to y_t^{(0)}$ as $\varepsilon \downarrow 0$ in the notation of Chapter 3. The diffusion limit in unscaled form (7.7) is then comparable to the general linearisation (3.3) of a stochastic differential equation (3.1), with $\varepsilon = 1/\sqrt{M}$. However, a key difference is that in the SDE linearisation procedure, we start from a continuous state-space process and arrive at a continuous state-space process in the limit. Whereas, in the diffusion limit, the converging process is on a *discrete* state space and the limit is on a continuous one. However, we can see that the diffusion limit also arises from a similar process: the unscaled diffusion limit (7.7) is a linearisation of (7.8) about the deterministic fluid limit (7.5). The nonlinear SDE (7.8) may serve as an alternative approximation of the density process (Allen, 2017). We have also proven a stronger notion of convergence than in the original work of Kurtz (1970, 1971), by establishing bounds on the expectation (thus implying convergence in rth mean) in Theorem 3.1. Regardless, our results and the diffusion limit for population processes are built around the same underlying linear stochastic differential equation, so our later practical results can be applied to characterise and approximate the solutions to these discrete models. In particular, we expect that our mixture model algorithm may find a place in models of moderate population size, where the population is large enough to be "reasonably" approximated by the continuous SDE equations but small enough to exhibit non-Gaussian behaviour.

We shall now demonstrate some of these calculations and postulate future applications of stochastic sensitivity and our mixture model algorithm. In Figure 7.2, we show histograms of 10000 realisations of the population-scaled SIR model for four different population sizes. We set the parameters $\beta = 1.2$ and $\gamma = 0.8$, initialised each process with 10% of the population infected and the remaining susceptible (so $X_0^{(M)} = (0.1M, 0.9M)^\top$) and show each realisation after $t = 5$ time units. We also compute the Gaussian solution to the diffusion limit (7.10), using the Mazzoni (2008) method for rapid computation of the moments of the solution, and show resulting density as contours on the joint histograms and by plotting the marginal PDFs on each axis. As the population size M increases, the Gaussian approximation provides a closer representation of the empirical distribution of the samples. This should be compared with Figures 4.2, 4.4 and 4.6, where we verified heuristically that the solutions to small-noise SDEs approached the Gaussian solutions of their respective linearisation. It is also evident in Figure 7.2 that the boundaries limit the validity of the Gaussian approximation: the density process is restricted to $[0, 1]$ with an absorbing state at 0 that results in the accumulation of probability mass seen in



(a) $M = 50$ (b) $M = 100$

(c) $M = 1000$ (d) $M = 10000$

Figure 7.2: Histograms of Monte Carlo simulations of the density process for the SIR model (with marginal plots on each axis), and the probability density function of the corresponding solution to the diffusion limit (7.10) plotted in red. The parameters are $\beta = 1.2$ and $\gamma = 0.8$, and each sample path is initialised with 10% of the population infected and simulated up to $t = 5$. The bins are chosen to reflect the state-space $\{0, 1/M, 2/M, \ldots, 1\}$ of the density process.

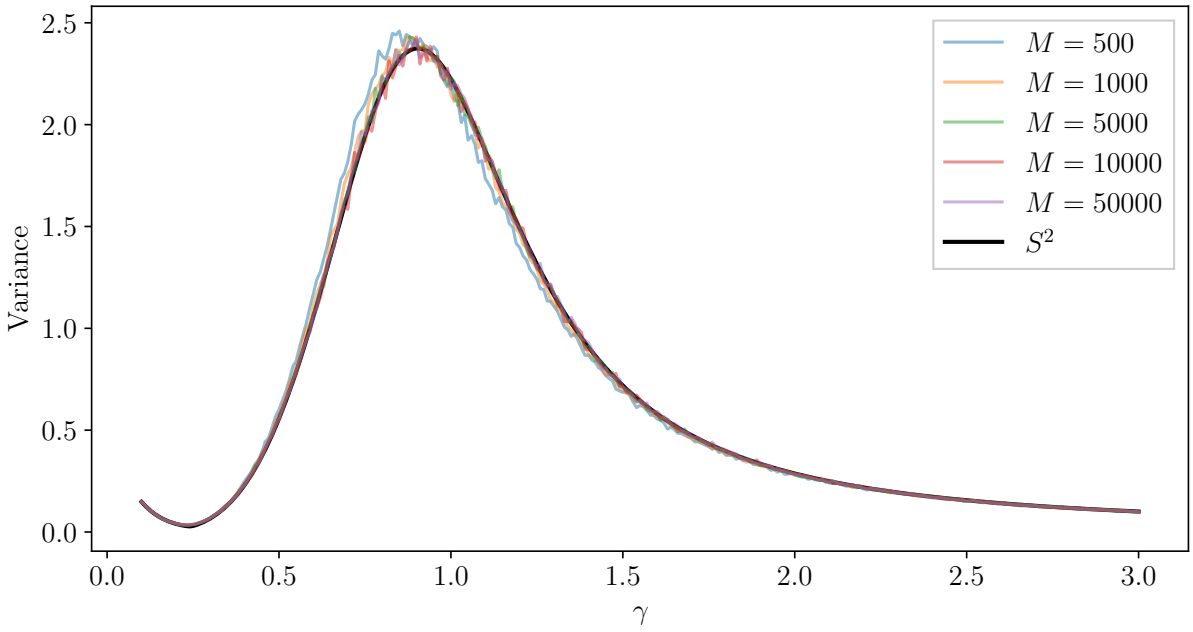


Figure 7.3: The stochastic sensitivity value (in black) of the SIR diffusion limit (7.10) at time $t = 5$ and from the fixed initial condition of 10% of the population infected, for $\beta = 1$ and varying γ. The operator norm of the sample covariance matrix (in various colours) for 1000 stochastic simulations of the discrete population process over the same time interval is included.

the histograms. Such boundary problems are common in epidemiological and biological models—e.g. the early spread of a disease through a population is often of interest—and the inability of the diffusion limit to account for these appropriately is a significant limitation. We discussed in Section 7.2 the boundary complication in our SDE linearisation framework and suggested that appropriately accounting for boundary conditions and absorbing states would be a valuable avenue for future work. Any developments in this direction, whether that be a modification to the linearised SDE itself or an adjustment to the solution PDF, would also be useful for the diffusion limit of a CTMC.

Next, we compute stochastic sensitivity for the SIR CTMC using the diffusion limit (7.10). We can compute the covariance matrix of the solution to any diffusion limit (7.6), which satisfies the covariance differential equation (3.9) from Chapter 3 with Q in place of u and g in place of σ. Then, the stochastic sensitivity value is calculated by taking the operator norm, using this matrix in place of Σ_0^t in Theorem 3.3. In Figure 7.3, we plot the computed stochastic sensitivity for the SIR model and compare the values with the corresponding empirical value (the operator norm of the sample covariance matrix) from 1000 realisations of the population process, for varying population size M. We have fixed $\beta = 1$ and varied γ. For each parameter value, we again initialised with 10% of the

Transition	Event	Rate λ_i
(1)	$(S, E) \to (S-1, E+1)$	$(\beta_I SI + \beta_H SH + \beta_F SF)/N$
(2)	$(E, I) \to (E-1, I+1)$	αE
(3)	$(I, H) \to (I-1, H+1)$	$\gamma_H \theta_1 I$
(4)	$(H, D) \to (H-1, D+1)$	$\gamma_{dh} \delta_2 H$
(5)	$(D) \to (D-1)$	$\gamma_f D$
(6)	$(I) \to (I-1)$	$\gamma_i (1-\theta_1)(1-\delta_1) I$
(7)	$(I, D) \to (I-1, D+1)$	$\delta_1 (1-\theta_1) \gamma_d I$
(8)	$(H) \to (H-1)$	$\gamma_{ih} (1-\delta_2) H$

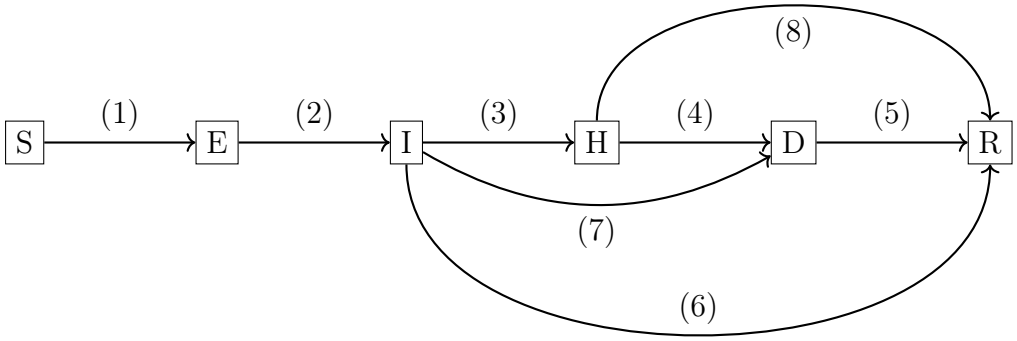


Figure 7.4: Transition probabilities of the Ebola model of Legrand et al. (2007).

population infected and considered the model up to $t = 4$. The variance is of the density process but can be rescaled to that of the original population process by multiplying by M^2. Even for a small population size, the stochastic sensitivity provides a reliable indication of the variation in the population process and, therefore, can serve as a tool in analysing these models. This is not dissimilar to the work of Pollett et al. (2010), who suggest that the covariance matrix of the solution to the diffusion limit measures the 'unexplained variation' in using the fluid limit alone to approximate the fully stochastic process. One advantage of stochastic sensitivity is that it provides a single number, by performing a dimension reduction, which avoids ambiguity in higher dimensions.

Finally, we will show the computation of the Gaussian approximation arising from the diffusion limit on a 5-dimensional model. These approximations become particularly useful in high dimensions, where the number of stochastic samples to make accurate inferences scales poorly. We consider a 5-dimensional model for the spread of Ebola, used in a case study by Legrand et al. (2007) on two particular outbreaks in the Democratic Republic of Congo and Uganda. This model is far more complicated than the SIR example; Figure 7.4 provide the transition events and a diagram corresponding to the

evolution of an individual through the stages of the infection. Similar to the SIR model, an individual in the population is in one of six stages: susceptible (S), exposed to the infection but not yet infectious (E), infectious to susceptible individuals (I), hospitalised (H), deceased with a traditional burial (D), and removed from the population and no longer infectious or susceptible (R). We choose this model specifically as the study by Legrand et al. (2007) provides values for all the involved parameters that are fitted using an actual outbreak of Ebola in 1995 in the Democratic Republic of Congo, immediately providing us with a physically relevant model. The population size is fixed at $M = 200000$ (Dowell et al., 1999) and the initial condition is $X_0 = (0, 0, 3, 0, 0)$ (Khan et al., 1999) in accordance with the 1995 outbreak. The model by Legrand et al. (2007) is then formulated as a 5-dimensional CTMC $Xt \equiv (S_t, E_t, I_t, H_t, D_t)^\top$, where each component tracks the number of individuals in each of the compartments and the number of removed individuals is $R_t = M - S_t - E_t - I_t - H_t - D_t$. Since we only provide a demonstration of the computation involved, rather than exploring this example in detail, further details are left to Appendix B.3. We take the parameter values from Legrand et al. (2007), who combined values estimated from other studies (Bwaka et al., 1999; Dowell et al., 1999; Khan et al., 1999; Ndambi et al., 1999; Rowe et al., 1999) with their own maximum likelihood estimates using morbidity data. These parameter values are provided in Figure B.1 in Appendix B.3.

The CTMC with the rates described in Figure 7.4 is density dependent, meaning that the fluid and diffusion limits hold. We provide the ODE and SDE in Appendix B.3, for which we can solve jointly by again using the Mazzoni method, which scales well with the increased dimension as the algorithm relies primarily on matrix calculations.

In Figure 7.5, we have generated 10000 realisations of the CTMC model after 20 weeks and plotted empirical histograms of the population-scaled counts. To visualise the 5-dimensional distribution of these realisations, we have plotted joint histograms of each pair of components on the off-diagonals. On the diagonal, we have included histograms of each component of the realisations. We also compute the Gaussian solution to the diffusion limit and overlay contours of the joint marginal PDFs corresponding to each pair of components, and the PDF itself on each single-variable marginal. Since the population size in this example is large ($M = 200000$), the Gaussian approximation is reasonable even in this high-dimensional setting. However, there are small departures from Gaussianity, including skewness in the single-variable marginals and nonlinear correlations in the joint density. The mixture model algorithm may have the potential to capture these features and provide a more accurate approximation. This would require further development of the algorithm, however, including appropriate handling of the boundary conditions (i.e. note the proximity of the component distributions to 0). Nonetheless, we expect that the mixture model will provide a computationally efficient alternative to stochastic simulation of high dimensional CTMC models, and this has much potential in this area.

We have only summarised the diffusion limit, drawn analogies with our work, and complemented this discussion with two examples. These connections are highly promising,

however, and suggest that stochastic sensitivity and our mixture model have applications across a far broader range of models than the stochastic differential equations we originally worked with. The mixture model in particular could provide an approximate method for solving complicated and high-dimensional stochastic models without simulation. Whereas the diffusion limit and the inference on the model it provides are emploted across applications (Pollett, 1990; Pollett et al., 2010, e.g.), these ideas are novel in the dynamical systems and stochastic differential equations settings. By establishing connections between our results and those for discrete stochastic models, we can provide tools from either field in new situations and contexts, bridging a gap between two otherwise disparate fields.

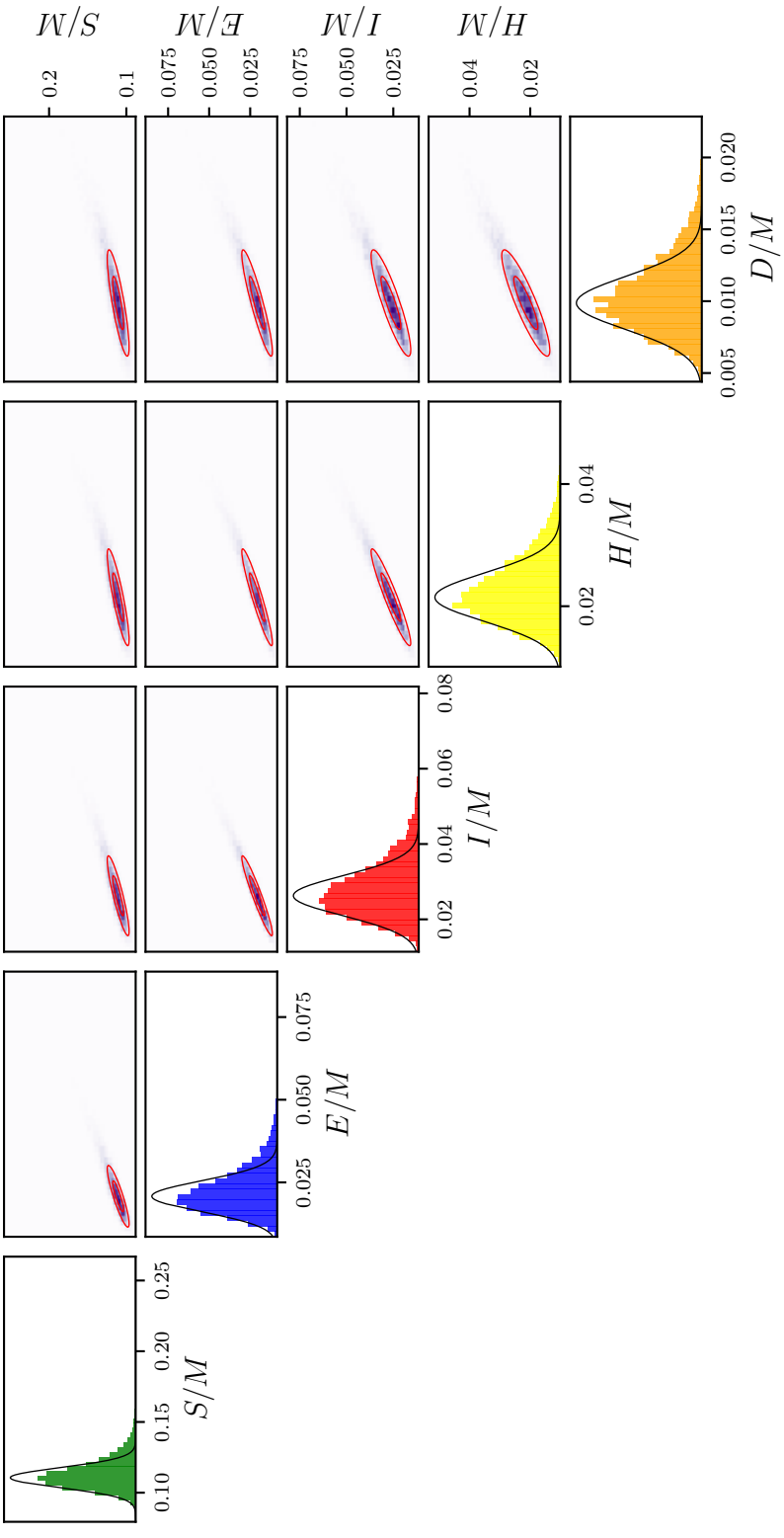


Figure 7.5: Histograms of 100000 realisations of the 5-dimensional Ebola model by Legrand et al. (2007), after 20 weeks. Each row and column corresponds to one of the 5 components of the process, with the off-diagonal plots showing the joint bivariate histograms of each pair of components and the diagonal entries showing the single-component marginals. Overlaid in red on each histogram are the contours (for the joint histograms) or probability density functions (for the single-component histograms) of the corresponding marginals of the Gaussian approximation that solves the diffusion limit.

Appendix A

Additional theoretical background

This appendix states several theoretical results from deterministic dynamical systems and stochastic calculus which are used throughout this thesis, and in particular the proofs presented in Section 3.3. These results are included for completeness, so we do not include proofs and instead refer the reader to other sources.

A.1 Deterministic results

An important result for establishing bounds on a function given an integral inequality is Grönwall's inequality.

Theorem A.1 (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. The integral form of this result was first stated and proven by Bellman (1943). □

A.2 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 particular use of in the proofs presented in Chapter 3. Let $(\Omega, \mathcal{F}, P)$ denote the probability space on which the Wiener process W_t is defined—the sample space Ω is typically the space of all continuous real-valued functions defined on $[0, T]$ (Kallianpur and Sundar, 2014). The Wiener process W_t has an associated natural filtration of $(\Omega, \mathcal{F}, P)$, which we denote $\mathcal{F}_t$. Let $\mathfrak{B}(A)$ generically denotes the Borel σ-algebra on the set $A \subseteq \mathbb{R}^q$. A random function $f : \Omega \times [0, T] \to \mathbb{R}$ is *Itô integrable* over the interval $[s, t] \subseteq [0, T]$ if (Øksendal, 2003; Kallianpur and Sundar, 2014)

(i) f is $\mathcal{F} \times \mathfrak{B}\big([0, T]\big)$-measurable,

(ii) f is $\mathcal{F}_t$-adapted, and

(iii)
$$\mathbb{E}\left[\int_s^t f(\omega, \tau)\, \mathrm{d}\tau\right] < \infty.$$

The first tool 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 A.2 (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 A.3 (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*

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

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 A.4 (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*

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

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

□

Appendix B

Details of population process models

In this appendix, we supplement the discussion on population processes and continuous time Markov chains in Section 7.7 by providing further details on the theoretical results by Kurtz (1970, 1971), the method for simulating from CTMCs, and the details of the 5-dimensional Ebola model by Legrand et al. (2007).

B.1 Coefficients of the fluid and diffusion limits

Let X_t denote the unscaled n-dimensional continuous-time Markov chain with state space $\mathcal{S} \subseteq \{0, 1, \ldots, M\}^n$, so that the process is evolving with a fixed population size M. Let $q\colon \mathcal{S} \times \mathcal{S} \to [0, \infty)$ denote the transition rates between two states of the Markov chain The process X_t is termed *density dependent* (in the sense of Kurtz (1970)) if

$$q(x, x + l) = Mf\left(\frac{x}{M}, l\right), \tag{B.1}$$

where f is a suitable function and $x, x + l \in \mathcal{S}$ for $l \in \mathbb{R}^n$. Let $Y_t^{(M)}$ describe the n-dimensional density process, i.e. $Y_t^{(M)} = X_t^{(M)}/M$. Theorem 3.1 of Kurtz (1970) establishes that in the large population limit $M \to \infty$, the density process $Y_t^{(M)}$ converges in probability to a deterministic trajectory $Y_t^{(\infty)}$ solving the ODE

$$\frac{\mathrm{d}Y_t^{(\infty)}}{\mathrm{d}t} = Q\left(Y_t^{(\infty)}\right), \quad Y_0^{(\infty)} = \lim_{M\to\infty} \frac{1}{M} X_0^{(M)}, \tag{B.2}$$

where

$$Q(y) = \sum_{\substack{l \in \mathbb{R}^n \\ l \neq 0,\, y + l \in \mathcal{S}}} l f(y, l).$$

For large M, the density process has small variation and is "close" to the deterministic solution to (B.2). Kurtz (1971) then established a stronger result, showing that the

variation of the density process about this deterministic limit is captured by an Itô diffusion. Define the scaled process

$$Z_t^{(M)} = \sqrt{M}\left(Y_t^{(M)} - Y_t^{(\infty)}\right),$$

then Theorem 3.5 of Kurtz (1971) proves that $Z_t^{(M)}$ converges in distribution (weakly) to an Itô diffusion $Z_t^{(\infty)}$ solving

$$\mathrm{d}Z_t^{(\infty)} = \nabla Q\left(Y_t^{(\infty)}\right) Z_t^{(\infty)}\,\mathrm{d}t + G\left(Y_t^{(\infty)}\right)\mathrm{d}W_t, \quad Z_0^{(\infty)} = 0, \tag{B.3}$$

where W_t is an n-dimensional Wiener process and the $n \times n$ diffusion matrix G is such that

$$\left[G(y)\,G(y)^\top\right]_{ij} = \sum_{\substack{l\in\mathbb{R}^n \\ l\neq 0,\, y+l\in\mathcal{S}}} l_i l_j f(y,l) \tag{B.4}$$

Any choice of diffusion matrix G such that (B.4) is satisfied will result in a statistically identical stochastic process $Z_t^{(\infty)}$.

B.2 Simulating from a continuous-time Markov chain

A sample path can be simulated from a continuous-time Markov chain using an algorithm first given by Gillespie (1977). Suppose we are given a CTMC defined on a state space $\mathcal{S}$ and with transition rates $q\colon \mathcal{S}\times\mathcal{S}\to[0,\infty)$. A single sample X from the CTMC at time T can be simulated with the following procedure (Gillespie, 1977):

1. Initialise $X = X_0$, the initial state of the process. If the initial state is uncertain, sample X from the initial state distribution.

2. Sample the time τ to the next transition event as

$$\tau \sim \mathrm{Exp}\left(\sum_{\substack{Y\in\mathcal{S} \\ Y\neq X}} q(X,Y)\right),$$

   where $\mathrm{Exp}(\lambda)$ denotes the exponential distribution with rate parameter λ. Set $t = t + \tau$.

3. If $t > T$, terminate. Otherwise, sample the next state from the set of possible transitions, where the probability of transitioning from X to a different state $Y \in \mathcal{S}\setminus\{X\}$ is given by

$$P(X\to Y) = \frac{q(X,Y)}{\sum_{\substack{Z\in\mathcal{S} \\ Z\neq X}} q(X,Z)}.$$

   Set X to this sampled state.

Parameter	Value	Source
M	200000	Dowell et al. (1999)
I_0	3	Khan et al. (1999)
α	1 per week	Bwaka et al. (1999), Dowell et al. (1999), and Ndambi et al. (1999)
γ_h	7/5 per week	Khan et al. (1999)
γ_i	7/10 per week	Dowell et al. (1999) and Rowe et al. (1999)
γ_d	35/48 per week	Khan et al. (1999)
θ_1	0.67	Khan et al. (1999)
δ_1	0.8	Khan et al. (1999)
δ_2	0.8	Khan et al. (1999)
β_I	0.588 per week	Legrand et al. (2007)
β_H	0.794 per week	Legrand et al. (2007)
β_F	7.653 per week	Legrand et al. (2007)
γ_{dh}	35/23 per week	Legrand et al. (2007)
γ_{ih}	7/5 per week	Legrand et al. (2007)

Figure B.1: Parameter values for the Ebola model, estimated from the 1995 outbreak in the Democratic Republic of Congo. This table is adapted from Tables 3 and 4 of Legrand et al. (2007), where values sourced directly from Legrand et al. (2007) have been estimated from morbidity data. See the original paper for the interpretations of the parameters.

4. Repeat Steps 2 and 3 until the sample path is terminated.

The result is a single sample X of the population process at time T. To generate N samples, the procedure is repeated N times. The sums are taken across all the possible transitions from the current state, and typically most of the corresponding rates are zero, so the sum does not need to be taken over the entire state space.

B.3 Details of 5-dimensional Ebola model

To conclude the discussion on population processes in Section 7.7, we considered the 5-dimensional model by Legrand et al. (2007) of the 1995 outbreak of Ebola in the Democratic In this appendix, we provide further details on the model. Figure B.1 provides parameter values for the model. These values are taken from a combination of previous studies Bwaka et al. (1999), Dowell et al. (1999), Khan et al. (1999), and Ndambi et al. (1999) and maximum likelihood estimates provided by Legrand et al. (2007). Let $Y_t^{(M)} = \left(S_t/M, E_t/M, I_t/M, H_t/M, D_t/M\right)^\top$ denote the proportion of individuals in each

stage after t weeks. The population process is density dependent, with

$$f\big((s,e,i,h,d),l\big) = \begin{cases} \beta_I si + \beta_H sh + \beta_F sd, & \text{if } l = (-1,1,0,0,0)\,, \\ \alpha e, & \text{if } l = (0,-1,1,0,0)\,, \\ \gamma_h \theta_1 i, & \text{if } l = (0,0,-1,1,0)\,, \\ \gamma_{dh} \delta_2 h, & \text{if } l = (0,0,0,-1,1)\,, \\ \gamma_f d, & \text{if } l = (0,0,0,0,-1)\,, \\ \gamma_i \left(1-\theta_1\right)\left(1-\delta_1\right) i, & \text{if } l = (0,0,-1,0,0)\,, \\ \delta_1 \left(1-\theta_1\right)\gamma_d i, & \text{if } l = (0,0,-1,0,1)\,, \\ \gamma_{ih}\left(1-\delta_2\right) h, & \text{if } l = (0,0,0,-1,0)\,, \\ 0, & \text{otherwise.} \end{cases}$$

The vector field of the fluid limit is

$$Q\Big(Y_t^{(\infty)}\Big) = \begin{bmatrix} -\beta_I Y_t^{(\infty,1)} Y_t^{(\infty,3)} - \beta_H Y_t^{(\infty,1)} Y_t^{(\infty,4)} - \beta_F Y_t^{(\infty,1)} Y_t^{(\infty,5)} \\ \beta_I Y_t^{(\infty,1)} Y_t^{(\infty,3)} + \beta_H Y_t^{(\infty,1)} Y_t^{(\infty,4)} + \beta_F Y_t^{(\infty,1)} Y_t^{(\infty,5)} - \alpha Y_t^{(\infty,2)} \\ \alpha Y_t^{(\infty,2)} - \left(\gamma_h \theta_1 + \gamma_i \left(1-\theta_1\right)\left(1-\delta_1\right) + \delta_1 \left(1-\theta_1\right)\gamma_d\right) Y_t^{(\infty,3)} \\ \gamma_h \theta_1 Y_t^{(\infty,3)} - \left(\gamma_{dh}\delta_2 + \gamma_{ih}\left(1-\delta_2\right)\right) Y_t^{(\infty,4)} \\ \gamma_{dh}\delta_2 Y_t^{(\infty,4)} + \delta_1 \left(1-\theta_1\right)\gamma_d Y_t^{(\infty,3)} - \gamma_f Y_t^{(\infty,5)} \end{bmatrix}.$$

so that the differential equation is

$$\frac{\mathrm{d}Y_t^{(\infty)}}{\mathrm{d}t} = Q\Big(Y_t^{(\infty)}\Big)\,, \quad Y_0^{(\infty)} = \frac{1}{200000}\left(0,0,3,0,0\right)^\top.$$

The diffusion limit is then

$$\mathrm{d}Z_t^{(\infty)} = \nabla Q\Big(Y_t^{(\infty)}\Big) Z_t^{(\infty)}\,\mathrm{d}t + g\Big(Y_t^{(\infty)}\Big)\,, \quad Z_0^{(\infty)} = 0,$$

where the 5×5 diffusion matrix g satisfies

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{11} = \beta_I Y_t^{(\infty,1)} Y_t^{(\infty,3)} + \beta_H Y_t^{(\infty,1)} Y_t^{(\infty,4)} + \beta_F Y_t^{(\infty,1)} Y_t^{(\infty,5)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{12} = \left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{21} = -\beta_I Y_t^{(\infty,1)} Y_t^{(\infty,3)} - \beta_H Y_t^{(\infty,1)} Y_t^{(\infty,4)} - \beta_F Y_t^{(\infty,1)} Y_t^{(\infty,5)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{22} = \beta_I Y_t^{(\infty,1)} Y_t^{(\infty,3)} + \beta_H Y_t^{(\infty,1)} Y_t^{(\infty,4)} + \beta_F Y_t^{(\infty,1)} Y_t^{(\infty,5)} + \alpha Y_t^{(\infty,2)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{23} = \left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{32} = -\alpha Y_t^{(\infty,2)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{33} = \alpha Y_t^{(\infty,2)} + \left(\gamma_H \theta_1 + \gamma_i (1-\theta_1)(1-\delta_1) + \delta_1 (1-\theta_1)\gamma_d\right) Y_t^{(\infty,1)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{34} = \left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{43} = -\gamma_H \theta_1 Y_t^{(\infty,1)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{35} = \left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{53} = -\delta_1 (1-\theta_1)\gamma_d Y_t^{(\infty,1)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{44} = \gamma_H \theta_1 Y_t^{(\infty,1)} + \left(\gamma_{dh}\delta_2 + \gamma_{ih}(1-\delta_2)\right) Y_t^{(\infty,4)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{45} = \left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{54} = -\gamma_{dh}\delta_2 Y_t^{(\infty,4)}$$

$$\left[g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top\right]_{55} = \delta_1 (1-\theta_1)\gamma_d Y_t^{(\infty,1)} + \gamma_{dh}\delta_2 Y_t^{(\infty,4)} + \gamma_f Y_t^{(\infty,5)}$$

with all other entries zero. There are many choices of g such that the product $gg^\top$ has these entries, but to compute the Gaussian approximation with Mazzoni's method, we only need to evaluate $g\left(Y_t^{(\infty)}\right) g\left(Y_t^{(\infty)}\right)^\top$ and therefore no not need to make such a choice.

Bibliography

A. Allawala and J. B. Marston (2016). "Statistics of the stochastically forced Lorenz attractor by the Fokker-Planck equation and cumulant expansions". In: *Physical Review E* 94.5, p. 052218. DOI: 10.1103/PhysRevE.94.052218.

L. J. Allen (2017). "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis". In: *Infectious Disease Modelling* 2.2, pp. 128–142. DOI: 10.1016/j.idm.2017.03.001.

W. Anderson and M. Farazmand (2024). "Fisher information and shape-morphing modes for solving the Fokker–Planck equation in higher dimensions". In: *Applied Mathematics and Computation* 467, p. 128489. DOI: 10.1016/j.amc.2023.128489.

W. J. Anderson (1991). *Continuous-Time Markov Chains*. Ed. by J. Gani and C. C. Heyde. Springer Series in Statistics. New York, NY: Springer New York. DOI: 10.1007/978-1-4612-3038-0.

D. Applebaum (2004). *Lévy Processes and Stochastic Calculus*. 1st. Cambridge Studies in Advanced Mathematics. Cambridge, United Kingdom: Cambridge University Press.

C. Archambeau, D. Cornford, M. Opper, and J. Shawe-Taylor (2007). "Gaussian Process Approximations of Stochastic Differential Equations". In: *Gaussian Processes in Practice*. PMLR, pp. 1–16.

L. Arnold (1998). *Random Dynamical Systems*. 1st ed. Springer Monographs in Mathematics. Berlin, Heidelberg: Springer.

V. I. Arnold (1973). *Ordinary differential equations*. Cambridge: MIT Press.

M. Bachar, J. Batzel, and S. Ditlevsen, eds. (2013). *Stochastic Biomathematical Models*. Vol. 2058. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-32157-3.

A. Badza, T. W. Mattner, and S. Balasuriya (2023). "How sensitive are Lagrangian coherent structures to uncertainties in data?" In: *Physica D: Nonlinear Phenomena* 444, p. 133580. DOI: 10.1016/j.physd.2022.133580.

S. Balasuriya (2017). "Stochastic uncertainty of advected curves in finite-time unsteady flows". In: *Physical Review E* 95.6, p. 062201. DOI: 10.1103/PhysRevE.95.062201.

— (2020a). "Stochastic approaches to Lagrangian coherent structures". In: *The Role of Metrics in the Theory of Partial Differential Equations* 85, pp. 95–105. DOI: 10.2969/aspm/08510095.

S. Balasuriya (2020b). "Stochastic Sensitivity: A Computable Lagrangian Uncertainty Measure for Unsteady Flows". In: *SIAM Review* 62, pp. 781–816. DOI: 10.1137/18M1222922.

— (2020c). "Uncertainty in finite-time Lyapunov exponent computations". In: *Journal of Computational Dynamics* 7.2, pp. 313–337. DOI: 10.3934/jcd.2020013.

S. Balasuriya and G. A. Gottwald (2018). "Estimating stable and unstable sets and their role as transport barriers in stochastic flows". In: *Physical Review E* 98.1, p. 013106. DOI: 10.1103/PhysRevE.98.013106.

S. Balasuriya, N. T. Ouellette, and I. I. Rypina (2018). "Generalized Lagrangian coherent structures". In: *Physica D: Nonlinear Phenomena* 372, pp. 31–51. DOI: 10.1016/j.physd.2018.01.011.

F. Balibrea-Iniesta, C. Lopesino, S. Wiggins, and A. M. Mancho (2016). "Lagrangian Descriptors for Stochastic Differential Equations: A Tool for Revealing the Phase Portrait of Stochastic Dynamical Systems". In: *International Journal of Bifurcation and Chaos* 26.13, p. 1630036. DOI: 10.1142/S0218127416300366.

R. Bellman (1943). "The stability of solutions of linear differential equations". In: *Duke Mathematical Journal* 10.4, pp. 643–647. DOI: 10.1215/S0012-7094-43-01059-2.

P. S. Berloff and J. C. McWilliams (2002). "Material Transport in Oceanic Gyres. Part II: Hierarchy of Stochastic Models". In: *Journal of Physical Oceanography* 32.3, pp. 797–830. DOI: 10.1175/1520-0485(2002)032<0797:MTIOGP>2.0.CO;2.

J. Berner, U. Achatz, L. Batté, et al. (2017). "Stochastic Parameterization: Toward a New View of Weather and Climate Models". In: *Bulletin of the American Meteorological Society* 98.3, pp. 565–588. DOI: 10.1175/BAMS-D-15-00268.1.

J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah (2017). "Julia: A Fresh Approach to Numerical Computing". In: *SIAM Review* 59.1, pp. 65–98. DOI: 10.1137/141000671.

Y. N. Blagoveshchenskii (1962). "Diffusion Processes Depending on a Small Parameter". In: *Theory of Probability and its Applications* 7.2, p. 17. DOI: 10.1137/1107013.

L. Blake, J. Maclean, and S. Balasuriya (2023). *The convergence of stochastic differential equations to their linearisation in small noise limits*. DOI: 10.48550/arXiv.2309.16334. arXiv: 2309.16334 [math].

A. Bracco, J. H. LaCasce, and A. Provenzale (2000). "Velocity Probability Density Functions for Oceanic Floats". In: *Journal of Physical Oceanography* 30.3, pp. 461–474. DOI: 10.1175/1520-0485(2000)030<0461:VPDFFO>2.0.CO;2.

M. Branicki and K. Uda (2021). "Lagrangian Uncertainty Quantification and Information Inequalities for Stochastic Flows". In: *SIAM/ASA Journal on Uncertainty Quantification* 9.3, pp. 1242–1313.

— (2023). "Path-Based Divergence Rates and Lagrangian Uncertainty in Stochastic Flows". In: *SIAM Journal on Applied Dynamical Systems*, pp. 419–482. DOI: 10.1137/21M1466530.

F. Brauer (2008). "Compartmental Models in Epidemiology". In: *Mathematical Epidemiology*. Ed. by F. Brauer, P. van den Driessche, and J. Wu. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, pp. 19–79. DOI: 10.1007/978-3-540-78911-6_2.

P. Brémaud (2020). *Probability Theory and Stochastic Processes*. 1st ed. Universitext. Springer.

S. L. Brunton and C. W. Rowley (2010). "Fast computation of finite-time Lyapunov exponent fields for unsteady flows". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20.1, p. 017503. DOI: 10.1063/1.3270044.

A. Budhiraja, E. Friedlander, C. Guider, et al. (2019). "Assimilating Data into Models". In: *Handbook of Environmental and Ecological Statistics*. 1st Edition. Chapman and Hall/CRC.

M. A. Bwaka, M. J. Bonnet, P. Calain, et al. (1999). "Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: clinical observations in 103 patients". In: *The Journal of Infectious Diseases*, S1–7. DOI: 10.1086/514308.

M. Capderou (2014). *Handbook of Satellite Orbits*. 1st ed. Springer Cham.

M. Collins (2007). "Ensembles and probabilities: a new era in the prediction of climate change". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365.1857, pp. 1957–1970. DOI: 10.1098/rsta.2007.2068.

Á. Correales and C. Escudero (2019). "Itô vs Stratonovich in the presence of absorbing states". In: *Journal of Mathematical Physics* 60.12, p. 123301. DOI: 10.1063/1.5081791.

C. J. Cotter and G. A. Pavliotis (2009). "Estimating eddy diffusivities from noisy Lagrangian observations". In: *Communications in Mathematical Sciences* 7.4, pp. 805–838. DOI: 10.4310/CMS.2009.v7.n4.a2.

R. Crocker (2021). "Lagrangian Coherent Data Assimilation for chaotic geophysical systems". Master of Philosophy. The University of Adelaide.

A. Dawson, T. N. Palmer, and S. Corti (2012). "Simulating regime structures in weather and climate prediction models". In: *Geophysical Research Letters* 39.21. DOI: 10.1029/2012GL053284.

A. Dawson and T. N. Palmer (2015). "Simulating weather regimes: impact of model resolution and stochastic parameterization". In: *Climate Dynamics* 44.7, pp. 2177–2193. DOI: 10.1007/s00382-014-2238-x.

D. del-Castillo-Negrete (1998). "Asymmetric transport and non-Gaussian statistics of passive scalars in vortices in shear". In: *Physics of Fluids* 10.3, pp. 576–594. DOI: 10.1063/1.869585.

K. J. DeMars, R. H. Bishop, and M. K. Jah (2013). "Entropy-Based Approach for Uncertainty Propagation of Nonlinear Dynamical Systems". In: *Journal of Guidance, Control, and Dynamics* 36.4, pp. 1047–1057. DOI: 10.2514/1.58987.

A. Denner, O. Junge, and D. Matthes (2016). "Computing coherent sets using the Fokker-Planck equation". In: *Journal of Computational Dynamics* 3.2, p. 163. DOI: 10.3934/jcd.2016008.

R. Ding and A. Mullhaupt (2023). "Empirical Squared Hellinger Distance Estimator and Generalizations to a Family of α-Divergence Estimators". In: *Entropy* 25.4, p. 612. DOI: 10.3390/e25040612.

P. D. Ditlevsen (1999). "Observation of α-stable noise induced millennial climate changes from an ice-core record". In: *Geophysical Research Letters* 26.10, pp. 1441–1444. DOI: 10.1029/1999GL900252.

F. Doglioni, R. Ricker, B. Rabe, and T. Kanzow (2021). *Sea surface height anomaly and geostrophic velocity from altimetry measurements over the Arctic Ocean (2011–2018)*. preprint. Oceanography – Physical. DOI: 10.5194/essd-2021-170.

T. Dombre, U. Frisch, J. M. Greene, et al. (1986). "Chaotic streamlines in the ABC flows". In: *Journal of Fluid Mechanics* 167, pp. 353–391. DOI: 10.1017/S0022112086002859.

S. F. Dowell, R. Mukunu, T. G. Ksiazek, et al. (1999). "Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidémies à Kikwit". In: *The Journal of Infectious Diseases*, S87–91. DOI: 10.1086/514284.

E.U. Copernicus Marine Service (CMEMS) (2020). *Global Ocean Gridded L4 Sea Surface Heights And Derived Variables Reprocessed 1993 Ongoing*. DOI: 10.48670/moi-00148.

L. Fang, S. Balasuriya, and N. T. Ouellette (2019). "Local linearity, coherent structures, and scale-to-scale coupling in turbulent flow". In: *Physical Review Fluids* 4.1, p. 014501. DOI: 10.1103/PhysRevFluids.4.014501.

— (2020). "Disentangling resolution, precision, and inherent stochasticity in nonlinear systems". In: *Physical Review Research* 2.2, p. 023343. DOI: 10.1103/PhysRevResearch.2.023343.

L. Fang and N. T. Ouellette (2021). "Assessing the information content of complex flows". In: *Physical Review E* 103.2, p. 023301. DOI: 10.1103/PhysRevE.103.023301.

F. Feppon and P. F. J. Lermusiaux (2018). "Dynamically Orthogonal Numerical Schemes for Efficient Stochastic Advection and Lagrangian Transport". In: *SIAM Review* 60.3, pp. 595–625. DOI: 10.1137/16M1109394.

M. I. Freidlin and A. D. Wentzell (1998). *Random Perturbations of Dynamical Systems*. 2nd ed. Grundlehren der mathematischen Wissenschaften. New York, NY: Springer.

P. K. Friz and N. B. Victoir (2010). *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511845079.

G. Froyland (2013). "An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems". In: *Physica D: Nonlinear Phenomena* 250, pp. 1–19. DOI: 10.1016/j.physd.2013.01.013.

C. W. Gardiner, A. S. Parkins, and P. Zoller (1992). "Wave-function quantum stochastic differential equations and quantum-jump simulation methods". In: *Physical Review A* 46.7, pp. 4363–4381. DOI: 10.1103/PhysRevA.46.4363.

D. T. Gillespie (1977). "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25, pp. 2340–2361. DOI: 10.1021/j100540a008.

G. A. Gottwald and I. Melbourne (2013). "Homogenization for deterministic maps and multiplicative noise". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469.2156, p. 20130201. DOI: 10.1098/rspa.2013.0201.

H. Guo, W. He, T. Peterka, et al. (2016). "Finite-Time Lyapunov Exponents and Lagrangian Coherent Structures in Uncertain Unsteady Flows". In: *IEEE transactions on visualization and computer graphics* 22.6, pp. 1672–1682. DOI: 10.1109/TVCG.2016.2534560.

S. Ha, J. Berner, and C. Snyder (2015). "A Comparison of Model Error Representations in Mesoscale Ensemble Data Assimilation". In: *Monthly Weather Review* 143.10, pp. 3893–3911. DOI: 10.1175/MWR-D-14-00395.1.

A. Hadjighasem, M. Farazmand, D. Blazevski, et al. (2017). "A critical comparison of Lagrangian methods for coherent structure detection". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.5, p. 053104. DOI: 10.1063/1.4982720.

G. Haller (2001). "Distinguished material surfaces and coherent structures in three-dimensional fluid flows". In: *Physica D: Nonlinear Phenomena* 149.4, pp. 248–277. DOI: 10.1016/S0167-2789(00)00199-8.

G. Haller, D. Karrasch, and F. Kogelbauer (2018). "Material barriers to diffusive and stochastic transport". In: *Proceedings of the National Academy of Sciences* 115.37, pp. 9074–9079. DOI: 10.1073/pnas.1720177115.

J. H. Hubbard and B. B. Hubbard (2009). *Vector calculus, linear algebra, and differential forms: a unified approach*. 4th ed. Ithaca, NY: Matrix Editions.

K. Itô (1944). "Stochastic integral". In: *Proceedings of the Imperial Academy* 20.8, pp. 519–524. DOI: 10.3792/pia/1195572786.

— (1946). "On a stochastic integral equation". In: *Proceedings of the Japan Academy* 22.2, pp. 32–35. DOI: 10.3792/pja/1195572371.

A. H. Jazwinski (2014). *Stochastic Processes and Filtering Theory*. Vol. 64. Mathematics in Science and Engineering. Burlington: Elsevier Science.

I. T. Jolliffe (2002). *Principal Component Analysis*. 2nd ed. Springer Series in Statistics. New York, NY: Springer.

S. Julier, J. Uhlmann, and H. Durrant-Whyte (2000). "A new method for the nonlinear transformation of means and covariances in filters and estimators". In: *IEEE Transactions on Automatic Control* 45.3, pp. 477–482. DOI: 10.1109/9.847726.

Y. Kabanov, R. Liptser, and J. Stoyanov (2006). *From Stochastic Calculus to Mathematical Finance*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-30788-4.

G. Kallianpur and P. Sundar (2014). *Stochastic analysis and diffusion processes*. First edition. Oxford graduate texts in mathematics 24. Oxford, United Kingdom: Oxford University Press.

I. Kamenkovich, I. I. Rypina, and P. Berloff (2015). "Properties and Origins of the Anisotropic Eddy-Induced Transport in the North Atlantic". In: *Journal of Physical Oceanography* 45.3, pp. 778–791. DOI: 10.1175/JPO-D-14-0164.1.

D. Kang and E. N. Curchitser (2013). "Gulf Stream eddy characteristics in a high-resolution ocean model". In: *Journal of Geophysical Research: Oceans* 118.9, pp. 4474–4487. DOI: 10.1002/jgrc.20318.

B. Kaszás and G. Haller (2020). "Universal upper estimate for prediction errors under moderate model uncertainty". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.11, p. 113144. DOI: 10.1063/5.0021665.

D. G. Kendall (1949). "Stochastic Processes and Population Growth". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 11.2, pp. 230–282.

A. S. Khan, F. K. Tshioko, D. L. Heymann, et al. (1999). "The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidémies à Kikwit". In: *The Journal of Infectious Diseases*, S76–86. DOI: 10.1086/514306.

P. E. Kloeden and E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. 1st. Applications of Mathematics. Berlin, Heidelberg: Springer.

M. V. Kulikova and G. Y. Kulikov (2014). "Adaptive ODE solvers in extended Kalman filtering algorithms". In: *Journal of Computational and Applied Mathematics*. Selected Papers from NUMDIFF-13 262, pp. 205–216. DOI: 10.1016/j.cam.2013.09.064.

T. G. Kurtz (1970). "Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes". In: *Journal of Applied Probability* 7.1, pp. 49–58. DOI: 10.2307/3212147.

— (1971). "Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes". In: *Journal of Applied Probability* 8.2, pp. 344–356. DOI: 10.2307/3211904.

K. Law, A. M. Stuart, and K. Zygalakis (2015). *Data Assimilation: A Mathematical Introduction*. Texts in applied mathematics volume 62. Cham Heidelberg New York Dordrecht London: Springer. DOI: 10.1007/978-3-319-20325-6.

L. Le Cam and G. Lo Yang (2000). *Asymptotics in Statistics*. Springer Series in Statistics. New York, NY: Springer New York. DOI: 10.1007/978-1-4612-1166-2.

F. LeGland and B. Wang (2002). "Asymptotic Normality in Partially Observed Diffusions with Small Noise: Application to FDI". In: *Stochastic Theory and Control*. Ed. by B. Pasik-Duncan. Lecture Notes in Control and Information Sciences. Berlin, Heidelberg: Springer, pp. 267–282. DOI: 10.1007/3-540-48022-6_19.

J. Legrand, R. F. Grais, P. Y. Boelle, et al. (2007). "Understanding the dynamics of Ebola epidemics". In: *Epidemiology and Infection* 135.4, pp. 610–621. DOI: 10.1017/S0950268806007217.

M. Leutbecher (2019). "Ensemble size: How suboptimal is less than infinity?" In: *Quarterly Journal of the Royal Meteorological Society* 145.S1, pp. 107–128. DOI: 10.1002/qj.3387.

M. Leutbecher, S.-J. Lock, P. Ollinaho, et al. (2017). "Stochastic representations of model uncertainties at ECMWF: state of the art and future vision". In: *Quarterly Journal of the Royal Meteorological Society* 143.707, pp. 2315–2339. DOI: 10.1002/qj.3094.

L. Li, B. Deremble, N. Lahaye, and E. Mémin (2023). "Stochastic Data-Driven Parameterization of Unresolved Eddy Effects in a Baroclinic Quasi-Geostrophic Model". In: *Journal of Advances in Modeling Earth Systems* 15.2, e2022MS003297. DOI: 10.1029/2022MS003297.

Y. Li (2019). "A data-driven method for the steady state of randomly perturbed dynamics". In: *Communications in Mathematical Sciences* 17.4, pp. 1045–1059. DOI: 10.4310/CMS.2019.v17.n4.a9.

R. Mannella (1999). "Absorbing boundaries and optimal stopping in a stochastic differential equation". In: *Physics Letters A* 254.5, pp. 257–262. DOI: 10.1016/S0375-9601(99)00117-6.

T. Mazzoni (2008). "Computational aspects of continuous–discrete extended Kalman-filtering". In: *Computational Statistics* 23.4, pp. 519–539. DOI: 10.1007/s00180-007-0094-4.

G. J. McLachlan, S. X. Lee, and S. I. Rathnayake (2019). "Finite Mixture Models". In: *Annual Review of Statistics and Its Application* 6.1, pp. 355–378. DOI: 10.1146/annurev-statistics-031017-100325.

I. Melbourne and A. M. Stuart (2011). "A note on diffusion limits of chaotic skew-product flows". In: *Nonlinearity* 24.4, p. 1361. DOI: 10.1088/0951-7715/24/4/018.

H. M. T. Menegaz, J. Y. Ishihara, G. A. Borges, and A. N. Vargas (2015). "A Systematization of the Unscented Kalman Filter Theory". In: *IEEE Transactions on Automatic Control* 60.10, pp. 2583–2598. DOI: 10.1109/TAC.2015.2404511.

I. Mezić, S. Loire, V. A. Fonoberov, and P. Hogan (2010). "A New Mixing Diagnostic and Gulf Oil Spill Movement". In: *Science* 330.6003, pp. 486–489. DOI: 10.1126/science.1194607.

L. Mitchell and G. A. Gottwald (2012). "Data Assimilation in Slow–Fast Systems Using Homogenized Climate Models". In: *Journal of the Atmospheric Sciences* 69.4, pp. 1359–1377. DOI: 10.1175/JAS-D-11-0145.1.

C. M. Mora, H. A. Mardones, J. C. Jimenez, et al. (2017). "A Stable Numerical Scheme for Stochastic Differential Equations with Multiplicative Noise". In: *SIAM Journal on Numerical Analysis* 55.4, pp. 1614–1649. DOI: 10.1137/140984488.

M. A. Muñoz (1998). "Nature of different types of absorbing states". In: *Physical Review E* 57.2, pp. 1377–1383. DOI: 10.1103/PhysRevE.57.1377.

R. Ndambi, P. Akamituna, M. J. Bonnet, et al. (1999). "Epidemiologic and clinical aspects of the Ebola virus epidemic in Mosango, Democratic Republic of the Congo, 1995". In: *The Journal of Infectious Diseases*, S8–10. DOI: 10.1086/514297.

T. Neckel and F. Rupp (2013). *Random Differential Equations in Scientific Computing*. De Gruyter Open Poland. DOI: 10.2478/9788376560267.

D. A. Nelson and G. B. Jacobs (2016). "High-order visualization of three-dimensional Lagrangian coherent structures with DG-FTLE". In: *Computers & Fluids*. 13th US-NCCM International Symposium of High-Order Methods for Computational Fluid Dynamics - A special issue dedicated to the 60th birthday of Professor David Kopriva 139, pp. 197–215. DOI: 10.1016/j.compfluid.2016.07.007.

J. N. Nielsen, H. Madsen, and P. C. Young (2000). "Parameter estimation in stochastic differential equations: An overview". In: *Annual Reviews in Control* 24, pp. 83–94. DOI: 10.1016/S1367-5788(00)90017-8.

B. Øksendal (2003). *Stochastic Differential Equations*. 6th. Universitext. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-14394-6.

M. J. Olascoaga, F. J. Beron-Vera, G. Haller, et al. (2013). "Drifter motion in the Gulf of Mexico constrained by altimetric Lagrangian coherent structures". In: *Geophysical Research Letters* 40.23, pp. 6171–6175. DOI: 10.1002/2013GL058624.

M. J. Olascoaga and G. Haller (2012). "Forecasting sudden changes in environmental pollution patterns". In: *Proceedings of the National Academy of Sciences* 109.13, pp. 4738–4743. DOI: 10.1073/pnas.1118574109.

T. N. Palmer (2019). "Stochastic weather and climate models". In: *Nature Reviews Physics* 1.7, pp. 463–471. DOI: 10.1038/s42254-019-0062-2.

J. B. Palter (2015). "The role of the Gulf Stream in European climate". In: *Annual Review of Marine Science* 7, pp. 113–137. DOI: 10.1146/annurev-marine-010814-015656.

Y.-H. Park (2004). "Determination of the surface geostrophic velocity field from satellite altimetry". In: *Journal of Geophysical Research* 109.C5, p. C05006. DOI: 10.1029/2003JC002115.

G. A. Pavliotis and A. M. Stuart (2008). *Multiscale Methods*. 1st ed. Vol. 53. Texts Applied in Mathematics. New York, NY: Springer New York. DOI: 10.1007/978-0-387-73829-1.

T. Peacock and J. O. Dabiri (2010). "Introduction to Focus Issue: Lagrangian Coherent Structures". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20.1, p. 017501. DOI: 10.1063/1.3278173.

C. Penland (2003). "A Stochastic Approach to Nonlinear Dynamics". In: *Bulletin of the American Meteorological Society* 84.7, pp. 43–52.

F. Perez-Cruz (2008). "Kullback-Leibler divergence estimation of continuous distributions". In: *2008 IEEE International Symposium on Information Theory*. Toronto, ON, Canada: IEEE, pp. 1666–1670. DOI: 10.1109/ISIT.2008.4595271.

L. Pichler, A. Masud, and L. A. Bergman (2013). "Numerical Solution of the Fokker–Planck Equation by Finite Difference and Finite Element Methods—A Comparative Study". In: *Computational Methods in Stochastic Dynamics: Volume 2*. Ed. by M. Papadrakakis, G. Stefanou, and V. Papadopoulos. Computational Methods in Applied Sciences. Dordrecht: Springer Netherlands, pp. 69–85. DOI: 10.1007/978-94-007-5134-7_5.

R. T. Pierrehumbert (1991). "Chaotic mixing of tracer and vorticity by modulated travelling Rossby waves". In: *Geophysical & Astrophysical Fluid Dynamics* 58.1-4, pp. 285–319. DOI: 10.1080/03091929108227343.

A. Pilipenko (2014). *An Introduction to Stochastic Differential Equations with Reflection*. Lectures in Pure and Applied Mathematics 1. Potsdam: Univ.-Verl.

P. K. Pollett (1990). "On a model for interference between searching insect parasites". In: *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics* 32.2, pp. 133–150. DOI: 10.1017/S0334270000008390.

P. K. Pollett, A. H. Dooley, and J. V. Ross (2010). "Modelling population processes with random initial conditions". In: *Mathematical Biosciences* 223.2, pp. 142–150. DOI: 10.1016/j.mbs.2009.11.008.

H. K. Preisler, A. A. Ager, B. K. Johnson, and J. G. Kie (2004). "Modeling animal movements using stochastic differential equations". In: *Environmetrics* 15.7, pp. 643–657. DOI: 10.1002/env.636.

C. Rackauckas and Q. Nie (2017a). "Adaptive methods for stochastic differential equations via natural embeddings and rejection sampling with memory". In: *Discrete and continuous dynamical systems. Series B* 22.7, pp. 2731–2761. DOI: 10.3934/dcdsb.2017133.

— (2017b). "DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia". In: *Journal of Open Research Software* 5.1, p. 15. DOI: 10.5334/jors.151.

S. Reich and C. J. Cotter (2015). *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107706804.

H. Risken (2012). *The Fokker-Planck Equation: Methods of Solution and Applications*. 1st ed. Springer Series in Synergetics. Berlin, Heidelberg: Springer.

A. Rößler (2010). "Runge-Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations". In: *SIAM Journal on Numerical Analysis* 48.3, pp. 922–952. DOI: 10.1137/09076636X.

A. K. Rowe, J. Bertolli, A. S. Khan, et al. (1999). "Clinical, virologic, and immunologic follow-up of convalescent Ebola hemorrhagic fever patients and their household contacts, Kikwit, Democratic Republic of the Congo. Commission de Lutte contre les Epidémies à Kikwit". In: *The Journal of Infectious Diseases*, S28–35. DOI: 10.1086/514318.

R. M. Samelson and S. Wiggins (2006). *Lagrangian Transport in Geophysical Jets and Waves: The Dynamical Systems Approach*. Vol. 31. Interdisciplinary Applied Mathematics. New York, NY: Springer.

J. M. Sancho, M. S. Miguel, S. L. Katz, and J. D. Gunton (1982). "Analytical and numerical studies of multiplicative noise". In: *Physical Review A* 26.3, pp. 1589–1609. DOI: 10.1103/PhysRevA.26.1589.

D. Sanz-Alonso and A. M. Stuart (2017). "Gaussian Approximations of Small Noise Diffusions in Kullback-Leibler Divergence". In: *Communications in Mathematical Sciences* 15.7, pp. 2087–2097. DOI: https://dx.doi.org/10.4310/CMS.2017.v15.n7.a13.

S. Särkkä and A. Solin (2019). *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge: Cambridge University Press. DOI: 10.1017/9781108186735.

S. C. Shadden, F. Lekien, and J. E. Marsden (2005). "Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows". In: *Physica D: Nonlinear Phenomena* 212.3, pp. 271–304. DOI: 10.1016/j.physd.2005.10.007.

B. W. Silverman (2017). *Density Estimation for Statistics and Data Analysis*. New York: Routledge. DOI: 10.1201/9781315140919.

R. L. Stratonovich (1966). "A New Representation for Stochastic Integrals and Equations". In: *SIAM Journal on Control* 4.2, pp. 362–371. DOI: 10.1137/0304028.

R. D. T. Strauss and F. Effenberger (2017). "A Hitch-hiker's Guide to Stochastic Differential Equations". In: *Space Science Reviews* 212.1, pp. 151–192. DOI: 10.1007/s11214-017-0351-y.

M. H. M. Sulman, H. S. Huntley, B. L. Lipphardt, and A. D. Kirwan (2013). "Leaving flatland: Diagnostics for Lagrangian coherent structures in three-dimensional flows". In: *Physica D: Nonlinear Phenomena* 258, pp. 77–92. DOI: 10.1016/j.physd.2013.05.005.

P. Sura (2003). "Stochastic Analysis of Southern and Pacific Ocean Sea Surface Winds". In: *Journal of the Atmospheric Sciences* 60.4, pp. 654–666. DOI: 10.1175/1520-0469(2003)060<0654:SAOSAP>2.0.CO;2.

P. Sura, M. Newman, C. Penland, and P. Sardeshmukh (2005). "Multiplicative Noise and Non-Gaussianity: A Paradigm for Atmospheric Regimes?" In: *Journal of the Atmospheric Sciences* 62.5, pp. 1391–1409. DOI: 10.1175/JAS3408.1.

C. Truesdell and W. Noll (2004). *The Non-Linear Field Theories of Mechanics*. Ed. by S. S. Antman. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-662-10388-3.

J. Uhlmann (1995). "Dynamic map building and localization : new theoretical foundations." PhD thesis. University of Oxford.

E. van Sebille, S. M. Griffies, R. Abernathey, et al. (2018). "Lagrangian ocean analysis: Fundamentals and practices". In: *Ocean Modelling* 121, pp. 49–75. DOI: 10.1016/j.ocemod.2017.11.008.

J. A. Viecelli (1998). "On the Possibility of Singular Low-Frequency Spectra and Lévy Law Persistence Statistics in the Planetary-Scale Turbulent Circulation". In: *Journal of the*

Atmospheric Sciences 55.5, pp. 677–689. DOI: 10.1175/1520-0469(1998)055<0677:OTPOSL>2.0.CO;2.

J. D. Wilson and B. L. Sawford (1996). "Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere". In: *Boundary-Layer Meteorology* 78.1, pp. 191–210. DOI: 10.1007/BF00122492.

Y. K. Ying, J. R. Maddison, and J. Vanneste (2019). "Bayesian inference of ocean diffusivity from Lagrangian trajectory data". In: *Ocean Modelling* 140, p. 101401. DOI: 10.1016/j.ocemod.2019.101401.

G. You and S. Leung (2021). "Computing the finite time Lyapunov exponent for flows with uncertainties". In: *Journal of Computational Physics* 425, p. 109905. DOI: 10.1016/j.jcp.2020.109905.

J. Zhai, M. Dobson, and Y. Li (2022). "A deep learning method for solving Fokker-Planck equations". In: *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*. PMLR, pp. 568–597.